**Minimal criteria for practical applicability of ML approaches for emotions classification by EEG in dataset DEAP**

Abstract

**Introduction**: The work is devoted to the analysis of the electroencephalogram of the brain (EEG) to determine the emotional state of a person.

**Aim**: This work aims to analyze and test SoTA approaches to emotion recognition by EEG.

**Results**: A selection of deep machine learning models with the high-quality indicators of emotion recognition by EEG as of 2021 was made. The results of their work were reproduced and analyzed.

**Conclusions:** We found that the high-quality indicators of emotion recognition in dataset DEAP don`t always guarantee the possibility of the practical application of the described approaches. The evaluation requires additional validation of the approach against several criteria. The minimum criteria for checking the practical applicability of the results of work on the recognition of emotions by EEG were formulated.

**Keywords**: EEG analysis, emotional states, deep learning, DEAP dataset, emotion recognition

**Introduction**

Emotion recognition is the definition of mental processes that reflect a subjective evaluative attitude to existing or possible situations and the objective world. Such a definition of the emotion recognition process can be given based on one, but not the only, definition of emotion.

Emotions are manifested through the expression of the human face, body movements, and gestures, respiratory rate, pulse, skin reactions, but one of the most interesting sources of data on emotions is the signal of the electroencephalogram (EEG) of the brain since it is directly related to the processes occurring in the human brain. In addition, the data obtained from the EEG is not subject to conscious control, unlike gestures, facial expressions, and postures. For some categories of people with physical disabilities, the EEG is one of the few sources of information about their emotional state.

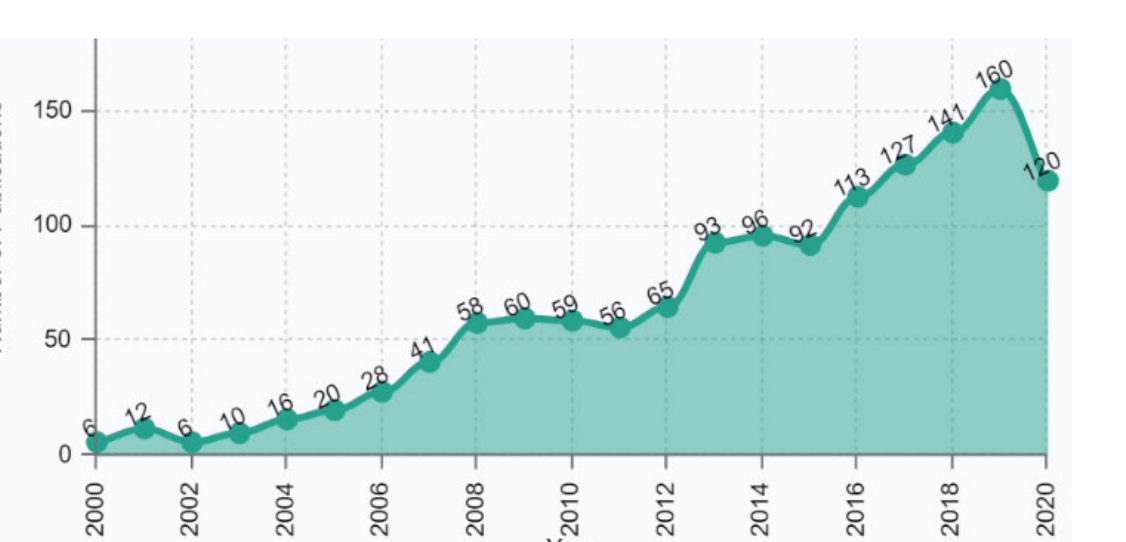


Figure 1: Number of publications on emotion recognition by EEG

In recent years, a significant amount of research has been conducted on the recognition of emotions by brain signals. With the development of artificial intelligence (AI) technologies, emotion recognition has become an integral part of research in the fields of neuroscience, computer science, cognitive sciences and medicine. This trend is well illustrated by the graph of the number of English-language publications (Fig.1) with the tags "EEG" and "Emotion recognition" [1]

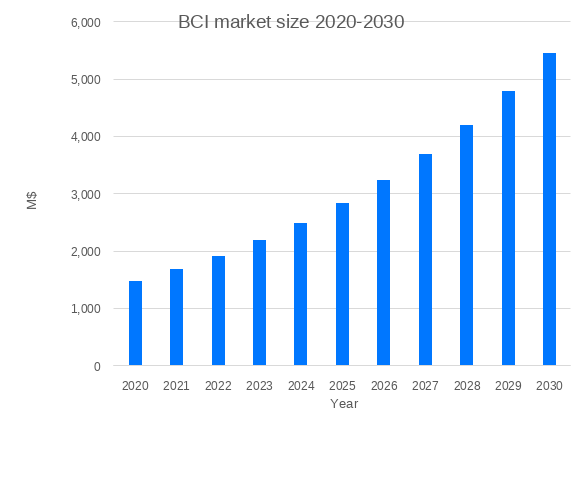


Figure 2: BCI Device Market size

The growth of the brain-computer interface (BCI) device market also confirms this trend. As of 2020, it amounts to $1.5 billion and continues to grow by 15% annually [2].

