ОРИГИНАЛ-МАКЕТ СТАТЬИ

(для статьи на русском языке)

**УДК** 621.389:57.087.1

Критерии практической применимости подходов ML для классификации эмоций по ЭЭГ в датасете DEAP

Екатерина Феликсовна Савкина, Александр Евгеньевич Соколов, Федор Андреевич Кравченко, Евгений Витальевич Шуранов

**Аннотация**

**Предмет исследования**: В статье рассмотрены подходы машинного обучения для анализа электроэнцефалограммы головного мозга (ЭЭГ) для определения эмоционального состояния человека. Целью работы является анализ и проверка SoTA-подходов распознавания эмоций по ЭЭГ.

**Методы**: В работе использованы методы глубокого машинного обучения для анализа ЭЭГ.

**Результаты**: Произведен выбор моделей с наивысшими показателями качества распознавания эмоций по ЭЭГ по состоянию на 2021 г. Воспроизведены результаты их работы. Произведен анализ примененных подходов. Разработаны минимальные критерии для оценки практической применимости результатов работ по распознаванию эмоций по ЭЭГ. Высокие показатели качества распознавания эмоций, приведенные в научных статьях, не гарантируют возможность практического применения описываемого подхода. Для оценки необходима дополнительная проверка подхода по ряду критериев.

**Практическая значимость:** Работа рассчитана на специалистов, начинающих исследование в области распознавания эмоций на основе данных ЭЭГ. Приведены критерии оценки практической применимости статей. Описаны подходы по улучшению метрик качества в датасете DEAP и даны рекомендации по улучшению разметки датасета.

**Ключевые слова**

ЭЭГ, анализ, эмоциональное состояние, глубокое обучение

**Благодарности**

Основные результаты статьи были сформированы в ходе исследований в выпускной работе студентов академии больших данных MADE Екатерины Савкиной и Александра Соколова.

Criteria for practical applicability of ML approaches for emotions classification by EEG in dataset DEAP

Ekaterina Feliksovna Savkina, Alexander Evgenievich Sokolov, Fedor Andreevich Kravchenko, Evgeniy Vitalievich Shuranov

**Abstract**

Subject of research: The article considers machine learning approaches for analyzing the brain electroencephalogram (EEG) to determine the emotional state of a person. The aim of this work is to analyze and test SoTA approaches to emotion recognition by EEG.

Methods: We used deep machine learning methods for EEG analysis.

Results: Models with the highest EEG emotion recognition quality scores as of 2021 were selected. The results of their work were reproduced. The analysis of the applied campaigns is made. The minimum criteria for assessing the practical applicability of the results of work on the recognition of emotions by EEG are developed, recommendations for conducting such studies and recommendations for preparing a dataset are given. The high quality indicators of emotion recognition given in scientific articles do not guarantee the possibility of practical application of the described approach. The evaluation requires additional validation of the approach against a number of criteria.

The practical significance. The work is designed for specialists starting research in the field of emotion recognition based on EEG data. The criteria for assessing the practical applicability of the articles are given. Approaches to improving quality metrics in the DEEP dataset are described and recommendations for improving the layout of the dataset are given.

**Keywords**

EEG, deap learning. DEAP dataset, emotion recognition

**Aknowledgements**

The main results of the article were formed during the research of the graduate work of students of the MADE Academy of Big Data Ekaterina Savkina and Alexander Sokolov.

1. **Введение**

Распознавание эмоций — это определение психических процессов, отражающих субъективное оценочное отношение к существующим или возможным ситуациям и объективному миру. Такое определение процессу распознавания эмоций можно дать, исходя из одного, но не единственного, определения эмоции.

Эмоции проявляются через выражение человеческого лица, движения тела и жесты, частоту дыхания, пульс, реакции кожи, но одним из наиболее интересных источников данных об эмоциях может служить сигнал электроэнцефалограммы (ЭЭГ) головного мозга, поскольку он непосредственно связан с процессами, происходящими в мозге человека. Кроме того, данные, получаемые из ЭЭГ, не поддаются сознательному контролю в отличие от жестов, выражения лица и позы. Если речь идет о некоторых категориях людей с физическими ограничениями, ЭЭГ – один из немногих источников информации об их эмоциональном состоянии.

За последние годы было проведено значительное количество исследований по распознаванию эмоций по сигналам мозга. С развитием технологий искусственного интеллекта (AI) распознавание эмоций стало неотъемлемой частью исследований в области нейробиологии, информатики, когнитивных наук и медицины. Этот тренд хорошо иллюстрирует график количества англоязычных публикаций (рис.1) с тегами «EEG» и «Emotion recognition» [1]

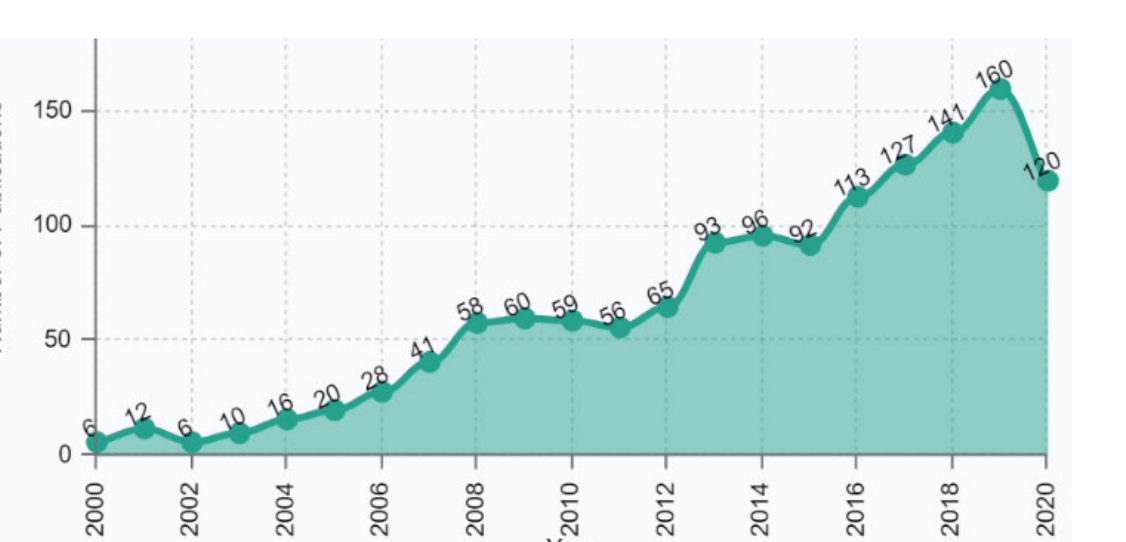


Рисунок 1: Количество публикаций по распознаванию эмоций по ЭЭГ

Рост рынка устройств brain computer interface (BIC) так же подтверждает эту тенденцию. По состоянию на 2020 год, его объем составляет 1,5 млрд долларов, и эта цифра продолжает ежегодно расти на 15% [2].

