Lurian Journal. 2023. Vol. 4. No. 1. P. 8–20
DOI 10.15826/Lurian.2023.4.1.1
УДК 612.821:612.014.42 + 612.825.8 + 612.825.26:159.953 + 159.953.4

# EmoEye: Eye-Tracking and Biometrics Database for Emotion Recognition

**Ekaterina O. Ivanina**
Neuro laboratory "NeuroLab",
Moscow, Russia
**Anna D. Tokmovtseva**
Neuro laboratory "NeuroLab",
Russian Orthodox University of Saint John the Divine,
Moscow, Russia
**Elizaveta V. Akelieva**
Neuro laboratory "NeuroLab",
Lomonosov Moscow State University,
Moscow, Russia

# EmoEye: айтрекер и биометрическая база данных для распознавания эмоций

**Екатерина Олеговна Иванина**
Нейролаборатория «НейроЛаб»,
Москва, Россия
**Анна Денисовна Токмовцева**
Нейролаборатория «НейроЛаб»,
Российский православный университет Святого Иоанна Богослова,
Москва, Россия
**Елизавета Владимировна Акельева**
Нейролаборатория «НейроЛаб»,
Московский государственный университет имени М. В. Ломоносова,
Москва, Россия

**To cite this article**: Ivanina, E. O., Tokmovtseva, A. D., & Akelieva, E. V. (2023). EmoEye: Eye-Tracking and Biometrics Database for Emotion Recognition. *Lurian Journal, 4*(1), pp. 8–20. doi: 10.15826/Lurian.2023.4.1.1

***Abstract.*** Emotion recognition using Machine Learning algorithms is often used both in science and commerce. Responding to the demand for deep learning techniques of automatic emotion detection using biological signals and our own business needs as a neuromarketing laboratory, we created a large dataset of eye tracking and biometrics data suitable for emotion recognition tasks. The EmoEye database sample consisted of 200 people (147 women, 49 men, 4 non-binary individuals; 27.46 ± 11.45 years old). Each respondent was asked to view 316 images from the Open Affective Standardized Image Set (OASIS) and rate them on arousal and valence scales from the Self-Assessment Manikin questionnaire. Eye tracking, galvanic skin response (GSR), and photoplethysmogram were recorded throughout the experiment. Demographic data was also collected for each respondent. The image ratings on the valence scale did not differ statistically from the standard ratings of the corresponding images for the original stimulus base. The overall distribution trends of ratings on both scales for different categories of images were similar for standard ratings and ratings obtained from our respondents. As a result of this study, a corpus of GSR, heart rate variability and eye movement reactions data (fixation coordinates; fixation duration; average pupil size for the right and left eye) was compiled and successfully trained on a multimodal neural network algorithm within our laboratory and is ready for further implementation.

***Keywords:*** *emotion recognition; deep learning; cross-cultural studies; eye-tracking; biometrics; galvanic skin response; heart rate*

***Аннотация.*** Распознавание эмоций с помощью алгоритмов машинного обучения часто используется как в науке, так и в коммерции. С помощью методов глубокого обучения для автоматического обнаружения эмоций с использованием биологических сигналов мы подготовили набор данных для айтрекинга и биометрических данных, подходящих для задач распознавания эмоций. Выборка базы данных EmoEye состояла из 200 человек (147 женщин, 49 мужчин, 4 небинарных индивидуума; 27,46 ± 11,45 лет). Каждому респонденту было предложено просмотреть 316 изображений из открытого аффективного стандартизированного набора изображений (OASIS) и оценить их по шкалам возбуждения и валентности из опросника «Манекен самооценки». Отслеживание взгляда, кожно-гальваническая реакция (GSR) и фотоплетизмограмма регистрировались на протяжении всего эксперимента. Также были собраны демографические данные по каждому респонденту. Оценки изображений по валентной шкале статистически не отличались от стандартных оценок соответствующих изображений для исходной базы стимулов. Общие тенденции распределения оценок по обеим шкалам для разных категорий изображений были одинаковыми для стандартных оценок и оценок, которые были даны нашими респондентами. В результате исследования в рамках нашей лаборатории были получены данные о GSR, вариабельности сердечного ритма и реакциях движения глаз (координаты фиксации; длительность фиксации; средний размер зрачка для правого и левого глаза), которые были успешно реализованы на основе мультимодального нейросетевого алгоритма и готовы к внедрению.