The price of user devices that allow processing EEG signals currently starts at $ 300 and will decrease with mass production, increasing their availability. Equipment manufacturers and marketers may be primarily interested in the results of research devoted to the analysis of a person's emotional state since such information can be used in such areas as:

* education
* entertainment and games
* increase productivity
* healthcare
* fitness
* mental practices

**Datasets**

As of 2021, Table 1 shows the most cited datasets for EEG emotion recognition that are in the public access.

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 |

As can be seen from Table 1, the DEAP dataset contains the largest amount of data for analysis. In addition, it is one of the earliest datasets in the field of EEG analysis, so this explains its popularity. The DEAP dataset is used in 60% of articles on the topic of determining emotions by EEG. For this reason, in this paper, the choice fell on the DEAP dataset for conducting research.

DEAP is a dataset created by Koelstra et al. [4] and contains 32 EEG channels, 8 channels of peripheral physiological signals from 32 healthy subjects. The signals are recorded when watching 40 one-minute music videos. After watching the clip, the subjects rated 4 parameters of each clip on a scale from 1 to 9: arousal, valence, dominance, and liking.

The EEG signals were sampled at a frequency of 512 Hz and then underwent preprocessing - lowering the sampling frequency to 128 Hz, filtering from 5 to 45 Hz, and removing artifacts associated with blinking and eye movement. The duration of the EEG recording of each experiment is 63 seconds. The first 3 seconds of it are allocated for fixing the initial state of the subject and 60 seconds of viewing of the music video.

**Review of existing literature**

As of 2021, quite high-quality indicators have been achieved in the recognition of emotions by EEG on the DEAP dataset. So Moon et al.[10] in his article reached 99.7% accuracy. At the same time, in the 2019 articles by Pandey and Seeja [11] and Chao et al.[12], the accuracy values were 68.28% and 62.5%, respectively. Table 1 provides an overview of the models and results achieved on the DEAP dataset as of 2021 according to [1**]**

Table 2. Quality metrics on the DEAP dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| №  п.п. | Auther | Разбиение | Year | Number of classes of emotions | Model | 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., [20] | - | 2018 | 2 | PCRNN | Val: 90.80,  Aro: 91.03 |
|  | Alhagry et al., [21] | + | 2017 | 3 | DNN | Val: 85.45  Aro: 85.65,  Lik: 87.99 |
|  | Liu et al., [22] | - | 2016 | 4 | BDAE | Val: 85.2,  Aro: 80.5,  Dom: 84.9,  Lik: 82.4 |
|  | Jinxiang Liao et al[22] | - | 2020 | 2 | CNN | Aro:89.68  Val:89.19 |

+ , - methods of splitting into a training and test dataset (see Fig. 7)

Based on the results of the review, it was decided to reproduce the results of models with the highest accuracy metric values. The following articles were selected:

* **model №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]
* **model №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]
* **model №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[23]

**1 Reproducing the results**

**1.1 Problem statement, choice of metrics and baseline**

Let's formulate a problem that was solved by the articles selected for reproduction. Only EEG signals is used for emotion recognition in this task. Only two emotion scales (Aurosal, Valence) is selected for recognition, represented by values from 1 to 9. Each of the emotions is divided into two classes with a threshold value of 5. Thus, the task of emotion recognition is reduced to the task of multi-label binary classification, and the emotions themselves are represented by two classes: LA/HA (low arousal/high arousal), LV/HV (low valence/ high valence).

To evaluate the quality, we selected not only the accuracy metric, as in the articles, but also f1-macro, which better reflects the quality of the classifier with an imbalance of classes. All calculations are made by us based on cross-validation on five folds. The values of the metrics are calculated individually for each subject (where possible) and are averaged over all subjects.

To assess the quality of emotion recognition, metrics are compared with reference values. According to the upper bound with the approaches having the highest indicators at the date of writing the article, and according to the lower bound with three baselines [4] - 1 - a random classifier choosing a class value with a probability of 0.5, 2 - with a classifier always choosing the most popular class (positive) and 3- a classifier choosing a class based on the proportion the popularity of the class. Metrics for these classifiers are evaluated on a test dataset that has a class imbalance similar to the DEAP dataset (\*\*\*\*).