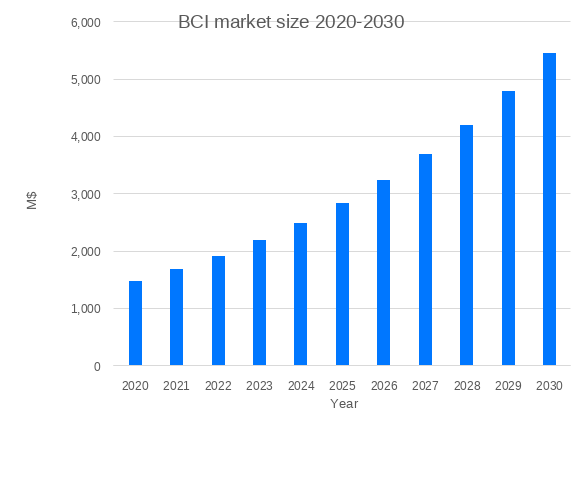


Рисунок 2: Объем рынка устройств BCI

Цена пользовательских устройств, позволяющих обрабатывать сигналы ЭЭГ на текущий момент начинается от 300$ и при массовом производстве будет снижаться, повышая их доступность.

В результатах исследований, посвященных анализу эмоционального состояния человека, в первую очередь могут быть заинтересованы производители оборудования и маркетологи, поскольку такая информация может быть использована в таких сферах, как:

* образование;
* развлечения и игры;
* здравоохранение;
* фитнес;
* ментальные практики.

1. **Данные**

В таблице 1 приведены наиболее цитируемые датасеты для распознавания эмоций по ЭЭГ, по состоянию на 2021 год, находящиеся в открытом доступе.

Как видно из таблицы 1, датасет DEAP содержит в себе наибольшее количество данных для анализа. Кроме того, он является одним из наиболее ранних датасетов в области анализа ЭЭГ, что объясняет его популярность. Датасет DEAP используют в 60% статей на тему определения эмоций по ЭЭГ. По этой причине для проведения исследований в рамках данной работы был выбран датасет DEAP.

DEAP — это датасет, созданный Коелстра и др [4]. Cодержит 32 канала ЭЭГ, 8 каналов периферических физиологических сигналов от 32 здоровых испытуемых. Сигналы были записаны при просмотре 40 одноминутных музыкальных клипов. После просмотра клипа испытуемые по шкале от 1 до 9 оценивали 4 каждый из клипов по 4 критериям: aurosal (интенсивность), valence (окраска), dominance (чувство контроля) и liking (понравилось/не понравилось). Сигналы ЭЭГ были дискретизированы с частотой 512 Гц и затем подверглись предобработке — понижению частоты дискретизации до 128Гц, фильтрации от 5 до 45Гц и удалению артефактов, связанных с морганием и движением глаз. Длительность записи ЭЭГ каждого эксперимента составляет 63 секунды, из которых первые 3 секунды отводятся для фиксации начального состояния испытуемого и 60 секунд – для непосредственного просмотра музыкального клипа.

Таблица 1. Характеристики датасетов с ЭЭГ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № п.п | Name | Year | Number of participants | Number of experiments per participant | Number of experiments total | Количество каналов ЭЭГ |
|  | DEAP[4] | 2012 | 32 | 40 Video clips | 1280 | 32 |
|  | SEED[5] | 2017 | 15 | 15 Video clips x 3 sessions | 675 | 15 |
|  | DREAMER[6] | 2015 | 23 | 18 Video clips | 414 | 14 |
|  | MAHOB-HCI[7] | 2011 | 27 | 20 Video clips and Pictures | 540 | 32 |
|  | AMIGOS[8] | 2017 | 40 | 20 Video clips | 800 | 14 |
|  | INTERFACES | 2019 | 43 | 15 Video clips | 643 | 32 |

1. **Обзор литературы и систематизация моделей**

При обзоре литературы и систематизации моделей в статьях обращали внимание на следующие моменты:

* используемый датасет;
* строится ли общая модель на всех испытуемых или отдельная для каждого;
* какая схема разделения на обучение и валидацию используется;
* как получается разметка для эмоционального состояния;
* какие признаки, полученные из ЭЭГ, идут на вход модели;
* что служит единицей обучения в модели;
* какие метрики используются для оценки модели.

По состоянию на 2021 год, при распознавании эмоций по ЭЭГ на датасете DEAP достигнуты довольно высокие показатели качества. Так Moon et al.[10] в своей статье достиг показателя 99.7% accuracy. В то же время в статьях 2019 года Pandey and Seeja [11] и Chao et al.[12] значение Accuracy составило 68.28% и 62.5% соответственно.

В таблице 2 приведен обзор моделей и достигнутых результатов на датасете DEAP по состоянию на 2021 г. по данным [1]

После анализа обзора, решено воспроизвести результаты моделей с наивысшими значениями метрики accuracy, были выбраны следующие статьи:

* модель **№**1:  
  «EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network» Heng Cui, Aiping Liu, Xu Zhang Xiang Chen Kongqiao Wang, Xun Chen School of Information Science and Technology, University of Science and Technology of China, Huami AI Research, Huami Corporation, China, 2020. 97.11% Accuracy [15]
* модель **№**2:  
  Convolution neural network approach for EEG-based emotion recognition using brian connectivity and its spatial information. Seong-Eun Moon Soobeom Jang Jong-Seok Lee School of Integrated Technology, Yonsei University Republic of Korea., 2018 CNN 99.7% Accuracy [10]
* модель **№3**:  
  Multimodal Physiological Signal Emotion Recognition Based on Convolutional Recurrent Neural Network Jinxiang Liaoa, Qinghua Zhong\*, Yongsheng Zhub and Dongli Caic School of Physics & Telecommunication Engineering, South China Normal University 2020 89.68% Accuracy[27]

