

Классификация мимики эмоциональных лиц людьми и нейронной сетью на материале экологически валидной базы изображений

А. В. Сладкоштитева^{1а}, А. С. Стародубцев¹, А. В. Петракова²

¹ Санкт-Петербургский государственный университет,

Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

² Национальный исследовательский университет «Высшая школа экономики»,

Российская Федерация, 101000, Москва, ул. Мясницкая, 20

Для цитирования: Сладкоштитева А. В., Стародубцев А. С., Петракова А. В. Классификация мимики эмоциональных лиц людьми и нейронной сетью на материале экологически валидной базы изображений // Вестник Санкт-Петербургского университета. Психология. 2025. Т. 15. Вып. 1. С. 116–133. <https://doi.org/10.21638/spbu16.2025.107>

В данной работе исследуется проблема универсальных мимических паттернов для валентности эмоций, эмоциональных семейств и культурно-специфических эмоций. Для решения этой проблемы проводилось множество исследований, но в них по результатам распознавания эмоций респондентами иногда делались выводы об универсальности выражения эмоций, а также использовались традиционные базы изображений, в которых не соблюдаются условия экологической валидности. Сравняются четыре группы теорий происхождения эмоций: теория универсальности базовых эмоций, теория культурной специфичности базовых эмоций, теория социального научения эмоциям и теория социального конструирования эмоций. Цель исследования: установить, имеет ли мимика людей специфические паттерны для положительной и отрицательной валентностей эмоций (степень положительной или отрицательной аффективной реакции), эмоциональных семейств (включает в себя базовую эмоцию и вариации базовой эмоции) и конкретных эмоций при использовании базы изображений, авторы которой стремились к экологической валидности: фотографируемым не задавалось стандарта паттернов мимики для демонстрации эмоций. Для независимого от респондентов способа рассмотрения процесса выражения используется нейронная сеть *teachable machine*. Использовалось гетерогенное трансферное обучение. В нашей работе респонденты и нейросеть классифицировали изображения эмоциональных лиц по 14 эмоциям. Респонденты чаще нейросети правильно классифицировали конкретные эмоции. И нейросеть, и люди более эффективно распознавали валентность и эмоциональные семейства, чем конкретные эмоции. Это может говорить о наличии специфических паттернов эмоциональных семейств и валентности эмоций. Возможна слабая специфичность на уровне конкретных эмоций. Результаты соответствуют теории базовых эмоций «New BET» А. Скарантино. Возможна интерпретация в русле теории социального научения эмоциям. Обсуждаются ограничения нейронной сети в распознавании всего диапазона вариаций прототипических выражений эмоций. Описываются дальнейшие возможные кросс-культурные исследования для уточнения результатов.

Ключевые слова: базовые эмоции, конструктивизм, классификация эмоций, выражение эмоций.

^а Автор для корреспонденции.

© Санкт-Петербургский государственный университет, 2025

Введение

Данная работа посвящена проблеме происхождения эмоций (Scarantino, 2015). Согласно концепции базовых эмоций П. Экмана (basic emotion theory of P. Ekman), эмоции являются биологически детерминированными процессами, универсальными во всех культурах (Ekman, 1973). В теории культурной специфичности базовых эмоций утверждается, что для разных культур характерны различные проявления базовых эмоций (Scherer, Wallbott, 1994). Противоположный взгляд на эмоции у сторонников социального научения эмоциям (social learning theory of emotion, см., например, (Bandura, Walters, 1977)). Они считают, что базовых эмоций не существует: есть культурно-специфические эмоции, которым обучает социум. Подход Л. Барретт (L. Barrett) отрицает существование эмоций как отдельных психических процессов (the theory of constructed emotion). Согласно данному подходу, эмоциональные состояния конструирует мозг: «Мозг предвидит входящие сенсорные сигналы, которые он реализует в виде предсказаний» (Barrett, Simmons, 2015, p. 1). Например, «радость», «грусть» — это понятия для совокупности случаев, которыми пользуется мозг для предсказаний о происходящем с гомеостазом тела и событий вокруг.

Концепцию базовых эмоций обосновывают данные о сходстве эмоциональных лиц людей в дописьменных и грамотных, а также в западных и восточных культурах (Ekman, 1999b). К примеру, люди из племени форе выбирали одно и то же выражение лица для разных историй-ситуаций (Ekman, Friesen, 1971). Однако последующее изучение дописьменных культур не получило столь однозначных результатов, см., например, (Gendron et al., 2020).

Вариативность мимических проявлений эмоций объясняет теория культурной специфичности базовых эмоций. У. Сато (W. Sato) и коллеги обнаружили отличия выражений лиц японцев от прототипических западных выражений (Sato et al., 2019). Анкетные исследования показали, что культура респондентов влияет на регуляцию эмоций (Scherer, Wallbott, 1994). Но все же эта теория, как и традиционная теория универсальных базовых эмоций, не объясняет существования множества слов для обозначения эмоций, не имеющих аналогов в других языках.

Теория социального научения справляется с феноменом существования множества слов для эмоций. Если эмоции порождены культурой, то самих эмоций может быть сколь угодно много. Исследования в русле теории социального научения показывают, что человек лучше распознает эмоции людей, принадлежащих к той же культуре, что и воспринимающий (Gendron et al., 2014). Исследования развития эмоций утверждают важность родительского и семейного воспитания для их формирования (Trommsdorff, 2006). Дж. Расселл (J. Russell) заключил, что даже внутри одной культуры существует вариативность в проявлении и распознавании эмоций (Russell, 1994).

Теория социального конструирования эмоций объясняет большую их вариативность тем, что специальных механизмов для эмоций нет вообще (Barrett, 2011). Эта концепция утверждает, что нет связи выражения лица и эмоции, которую испытывает человек (Barrett, Simmons, 2015). По лицевым паттернам без контекста тяжело определить даже вариативность проявления базового аффекта по двум шкалам — валентности (положительная и отрицательная) и уровню возбуждения

(низкий и высокий) (Leon et al., 2004). Навык лучше распознавать эмоции своей культуры объясняется тем, что понятие эмоции содержит диапазон случаев ее проявления и в рамках этого диапазона люди могут верно называть эмоции, предполагаемые экспериментатором.

Теория базовых эмоций П. Экмана допускает вариативность мимики одной и той же эмоции внутри культуры только на уровне одного прототипа. Существует базовая эмоция и ее незначительные вариации — субэмоции, которые формируются путем некоторого изменения первоначального прототипа. Экман описывал возможную вариативность внутри эмоционального семейства (базовая эмоция и субэмоции) влиянием культуры, интенсивности эмоции, контроля над проявлениями эмоции со стороны человека (Ekman, 1992). Вариативность с возможностью существования нескольких прототипов учитывает новая теория базовых эмоций (New BET) (Scarantino, 2015). Теория предполагает наличие нескольких механизмов, специфических для эмоций. Проявление в мимике базовой эмоции — это набор гибких тенденций реагирования, адаптируемых к текущему контексту, а не жесткая однозначная программа. Не существует единого прототипа эмоции, как у Экмана. Мы придерживаемся подхода А. Скарантино, но предполагаем, что на процесс распознавания эмоций людьми могут влиять социальные факторы. Также мы считаем необходимым учитывать критику симулированных актерских эмоций в наборах изображений и пользоваться стимулами, которые разрабатывались с учетом экологической валидности (Grühn, Sharifian, 2016).