Таблица 2[4]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Classifier | Aurosal | | Valence | |
| Accuracy | F1 | Accuracy | F1 |
| 1 | Random | 0.500 | 0.483 | 0.500 | 0.494 |
| 2 | Majority class | 0.644 | 0.483 | 0.586 | 0.368 |
| 3 | Proportional to class’s quantities | 0.562 | 0.500 | 0.525 | 0.500 |

* 1. **model №1**

The model, named by the authors of RA-CNN, was developed by Heng Cui, Aiping Liu, Xu Zhang, Xiang Chen, Kongqiao Wang, Xun Chen in 2020. The input of the model receives samples of the original EEG signal with a duration of 1 second (32 channels of 128 samples each). Next, the samples undergo a spatial transformation. A sample with a dimension of 32x128 is converted into an array of 9x9x128, where a 9 x 9 matrix reflects the spatial location of the electrodes on the scalp. Missing values in the three-dimensional array are filled with zeros (see Figure 3)

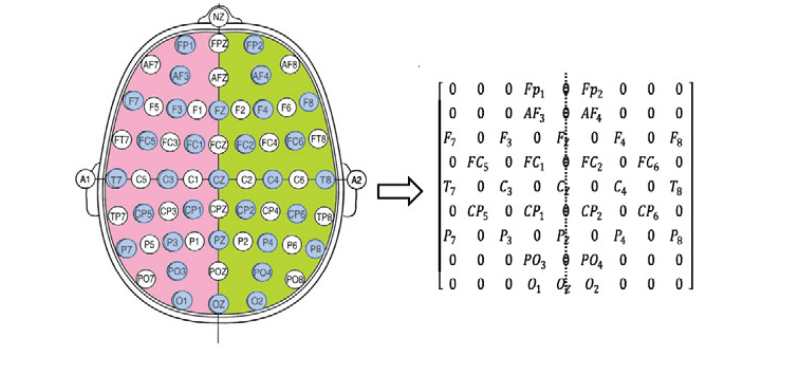


Figure 3 Spatial coding

An array with a dimension of 9x9x128 is fed to the input of a convolutional neural network consisting of four blocks: three convolutional and one classifier - Fig. 4