Таблица 2. Метрики качества на датасете DEAP

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| №  п.п. | Автор | Разбиение | Год | Классы эмоций | Модель | Accuracy  % |
|  | Cimtay and Ekmekcioglu, [13] | + | 2020 | 2 | CNN | 72.81 |
|  | Wang et al. [14] | ? | 2020 | 3(P, N, Neu) | CNN | 82.84 |
|  | Cui et al., [15] | ? | 2020 | 2 | RA-CNN | Val: 96.65  Aro: 97.11 |
|  | Hassan et al., [16] | + | 2019 | 5 | FC-SVM | 89.53 |
|  | Islam and Ahmad, [17] | ? | 2019 | 2 | CNN | Val: 81.51  Aro: 79.42 |
|  | Pandey and Seeja, [11] | + | 2019 | 4 | DNN | Val:62.50  Aro: 61.25 |
|  | Chen et al. [18] | ? | 2019 | 2 | CV-CNN' | 85.57 |
|  | Yang et al., [19] | + | 2019 | 2d | MC-CNN | Val: 81.4  Aro: 80.5 |
|  | Chao et al., [12] | ? | 2019 | 2 | CDL | Val: 68.28,  Aro: 66.73  Dom: 67.25 |
|  | Moon et al., [10] | ? | 2018 | 2 | CNN | 99.72 |
|  | Yang et al., [24] | ? | 2018 | 2 | PCRNN | Val: 90.80,  Aro: 91.03 |
|  | Alhagry et al., [31] | + | 2017 | 3 | DNN | Val: 85.45  Aro: 85.65,  Lik: 87.99 |
|  | Liu et al., [32] | ? | 2016 | 4 | BDAE | Val: 85.2,  Aro: 80.5,  Dom: 84.9,  Lik: 82.4 |
|  | Jinxiang Liao et al[27] | ? | 2020 | 2 | CNN | Aro:89.68  Val:89.19 |

+ независимое разбиение на обучающий и тестовый датасет,

? - зависимое разбиение, или способ разбиения не обозначен однозначно

(способы разбиения будут рассмотрены позже, в пункте 5, рис. 6, 7)

1. **Воспроизведение результатов**

**4.1 Постановка задачи, выбор метрик и бейзлайн**

Сформулируем задачу, которая решается выбранными для воспроизведения моделей статьями. Для распознавания эмоций в этой задаче используются сигналы ЭЭГ. Выбранные эмоции (Aurosal, Valence), представлены значениями от 1 до 9. Каждая из эмоций делится на два класса с пороговым значением равным 5. Таким образом, задача распознавания эмоций сведена к задаче multi-label бинарной классификации, а сами эмоции представляются двумя классами: LA/HA (low aurosal/high aurosal), LV/HV (low valence/ high valence).

Для оценки качества нами выбрана не только метрика accuracy, как в статьях, но и f1-macro, которая лучше отражает качество классификатора при дисбалансе классов. Все расчеты произведены на основе кросс-валидации на пяти фолдах. Значения метрик рассчитываются индивидуально для каждого испытуемого (там, где это возможно), а затем усредняются по всем испытуемым.

Для оценки качества распознавания эмоций метрики сравниваются с опорными значениями. По верхней границе – с подходами, имеющими наивысшие показатели на дату написания статьи, а по нижней границе – с тремя безлайнами [4]:

* случайным классификатором, выбирающим значение класса с вероятностью 0.5
* классификатором, выбирающим всегда наиболее популярный класс (положительный)
* классификатором, выбирающим класс на основе пропорции популярности класса.

Оценка метрик (табл. 3) для данных классификаторов производится на тестовой выборке, имеющей дисбаланс классов, аналогичный датасату DEAP.

Таблица 3. Метрики качества для бейзлайн классификаторов [4].

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п.п. | Тип выборки | Aurosal | | Valence | |
| Acc | F1 | Acc | F1 |
| 1 | Случайная | 0.500 | 0.483 | 0.500 | 0.494 |
| 2 | Наиболее популярный класс | 0.644 | 0.483 | 0.586 | 0.368 |
| 3 | Пропорционально популярности | 0.562 | 0.500 | 0.525 | 0.500 |

* 1. **модель №1**

Модель, названная авторами RA-CNN, разработана Heng Cui, Aiping Liu , Xu Zhang, Xiang Chen, Kongqiao Wang, Xun Chen в 2020 году. На вход модели поступают семплы исходного сигнала ЭЭГ длительностью 1 сек (32 канала по 128 отсчетов). Далее значение выборки подвергается пространственному преобразованию. Выборка размерностью 32х128 преобразуется в массив 9х9х128, где матрица 9 x 9 отражает пространственное расположение электродов на коже головы. Недостающие значения в трехмерном массиве заполняются нулями (см. рис 3)

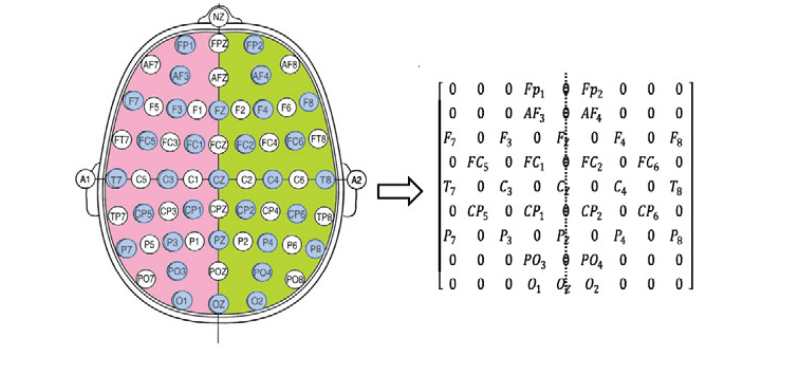


Рисунок 3 Пространственное кодирование

Массив размерностью 9х9х128 подается на вход сверточной нейронной сети, состоящей из четырех блоков: трех сверточных и одного классификатора (рис. 4).

Особенность подхода авторов статьи – применение для каждого из участников эксперимента отдельной модели.

Данные формируются из односекундных семплов из ЭЭГ испытуемого. Таким образом, общий размер набора данных составляет 60 \* 40 = 2400 семплов для одного испытуемого. Обучающий и тестовый наборы данных получаются применением кросс-валидации для данного объединенного набора семплов.

При воспроизведении модели RA-CNN, мы получили метрики (табл. 5), аналогичные метрикам авторов.