Настоящее исследование

Выражение эмоций — процесс внешнего проявления через различные каналы, такие как мимика, жесты и т. д. Распознавание эмоций — процесс интерпретации внешнего проявления эмоций других людей. На наш взгляд, в исследованиях эмоций авторы иногда распространяют выводы о распознавании эмоциональных экспрессий на объяснение процесса выражения (Ekman, Friesen, 1971; Ekman, 1999b). В данном исследовании мы предполагаем, что социальные факторы в большей степени влияют именно на распознавание эмоции, а не на выражение. Наличие в языке разных слов для обозначения эмоций может являться лингвистической ловушкой, которая может помешать определить базовую эмоцию, потому что слова-эмоции могут обозначать разные моменты процесса одной и той же эмоции, они могут быть связаны с целью, причинами, контекстом, присущим конкретной культуре (Oatley, Johnson-Laird, 2014). Однако выражение лица у разных людей в рамках одной и той же эмоции может иметь схожие паттерны. Существуют способы разделить процессы распознавания и выражения эмоций: изучение приматов, электромиография и другие физиологические методы и т. д. Тем не менее данные методы подвергаются критике, см., например, (Barrett et al., 2019). В данной работе мы предпринимаем попытку разделить данные процессы, сравнивая результаты классификации эмоциональных лиц нейросетью и респондентами. Нейросеть имеет свои особенности получения сигнала и свои ошибки, не тождественные особенностям человеческого распознавания. Она опирается именно на визуализационный материал и — в меньшей степени — на социальные факторы (нейронная сеть будет объективно искать общие паттерны в изображениях, опираясь только на обучение

категориям и предоставленный визуализационный материал, но не на свой прошлый опыт использования слов для эмоций, в отличие от людей). Мы предположили, что нейронная сеть, которой не свойственны ошибки распознавания, характерные для человека, может являться подходящим способом исследования выражения эмоций. Мы использовали нейронную сеть на базе данных, авторы которой стремятся учитывать экологическую валидность (Петракова и др., 2024). Фотографируемым не было задано стандартов мимических выражений эмоций. Также в целях избегания демонстрирования эмоции по принятому шаблону сама эмоция не называлась, вместо этого описывалась ситуация, в которой данная эмоция должна проявиться. Так как база содержит только изображения лиц из российской выборки и респондентами являются также россияне, данное исследование позволит решить не широкий вопрос межкультурной универсальности проявлений конкретной эмоции, а более узкий — есть ли универсальность или разнообразие паттернов эмоции внутри одной культуры.

Ранее на материале эмоциональных лиц проводились исследования с использованием нейросетей. В предыдущих исследованиях нейронным сетям заранее задавались области изображений, по которым нужно определять эмоцию (Рюмина, Карпов, 2020). Они применялись как инструмент для изучения распознавания эмоций, включали наборы изображений, соответствующие классическим стандартам изображения эмоции по шаблону, заданному исследователем. В разработанный нами набор изображений входят лица моделей, которым не давалось инструкций по поводу того, как правильно демонстрировать эмоции через мимику. Фотографируемые сами определяли, как им лучше показать эмоцию. Это позволит изучить мимические паттерны, более приближенные к реальности, чем в традиционных датасетах изображений, созданных по системе FACS.

Основная цель исследования — установить, имеет ли мимика людей специфические паттерны для положительной и отрицательной валентностей эмоций, а также эмоциональных семейств и специфических эмоций. Аспекты научной новизны нашего исследования:

- использование новой базы экологически валидных изображений эмоциональных лиц;
- использование нейросети как инструмента для изучения выражения эмоций;
- проверка предположения о том, что социальные факторы в большей мере влияют на распознавание, чем на выражение эмоций.

Зависимые переменные — успешность классификации эмоций на уровнях точной категоризации, эмоционального семейства и валентности, измеряемая процентом правильных ответов. Уровень точной категоризации предполагает определение испытуемым в точности той эмоции, которая предполагалась авторами базы изображений. Эмоциональное семейство включает в себя базовую эмоцию и субэмоции (вариации базовой эмоции). Валентность эмоции — степень положительной или отрицательной аффективной (эмоциональной) реакции.

Были выдвинуты следующие гипотезы:

- классификация эмоций испытуемыми выше случайного будет наблюдаться начиная с уровня эмоционального семейства (исходя из теории базовых эмоций мы предполагаем, что у небазовых эмоций нет специфических паттернов выражения, а значит, их невозможно успешно распознать только по мимике);

- нейронная сеть будет точнее человека категоризировать эмоции на уровне эмоционального семейства и валентности (так как нейронная сеть в меньшей степени подвержена влиянию слов в языке для обозначения эмоций);
- с возрастанием интенсивности эмоций будет возрастать точность их категоризации.

Методы

Выборка. 33 человека с нормальным или скорректированным до нормального зрением (18–44 года, $M = 23$, $SD = 5$), 5 мужчин. Участники являлись студентами СПбГУ и другими добровольцами. При определении размера выборки мы полагались на количество респондентов, участвующих в подобных исследованиях распознавания эмоциональных стимулов (Барабанщиков и др., 2018).

Критерий отсева испытуемых. Время прохождения менее 40 минут — в этом случае возникли сомнения во вдумчивом прохождении эксперимента (отсеян 1 испытуемый).

Стимульный материал. Использовался набор изображений А. В. Петраковой и коллег (Петракова и др., 2024). Было использовано 248 цветных изображений эмоциональных лиц людей с выражением 14 различных эмоций (см. их подробное описание в: (Барабанщиков, Суворова, 2020)) в трех вариантах интенсивности (малая, средняя и большая): гордость, злость, радость, раздражение, веселье, отвращение, удовольствие, грусть, облегчение, отчаяние, интерес, страх, удивление, тревога. Использовались фотографии только взрослых людей. Разрешение фотографий — 682×1024 px. Использован метод пакетной обработки в программе Adobe Photoshop 24.7.1. Модель отделялась от фона, фон для стандартизации заменялся на однотонный белый, так как оригинальные фотографии содержали на фоне полотно при разном освещении. Вручную правились детали, неправильно обработанные автоматически. Применялась цветокоррекция — выравнивался цвет изображений.

Аппаратура. Стимулы предъявлялись на LCD-дисплее стационарного компьютера. Характеристики монитора: диагональ 24' (61 см); ширина экрана 53 см; высота экрана: 29,5 см; разрешение: 1920×1080 (16:9); частота обновления экрана 60 Гц.