***Ключевые слова:*** *распознавание эмоций; глубокое обучение; кросс-культурные исследования; айтрекинг; биометрия; кожно-гальваническая реакция; частота сердечных сокращений*

## Introduction

Emotion recognition is one of the three key elements of Affective computing besides emotion classification and modulation of human affective states with computerized techniques (Picard, 1995). Common algorithms for emotion recognition and detection employ large amounts of data to determine emotions and make predictions about the emotional states of new same-type data.

Automatic emotion detection is often used to evaluate advertisements and media content, as the traditional marketing methods such as questionnaires or customer interviews could be affected by social biases (Fisher, 2000; Larson, 2019). Neuromarketing methods such as electroencephalography (EEG) or eye-tracking help to resolve the problem of socially desirable behavior by providing an additional measure of emotional states, based on psychophysiological reactions to the viewed content (Ariely & Berns, 2010; Jordão, De Souza, De Oliveira, & Giraldi, 2017; Ouazzani Touhami et al., 2011).

Current datasets of psychophysiological data on which one could base a machine learning algorithm of affect recognition are numerous, however they suffer from several issues which limit their applicability. The first issue is a so-called "representational harm," when datasets for machine learning tend to be affected by social biases or overrepresent only one group of people (Paullada, Raji, Bender, Denton, & Hanna, 2021). Russian sample is often underrepresented in cross-cultural studies of emotions (Pogosyan & Engelmann, 2011; Wierzbicka, 1998), so it is up to no surprise that there is a limited number of available datasets with Russian sample. That circumstance limits the future product's perspectives on the market, as end-to-end local initiatives tend to be given more favor, and neglects cultural differences in emotional expression. The second issue is that datasets often have sample sizes smaller than 100 or even 50 participants: DEAP — 32 subjects (Koelstra et al., 2012), AMIGOS — 40 subjects (Miranda-Correa, Abadi, Sebe, & Patras, 2021), DREAMER — 23 subjects (Abadi et al., 2015; Katsigiannis & Ramzan, 2018). It is possible to overcome such limitations using Nested Cross Validation, however small sample sizes still give a bad approximation of true randomness and could be hard to train and split (Vabalas, Gowen, Poliakoff, & Casson, 2019).

Thus, we decided to collect our own dataset for emotion classification, based on psychophysiological reactions to affective images. We implement eye-tracking, pupillometry and biometrics (GSR, heart rate) methods in our research for following reasons:

(1) Eye-tracking is a popular neuromarketing tool to capture the specifics of visual behavior towards a stimulus (Lim, Mountstephens, & Teo, 2020), patterns of which could relate to expression of affect (Alshehri & Alghowinem, 2013; Lim et al., 2020; Roux, Brunet-Gouet, Passerieux, & Ramus, 2016).
(2) Pupil size, GSR and heart rate all serve as biomarkers of arousal and show an activation level of the autonomic nervous system (Bradley, Miccoli, Escrig, & Lang, 2008; Flykt, 2005; Wu, Liu, & Hao, 2010).

*Table* describes current uni- and multimodal databases with eye-tracking and their availability. Emotional recognition through eye-tracking is quite a new field of research, as all of the datasets presented in the table were collected during the last 3 years. Selected datasets suffer from aforementioned issues typical for psychophysiological databases: 3 of 4 datasets have less than 50 samples, only 2 of 4 are available for commercial studies, and none include Russian participants.