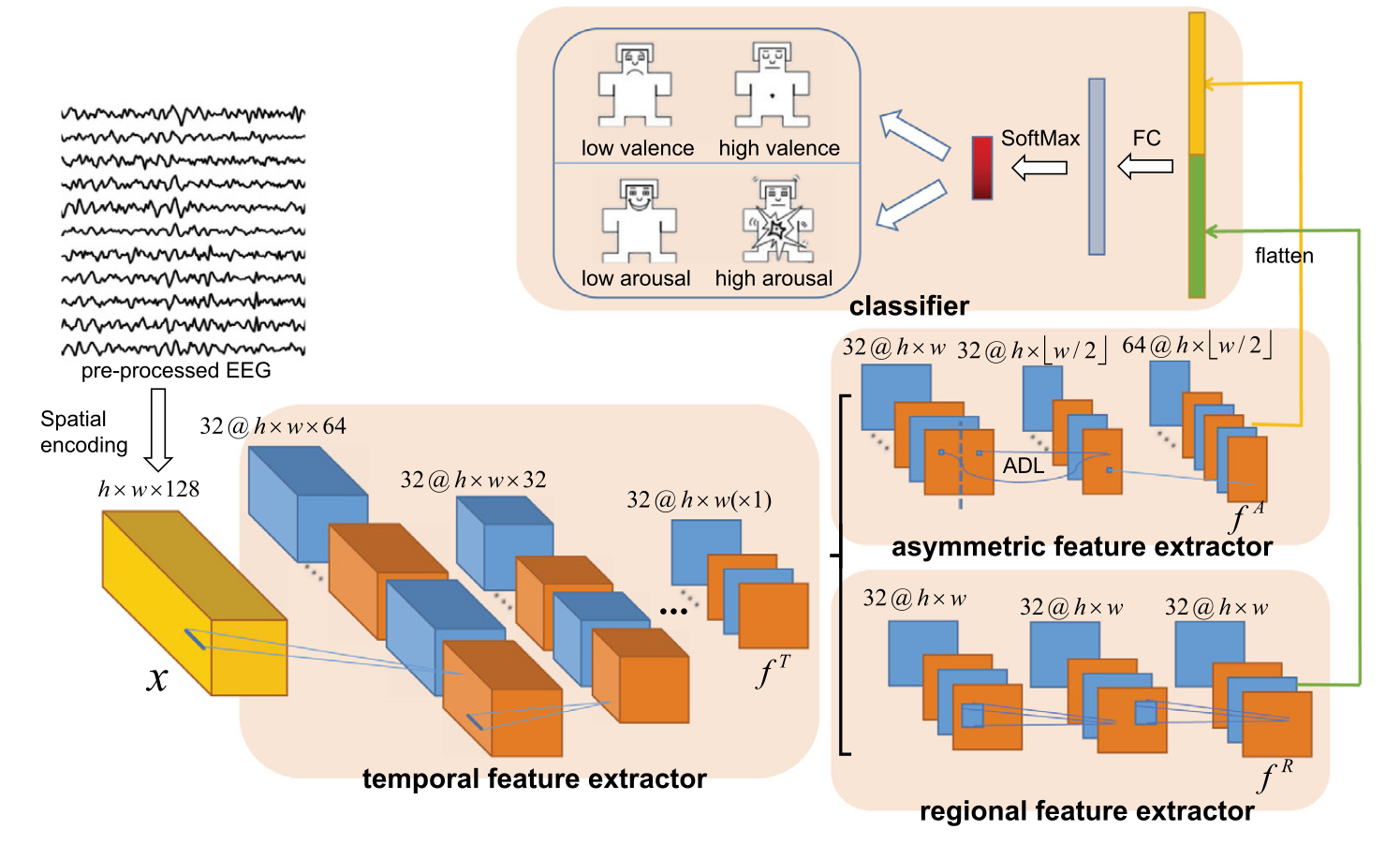


Рисунок 4: Архитектура нейронной сети RA-CNN

The authors' approach is to use the separate model for each of the participants in the experiment. The data is formed from one-second samples from the subject's EEG. Thus, the total size of the data set is 60 \* 40 = 2400 samples for one subject. Training and test data sets are obtained by using cross-validation for a given Table 4 Reprodiced results:

set of samples. When reproducing the RA-CNN model, we obtained metrics similar to those of the authors.

Table 4 Reprodiced results:

|  |  |  |  |
| --- | --- | --- | --- |
| № |  | Arousal  % | Valence  % |
|  | Original results:  accuracy | 97.1 | 95.55 |
|  | Results reproduced:  accuracy  f1 | 97.5  97.2 | 97.5  97.3 |
|  |  |  |  |

* 1. **Model №2**

In the second of the selected models, developed by Seong-Eun Moon, Soobeom Jang, and Jong-Seok Lee in 2018, EEG signal filtering is used to extract features using ten frequency bands delta (0-3 Hz), theta (4-7 Hz), low alpha (8-9.5 Hz), high alpha (10.5-12 Hz), alpha (8-12 Hz), low beta (13-16 Hz), mid beta (17-20 Hz), high beta (21-29 Hz), beta (13-29 Hz), and gamma (30-50 Hz). The article discusses several implementations of the model: spectral powers in each range (PSD) are calculated, Pearson correlation coefficients (PCC), phase-locking value (PLV) and phase lag index (PLI) are calculated for each pair of electrodes. An array of 32x32x10 is formed from the received values. For the PSD case, 32x32 is a spatial matrix representing the surface of the head. For the other models, it is a matrix of pairwise coefficients for each pair of channels. The array is fed to a convolutional neural network consisting of sequentially connected blocks of a 2d convolutional layer with a 3x3 filter, a RELU activation layer and a 2d max pool layer with a 2x2 filter. The classifier is a fully connected layer with a hidden dimension of 256.

To form a training and test data set, the original EEG signal is split into 3-second samples with an overlap of 2.5 seconds. Thus, 115 segments are obtained from one experiment, which is randomly distributed into training and test samples using cross-validation. The authors use one model for all subjects.

To confirm the claimed characteristics, variants of the model were implemented, with PCC and PSD methods for extracting features from the EEG. The experimental data confirmed the characteristics of the model.

Table 5 Reproduced results from model 2:

|  |  |  |  |
| --- | --- | --- | --- |
| № |  | Valence, PSD  % | Valence, PCC  % |
|  | Original resuts:  accuracy | 80.86% | 94.44% |
|  | Results reproduced:  accuracy  f1 | 78.93%  76,88% | 95.27%  94.85% |

As can be seen from Table 5, it was possible to achieve comparable metrics with those given in the original article.

* 1. model №3

The third model [23] developed by Jinxiang Liaoa, Qinghua Zhong, Yongsheng Zhub and Dongli Cai in 2020 is similar to the first model in terms of extracting features from the EEG. Spatial coding on the surface of the head and sampling with a duration of one second is also used. The part of the article considering only the EEG signal was selected for reproduction. As a signal preprocessing, z-score standardization is used and the average value of the first three seconds of the experiment is subtracted from the rest of the data. In this model, the input matrix of dimension [9x9x128] is 128 2d matrices, which are fed to 128 parallel convolutional subnets. Each subnet consists of three sequentially connected blocks consisting of a 2d convolutional layer and a batch normalization layer. Next, the data of all subnets are concatenated and sent to a similar convolutional segment.

