Traumatology and Orthopaedics of Kazakhstan
ISSN - e2789-9640

Original article

EEG-Based Acute Stress Detection for Post-Trauma Rehabilitation and Biofeedback: Limits of Cross-Dataset Generalisation

Anar Mengdigali [1*], Medet Mukushev [2], Beibit Abdikenov [3], Almira Kustubayeva [4], Altyngul Kamzanova [5], Manzura Zholdassova [6]

Received: August 12, 2026
Revised: September 17, 2026
Accepted: September 29, 2026
Published: October 28, 2026

Citation: Anar Mengdigali, Medet Mukushev, Beibit Abdikenov, Almira Kustubayeva, Altyngul Kamzanova, Manzura Zholdassova. EEG-based acute stress detection for post-trauma rehabilitation and biofeedback: limits of cross-dataset generalisation. Trauma & Ortho Kaz, 2026, 77 (5), jto066. https://doi.org/10.52889/1684-9280-2026-77-5-jto066

This work is licensed under a Creative Commons Attribution 4.0 International License

[1] Researcher, School of Artificial Intelligence and Data Science, Astana IT University, Astana, Kazakhstan.
E-mail: 242886@astanait.edu.kz
[2] Assistant professor, School of Creative Industry, Astana IT University, Astana, Kazakhstan.
E-mail: m.mukusev@astanait.edu.kz
[3] Director, Science and Innovation Center "Artificial Intelligence", Astana IT University, Astana, Kazakhstan.
E-mail: beibit.abdikenov@astanait.edu.kz
[4] Director, Brain Institute, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
E-mail: almkusto@kaznu.kz
[5] Postdoctoral Researcher, Brain Institute, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
E-mail: kamzanova.altyngul@kaznu.kz
[6] Researcher, Brain Institute, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
E-mail: manzur777@gmail.com

*** Corresponding author:** 242886@astanait.edu.kz

Abstract

Psychological stress after injury is frequent in trauma patients and is associated with poorer recovery. EEG-based biofeedback could help monitor and reduce it during rehabilitation, but only if the stress detector works for new patients and new hardware. Published EEG classifiers of acute stress are usually trained and tested on a single dataset, and accuracies above 90% are common. We asked how much of this performance holds for participants the model has never seen, and for a dataset recorded with other hardware and another stress protocol.

Methods. Two open datasets of acute cognitive stress in healthy volunteers were processed with one pipeline: SAM40 (40 participants; Stroop, arithmetic and mirror-image tasks versus relaxation; 32-channel Emotiv, 128 Hz) and EEGMAT (36 participants; serial subtraction versus rest; 19-channel clinical system, 500 Hz). Four compact classifiers (logistic regression, random forest and RBF-SVM on band power, and a Riemannian tangent-space classifier) were evaluated with leave-one-subject-out (LOSO) validation, with cross-corpus transfer in both directions on ten shared electrodes, and with two unsupervised alignment methods, Riemannian re-centring and CORAL.

Results. Within SAM40, the best LOSO macro-F1 for one stressor against relaxation was 70.3% (best AUROC 0.804). Cross-corpus transfer lowered macro-F1 significantly in 22 of 24 model–direction–stressor settings (paired Wilcoxon test, $p < 0.01$), by up to 37.6 percentage points. Transfer was asymmetric: classifiers trained on EEGMAT reached an AUROC below 0.5 on SAM40 in 9 of 12 settings, so the learned decision direction was reversed.

Conclusions. Stress labels defined by the task carry protocol-specific information that does not survive a change of dataset. Cross-corpus testing should come before any claim that an EEG stress detector is ready for post-trauma rehabilitation or biofeedback use; clinical application will require per-user calibration and validation in patient populations.

Keywords: post-traumatic stress, trauma rehabilitation, biofeedback, EEG, acute psychological stress, cross-corpus generalisation.

1. Introduction

Prolonged psychological stress contributes to cardiovascular, metabolic and mental disorders [1]. After physical injury the problem is especially visible: a meta-analysis of 27 studies found post-traumatic stress disorder (PTSD) in about one in four patients after acute injury of the limbs or pelvis [2], and in a cohort of 580 injured patients half met screening criteria for PTSD about a year after injury [3]. EEG neurofeedback, a form of biofeedback, reduced PTSD symptoms in a randomised controlled trial [4], which makes EEG-based stress monitoring a candidate tool for post-trauma care. Interventions that respond to stress as it happens, for example biofeedback-guided relaxation, need a measurement of the user's state that is objective and continuous. EEG is a natural candidate: it records cortical activity directly, with millisecond resolution, and wearable headsets have made it inexpensive [5].