Обучение нейросети. Мы сохраняем фотографии для нейросети в том виде, в котором они предъявлялись испытуемым. Используется нейронная сеть *teachable machine* (Carney et al., 2020), обученная на датасете изображений ImageNet (Deng et al., 2009) с более чем тысячей классами изображений. Модель является сверточной, так как такая архитектура нейронной сети подходит для распознавания паттернов изображения. Для обучения нейросети используется трансферное обучение. Модель, ранее обученная на других изображениях, дообучалась на изображениях датасета, который использовался для эксперимента с людьми (для обучения нейросети использовались другие 717 стандартизированных фотографий с лицами людей, которые не вошли в тестировочную выборку). Используется гетерогенное обучение — нейросеть первоначально обучена на других классах, не являющихся эмоциями. Таким образом минимизируется вероятность негативного переноса, который возник бы при обучении тем же классам в прошлом. Фотографии таких же классов в прошлом обучении могут негативно сказаться на обучении новым данным. Нейросеть обучалась

с параметрами в 40 эпох (1 эпоха означает, что модель ознакомилась с каждым изображением хотя бы раз), размером пакета 16 (набор образцов, использованных в одной итерации при обучении), подходящим для обучения имеющимся количеством изображений, скорости обучения 0,001. Критерием обучения нейросети (функция потерь) была Cross-Entropy Loss, традиционно используемая для нейросетей при классификации. В приложении представлена точность при обучении на каждый класс (см. Приложение, табл.), матрица ошибок при обучении (Приложение, рис. П1), точность на каждую эпоху при обучении (Приложение, рис. П2), потеря на каждую эпоху (это показатель эффективности обучения, измеряемый как разница между прогнозами модели и фактическими данными на каждой эпохе обучения с учетом уверенности модели в ответе) (Приложение, рис. П3).

Процедура. Перед экспериментом в инструкции испытуемому были показаны определения всех 14 эмоций. Эмоции были причислены к эмоциональным семействам и валентности, согласно П. Экману об эмоциональных семействах: к семейству «удовольствие» относятся гордость, радость, веселье, удовольствие, облегчение, интерес, удивление, к семейству агрессии — злость и раздражение, в семействе отвращения только одна одноименная эмоция, к семейству «грусть» относятся грусть и отчаяние, к семейству страх — страх и тревога. К положительной валентности относится только семейство удовольствия, остальные семейства относятся к отрицательной валентности.

Участники проходили исследование индивидуально и очно, стимулы предъявлялись в Google-формах. Ставилась задача категоризации эмоциональных лиц по 14 эмоциям, определения которых были перед испытуемыми во время исследования. Каждая фотография предъявлялась на отдельной странице. Варианты ответов были перемешаны случайным образом для каждой фотографии. Нейросеть категоризовала изображения на те же эмоции. Не было ограничений по времени ответа.

Анализ данных

Анализ данных проводился в программной среде RSTUDIO 2023.09.1 Build 494. Производилось усреднение по стимулам. Для гипотез о сравнении точности категоризации эмоций, классификации эмоционального семейства и валентности использовался однофакторный дисперсионный анализ. Для гипотез сравнения случайности правильного ответа и правильных ответов респондентов и нейросети использовался односторонний критерий Стьюдента. В случае невозможности его применения к эмоциональным семействам использовался G-тест (логарифмический критерий правдоподобия). Ответ засчитывался как верный, если в случае точной категоризации совпадал с предполагаемым ответом; в случае семейств ответ считался правильным, если ответ, данный испытуемым или нейросетью, совпадал с семейством эмоции, действительно изображенной на фото (например, если вместо радости ответил, что это веселье); в случае валентности ответ считался правильным, если ответ, данный испытуемым или нейросетью, совпадал с валентностью эмоции, действительно изображенной на фото (например, вместо страха ответил, что это отвращение).

Для поиска различий результатов точности нейросети и людей при категоризации от случайного было рассчитано процентное соотношение правильности

ответов (табл. 1). Также с помощью одностороннего критерия Стьюдента были проведены сравнения вероятности случайного угадывания и доли правильных ответов на уровнях описания эмоции: точная категоризация и валентность.

Таблица 1. Процент распознавания эмоций людьми и нейросетью, %

Параметр	Случайное угадывание	Нейросеть	Респонденты (среднее)
Эмоция	7	11	22
Валентность эмоции	50	62	73

При сравнении эффективности нейронной сети и респондентов было выявлено нарушение нормальности распределения по критерию Шапиро — Уилка в выборке при усреднении правильности ответов по каждой эмоциональной категории для нейронной сети ($W = 0,8$, $p = 0,004$), по эмоциональным семействам и валентности нарушений нормальности и гомогенности дисперсий не выявлено. Для респондентов нарушений нормальности и гомогенности дисперсий на всех уровнях описания эмоции не выявлено. Так как, по мнению некоторых исследователей, нормальность распределения не является обязательным требованием для использования дисперсионного анализа (Федосеева и др., 2015), для сравнения правильности ответов по каждой эмоциональной категории в зависимости от проходящего эксперимент (нейросеть или респонденты) был проведен однофакторный дисперсионный анализ ANOVA для двух независимых выборок.

Результаты

Эффективность классификации, по сравнению со случайностью. Не обнаружено значимых различий ($t = 1,0239$, $df = 13$, $p = 0,32$) между случайной вероятностью правильного ответа (0,07) и долей правильных ответов нейросети ($M = 0,11$; $SD = 0,13$) при точной категоризации. Обнаружены значимые различия ($t = 5,3568$, $df = 13$, $p < 0,001$) между случайной вероятностью правильного ответа и долей правильных ответов людей ($M = 0,22$; $SD = 0,1$) при точной категоризации. Обнаружены значимые различия ($t = 2,2808$, $df = 13$, $p < 0,05$) между случайной вероятностью правильного ответа (0,5) и долей правильных ответов нейросети для валентности ($M = 0,62$, $SD = 0,2$). Также обнаружены значимые различия ($t = 6,3094$, $df = 13$, $p < 0,001$) между случайной вероятностью правильного ответа и долей правильных ответов людей для валентности ($M = 0,73$, $SD = 0,14$).

Эффективность респондентов статистически значимо отличается от случайной ($G = 471$, $df = 13$, $p < 0,001$). Эффективность нейронной сети статистически значимо отличается от случайной ($G = 1142,5$, $df = 13$, $p < 0,001$). Направление и значение отклонений для каждой эмоции у респондентов см. на рис. 1, у нейронной сети — на рис. 2. Процентное сравнение правильных ответов и случайной вероятности правильного ответа для каждого семейства см. в табл. 2.