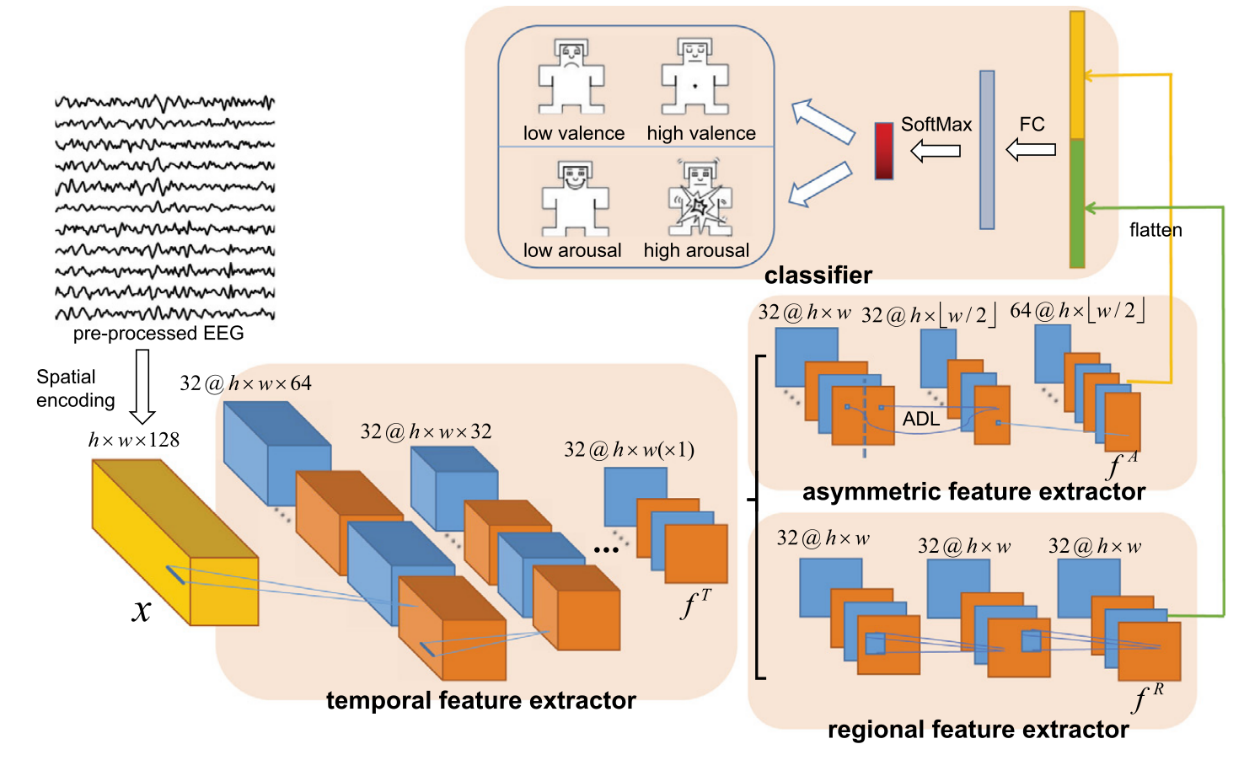


Рисунок 4: Архитектура нейронной сети RA-CNN

Таблица 5 Результаты воспроизведения модели №1

|  |  |  |  |
| --- | --- | --- | --- |
| №  п.п. |  | Arousal  % | Valence  % |
|  | Данные авторов  accuracy | 97.1 | 96.6 |
|  | Воспроизведение  accuracy  f1 | 97.4  97.1 | 97.5  97.2 |

* 1. **модель №2**

Во второй из выбранных моделей, разработанной Seong-Eun Moon, Soobeom Jang и Jong-Seok Lee в 2018 году, для извлечения признаков применяется фильтрация сигнала ЭЭГ по десяти частотным полосам delta (0-3 Гц), theta (4-7Гц), low alpha (8-9.5 Гц), high alpha (10.5-12 Гц), alpha (8-12 Гц), low beta (13-16 Гц), mid beta (17-20 Гц), high beta (21-29 Гц), beta (13-29 Гц), gamma (30-50 Гц). В статье рассматриваются несколько вариантов реализации модели. Рассчитываются спектральные мощности в каждом диапазоне (PSD); для каждой пары электродов вычисляются коэффициенты корреляции Пирсона (PCC), phase locking value (PLV) и phase lag index (PLI). Из полученных значений формируется массив 32х32х10. Для случая PSD, 32x32 – это пространственная матрица, отображающая поверхность головы. Для остальных моделей – это матрица попарных коэффициентов для каждой пары каналов. Массив подается на сверточную нейронную сеть, состоящую из последовательно соединенных блоков из 2d-сверточного слоя с ядром 3х3, активационного слоя ReLU и 2d-maxpool слоя с ядром 2х2. Классификатор представляет собой полно связный слой со скрытой размерностью 256. Для формирования обучающего и тестового набора данных исходный сигнал ЭЭГ разбивается с перекрытием 2.5 секунды на 3х секундные семплы. Таким образом, из одного эксперимента получается 115 сегментов. Общие данные случайным образом с использованием кросс-валидации распределяются на обучающую и тестовую выборки. В подходе авторов используется одна модель на всех испытуемых.

Для подтверждения заявленных характеристик были реализованы варианты модели, с PCC и PSD-способами извлечения признаков из ЭЭГ. Данные проведенных экспериментов подтвердили характеристики модели.

Таблица 6 Результаты воспроизведения модели №2:

|  |  |  |  |
| --- | --- | --- | --- |
| №  п.п. |  | Valence, PSD  % | Valence, PCC  % |
|  | Данные авторов  accuracy | 80.86% | 94.44% |
|  | Воспроизведение  accuracy  f1 | 78.93%  76,88% | 95.22%  94.85% |

* 1. **модель №3**

Третья модель [27], разработанная Jinxiang Liaoa, Qinghua Zhong, Yongsheng Zhub и Dongli Cai в 2020 году, аналогична первой модели в части извлечения признаков из ЭЭГ. Здесь также используется пространственное кодирование по поверхности головы и семплирование длительностью в одну секунду. Для воспроизведения была выбрана часть статьи, рассматривающая только ЭЭГ сигнал.

В качестве предобработки сигнала используется z-score стандартизация и вычитание среднего значения первых трех секунд эксперимента из остальных данных.