Published EEG stress classifiers look impressive. Many report accuracies above 90% [5], and on the SAM40 dataset a feature-selection pipeline with an LSTM classifier reached 99.67% [6]. Nearly all such figures come from one dataset, and protocols in which epochs of the same participant or recording fall into both the training and the test set are still common. Protocols of this kind reward models that memorise participant- and session-specific properties of the signal, a problem documented across EEG decoding [7, 8] and rooted in the large variability of EEG between individuals [9]. In our earlier study of EEG-based fatigue, strict leave-one-subject-out (LOSO) validation held the best compact network to 74.0 ± 11.3% accuracy [10].

A detector meant for practical use faces a harder test than a new participant: a new headset, a different montage and a different way of inducing stress. In emotion recognition this cross-corpus setting is well studied. Models trained on SEED [11] and tested on DEAP [12], or the reverse, lose much of their accuracy, and domain adaptation has been compared systematically on this pair [13]. For acute psychological stress the picture is thinner. Agnelli et al. [14] compared stress detection across tasks and subjects on two datasets, and Vos et al. [15] applied a stress model trained on EEGMAT to SAM40 as part of a study on synthetic EEG. We are not aware of a benchmark that combines subject-independent validation, transfer in both directions between stress corpora, and domain adaptation.

Domain adaptation methods for EEG fall into a few families. Statistical alignment includes Riemannian re-centring of covariance matrices [16], which builds on Riemannian classifiers for brain–computer interfaces [17], and correlation alignment (CORAL) [18]. Adversarial methods learn features that a domain discriminator cannot separate [19]. Alignment methods are attractive for a first benchmark because they need no labels from the target dataset and add almost no computational cost.

The aim of this study was to determine whether EEG classifiers of acute stress keep their performance on new participants and on a new dataset, as a prerequisite for their use in stress monitoring and biofeedback during post-trauma rehabilitation. To this end, we report a benchmark for two open stress datasets, SAM40 and EEGMAT. We apply one preprocessing pipeline to both, evaluate four compact classifiers at three levels (within-corpus LOSO, cross-corpus transfer, and transfer with unsupervised alignment) and examine what the classifiers rely on. We test four hypotheses:

H1 Cross-corpus performance is significantly lower than within-corpus LOSO performance.

H2 Transfer is better when source and target use the same type of stressor.

H3 Unsupervised alignment recovers part of the cross-corpus loss.

H4 Feature importance shows whether a classifier depends on cortical activity or on artefacts such as eye movements.

2. Materials and methods

Datasets

SAM40 [20] contains recordings from 40 participants made with a 32-channel Emotiv Epoc Flex headset at 128 Hz. Each participant performed a Stroop colour–word test, a mental arithmetic task and a mirror-image recognition task, and also sat through a relaxation condition; each condition lasted 25 s and was repeated three times. We used the raw recordings rather than the authors' cleaned versions so that both corpora passed through exactly the same processing. Relaxation was labelled 0 and every stressor 1.

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

EEGMAT [21], distributed through PhysioNet [22], contains recordings from 36 participants made with a clinical 19-channel system (10–20 montage, ear reference) at 500 Hz. Each participant has a 182-s segment of rest recorded before the task and a 62-s segment recorded during serial subtraction. Rest was labelled 0 and arithmetic 1. Table 1 summarises both corpora. Both corpora were recorded in healthy volunteers exposed to laboratory cognitive stressors, and no trauma patients were included. We are not aware of open EEG stress datasets recorded in trauma patients; a detector that does not transfer between two healthy cohorts is unlikely to work in a clinical one, so this comparison is a necessary first step.

Table 1 - The two corpora after harmonisation. Epochs are non-overlapping 3-s windows

	SAM40	EEGMAT
Participants	40	36
Device, channels	Emotiv Epoc Flex, 32	Clinical system, 19
Sampling rate	128 Hz	500 Hz (resampled to 128 Hz)
Stressor	Stroop, arithmetic, mirror image	Serial subtraction
Control condition	Relaxation, 3 × 25 s	Rest, 182 s
Epochs (control / stress)	960 / 2880 (960 per stressor)	2124 / 720

Preprocessing

The pipeline follows our earlier study [10] and was implemented with MNE-Python [23]. For each recording we (i) selected channels, (ii) applied a common average reference, (iii) band-pass filtered between 1 and 50 Hz with a zero-phase fourth-order Butterworth filter, (iv) removed 50-Hz line noise with a second-order Butterworth notch, (v) cut the signal into non-overlapping 3-s epochs (384 samples at 128 Hz) and (vi) z-scored each channel with the mean and standard deviation of that participant's epochs. Filtering was applied to the continuous recording, before epoching, to avoid edge effects. Step (vi) uses no labels and is applied in the same way to training and test participants. EEGMAT was resampled to 128 Hz with an anti-aliasing filter before step (iii).

For cross-corpus experiments both datasets were reduced to ten electrodes present in both montages and on 14-channel Emotiv EPOC headsets: F3, F4, F7, F8, T7, T8, P7, P8, O1 and O2 (EEGMAT labels T3, T4, T5 and T6 were mapped to T7, T8, P7 and P8). Channels were selected before re-referencing, so the average reference was computed over the same ten sites in both corpora. Within-corpus SAM40 analyses used all 32 channels.