*Table*
**Eye-tracking affective datasets**

| Database | Year | Stimuli | Number of subjects | Participants' country of residence | Psychological measure | Availability status |
|---|---|---|---|---|---|---|
| eSEE-d (Skaramagkas et al., 2023) | 2023 | Video | 48 | Greece | Arousal, valence, 4-word differential emotions scale | Available for commercial and non-commercial research |
| VREED (Tabbaa et al., 2021) | 2021 | VR | 34 | UK | Arousal, valence | Available for non-commercial research |
| ForDigitStress (Heimerl et al., 2023) | 2023 | Job Interview | 40 | Germany | Stress and occurred emotions (e. g. shame, anger, anxiety, surprise) | Not available |
| EyeT4Empathy (Lencastre et al., 2022) | 2022 | Structureless images, gaze typing | 60 | Norway | Empathy | Available for commercial and non-commercial research |

In the literature, various theories of emotions are used to distinguish affective states in data. These theories can be categorized as either discrete, where emotions are defined as limited categories, or dimensional, where emotions exist in continuum among 2 or more qualities. Dimensional models could overcome some of the challenges imposed by discrete models and introduce a more personalized approach to emotional classification (Thanapattheerakul, Mao, Amoranto, & Chan, 2018). One of the models widely employed in emotion detection is Pleasure-Arousal-Dominance (*PAD*) (Mehrabian, 1996), which covers three dimensional spaces: valence, representing positive, negative, or neutral states; arousal, measuring the level of physiological alertness; dominance, representing inner or outer source of emotion. Another commonly used theory, the circumplex theory of Russell (Posner, Russell, & Peterson, 2005), allows categorizing complex emotions, using a simpler, two-dimensional (valence and arousal) quadrant space. In our research, we decided to follow Russel's model and determined to obtain a distribution of affective states covering the maximum of circumplex space.

Thus, we present an initial step of the EmoEye project — EmoEye dataset of affective psychophysiological reactions, which will include behavioral measures of affect according to Russel's model of emotions and eye-tracking and biometrics data.

## Methods

In studies devoted to the development of emotion recognition algorithms, the following design is most often used:

- signal recording and data collection;
- identification of the main features;
- automated classification checking, using Machine Learning and / or Deep Learning algorithms.

In this study, a similar experimental plan was applied.

We collected 200 recordings of affective physiological reactions to the images. Participants (147 women, 49 men, 4 non-binary individuals; 27.46 ± 11.45 years old) were viewing pictures and assessing them for the period from 40 minutes to 1 hour. We collected eye movements, GSR and heart rate by the Gazepoint GP3 eye-tracker with Biometrics kit, with a sampling rate of 60 Hz. Participants proceeded to the main task (picture assessment) after completing 9-point calibration.

We used the *OASIS* (Open Affective Standardized Image Set) (Kurdi, Lozano, & Banaji, 2017) as a basis for our stimuli. OASIS stimuli base overcomes many faults typical for this kind of stimuli sets: it contains many images which cover various topics (objects, people, scenes and animals) and fall into all parts of the vector space (including moderate and strong arousal, neutral values), the evaluation of images in the OASIS dataset involved a broad sample of adult participants and not just a narrow circle of respondents. OASIS also has no copyright restrictions.

Each picture from OASIS was shown for 10 seconds, followed by a non-verbal *SAM* (Self-Assessment Manikin) questionnaire (Bradley & Lang, 1994), timing on which was controlled by the participant. The scores assigned in the SAM scale can easily be translated into the coordinate system of Russell's theory of emotions. Thus, having a certain score on this scale for any unit of content, we can determine the exact emotional state by correlating the valence and arousal scores with the areas of the circle. The SAM questionnaire has variations with 5, 7, and 9-point scales, and we used a 7-point scale to match the results with OASIS scores. This questionnaire was chosen due to its nonverbal nature and short format, as many domestic and foreign questionnaires have lengthy descriptions of questions and are designed to assess long-term emotional states, which was not our goal.

We obtained 9 variables: fixation coordinates (*X* and *Y*, as a fraction of screen size from 0.00 to 1.00), start of the fixation (*s*), fixation duration (*s*), fixation *ID*, left pupil diameter (*mm*), right pupil diameter (*mm*), galvanic skin response (*k*Ω), heart rate (*bpm*) from which we predicted arousal and valence dimensions, ranging from 1 to 7. In order to omit class imbalance, we separated two dimensions in separate datasets — one balanced for arousal and another balanced for valence — and trained two classifiers for arousal and valence respectively. Each of the classifiers solves the problem of classifying into 7 classes.

To preserve the temporal structure of the data, missing parts were linearly interpolated (*Fig. 1*). More complex and accurate interpolation methods (such as spline interpolation) in our case could behave unpredictably due to heavy noise.