Сравнение эффективности нейронной сети и респондентов. Выявлены отличия между долей правильных ответов респондентов и нейронной сети в точной

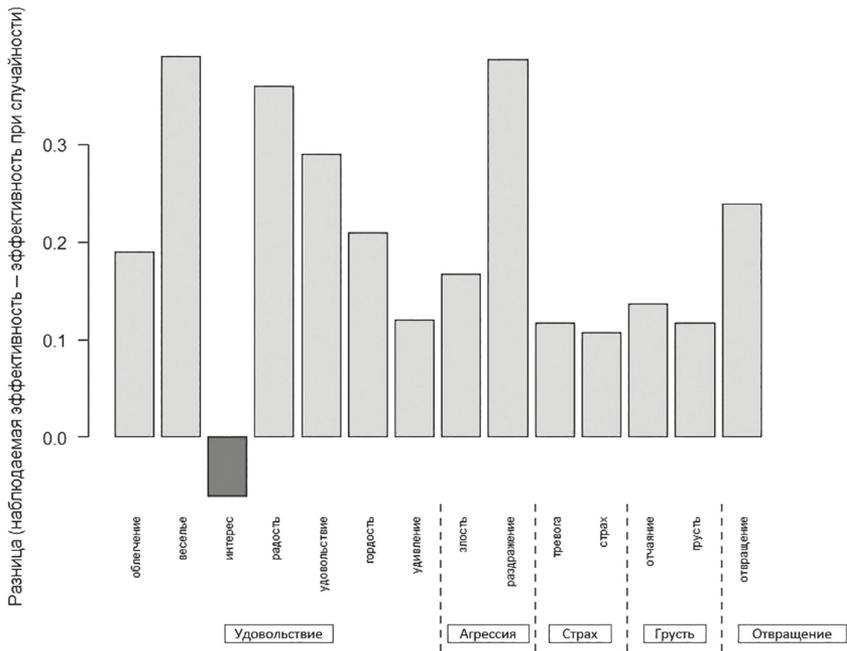


Рис. 1. Разница между наблюдаемой долей правильных ответов и долей случайных правильных ответов у респондентов

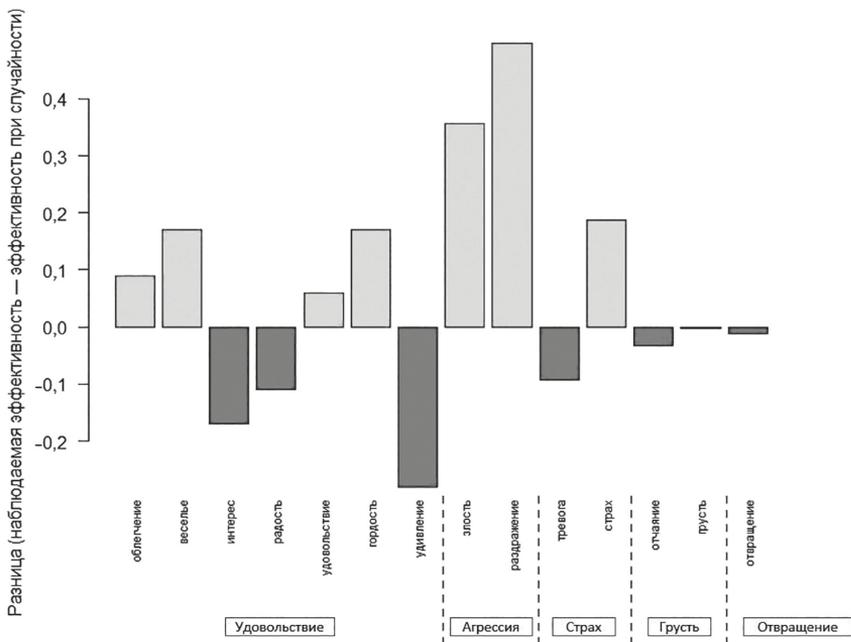


Рис. 2. Разница между наблюдаемой долей правильных ответов и долей случайных правильных ответов у нейронной сети

Таблица 2. Процент правильных ответов в сравнении с эффективностью при случайном угадывании эмоционального семейства для каждой эмоции, %

Эмоциональное семейство и его составляющие	Ожидаемая эффективность при случайном угадывании	Наблюдаемая эффективность нейронной сети	Наблюдаемая эффективность респондентов
<i>Удовольствие</i>	50	49	69
облегчение		59	52
веселье		67	89
интерес		<u>33</u>	<u>44</u>
радость		<u>39</u>	86
удовольствие		56	79
гордость		67	71
удивление		<u>22</u>	62
<i>Агрессия</i>	14,29	57	42
злость		50	31
раздражение		64	53
<i>Отвращение</i>	7,13	<u>6</u>	31
<i>Грусть</i>	14,29	14	26
отчаяние		<u>11</u>	28
грусть		17	23
<i>Страх</i>	14,29	19	26
тревога		5	26
страх		33	25

Примечания: Курсивом в левом столбце выделено эмоциональное семейство; эффективность ниже случайной выделена подчеркиванием.

категоризации ($F(1, 26) = 5,11, p < 0,05, \eta^2 = 0,16$). Респонденты ($M = 0,22; SD = 0,1$) справляются с задачей категоризации эмоций статистически лучше, чем нейронная сеть ($M = 0,11; SD = 0,13$). Сравнение правильности ответов для изображений, правильно определенных в эмоциональное семейство, показало незначимые различия между точностью ответов респондентов и нейронной сетью ($F(1, 26) = 1,95, p = 0,17, \eta^2 = 0,07$): нейронная сеть ($M = 0,38, SD = 0,23$), респонденты ($M = 0,5, SD = 0,24$). Сравнение правильности ответов по категориям валентности эмоций показало незначимые отличия между точностью ответов респондентов и нейронной сетью ($F(1, 26) = 2,98, p = 0,1, \eta^2 = 0,1$): нейронная сеть ($M = 0,62, SD = 0,2$), респонденты ($M = 0,73, SD = 0,14$) (рис. 3). Дополнительная проверка непараметрическим критерием Манна — Уитни показала такие же результаты: значимые отличия эффективности на уровне точной категоризации ($U = 42, p < 0,05$), незначимые отличия на уровне семейства ($U = 73, p = 0,26$), незначимые отличия на уровне валентности ($U = 63, p = 0,11$).

Основные результаты. Респонденты лучше случайного категоризируют изображения по 14 эмоциям, по семействам и по валентности. Нейросеть лучше случайного категоризирует изображения по валентности и по эмоциональным семействам.