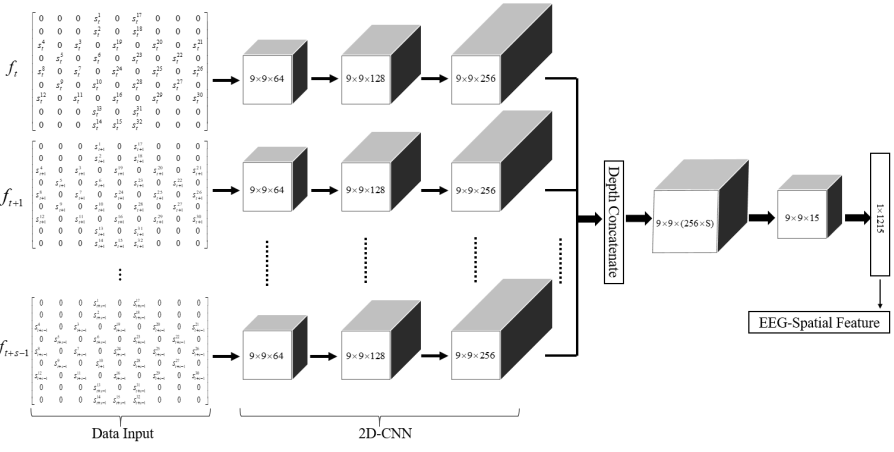
В данной модели (рис. 5) 128 2d-матриц размерностью [9x9x128], которые подаются на 128 параллельных сверточных подсетей. Каждая подсеть состоит из трех последовательно соединенных блоков, состоящих из 2d-сверточного слоя и слоя batch normalization. Далее данные всех подсетей конкатенируются и поступают на аналогичный сверточный сегмент.

Таблица 7 Результаты воспроизведения модели №3

Рисунок 5: Рисунок 5. Архитектура нейронной сети

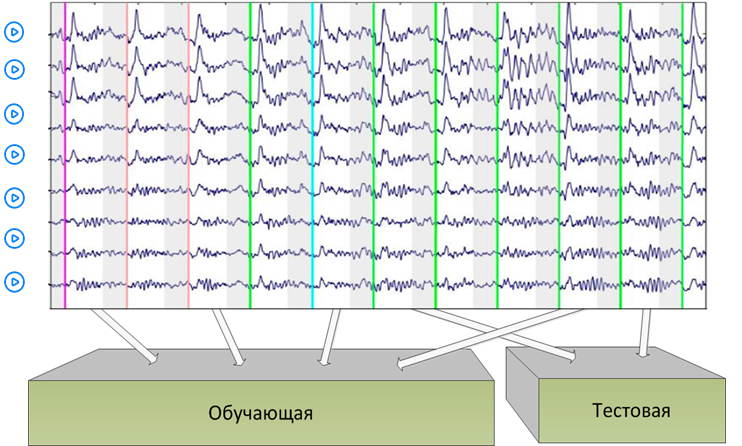
|  |  |  |  |
| --- | --- | --- | --- |
| №  п.п. |  | Arousal  % | Valence  % |
|  | Данные авторов  accuracy | 89.68% | 89.19% |
|  | Воспроизведение  accuracy  f1 | 91,54%  89,86% | 91,39%  90,42% |

В результате воспроизведения получили метрики (табл. 7), аналогичные метрикам автора

1. **Анализ полученных результатов**

Проверка подходов, предложенных в [15, 10, 23], показала, что c их помощью действительно можно достичь заявленных метрик. В то же время, как эти, так и подходы глубокого обучения во многих других статьях с высокими метриками accuracy, используют явно или неявно прописанный, одинаковый способ формирования обучающего и тестового набора данных. При этом способе семплированные данные объединяются и затем кросс-валидацией случайно разбиваются на тестовую и обучающую выборки. При таком способе данные из одного и того же эксперимента попадают и в тестовый, и в обучающий набор (рис. 6)

Мы предположили, что подобная практика подменяет задачу выявления значимых признаков для распознавания эмоций в ЭЭГ на задачу выявления признаков схожести семплов внутри одного эксперимента и, как следствие, определение разметки.

**

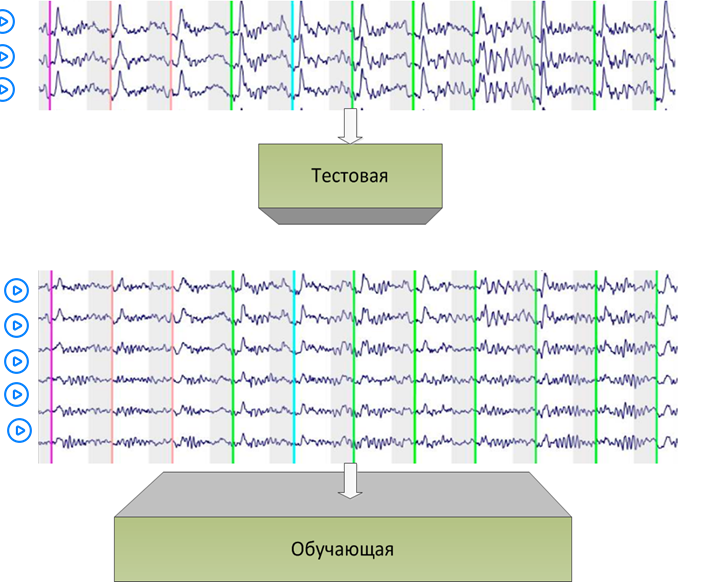
**

Рисунок 7. Разбиение, где обучающие и тестовые наборы независимы (разбиение (+) в таблице 2)

Рисунок 6. Разбиение, где обучающие и тестовые наборы зависимы

Для проверки этого предположения все три модели были проверены в условиях, исключающих попадание в тестовый и обучающий набор данных одного и того же эксперимента (см. рис 7). Для моделей, отдельных для каждого испытуемого это может быть кросс-валидационное разбиение по экспериментам. Что касается общих для всех испытуемых моделей, можно получить один обучающий набор объединением обучающих наборов (полученных кросс валидацией по экспериментам) для каждого испытуемого. Второй способ для общих моделей – кросс-валидация по испытуемым: все эксперименты части испытуемых попадают в обучающий набор, а эксперименты оставшихся испытуемых – в тестовый набор.

Таблица 8. Сравнение результатов модели для зависимого и независимого разбиения, accuracy

|  |  |  |  |
| --- | --- | --- | --- |
| № п.п. | Тип модели | Зависимое разбиение | Независимое разбиение |
|  | Наиболее популярный класс |  | 64,40% |
|  | №1 | 97.5% | 63,7% |
|  | №2 | 94.30% | 57,8% |
|  | №3 | 91,39% | 64,3% |

Результаты экспериментов с “независимой” кросс-валидацией на пяти фолдах приведены в таблице 8. Из нее видно, что ни одна из предложенных моделей на независимых наборах данных не смогла превзойти по accuracy классификатор выбора наиболее распространенного класса.