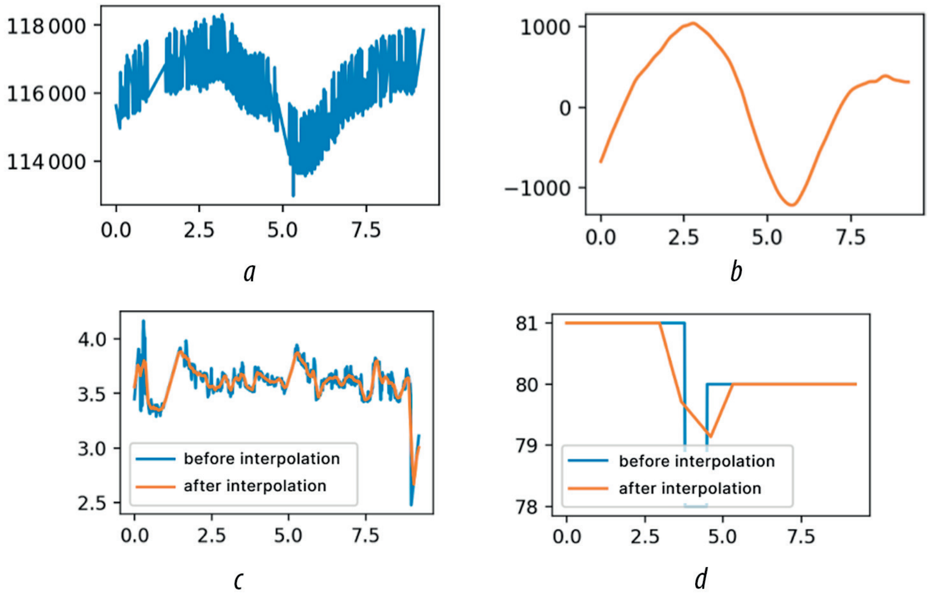


*Fig. 1.* An example of linear interpolation: *a* — GSR before linear interpolation; *b* — GSR after linear interpolation; *c* — pupil size before and after linear interpolation; *d* — heart rate before and after linear interpolation

From the tabular data, we limited our feature number to 5 variables: fixation coordinates (*X* and *Y*), averaged pupil diameter, GSR, and HR, as duration and start of fixation data did not show any valuable contribution for the emotion detection. We did not include saccades in our analysis as their correct recording is below our eye-tracker capabilities. Moreover, inclusion of saccade metrics, namely duration and magnitude, further showed no enhancement to the predictive ability of the probed deep learning model. Left and right pupil diameters were averaged and combined, in line with other similar studies not differentiating between left and right pupils in emotion detection tasks (Ren, Barreto, Gao & Adjouadi, 2012). Fixation Ids was kept in analysis to draw fixation and saccade series without directly feeding it into the classifier.

Left and right eye pupil diameters were interpolated, averaged and smoothed with a moving window of 10 time points to remove the noise. GSR signal was linearly interpolated, smoothed with a moving window of 100 time points to ensure that the bends do not have too much impact on the filtration, and filtered with a 4th order Butterworth bandpass filter in the range of 0.1–3 *Hz*. *HR* was smoothed with a moving window of 100 time points.

Preprocessed record was cut into "epochs" of about 5 seconds each (300 time points) immediately after the picture appeared on the screen. As a result, we obtained epochs of dimension 300 × 5 (300 time points, 5 variables). Data was normalized by calculating the mean and standard deviation for all people and using them in normalization of all epochs to ensure the quality of classification.

The final model inputs three types of data: (a) a picture with scan paths (190 × 190), (b) a time series (300 × 5), and (c) an image that was shown to a respondent (200 × 250 × 3). The scan paths were analyzed using convolutional neural network architecture — convolutional (5 × 5 convolutional core) and maxpooling (2 × 2) layers with added dropouts ($p = .5$) and batch-normalizations.

Time series were analyzed by recurrent neural network architecture. It first produces two-time convolutions and then after maxpooling and Flatten layer LSTM layer is applied. The output of the LSTM layer is directed to the full-linked neuron layer and the batch-normalization layer. The layers are interspersed with dropout layers.