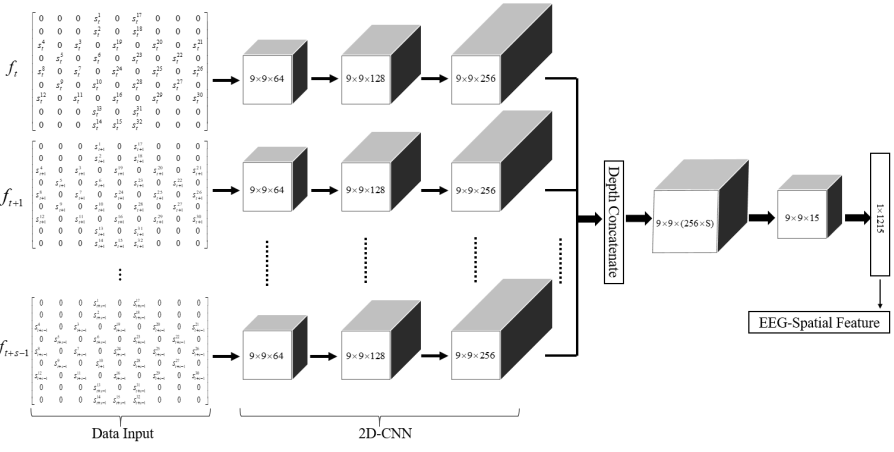


Рисунок 5: Архитектура нйронной сети

As a result of reproduction, metrics similar to the author's metrics were obtained

Table 6 Reproduced results from model 3:

|  |  |  |  |
| --- | --- | --- | --- |
| № |  | Arousal  % | Valence  % |
|  | Original results:  accuracy | 89.68% | 89.19% |
|  | Results reproduced  accuracy  f1 | 91,54%  89,86% | 91,39%  90,42% |

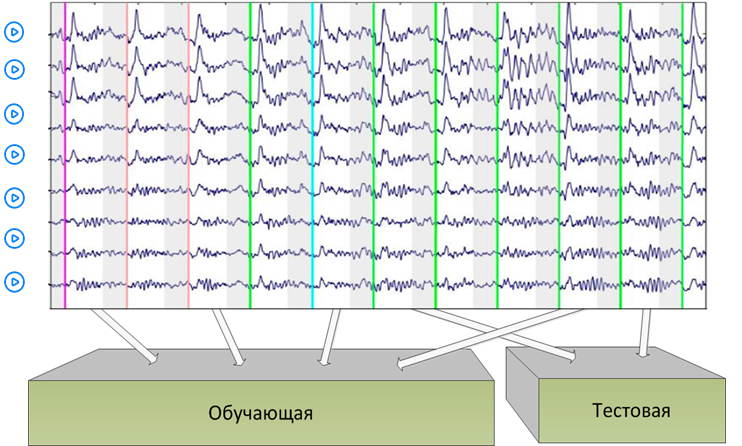
**Results**

Verification of the approaches proposed in [15, 10, 23] showed that with it is possible to achieve the stated metrics. At the same time, both these and deep learning approaches in many other articles with high accuracy metrics use an explicitly or implicitly prescribed, identical way of forming a training and test data set. According to this approach, the sampled data is added to one pile and then randomly split into test and training samples by cross-validation. With this method, data from the same experiment fall into both the test and the training set (Fig. 6).

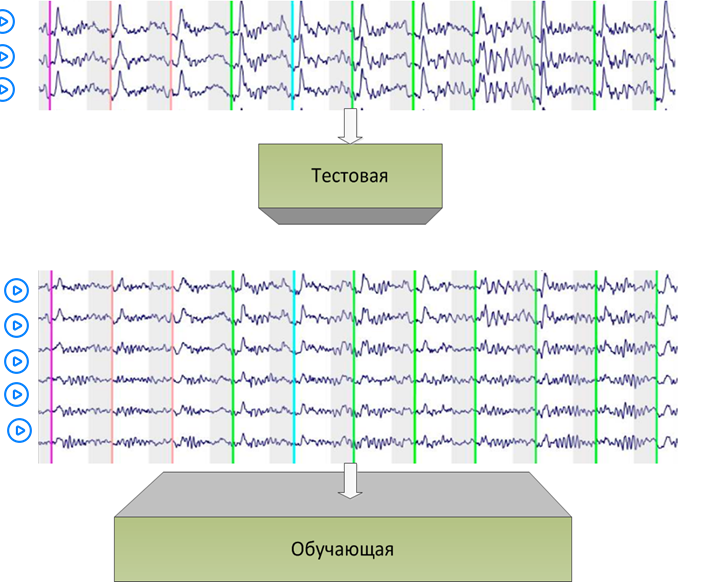
In our opinion, this practice makes it possible to replace the task of identifying qualitative signs of emotion recognition in the EEG with the task of identifying a specific experiment and identifying signs of similarity of samples within one experiment and, as a result, determining the markup.