Респонденты лучше справляются с заданием точной категоризации, чем нейронная сеть. Сравнение эффективности нейронной сети и респондентов на уровнях

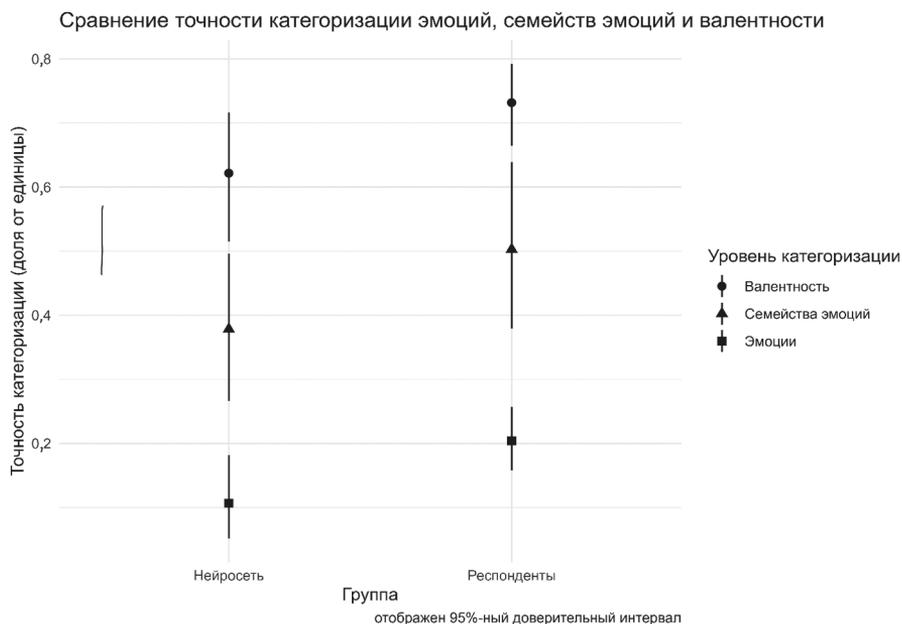


Рис. 3. Доля правильных ответов нейросети и человека на трех уровнях категоризации эмоций (точка обозначает среднюю правильность ответов)

категоризации по эмоциональным семействам и валентности не показало значимых отличий.

Дополнительный анализ сравнения успешности классификации эмоций, в зависимости от интенсивности, на всех уровнях описания эмоций показал незначимые результаты.

Обсуждение результатов и выводы

Эффективность классификации в нашем исследовании даже на уровне валентности была меньше, чем в подобных исследованиях с изображениями, изготовленными по инструкциям FACS (Dailey et al., 2002). Более низкая эффективность классификации и респондентами, и нейронной сетью может объясняться использованием экологического датасета. Датасеты со симулированными по заготовке фотографа эмоциями в исследованиях могут завышать эффективность классификации эмоций респондентами (Королькова, Лободинская, 2022). Это может быть связано с тем, что реальная вариативность проявлений мимики одной эмоции шире, чем считается в версии теории базовых эмоций Экмана.

Эффективная классификация эмоций на уровне валентности говорит о наличии паттернов в выражении лица, разделяющих два полюса шкалы валентности: позитивный и негативный аффект (Barrett, 1998). Барретт отмечает, что нет четких лицевых паттернов даже для разных валентностей. Наш эксперимент показал обратное при условии, что подсказок контекста ситуации не было.

Мы можем наблюдать эффективное отнесение эмоций к категориям семейств нейросетью (за исключением нескольких эмоций, см. рис. 2) и людьми (см. рис. 1). Мы можем увидеть положительные отклонения от случайного при отнесении эмоций к семействам. Интерпретируя положительные отклонения обоих способов категоризации (людей и нейросети), можно сказать о существовании лицевых паттернов для семейств, описанных сторонниками базовых эмоций (Ekman, 1992).

Респонденты справляются также с определением точной категоризации. С одной стороны, наблюдается успешная классификация категорий эмоций более узких, чем эмоциональные семейства, чего не должно наблюдаться в случае теории Экмана (Ekman, 1999a). С другой стороны, эффективность классификации этих эмоций у респондентов ниже, чем в исследованиях классификации базовых эмоций (Dailey et al., 2002; Ekman, 1971). Нейронная сеть не справилась с классификацией точных категорий эмоций, в отличие от людей. Вероятно, особенности распознавания эмоций людьми помогли им опознать эмоции за счет опыта взаимодействия с такими категориями эмоций в жизни, что несопоставимо с тем количеством фотографий, которые получила нейронная сеть для обучения. Особенности людей, которые могут стать причиной ошибок для определения эмоций между разными культурами, способствуют более успешной категоризации внутри культуры (Gendron et al., 2014): все модели из разработанного нами набора изображений принадлежат к культуре России. Учитывая неэффективность нейронной сети и малую эффективность людей в классификации эмоций на уровне точной категоризации, мы можем говорить о малой специфичности выражений лица для небазовых эмоций. Вероятно, узкие категории эмоций могут иметь свои особенности выражения, имея при этом признаки своего эмоционального семейства (Ekman, 1992).

Вопреки нашим ожиданиям, нейронная сеть на уровне точной категоризации эмоций справилась хуже респондентов и не лучше на уровнях эмоционального семейства и валентности. Слова для эмоций, которые, как мы предполагали, будут лишь мешать определению реальной эмоции у людей (Oatley, Johnson-Laird, 2014), оказались в действительности помощниками для нахождения более специфичных выражений, чем эмоциональные семейства. Многолетнее знакомство с этими словами, принятыми в данной культуре, способствует успешности классификации для людей, но не для нейронной сети, имеющей, по сравнению с людьми, небольшой опыт знакомства с эмоциями при обучении. Даже на уровне эмоциональных семейств нейронная сеть справлялась хуже людей при категоризации, хотя в предыдущих исследованиях базовых эмоций она успешно категоризировала эмоциональные лица с изображений, разработанных по правилам FACS (Dailey et al., 2002; Badrullhisham, Mangshor, 2021). Это может быть следствием того, что внутри одной эмоции большой диапазон для вариации, что соответствует теории New BET (Scarantino, 2015). Нейронная сеть, имеющая стандарт эмоции, может не справляться со всей вариативностью лицевых паттернов, действительно ей присущей (Kretschmar et al., 2003). Таким образом, изменчивость внутри одной категории гораздо больше, чем это предполагает теория Экмана. Успешность классификации валентности свидетельствует о более четких разграничительных сигналах для базовых негативного и позитивного аффектов, чем для базовых эмоций с их вариациями. Результаты исследования согласуются с теориями о социальном научении эмоциям и теории New BET. Учитывая низкую эффективность определения эмоций

при точной категоризации, существование схожих паттернов мимики для всех слов-эмоций остается под вопросом, что ослабляет позиции теории социального научения эмоциям. Еще П. Экман отмечал возможное сходство паттернов для субэмоций, входящих в семейства эмоций (Ekman, 1992). New BET не отрицает большого диапазона вариаций внутри семейства (Scarantino, 2015), поэтому интерпретация в русле данного подхода кажется нам самой подходящей для наших результатов. Однако альтернативную версию социального научения эмоциям также нельзя игнорировать.

Незначимые результаты при изучении влияния фактора интенсивности на успешность классификации эмоций наблюдались на всех трех уровнях описания эмоций — точной категоризации, эмоционального семейства и валентности. Такое могло возникнуть вследствие маленького эффекта на всех трех уровнях описания эмоций (ниже 0,1).