Images were analyzed using the pre-trained VGG16 neural network. The SlicingOpLambda and TFOpLambda layers were responsible for image preprocessing for VGG16. The weights of VGG16 are frozen and do not change during training. The GlobalAveragePooling2D layer computes the average values for each of the channels of the final VGG16 convolution. The results of this layer were forwarded to the fully connected layers and the batch-normalization layer. Dropouts were also involved.

The outputs from the last layers of each of the three neural network parts are merged (concatenated) into one. Then there are several fully connected layers with dropouts and a final layer giving out seven probabilities of belonging to each class (softmax activation, the sum of probabilities is reduced to 1).

## Results

The average scores for valence, arousal were calculated for each image (*Fig. 2*, right). The scores on the valence scale for different image categories were not statistically significantly different from the standard scores of the corresponding images for the original stimulus base. However, each image category has lower scores on the arousal scale in our dataset (*Wilcoxon test* for paired samples with Bonferroni multiple comparisons correction, $p < .0001$) with on average .84 point higher assessments on the arousal scale for the standard scores in the original OASIS dataset than for assessments collected from our participants. Nevertheless, the general tendencies between categories did not differ, for example, animal category was evaluated as having highest points on the arousal scale in both standard scores and assessments collected from our participants, and object category was evaluated as having the lowest arousal scores in both assessments. The general order for categories on the arousal scale from lowest scores to highest scores is the following — object, person, scene, animal (*Manna-Whitney test* applied for neighboring combinations of categories, $p < .01$).

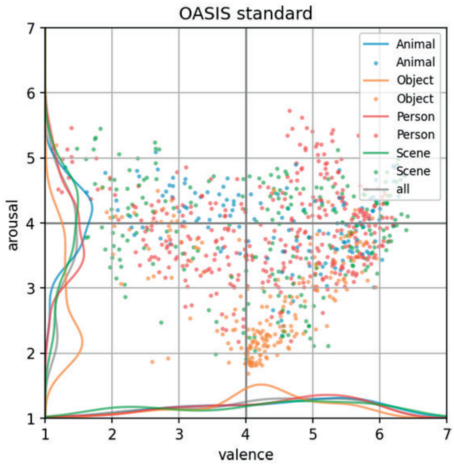


*Fig. 2.* Average scores for all pictures from different categories in the original stimulus base (left) and in the chosen subset of images used in our research (right). Each point represents the average score for a picture, color of the point represents the picture category

For the neural network training, we performed a downsampling procedure to equalize the number of samples from different classes. The resulting downsampled dataset was divided into training and testing sets (75 / 25 split). The resulting average accuracies for 7-class arousal and valence prediction were 29.5 % and 27.5 % with the higher accuracies up to 50 % for the boundary cases (*Fig. 3*). Note that random guess accuracy is 14.3 %.

## Discussion

In our first stage of EmoEye project we collected a corpus of GSR, pulse, and oculomotor responses (fixation coordinates; fixation duration; mean pupil size for the right and left eyes) data that can be applied to emotion detection research.