To verify this assumption, all three models considered were tested under conditions that exclude the same experiment from entering the test and training data sets (see Figure 7). For models that are separate for each subject, this can be a cross-validation breakdown by experiment. For models common to all subjects, it is possible to obtain a training set by combining training sets (obtained by cross-validation of experiments) for each subject. The second method for general models is cross-validation by subjects: all experiments of a part of the subjects fall into the training set, and the experiments of the remaining subjects fall into the test set.

The results of experiments with cross-validation on 5 folds according to the experiments of the subjects are shown in Table 7. It shows that none of the proposed models on independent data sets could surpass the accuracy baseline of the classifier for choosing the most common class.

Рисунок 6: Существующий способ распределение на тестовую и обучающую выборки

Partitioning, where the training and test sets are dependent (partitioning (–) in Table 2)

Рисунок 7: Предложенный способ распределения на тестовую и обучающую выборки

Partitioning, where the training and test sets are independent (partitioning (+) in Table 2)

Table 7. Model validation results

|  |  |  |  |
| --- | --- | --- | --- |
| № п.п. | Model | Dependent partioning | Independent partioning |
|  | Majority class classifier |  | 64,40% |
|  | Model №1 | 97.5% | 63,7% |
|  | Model №2 | 94.30% | 56,1% |
|  | Model №3 | 91,39% | 64,3% |

**Подходы и рекомендации к датасету DEAP**

Некоторые статьи производят эксперименты сразу на нескольких датасатах (SEED, MANHOB). Метрики на датасете DEAP чаще всего уступают метрикам на других датасетах. Ряд авторов считает, что сложности с распознаванием эмоций в датасете DEAP связаны с субъективностью разметки эмоций испытуемыми. Напомним, что испытуемых просили оценить свою эмоцию по 9-балльной системе. Мы предположили, что нейтральные эмоции в диапазоне оценки 3.5 – 5.5 могут оцениваться менее точно и предложили предложили из датасета эксперименты с нейтральной оценкой.

В качестве признаков были выбраны перспективные с точки зрения ряда статей признаки:

* DE – дифференциальная энтропия
* DASM – разность DE в симметричных каналах справа-слева
* RASM - отношение DE в симметричных каналах справа-слева
* DCAU – разность DE в симметричных каналах для фронтальных – затылочных электродов

Данные признаки были взяты в четырех основных частотных диапазонах. В качестве классификатора были выбран бустинг решающих деревьев.

Исключение нейтральных эмоций было произведено таким образом, чтобы не увеличить дисбаланс классов (63 % позитивный класс, 37% негативный класс)

Приведем результаты для различных признаков в различных диапазонах, полученные до исключения и после исключения экспериментов с нейтральными эмоциями

До исключения нейтральных эмоций

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 |

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

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 |

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 полученных данных видно, что метрики повышаются с увеличением частоты дипазона, и гамма и бета диапазоны более информативны для распознавания эмоций, чем тета и альфа. Данные выводы согласуются со статьей [].

Также была рассмотрена комбинация всех признаков DE, DASM, RASM, DCAU полностью и с отобранными признаками методом отбора mRMR, который максимизирует число коррелирующих с таргетом признаков и минимизирует количество коллинеарных признаков.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Valance | | Arousal | |
| Accuracy | F1 | Accuracy | F1 |
| C нейтральными, все признаки | 66.3 | 57.6 | 66.3 | 52.7 |
| Без нейтральных, все признаки | 73.6 | 63.0 | 72.4 | 57.3 |
| C нейтральными, все признаки, отобранные признаки (100, 50, 30, 10) | (66.4, 67.9, 66.7 | 56.9, 58.3, 57.2 | ( 64.6, 63.5, | ( 51.1, 50.4) |
| Без нейтральных, отобранные признаки (150, 140, 120, 100, 50, 30, 10) | (75.2, 76.0, 75.8, 74.2, 72.7, 71.9, \_) | (65.5, 66.4, 65.5, 64.7\_,62.7, 61.7, \_) | (72.0, 72.5, 71.8\_,73.6, 71.9, \_) | (56.1, 56.5, 56.4, \_, 57.0, 54.9, \_) |