Ограничения и перспективы исследования. В связи с тем, что нейронная сеть могла получить плохие результаты вследствие недостаточной обученности или влияния нерелевантных параметров изображений на категоризацию, требуются дальнейшие исследования, учитывающие текущие иррелевантные факторы обучения и тестировочной серии. Teachable machine имеет ограничения для научных исследований — инструмент не позволяет построить сложную модель со специфическими настройками. Для последующих исследований требуется использование нейронной сети, позволяющей поддерживать более сложную архитектуру для обучения. Для корректного проведения статистического анализа сравнения со случайностью тестом Стьюдента на уровне эмоциональных семейств требуется изменить дизайн эксперимента, подобрав семейства так, чтобы соблюдался принцип равной вероятности случайного угадывания семейства. Так как наше исследование проводилось только на российской выборке, мы не можем распространять выводы на кросс-культурные общности. Для уточнения и расширения теоретических результатов возможны кросс-культурные исследования.

Выводы

Нейронная сеть и респонденты успешно справляются с категоризацией эмоций на уровне семейств и валентности, что говорит о существовании специфических мимических паттернов для этих категорий. Респонденты также выше случайного распознают эмоции на уровне конкретных их проявлений. На уровне категоризации конкретных эмоций респонденты справляются лучше нейронной сети. Не обнаружено статистически значимых различий в эффективности на уровнях эмоциональных семейств и валентности. Теория New BET объясняет низкую эффективность людей и нейронной сети на уровне конкретных эмоций в совокупности с высокой эффективностью на уровнях эмоционального семейства и валентности. Успешность человека в сравнении с нейронной сетью говорит о наличии возможных паттернов мимики для культурно специфических эмоций. Однако, сравнение с похожими исследованиями категоризации базовых эмоций говорит о малой специфичности выражения лица для небазовых эмоций.

Литература

- Барбанищikov В. А., Королькова О. А., Лободинская Е. А. Распознавание эмоций в условиях ступенчатой стробоскопической экспозиции выражений лица // Экспериментальная психология. 2018. Т. 11, № 4. С. 50–69. <https://doi.org/10.17759/exppsy.2018110405>
- Барбанищikov В. А., Суворова Е. В. Оценка эмоционального состояния человека по его видеоизображению // Экспериментальная психология. 2020. Т. 13, № 4. С. 4–24. <https://doi.org/10.17759/exppsy.2020130401>
- Королькова О. А., Лободинская Е. А. База видеоизображений естественных эмоциональных экспрессий: восприятие эмоций и автоматизированный анализ мимики лица // Оптический журнал. 2022. Т. 89, № 8. С. 97–103. <https://doi.org/10.17759/exppsy.2021140401>
- Петракова А. В., Лебедева Е. И., Кузьмина Ю. В., Юрчик Е. Н. Опыт создания российской базы лиц, изображающих различные эмоции: первый этап // Психология. Журнал Высшей школы экономики. 2024. Т. 21, № 2. С. 423–431. <https://doi.org/10.17323/1813-8918-2024-2-423-431>
- Рюмина Е. В., Карпов А. А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20, № 2. С. 163–176. <https://doi.org/10.17586/2226-1494-2020-20-2-163-176>
- Федосеева Е. В., Терехова В. А., Цесаренко О. В., Гладкова М. М. Обработка результатов токсикологических исследований в статистической программе R // Принципы экологии. 2015. № 3 (15). С. 12–26. <https://doi.org/10.15393/j1.art.2015.4381>
- Badrulhisham N. A. S., Mangshor N. N. A. Emotion recognition using convolutional neural network (CNN) // Journal of Physics: Conference Series. IOP Publishing. 2021. Vol. 1962, no. 1. P. 012040. <https://doi.org/10.1088/1742-6596/1962/1/012040>
- Bandura A., Walters R. H. Social learning theory. Englewood Cliffs: Prentice hall, 1977. Vol. 1. P. 141–154.
- Barrett L. F. Constructing emotion // Psihologijske teme. 2011. Vol. 20, no. 3. P. 359–380.
- Barrett L. F. Discrete emotions or dimensions? The role of valence focus and arousal focus // Cognition & Emotion. 1998. Vol. 12, no. 4. P. 579–599. <https://doi.org/10.1080/026999398379574>
- Barrett L. F., Adolphs R., Marsella S., Martinez A. M., Pollak S. D. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements // Psychological science in the public interest. 2019. Vol. 20, no. 1. P. 1–68. <https://doi.org/10.1177/1529100619832930>
- Barrett L. F., Simmons W. K. Interoceptive predictions in the brain // Nature reviews neuroscience. 2015. Vol. 16, no. 7. P. 419–429. <https://doi.org/10.1038/nrn3950>
- Carney M., Webster B., Alvarado I., Phillips K., Howell N., Griffith J., Jongejan J., Pitaru A., Chen A. Teachable machine: Approachable Web-based tool for exploring machine learning classification // Extended abstracts of the 2020 CHI conference on human factors in computing systems. 2020. P. 1–8. <https://doi.org/10.1145/3334480.3382839>
- Dailey M. N., Cottrell G. W., Padgett C., Adolphs R. EMPATH: A neural network that categorizes facial expressions // Journal of cognitive neuroscience. 2002. Vol. 14, no. 8. P. 1158–1173. <https://doi.org/10.1162/089892902760807177>
- Deng J., Dong W., Socher R., Li L. J., Li K., Fei-Fei L. Imagenet: A large-scale hierarchical image database // 2009 IEEE conference on computer vision and pattern recognition. 2009. P. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Ekman P. An argument for basic emotions // Cognition & Emotion. 1992. Vol. 6, no. 3–4. P. 169–200. <https://doi.org/10.1080/02699939208411068>
- Ekman P. Basic emotions // Handbook of cognition and emotion. Chichester: Wiley, 1999a. P. 16.
- Ekman P. Facial expressions // Handbook of cognition and emotion. Chichester: Wiley, 1999b. P. 320.
- Ekman P. Universal facial expressions in emotion // Studia Psychologica. 1973. Vol. 15, no. 2. P. 140–147.
- Ekman P. Universals and cultural differences in facial expressions of emotion // Nebraska symposium on motivation. Lincoln: University of Nebraska Press, 1971. P. 207–283.
- Ekman P., Friesen W. V. Constants across cultures in the face and emotion // Journal of personality and social psychology. 1971. Vol. 17, no. 2. P. 124. <https://doi.org/10.1037/h0030377>
- Gendron M., Hoemann K., Crittenden A. N., Mangola S. M., Ruark G. A., Barrett L. F. Emotion perception in Hadza hunter-gatherers // Scientific reports. 2020. Vol. 10, no. 1. P. 3867. <https://doi.org/10.1038/s41598-020-60257-2>