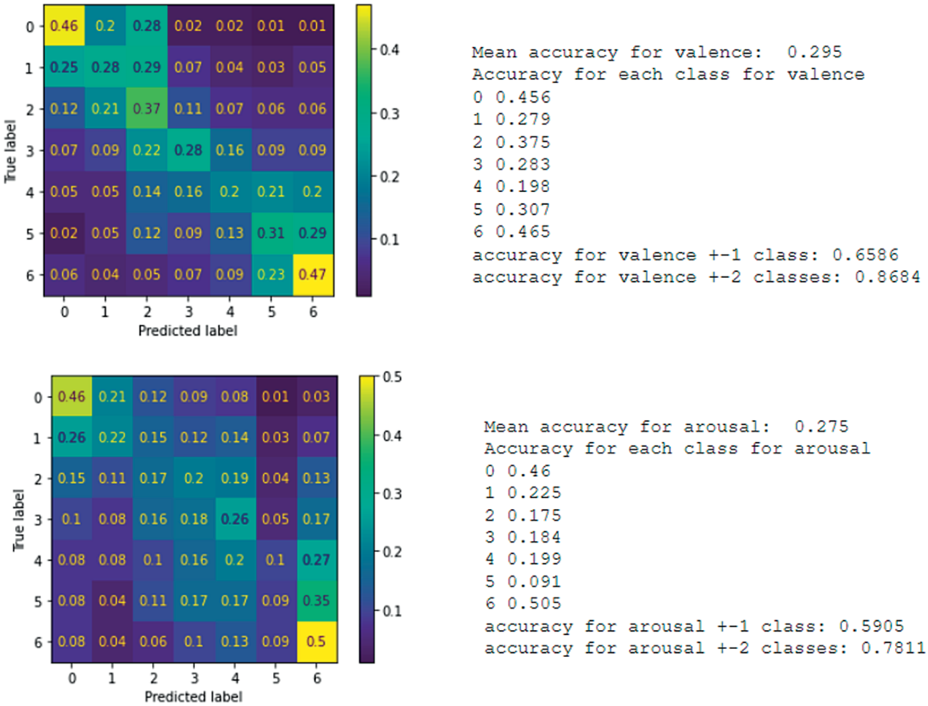


*Fig. 3.* Confusion matrices with the accuracies on the testing set for the valence (top) and arousal scales (bottom)

The distribution of subjects' behavioral responses to the images did not differ much from the OASIS range of valence and arousal ratings and preserved the tendencies shown in the original stimuli base. For instance, the category of objects was assessed as neutral (having medium scores in valence and low in arousal) more often than the other categories, which is also true for the OASIS. Thus, our recordings could be divided into 3 separate groups: negative, neutral, and positive reactions, making it suitable for emotion detection tasks. However, we noticed that Russian participants were more constricted in their evaluations of the pictures. It was less likely that the picture would have been given a maximum value on the scale (1 or 7). This decrease in extremity of the answers could be explained as, firstly, for SAM assessment as a separate instance our sample size was smaller than the typical number of responses for machine learning data (Brereton, 2006; Kyriazos, 2018), secondly, Russian people tend to suppress their emotional expression (Wierzbicka, 1998), which could also affect the evaluation process. Our "average" respondent also differed from the Amazon MTurk volunteers in OASIS who were white, highly educated, Christian males and mean age was 36.63 years (*SD* = 11.91) (Kurdi et al., 2017). Most of our sample consisted of younger (mean age — 27.46) Christian or Atheist white female students with graduate or undergraduate degrees. Nevertheless, we obtained similar results for valence and arousal scores with different demographics, which ensures the applicability of our research.

Multimodal deep learning model created by us was able to achieve accuracy several times higher than the random guess. The prediction accuracies were higher for the boundary cases that seems reasonably correct as the boundary cases should produce more prominent physiologic response. We suppose that the model may be improved in the future with fine-tuning of architecture and fine adjustment of the training procedure but obtained accuracies already show that the data itself are meaningful and have potential for exploratory and commercial purposes.

One of our main limitations is that a significant percentage of data in the records was subjected to interpolation due to heavy noise. On average, ~41 % (from 17 to 87 %) of rows were missing in the tables. In most cases, rows were skipped during saccades. Overall, heart rate turned out to be the most problematic of all variables — almost all subjects in this variable had many outliers, but it also happened that this variable was almost unchanged, showing one number for almost the entire record. The second most problematic variable was GSR — it also often had many spiking values that exceeded the mean by several orders of magnitude.

In addition, the gender distribution of the database was predominantly female. As it could impose some issues regarding facial stimuli (Krumhuber, Skora, Küster, & Fou, 2016), it is not uncommon in eye-tracking laboratory studies (Alshehri & Alghowinem, 2013).

## Conclusions

Creating a specific emotion is a primary goal of any advertisement. Traditional marketing and neuromarketing companies seek nuanced and cutting-edge solutions to detect and assess emotional states on cognitive, behavioral, and neural levels.

In this paper, we present our own psychophysiological database for automatic emotion detection probed by a deep-learning algorithm. By creating a novel database with Russian participants, we tried to ensure cultural sensitivity of the future analysis and provide representation of Slavic cultures in the machine learning field.

Our database showed the same tendencies for valence and arousal assessments as in the OASIS, dividing the sample by negative, positive and neutral reactions with high and low arousal.