1. **Улучшение качества распознавания эмоций на датасете DEAP**

**6.1. Описание модели**

Метрики качества, получаемые при распознавании эмоционального состояния человека на датасете DEAP, оказываются ниже, чем получаемые по той же методике на других датасетах. Это показано в статьях [19, 20, 21, 22, 23]. В ряде статей [19, 20, 21, 23] качество распознавания на датасете DEAP находится на уровне предсказания наиболее популярного класса, тогда как на других датасетах те же модели дают значимый результат. Авторы [20] объясняют эту особенность недостатками разметки датасета DEAP.

На наш взгляд, снижение метрик обусловлено отнесением эмоций, оценки которых по шкале от 1 до 9 были близки к середине диапазона, к неверной категории из-за субъективности оценки эмоций (и особенно не сильно выраженных). Поэтому эмоции с оценкой от 3.5 до 5.5 мы выделили в категорию «нейтральных» и провели ряд экспериментов для оценки влияния этой категории на качество распознавания.

В проведенных экспериментах в качестве признаков были выбраны следующие параметры в четырех диапазонах частот:

* DE – дифференциальная энтропия;
* DASM – разность DE в симметричных каналах в правом и левом полушарии;
* RASM - отношение DE в симметричных каналах в правом и левом полушарии;
* DCAU – разность DE в симметричных каналах для фронтальных – затылочных электродов.

По мнению [19, 24, 25], подобный набор признаков наиболее перспективен для распознавания эмоционального состояния по ЭЭГ.

Также мы исследовали, являются ли признаки, важные в анализе состояний погружения и концентрации согласно статьям [29, 30], существенными для анализа эмоций. В частности, относительные изменения отношений спектральных мощностей Theta/Alpha, Beta/Theta, (SMR + Mid-β)/Theta, по сравнению с состоянием покоя.

Классификация осуществлялась с помощью модели, составленной на основе бустинга решающих деревьев LightGBM. Был использован общий классификатор для всех испытуемых. Эксперименты каждого испытуемого делились на обучающие и тестовые, а затем объединялись в общие обучающий и тестовый наборы. Метрики accuracy и f1-macro рассчитывались для каждого испытуемого отдельно и затем усреднялись. Единица обучения – запись ЭЭГ для отдельного эксперимента.

* 1. **Признаки DE, DASM, RASM, DCAU**

Для оценки влияния нейтральных эмоций приведем метрики, полученные на основе признаков DE, DASM, RASM, DCAU в различных спектральных диапазонах theta (4-7Гц), alpha (8-12 Гц), beta (13-29 Гц), и gamma (30-50 Гц) отдельно для каждой группы. Метрики получены до исключения и после исключения экспериментов с нейтральными эмоциями.

До исключения нейтральных эмоций

Таблица 9. Valence

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Признак/Диапазон | teta | | alpha | | beta | | gamma | |
| Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| DE | 62.1 | 53.3 | 64.5 | 55.4 | 64.4 | 56.0 | 63.2 | 53.5 |
| DASM | 61.7 | 52.2 | 64.0 | 55.2 | 64.6 | 55.7 | 64.2 | 55.7 |
| RASM | 64.1 | 54.1 | 64.6 | 55.0 | 64.6 | 54.6 | 63.2 | 54.6 |
| DCAU | 62.2 | 53.3 | 61.9 | 52.3 | 62.2 | 53.6 | 64.6 | 56.2 |

Таблица 10. Arousal

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Признак/Диапазон | teta | | alpha | | beta | | gamma | |
| Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| DE | 60.5 | 47.5 | 60.7 | 47.3 | 63.8 | 50.8 | 64.4 | 51.6 |
| DASM | 61.1 | 48.3 | 61.2 | 48.1 | 62.1 | 49.4 | 63.5 | 50.9 |
| RASM | 62.1 | 49.1 | 61.1 | 47.7 | 65.7 | 52.3 | 65.3 | 52.7 |
| DCAU | 59.2 | 46.7 | 62.1 | 49.1 | 63.6 | 50.3 | 63.1 | 50.2 |

После исключения нейтральных эмоций

Таблица 11. Valence

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Признак/Диапазон | teta | | alpha | | beta | | gamma | |
| Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| DE | 66.6 | 54.6 | 67.8 | 57.0 | 70.8 | 61.1 | 70.1 | 60.1 |
| DASM | 70.4 | 60.8 | 69.8 | 58.5 | 71.7 | 61.4 | 68.4 | 56.5 |
| RASM | 68.5 | 57.6 | 68.6 | 56.9 | 68.7 | 57.8 | 69.3 | 58.9 |
| DCAU | 66.3 | 55.3 | 67.1 | 55.7 | 68.5 | 57.6 | 68.2 | 57.6 |

Таблица 12. Arousal

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Признак/Диапазон | teta | | alpha | | beta | | gamma | |
| Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| DE | 68.6 | 53.6 | 70 | 54.8 | 70.2 | 55.9 | 70.0 | 55.7 |
| DASM | 66.8 | 50.5 | 69.6 | 53.8 | 71.1 | 56.9 | 69.8 | 56.5 |
| RASM | 68.7 | 53.1 | 68.9 | 53.9 | 70.1 | 54.5 | 71.7 | 58.0 |
| DCAU | 67.1 | 51.5 | 67.6 | 52.7 | 69.9 | 548 | 70.4 | 55.0 |

Из таблиц 9-12 видно, что метрики повышаются с увеличением частоты диапазона, и гамма и бета диапазоны более информативны для распознавания эмоций, чем тета и альфа. Данные выводы согласуются со статьей [19].

Также были рассмотрены сразу все признаки DE, DASM, RASM, DCAU во всех каналах с исключенными нейтральными эмоциями и без их исключения.

Таблица 14. Метрики, полученные на всей совокупности признаков DE, DASM, RASM, DCAU

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Valance | | Arousal | |
| Accuracy | F1 | Accuracy | F1 |
| C нейтральными | 69.1 | 60.2 | 64.9 | 50.8 |
| Без нейтральных | 75.7 | 66.1 | 73.6 | 57.4 |

Путем поочередного исключения признаков групп DE, DASM, RASM, DCAU из всего набора признаков (без нейтральных эмоций), было найдено, что лучшие результаты дает комбинация признаков DE, DASM для valence (accuracy = 75.8, f1 = 66.3) и DE, RASM (accuracy = 74.0, f1 = 58.4 ) для arousal.