Для случая отбора признаков для датасета без нейтральных эмоций приведем первые 20 отобранных признаков в порядке убывания важности.

|  |  |
| --- | --- |
| Valence | ['RASM\_3\_5', 'RASM\_3\_8', 'RASM\_3\_9', 'RASM\_3\_2', 'RASM\_3\_6', 'RASM\_3\_12', 'RASM\_3\_11', 'RASM\_3\_3', 'RASM\_3\_7', 'RASM\_3\_10', 'DASM\_3\_4', 'RASM\_3\_1', 'DCAU\_3\_6', 'DE\_3\_0', 'DE\_0\_29', 'RASM\_3\_4', 'DE\_3\_3', 'DE\_3\_6', 'RASM\_1\_11, 'RASM\_1\_10' ['RASM\_3\_10', 'RASM\_3\_4', 'RASM\_3\_7', 'RASM\_3\_9', 'RASM\_3\_6', 'RASM\_3\_11', 'RASM\_3\_12', 'RASM\_3\_5', 'RASM\_3\_3', 'RASM\_3\_8', 'RASM\_3\_1', 'DE\_1\_24', 'RASM\_3\_2', 'DE\_2\_29', 'DCAU\_1\_1', 'DE\_1\_31', 'DASM\_1\_11', 'DE\_1\_0', 'DE\_2\_13'] |
| Arousal | ['RASM\_3\_5', 'RASM\_3\_7', 'DASM\_1\_0', 'RASM\_1\_8', 'RASM\_3\_6', 'RASM\_3\_2', 'RASM\_3\_11', 'RASM\_3\_12', 'RASM\_3\_10', 'RASM\_3\_9', 'RASM\_3\_8', 'DASM\_3\_1', 'DE\_2\_14', 'RASM\_3\_4', 'DASM\_2\_6', 'RASM\_3\_3', 'RASM\_1\_0', 'DE\_3\_7', 'DASM\_1\_1', 'DE\_0\_3', 'RASM\_3\_1', 'DASM\_1\_6', 'DE\_1\_11', 'DASM\_3\_7', 'RASM\_0\_8', 'DE\_2\_24', 'DASM\_3\_8', ' |

Видим, что наиболее важными признаками получились признаки RASM в гамме-диапазоне

Таким образом, исключение записей с нейтральными эмоциями дало прирост качества классификатора для valence с (accuracy = 67.9, f1 = 58.3 ) до (accuracy = 76.0, f1 = 66.4 ), для arousal с (accuracy = 66.9, f1 = 52.7 ) до (accuracy = 73.6, f1 = 57.0 ),

Хочется отметить, что как один из способов решения проблемы субъективности и неточности оценок испытуемыми своих эмоций в некоторых датасетах, например датасете SEED, испытуемых не просят оценивать свои эмоции, а разметка одинаковая для всех испытуемых. Так же в датасете DEAP для некоторых экспериментов дается сопутствующее эксперименту видео, по которому так же можно было бы сделать объективную оценку эмоции испытуемого.

1. **Conclusions**

**Из статьи должно быть понятно**

1. **Одна модель на человека или общая для всех**
2. **Каким образом производится разбиение на тест и трейн**
3. **Что является единицей обучения**
4. **Каким образом получены метки. Как производится дискретизация на метки, удаляются ли нейтральные эмоции, какие пороги**
5. **Какой дисбаланс классов при подсчете метрик**
6. **Каким образом подсчитываются метрики (усредняется по метрикам людей) или метрики по общему тесту, если общий на всех классификатор**
7. **F1-macro кроме accuracy, который лучше отражает метрики для дисбаланса классов**
8. **Если метод отбора фич, то должен делаться только по трейну, не по всей совокупности**
9. **Если применяется кросс-фолд, то он случайный или стратифицированный**

When searching for a model or approach to EEG analysis for the practical application of knowledge, the main source of information is scientific articles published in relevant publications. As shown in this paper, the high metrics given in the articles do not always guarantee their practical value. In this section, we present the minimum criteria by which it is possible to decide on the appropriateness of further analysis of the approach given in this article. The analysis of scientific articles given in section 3 shows that often articles with high-quality metrics do not meet the above criteria.

For this reason, the recommendations below may be useful both to other researchers conducting EEG analysis on data from the DEAP dataset or similar and to parties interested in the practical application of EEG analysis techniques.