Final algorithm for automatic classification was based on Convolutional (CNN) and Long Short-term memory (LSTM) layers using three types of inputs — scanpaths, GSR and HR signal values and images. Maximum accuracy for 3 classes was 87 % for valence dimension and 78 % for arousal.

We propose affective psychophysiological dataset for deep learning emotion detection tasks. We tried to ensure generalizability and sufficiency of our data by collecting the largest by our estimation sample of affective psychophysiological reactions.

This study faced some limitations, including an unequal number of men and women in the sample, noisy data and inefficiency of some of the OASIS images to elicit emotional response due to cultural differences. In the future work, we hope to emphasize the cultural sensitivity of the analysis by including culturally relevant images of Russian nature and culture.

## References

Abadi, M. Kh., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing, 6*(3), 209–222. https://doi.org/10.1109/TAFFC.2015.2392932

Alshehri, M., & Alghowinem, Sh. (2013). An exploratory study of detecting emotion states using eye-tracking technology. *Science and Information Conference* (pp. 428–433). London, UK: IEEE.

Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscienc*e, 11(4), 284–292. https://doi.org/10.1038/nrn2795

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005–7916(94)90063–9

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602–607. https://doi.org/10.1111/j.1469–8986.2008.00654.x

Brereton, R. G. (2006). Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TrAC Trends in Analytical Chemistry, 25*(11), 1103–1111. https://doi.org/10.1016/j.trac.2006.10.005

Fisher, R. J. (2000). The future of social-desirability bias research in marketing. *Psychology and Marketing, 17*(2), 73–77. https://doi.org/10.1002/(SICI)1520–6793(200002)17:2<73:: AID-MAR1>3.0.CO;2-L

Flykt, A. (2005). Visual search with biological threat stimuli: Accuracy, reaction times, and heart rate changes. *Emotion,* 5(3), 349–353. https://doi.org/10.1037/1528–3542.5.3.349

Heimerl, A., Prajod, P., Mertes, S., Baur, T., Kraus, M., Liu, A., … Becker, L. (2023). ForDigitStress: A multi-modal stress dataset employing a digital job interview scenario [arXiv: 2303.07742]. *arXiv*. https://doi.org/10.48550/arXiv.2303.07742

Jordão, I. L. D. S., De Souza, M. T., De Oliveira, J. H. C., & Giraldi, J. D. M. E. (2017). Neuromarketing applied to consumer behaviour: An integrative literature review between 2010 and 2015. *International Journal of Business Forecasting and Marketing Intelligence, 3*(3), 270–288. https://doi.org/10.1504/ijbfmi.2017.085371

Katsigiannis, S., & Ramzan, N. (2018). 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, 22*(1), 98–107. https://doi.org/10.1109/JBHI.2017.2688239

Koelstra, S., Muhl, Ch., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., … Patras, I. (2012). DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing,* 3(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2016). A review of dynamic datasets for facial expression research. *Emotion Review,* 9(3), 280–292. https://doi.org/10.1177/1754073916670022

Kurdi, B., Lozano, Sh., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods, 49*(2), 457–470. https://doi.org/10.3758/S13428-016-0715-3

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology, 9*(8), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research, 61*(5), 534–547. https://doi.org/10.1177/1470785318805305

Lencastre, P., Bhurtel, S., Yazidi, A., e Mello, G. B. M., Denysov, S., & Lind, P. G. (2022). EyeT4Empathy: Dataset of foraging for visual information, gaze typing and empathy assessment. *Scientific Data,* 9, 1–8. https://doi.org/10.1038/s41597-022-01862-w

Lim, J. Z., Mountstephens, J., & Teo, J. (2020). Emotion recognition using eye-tracking: Taxonomy, review and current challenges. *Sensors, 20*(8), 2384. https://doi.org/10.3390/s20082384

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology,* 14, 261–292. https://doi.org/10.1007/BF02686918

Miranda-Correa, J. A., Abadi, M. Kh., Sebe, N., & Patras, I. (2021). AMIGOS: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing, 12*(2), 479–493. https://doi.org/10.1109/TAFFC.2018.2884461