Таким образом, исключение записей с нейтральными эмоциями дало прирост качества классификатора для valence с (accuracy = 69.1, f1 = 60.2 ) до (accuracy = 75.7, f1 = 66.1 ), для arousal с (accuracy = 64.9, f1 = 50.8) до (accuracy = 73.6, f1 = 57.4 ).

**6.3 Признаки Theta/Alpha, Beta/Theta, (SMR + Mid-β)/Theta** вычисляются следующим образом: для каждого эксперимента рассчитывается данный индекс сначала для первых трех секунд сигнала (базовый), и затем он же – для оставшихся 60 секунд записи ЭЭГ. Находится относительное изменение индекса по отношению к базовому. Theta (4-8 Гц), Alpha (8-12 Гц), Beta (11-30 Гц), SMR (12 – 15 Гц), Mid-β (15- 20 Гц) – спектральные мощности в соответствующих диапазонах

Результаты, полученные на каждой группе признаков по всем каналам Theta/Alpha, Beta/Theta, (SMR + Mid-β)/Theta и по всем признакам сразу, выведены в таблице 13.

Таблица 13. Признаки, перспективные для оценки состояний погружения и концентрации в оценке эмоционального состояния (без нейтральных эмоций)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Признак | Valence | | Arousal | |
| Acc | F1 | Acc | F1 |
| Theta/Alpha | 0.62 | 0.52 | 0.62 | 0.48 |
| Beta/Theta | 0.60 | 0.48 | 0.65 | 0.52 |
| (SMR + Mid-β)/Theta | 0.62 | 0.52 | 0.62 | 0.48 |
| Theta/Alpha + Beta/Theta + (SMR + Mid-β)/Theta | 0.62 | 0.50 | 0.63 | 0.48 |

На основе полученных результатов можно сделать вывод, что признаки Theta/Alpha, Beta/Theta, (SMR + Mid-β)/Theta не дают значимого эффекта в оценке эмоционального состояния, по сравнению с бейзлайн-классификаторами.

**6.4 Выводы**

Исключение “нейтральных” эмоций привело к улучшению метрик. Это свидетельствует о том, что для улучшения качества классификации необходимо использовать данные экспериментов, в которых оценки испытуемых и по параметру arousal, и по параметру valence близки либо к максимальным, либо к минимальным значениям.

Хочется отметить подход к проведению эксперимента, предложенный авторами датасета SEED [28]. В нем испытуемым предлагались к просмотру фрагменты фильмов, подобранные таким образов, чтобы вызывать однозначные, заранее определенные эмоциональные реакции, т.е. оценку эмоциональной окраски видеоряда давал экспериментатор, а не испытуемый. Подобный подход позволил существенно улучшить качество распознавания эмоций, что подтверждается результатами, приведенными в [19, 20, 22, 23], где исследовались данные как датасета SEED, так и датасета DEAP.

**7. Заключение**

При поиске модели или подхода к анализу ЭЭГ для практического применения основным источником информации являются опубликованные научные статьи.

Как показано в данной работе, высокие метрики, приводимые в статьях, не всегда гарантируют их практическую ценность. В этом разделе мы приводим минимальные критерии, по которым можно принять решение о целесообразности применения подхода, использованного в той или иной статье. Анализ научных статей, приведенный в разделе 4 показывает, что часто работы с высокими метриками качества не удовлетворяют этим критериям.

В настоящей работе показано, что наибольшее влияние на метрики качества оказывает методика разбиения датасета на тестовую и обучающую выборки. Попадание в тестовую и обучающую выборки семплов, выделенных из ЭЭГ в рамках одного и того же эксперимента, может привести к необоснованному повышению метрик. При этом некоторые виды предобработки данных, такие как: вычитание какого-либо значения из всех семплов выборки [27], перекрытие семплов [10] – только усиливают этот эффект. Таким образом, признаки в обучающей и тестовой выборках должны быть рассчитаны на основе данных ЭЭГ различных экспериментов. Способ разбиения в статье должен быть явно указан.

Другим важным критерием являются приведенные в статье F1-macro метрики качества классификации. Как показано в разделе 4, несбалансированность классов в датасете позволяет получать относительно высокие показатели accuracy при классификации (порядка 60-70%) за счет предсказания только наиболее популярного класса.

Описаны подходы по улучшению метрик качества в датасете DEAP. В силу субъективности оценки эмоций испытуемыми, исключение экспериментов, которые вызывают нейтральные, неоднозначные эмоции для конкретного испытуемого, приводит к улучшению метрик качества. Возможно, стоит при создании нового датасета делать разметку не на основе субъективных оценок испытуемых, а используя заранее полученную классификацию стимула, как в датасете SEED.

Увеличение метрик было получено на основе признаков DE, DASM, RASM, DCAU, перспективных по мнению ряда статей [19, 24, 25]. Признаки, значимые в оценке состояний погружения и концентрации Theta/Alpha, Beta/Theta, (SMR + Mid-β)/Theta, не дали значимого эффекта.

Сформулированные выводы и рекомендации могут быть полезны специалистам, начинающим исследования в области распознавания эмоций на основе данных ЭЭГ.