As shown in this paper, the method of splitting the dataset into test and training samples has the greatest impact on quality metrics. Getting into the test and training samples of features isolated from the EEG of the same experiment can lead to an unjustified increase in metrics. At the same time, some types of data preprocessing, such as min-max scaling, subtracting any value from all samples of the sample [xx, xx, xx, xx], overlapping samples [xx, xxx, xx, xx] only enhances this effect. Thus, the signs in the training and test samples should be calculated based on EEG data from various experiments. The method of splitting in the article must be explicitly specified.

The second important criterion is the introduction of the F1-macro classification quality metric in the article. As shown in section 4, the imbalance of classes in the dataset makes it possible to obtain relatively high accuracy rates in classification (about 60-70%) by predicting only the most popular class.

Recommendations are made to improve quality metrics in the DEAP dataset. The assumption is formulated about the inaccuracy of subjective assessments of emotions by the subjects and the ways to overcome this problem are indicated.

1. **Краткая справка об участниках проекта**

**Шуранов Евгений**

к.т.н. Руководитель лаборатории по искусственному интеллекту Huawei специализирующейся на обработке речевых сигналов. Доцент ИТМО. Практика по ЦОС , курс по обработке аудио сигналов. Более 30 статей в том числе ВАК

**Кравченко Федор**

Senior software engineer — Huawei ,?????

**Савкина Екатерина**

студент MADE, разработчик ПО в «?????»

**Соколов Александр**

к.т.н. Студент MADE, Machine learning engineer в «6th Grain»

1. **Библиография**
2. Md. Rabiul Islam et al «Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques» IEEE Access · June 2021 Md. Rabiul Islam et. Al
3. Данные по исследованию рынка BCI
4. T. B. Alakus, M. Gonen, and I. Turkoglu, ‘‘Database for an emotionrecognition system based on EEG signals and various computer games—GAMEEMO,’’ Biomed. Signal Process. Control, vol. 60, Jul. 2020,
5. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi,T. Pun, A. Nijholt, and I. Patras, ‘‘DEAP: A database for emotion analysis;using physiological signals,’’ IEEE Trans. Affect. Comput., vol. 3, no. 1,
6. R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, ‘‘Differential entropy feature for EEG-based emotion classification,’’ in Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER), Nov. 2013, pp. 81–84, doi: 10.1109/NER.2013.6695876.
7. S. Katsigiannis and N. Ramzan, ‘‘DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost offthe-shelf devices,’’ IEEE J. Biomed. Health Informat., vol. 22, no. 1,
8. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, ‘‘A multimodal database for affect recognition and implicit tagging,’’ IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 42–55, Jan. 2012, doi: 10.1109/T-AFFC.2011.25.
9. J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, ‘‘AMIGOS: A dataset for affect, personality and mood research on individuals and groups,’’ 2017, arXiv:1702.02510.[Online]. Available:<http://arxiv.org/abs/1702.02510>
10. R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, ‘‘ASCERTAIN: Emotion and personality recognition using commercial sensors,’’ IEEE Trans. Affect. Comput., vol. 9, no. 2, pp. 147–160,
11. S.-E. Moon, S. Jang, and J.-S. Lee, ‘‘Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2018, pp. 2556–2560, doi:10.1109/ICASSP.2018.8461315.
12. P. Pandey and K. R. Seeja, ‘‘Subject independent emotion recognition from EEG using VMD and deep learning,’’ J. King Saud Univ.-Comput. Inf. Sci., Nov. 2019, doi:10.1016/j.jksuci.2019.11.003.
13. H. Chao, L. Dong, Y. Liu, and B. Lu, ‘‘Emotion recognition from multiband EEG signals using CapsNet,’’ Sensors, vol. 19, no. 9, p. 2212,May 2019, doi: 10.3390/s19092212.
14. Y. Cimtay and E. Ekmekcioglu, ‘‘Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition,’’ Sensors, vol. 20, no. 7, p. 2034, Apr. 2020, doi: 10.3390/s20072034
15. F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, ‘‘Emotion recognition with convolutional neural network and EEG-based EFDMs,’’ Neuropsychologia, vol. 146, Sep. 2020, Art. no. 107506, doi:
16. H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, ‘‘EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network,’’ Knowl.-Based Syst., vol. 205, Oct. 2020
17. M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino,‘‘Human emotion recognition using deep belief network architecture,’’ Inf. Fusion, vol. 51, pp. 10–18, Nov. 2019, doi:10.1016/j.inffus.2018.10.009.
18. M. R. Islam and M. Ahmad, ‘‘Virtual image from EEG to recognize appropriate emotion using convolutional neural network,’’ in Proc. 1St Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT), May 2019, pp. 1–4, doi: 10.1109/ICASERT.2019.8934760.
19. J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, ‘‘Accurate EEG-based emotion recognition on combined features using deep convolutional neural