Ouazzani Touhami, Z., Benlafkih, L., Jiddane, M., Cherrah, Y., El Malki, H. O., & Benomar, A. (2011). Neuromarketing: Where marketing and neuroscience meet. *Revue Neurologique, 167*(2), 135–140. https://doi.org/10.1016/j.neurol.2010.07.025

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns, 2*(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Picard, R. W. (1995). Affective computing. *M. I. T Media Laboratory Perceptual Computing Section Technical Report,* 321, 1–16. Retrieved from https://affect.media.mit.edu/pdfs/95.picard.pdf

Pogosyan, M., & Engelmann, J. B. (2011). Cultural differences in affect intensity perception in the context of advertising. *Frontiers in Psychology, 2*, 313. https://doi.org/10.3389/fpsyg.2011.00313

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Ren, P., Barreto, A., Gao, Y., & Adjouadi, M. (2012). Comparison of the use of pupil diameter and galvanic skin response signals for affective assessment of computer users. *Biomedical Sciences Instrumentation, 48*, 345–350.

Roux, P., Brunet-Gouet, E., Passerieux, C., & Ramus, F. (2016). Eye-tracking reveals a slowdown of social context processing during intention attribution in patients with schizophrenia. *Journal of Psychiatry and Neuroscience, 41*(2), E13–E21. https://doi.org/10.1503/jpn.150045

Skaramagkas, V., Ktistakis, E., Manousos, D., Kazantzaki, E., Tachos, N. S., Tripoliti, E., … Tsiknakis, M. (2023). eSEE-d: Emotional State Estimation Based on Eye-Tracking Dataset. *Brain Sciences, 13*(4), 589. https://doi.org/10.3390/brainsci13040589

Tabbaa, L., Searle, R., Bafti, S. M., Hossain, M. M., Intarasisrisawat, J., Glancy, M., & Ang, Ch. S. (2021). VREED: Virtual Reality Emotion Recognition Dataset using eye tracking and physiological measures. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(4), 1–20. https://doi.org/10.1145/3495002

Thanapattheerakul, Th., Mao, K., Amoranto, J., & Chan, J. H. (2018). Emotion in a century: A review of emotion recognition. *IAIT 2018: Proceedings of the 10th International Conference on Advances in Information Technology*, 17, 1–8. http://dx.doi.org/10.1145/3291280.3291788

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE, 14*(11), e0224365. https://doi.org/10.1371/JOURNAL.PONE.0224365

Wierzbicka, A. (1988). Russian Emotional Expression. *Ethos,* 26(4), 456–483. Retrieved from 10.1525/eth.1998.26.4.456

Wu, G., Liu, G., & Hao, M. (2010). The analysis of emotion recognition from GSR based on PSO. *International Symposium on Intelligence Information Processing and Trusted Computing* (pp. 360–363). Huanggang: IEEE. https://doi.org/10.1109/IPTC.2010.60

*Original manuscript received February 01, 2023*
*Revised manuscript accepted February 20, 2023*

**About the authors:**

**Ivanina Ekaterina O.,** Head of Research, Neuro laboratory "NeuroLab", Moscow, Russia; https://orcid.org/0000-0002-4488-3015; ekaterina@leo-neurolab.ru

**Tokmovtseva Anna D.,** Research Intern, Neuro laboratory "NeuroLab"; Student, Russian Orthodox University of Saint John the Divine, Moscow, Russia; profiguit@gmail.com

**Akelieva Elizaveta V.,** Research Intern, Neuro laboratory "NeuroLab"; Graduate Student, Department of Philosophy, Lomonosov Moscow State University, Moscow, Russia; akelieva_elizaveta@mail.ru

**Об авторах:**

**Иванина Екатерина Олеговна,** главный психолог, нейролаборатория «НейроЛаб», Москва, Россия; https://orcid.org/0000-0002-4488-3015; ekaterina@leo-neurolab.ru

**Токмовцева Анна Денисовна,** ассистент психолога, нейролаборатория «НейроЛаб»; студент, Российский православный университет Святого Иоанна Богослова, Москва, Россия; profiguit@gmail.com

**Акельева Елизавета Владимировна,** ассистент психолога, нейролаборатория «НейроЛаб»; магистр, факультет философии, Московский государственный университет имени М. В. Ломоносова, Москва, Россия; akelieva_elizaveta@mail.ru