- Gendron M., Roberson D., van der Vyver J. M., Barrett L. F. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture // *Emotion*. 2014. Vol. 14, no. 2. P. 251. <https://doi.org/10.1037/a0036052>
- Grühn D., Sharifian N. Lists of emotional stimuli // *Emotion measurement*. Sawston: Woodhead Publishing, 2016. P. 145–164. <https://doi.org/10.1016/B978-0-08-100508-8.00007-2>
- Kretschmar R., Karayiannis N. B., Eggimann F. Handling class overlap with variance-controlled neural networks // *Proceedings of the International Joint Conference on Neural Networks*, 2003. New York: IEEE, 2003. Vol. 1. P. 517–522. <https://doi.org/10.1109/IJCNN.2003.1223400>
- Leon E., Clarke G., Callaghan V., Sepulveda F. Real-time detection of emotional changes for inhabited environments // *Computers & Graphics*. 2004. Vol. 28, no. 5. P. 635–642. <https://doi.org/10.1016/j.cag.2004.06.002>
- Oatley K., Johnson-Laird P. N. The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction // *Striving and feeling*. Hove: Psychology Press, 2014. P. 363–393.
- Russell J. A. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies // *Psychological Bulletin*. 1994. Vol. 115, no. 1. P. 102. <https://doi.org/10.1037/0033-2909.115.1.102>
- Sato W., Hyniewska S., Minemoto K., Yoshikawa S. Facial expressions of basic emotions in Japanese lay-people // *Frontiers in psychology*. 2019. Vol. 10. P. 420264. <https://doi.org/10.3389/fpsyg.2019.00259>
- Scarantino A. Basic emotions, psychological construction, and the problem of variability // *The psychological construction of emotion*. New York: The Guilford Press, 2015. P. 334–376. <https://doi.org/10.1007/s13164-020-00492-8>
- Scherer K. R., Wallbott H. G. Evidence for universality and cultural variation of differential emotion response patterning // *Journal of personality and social psychology*. 1994. Vol. 66, no. 2. P. 310. <https://doi.org/10.1037/0022-3514.66.2.310>
- Trommsdorff G. Development of emotions as organized by culture // *ISSBD Newsletter*. 2006. Vol. 49, no. 1. P. 1–4.

Статья поступила в редакцию 23 сентября 2024 г.;
рекомендована к печати 18 ноября 2024 г.

Контактная информация:

Сладкоштиева Анастасия Владимировна — магистр; <https://orcid.org/0009-0009-1162-9932>,
st076509@student.spbu.ru

Стародубцев Алексей Сергеевич — канд. психол. наук; <https://orcid.org/0000-0001-9322-6911>,
fleksbr@yandex.ru

Петракова Анастасия Владимировна — канд. психол. наук; <https://orcid.org/0000-0001-9708-5693>,
anastasiya-petrakova@yandex.ru

Classification of facial expressions of emotional faces by people and a neural network based on an ecologically valid image database

A. V. Sladkoshtieva^{1a}, A. S. Starodubtsev¹, A. V. Petrakova²

¹ St. Petersburg State University,
7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

² HSE University,
20, ul. Myasnitskaya, Moscow, 101000, Russian Federation

For citation: Sladkoshtieva A. V., Starodubtsev A. S., Petrakova A. V. Classification of facial expressions of emotional faces by people and a neural network based on an ecologically valid image database. *Vestnik of Saint Petersburg University. Psychology*, 2025, vol. 15, issue 1, pp. 116–133. <https://doi.org/10.21638/spbu16.2025.107> (In Russian)

^a Author for correspondence.

This paper investigates the problem of universal facial expression patterns for emotion valences, emotion families, and culturally specific emotions. Many studies have been conducted to address this problem, but they sometimes infer the universality of emotion expression from respondents' recognition of emotions and use traditional image databases that do not meet the conditions of ecological validity. Four groups of theories about the origin of emotions are compared: the theory of universality of basic emotions, the theory of cultural specificity of basic emotions, the theory of social learning of emotions, and the theory of social construction of emotions. Purpose of the study: to determine whether people's facial expressions have specific patterns for positive and negative valences of emotion (degree of positive or negative affective response), emotion families (includes basic emotion and variations of basic emotion), and specific emotions using an image database whose authors strived for ecological validity: subjects were not given a standard of facial expression patterns to demonstrate emotions. A teachable machine neural network was used to examine the expression process in a respondent-independent manner. Heterogeneous transfer learning was used. In our work, respondents and the neural network categorized images of emotional faces into 14 emotions. Respondents were more likely than the neural network to correctly classify specific emotions. Both the neural network and humans were more efficient at recognizing valence and emotion families than recognizing specific emotions. This may suggest specific patterns of emotion families and emotion valence. Weak specificity at the level of specific emotions is possible. The results are consistent with Scarantino's "New BET" theory of basic emotions. An interpretation in line with the social learning theory of emotion is possible. The limitations of the neural network in recognizing the full range of variations in prototypical emotion expressions are discussed. Further possible cross-cultural studies to refine the results are described.

Keywords: origin of emotion, basic emotions, constructivism, emotion classification, emotional expression.

References

- Akers, R. L., Jennings, W. G. (2015). Social learning theory. *The handbook of criminological theory*, 230–240.
- Badrulhisham, N. A. S., Mangshor, N. N. A. (2021). Emotion recognition using convolutional neural network (CNN). *Journal of Physics: Conference Series, IOP Publishing*, 1962, 1, 012040.
- Barabanshnikov, V. A., Korol'kova, O. A., Lobodinskaja, E. A. (2018). Emotion recognition under staggered stroboscopic exposure of facial expressions. *Eksperimental'naja psikhologija*, 11 (4), 50–69. (In Russian)
- Barabanshnikov, V. A., Suvorova, E. V. (2020). Estimation of a person's emotional state from his video image. *Eksperimental'naja psikhologija*, 13 (4), 4–24. (In Russian)
- Barrett, L. F. (2011). Constructing emotion. *Psihologijske teme*, 20 (3), 359–380.
- Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, 12 (4), 579–599.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20 (1), 1–68.
- Barrett, L. F., Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16 (7), 419–429.
- Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., Chen, A. (2020). Teachable machine: Approachable Web-based tool for exploring machine learning classification. In: *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–8).
- Dailey, M. N., Cottrell, G. W., Padgett, C., Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of cognitive neuroscience*, 14 (8), 1158–1173.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).

- Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In: *Nebraska symposium on motivation* (pp. 207–283). Lincoln, University of Nebraska Press.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6 (3–4), 169–200.
- Ekman, P. (1999a). Basic emotions. In: *Handbook of cognition and emotion* (p. 16). Chichester: Wiley.
- Ekman, P. (1999b). Facial expressions. In: *Handbook of cognition and emotion* (p. 320). Chichester: Wiley.
- Ekman, P., Keltner, D. (1970). Universal facial expressions of emotion. *California mental health research digest*, 8 (4), 151–158.
- Ekman, P., Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17 (2), 124.
- Fedoseeva, E. V., Terekhova, V. A., Tsesarenko, O. V., Gladkova, M. M. (2015). Processing the results of toxicological studies in a statistical program R. *Printsipy ekologii*, 3 (15), 12–26. (In Russian)
- Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., Barrett, L. F. (2020). Emotion perception in Hadza hunter-gatherers. *Scientific reports*, 10 (1), 3867.
- Gendron, M., Roberson, D., van der Vyver, J. M., Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, 14 (2), 251.
- Grühn, D., Sharifian, N. (2016). Lists of emotional stimuli. In: *Emotion measurement* (pp. 145–164). Sawston, Woodhead Publishing.
- Korol'kova, O. A., Lobodinskaia, E. A. (2022). A video database of natural emotional expressions: Emotion perception and automated analysis of facial expressions. *Opticheskii zhurnal*, 89 (8), 97–103. (In Russian)
- Kretschmar, R., Karayiannis, N. B., Eggmann, F. (2003). Handling class overlap with variance-controlled neural networks. In: *Proceedings of the International Joint Conference on Neural Networks* (vol. 1, pp. 517–522). New York: IEEE.
- Leon, E., Clarke, G., Callaghan, V., Sepulveda, F. (2004). Real-time detection of emotional changes for inhabited environments. *Computers & Graphics*, 28 (5), 635–642.
- Oatley, K., Johnson-Laird, P. N. (2014). The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction. In: *Striving and feeling* (pp. 363–393). Hove, Psychology Press.
- Petrakova, A. V., Lebedeva, E. I., Kuz'mina, Iu. V., Iurchik, E. N. (2024). Experience of creating a Russian database of faces depicting different emotions: The first stage. *Psikhologiya. Zhurnal Vysshei shkoly ekonomiki*, 21 (2), 423–431. (In Russian)
- Rjumina, E. V., Karpov, A. A. (2020). Analytical review of methods for recognizing emotions from human facial expressions. *Nauchno-tehnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 20 (2), 163–176. (In Russian)
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115 (1), 102.
- Sato, W., Hyniewska, S., Minemoto, K., Yoshikawa, S. (2019). Facial expressions of basic emotions in Japanese laypeople. *Frontiers in Psychology*, 10, 259.
- Scarantino, A. (2015). Basic emotions, psychological construction, and the problem of variability. In: *The psychological construction of emotion* (pp. 334–376). New York: The Guilford Press.
- Scherer, K. R., Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66 (2), 310.
- Trommsdorff, G. (2006). Development of emotions as organized by culture. *ISSBD Newsletter*, 49, 1, 1–4.

Received: September 23, 2024

Accepted: November 18, 2024

Authors' information:

Anastasija V. Sladkoshtieva — Master; <https://orcid.org/0009-0009-1162-9932>,
st076509@student.spbu.ru

Aleksej S. Starodubtsev — PhD in Psychology; <https://orcid.org/0000-0001-9322-6911>,
fleksbr@yandex.ru

Anastasija V. Petrakova — PhD in Psychology; <https://orcid.org/0000-0001-9708-5693>,
anastasiya-petrakova@yandex.ru

Точность при обучении на каждый класс

Класс	Точность	Образцы
Гордость	0,63	8
Злость	0,56	9
Радость	0,38	8
Раздражение	0,25	8
Веселье	0,50	8
Отвращение	0,38	8
Удовольствие	0,38	8
Грусть	0,44	9
Облегчение	0,25	8
Отчаяние	0,38	8
Интерес	0,50	8
Страх	0,25	8
Удивление	0,38	8
Тревога	0,38	8

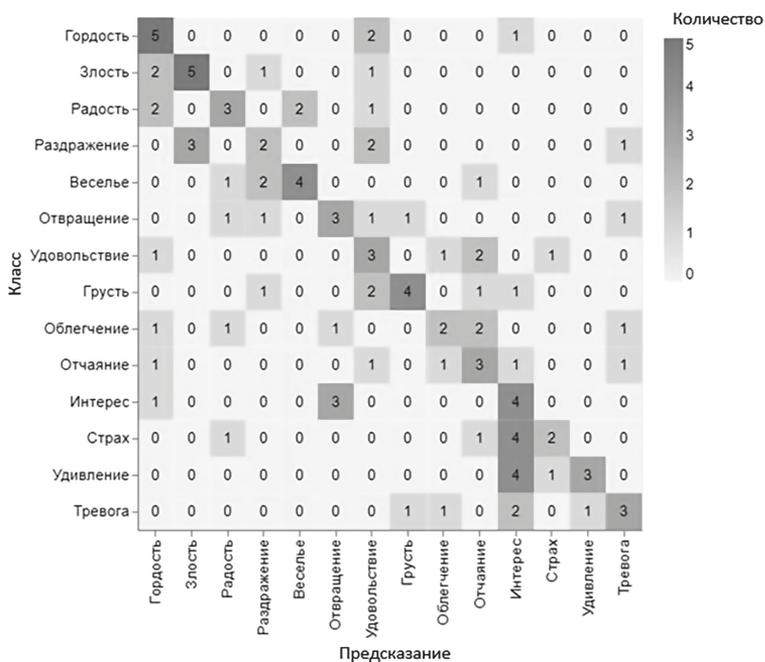


Рис. III. Матрица ошибок

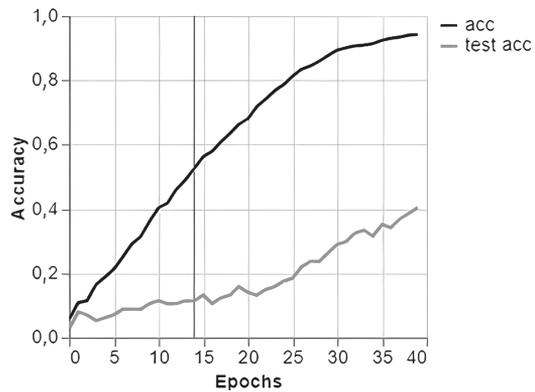


Рис. 12. Точность на каждую эпоху при обучении (тестовая точность обозначена более светлым цветом)

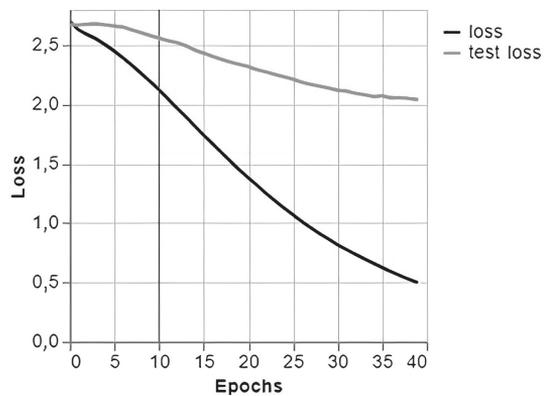


Рис. 13. Потеря на каждую эпоху (тестовая потеря обозначена более светлым цветом)