**Литература**

1. Islam M. R. et al., Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques // IEEE Access 2021. V.9. P. 94601-94624 doi: 10.1109/ACCESS.2021.3091487.
2. Rachita R., Shreyas W. Brain Computer Interface Market by Component (Hardware, Software), Type (Invasive, Non Invasive, Partially Invasive), and Application (Healthcare, Communication & Control, Entertainment & Gaming, Smart Home Control, and Others): Global Opportunity Analysis and Industry Forecast, 2021-2030 [Электронный ресурс] – Режим доступа: <https://www.alliedmarketresearch.com/brain-computer-interfaces-market>, свободный. Яз. англ. (дата обращения 01.10.2021)
3. Alakuş T., Gonen M., Turkoglu I. Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO // Biomedical Signal Processing and Control. 2020. V. 60. P.101951. doi:10.1016/j.bspc.2020.101951
4. Koelstra S. et al. DEAP: A database for emotion analysis using physiological signals // IEEE Transactions on Affective Computing. 2012. V.3. N.1. P.18-31 doi: 10.1109/T-AFFC.2011.15
5. Duan R.-N., Zhu J.-Y., Lu B.-L., Differential entropy feature for EEG-based emotion classification // Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER), 2013. P. 81–84, doi: 10.1109/NER.2013.6695876.
6. Katsigiannis S., Ramzan N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices // IEEE Journal of Biomedical and Health Informatics. 2018. V.22. P.98-107. doi: 10.1109/JBHI.2017.2688239
7. Soleymani M., Lichtenauer J., Pun T., Pantic M. A multimodal database for affect recognition and implicit tagging // IEEE Transactions on Affective Computing. 2012. V.3. N.1. P. 42–55. doi: 10.1109/T-AFFC.2011.25.
8. Miranda-Correa J.A., Abadi M.K., Sebe N., Patras I. AMIGOS: A dataset for affect, personality and mood research on individuals and groups [Электронный ресурс] – режим доступа: <http://arxiv.org/abs/1702.02510>, свободный. Яз. англ. (дата обращения 01.10.2021) doi: 10.48550/ARXIV.1702.02510
9. Subramanian R., Wache J., Abadi M. K., Vieriu R. L., Winkler S., Sebe N. ASCERTAIN: Emotion and personality recognition using commercial sensors // IEEE Transactions on Affective Computing. 2018. V.9. N.1. P.147-160 doi: 10.1109/TAFFC.2016.2625250
10. Moon S.-E., Jang S., Lee J.-S., ‘Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information // Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) . 2018. P. 2556–2560. doi: 10.1109/ICASSP.2018.8461315
11. Pandey P., Seeja K. R. Subject independent emotion recognition from EEG using VMD and deep learning // Journal of King Saud University - Computer and Information Sciences. 2019. doi:10.1016/j.jksuci.2019.11.003
12. Chao H., Dong L., Liu Y., Lu B. Emotion recognition from multiband EEG signals using CapsNet // Sensors. 2019. V.19. N.9. doi: 10.3390/s19092212
13. Cimtay Y., Ekmekcioglu E. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition // Sensors. 2020. V.20. N.7. doi: 10.3390/s20072034
14. Wang F., Wu S., Zhang W., Xu Z., Zhang Y., Wu C., Coleman S.Emotion recognition with convolutional neural network and EEG-based EFDMs // Neuropsychologia. 2020. V. 146. doi: 10.1016/j.neuropsychologia.2020.107506
15. Cui H., Liu A., Zhang X., Chen X., Wang K., Chen X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network // Knowledge-Based Systems. 2020. V.205. doi: 10.1016/j.knosys.2020.106243
16. Hassan M., Alam M., Uddin M., Huda S., Almogren A., Fortino G. Human emotion recognition using deep belief network architecture // Information fusion. 2019. V.51. P.10-18. doi: 10.1016/j.inffus.2018.10.009
17. Islam R., Ahmad M. Virtual image from EEG to recognize appropriate emotion using convolutional neural network // Proc. 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). 2019. P. 1–4. doi: 10.1109/ICASERT.2019.8934760
18. Chen X., Zhang P., Mao Z., Huang Y., Jiang D., Zhang Y. Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks // IEEE Access. 2019. V.PP. P.1-1. doi: 10.1109/ACCESS.2019.2908285
19. Li P., Liu H., Si Y., Li C., Li F., Zhu X., Huang X., Zeng Y., Yao D., Zhang Y., Xu P. EEG based emotion recognition by combining functional connectivity network and local activations // IEEE Trans Biomed Eng. V.66, N.10. doi: 10.1109/TBME.2019.2897651
20. Xiang L., Dawei S., Zhang P., Zhang Y., Yuexian H., Bin H. Exploring EEG Features in Cross-Subject Emotion Recognition // Frontiers in Neuroscience. 2018. V.12. P. 162. doi: 10.3389/fnins.2018.00162
21. Nakisa B., Rastgoo N., Tjondronegoro D., Chandran V. Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors // Expert Systems with Applications. 2017. V. 93. P.143–155 doi:10.1016/j.eswa.2017. 09.062
22. Cimtay Y., Ekmekcioglu E. Investigating the Use of Pretrained Convolutional Neural Network on Cross-Subject and Cross-Dataset EEG Emotion Recognition // Sensors. 2020. V.20. N.7. doi: 10.3390/s2007203
23. Qing C., Qiao R., Xu X., Cheng Y. Interpretable emotion recognition using EEG signals // IEEE Access. 2019. V. 7. doi: 10.1109/ACCESS.2019.2928691
24. Yang Y., Wu J., Zheng W.-L., Lu B.-L. EEG-based emotion recognition using hierarchical network with subnetwork nodes // IEEE Transactions on Cognitive and Developmental Systems. 2018. V. 10. N. 2. P. 408–419. doi: 10.1109/TCDS.2017. 2685338
25. Zheng W.-L., Lu B.-L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks // IEEE Transactions on Autonomous Mental Development. 2015. V. 7. N. 3. P. 162–175. doi: 10.1109/TAMD. 2015.2431497
26. Alazrai R., Homoud R., Alwanni H., Daoud M. EEG-based emotion recognition using quadratic time-frequency distribution //Sensors. 2018. V.18. N.8. doi: 0.3390/s18082739
27. Liao J., Zhong Q.; Zhu Y., Cai D. Multimodal Physiological Signal Emotion Recognition Based on Convolutional Recurrent Neural Network // IOP Conference Series: Materials Science and Engineering. 2020. V.783. N.3. doi: 10.1088/1757-899X/782/3/032005
28. Zheng W-L., Liu W., Lu B-L., Cichocki A. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions // IEEE Transactions on Cybernetics. 2019. V. 49. P.1-13. doi: 10.1109/TCYB.2018.2797176
29. [McMahan T.,Parberry I., Parsons T.](https://www.sciencedirect.com/science/article/pii/S2351978915003777#!), Evaluating Player Task Engagement and Arousal Using Electroencephalography // Proc. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015). 2015 V.3
30. Lim S, Yeo M, Yoon G. Comparison between Concentration and Immersion Based on EEG Analysis. // Sensors (Basel). 2019. V.19. N.7 doi: 10.3390/s19071669

1. Alhagry S, Aly A, "Emotion recognition based on EEG using LSTM recurrent neural network" Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 355–358, 2017, doi:10.14569/ijacsa.2017.081046
2. Liu W, Zheng W, and Lu B, "Emotion recognition using multimodal deep learning," in Proc. Int. Conf. Neural Inf. Process., in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9948, 2016, pp. 521–52 9, doi: 10.1007/978-3-319-46672-9\_58