Features and classifiers

Band power was computed from the Welch spectrum (256-sample Hann windows, 50% overlap), log-transformed and averaged within the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (13–30 Hz) bands, giving four features per channel. Three classifiers used these features: logistic regression, a random forest with 300 trees and an RBF support vector machine (C = 3). All three used balanced class weights and a standard scaler fitted on the training data only. The fourth classifier worked on the signal itself: an OAS-shrinkage covariance matrix was estimated for each epoch, projected to the log-Euclidean tangent space and passed to logistic regression [17]. Shrinkage is needed because the average reference makes the sample covariance rank-deficient.

Unsupervised domain alignment

Riemannian re-centring (RA). For participant s with Riemannian mean covariance Ms, every epoch covariance C is replaced by

$$\mathbf{C}' = \mathbf{M}_s^{-1/2}\,\mathbf{C}\,\mathbf{M}_s^{-1/2}, \qquad (1)$$

which moves each participant, and therefore each corpus, to the identity matrix [16]. The re-centred matrices were projected to the tangent space at their Riemannian mean, which is close to the identity, and classified with logistic regression. No labels are involved.

CORAL. Band-power features were first standardised within each corpus. Source features Xs were then whitened with their own covariance and re-coloured with the covariance of the unlabelled target features [18],

$$\mathbf{X}_s^{\sim} = \mathbf{X}_s\,(\mathbf{C}_s + \mathbf{I})^{-1/2}\,(\mathbf{C}_t + \mathbf{I})^{1/2}, \qquad (2)$$

after which the classifier was trained on $\tilde{X}$s and applied to the target.

Evaluation

Figure 1 shows the three evaluation levels. At level 1 (within-corpus LOSO) a model was trained on all but one participant and tested on the remaining one. At level 2 (cross-corpus) a model was trained on every participant of corpus A and tested on each participant of corpus B, which played no part in training. Level 3

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

repeated level 2 with an alignment step that could see unlabelled data from corpus B.

We report macro-F1 as the primary metric because it is insensitive to the 1:3 and 3:1 class ratios, together with Matthews correlation coefficient (MCC), the area under the ROC curve (AUROC) and accuracy. All metrics were computed per test participant and summarised as mean ± standard deviation.

The generalisation gap is the level-1 macro-F1 on corpus B minus the level-2 macro-F1 on B. Because both values are available for each participant of B, the gap was tested with a paired Wilcoxon signed-rank test. The same test compared aligned and unaligned models. The random seed was fixed at 42.

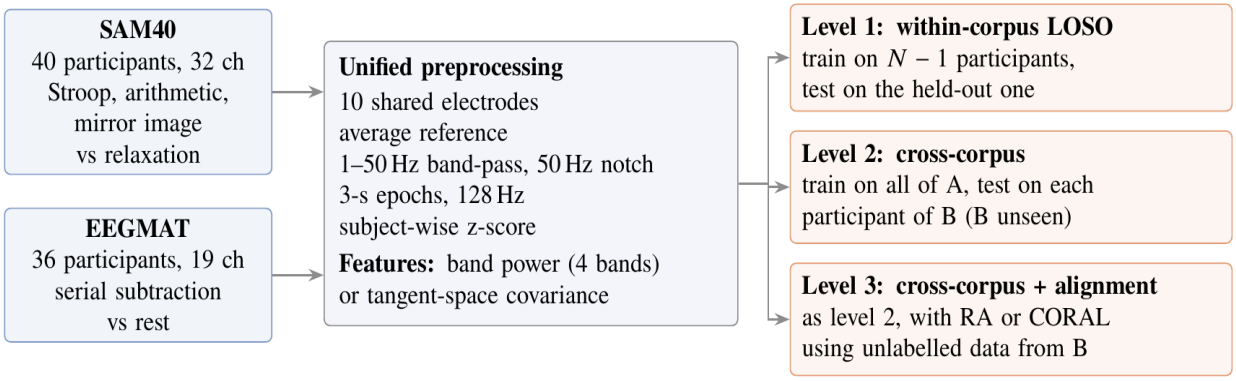


Figure 1 - Study design. Both corpora pass through the same pipeline. Generalisation is measured at three levels, and the generalisation gap is the difference between levels 1 and 2 on the same target participants

Feature importance and ocular control

To see which features the classifiers used, a random forest with 500 trees was fitted to all SAM40 epochs of one contrast, and impurity-based importances were summed over bands (channel importance) and over channels (band importance). As a control for eye movements, the all-stressor LOSO analysis was repeated after independent component analysis: for each participant, FastICA with 15 components was fitted to the filtered recordings, and components correlated with the frontopolar channels Fp1 and Fp2 were removed before epoching.

3. Results

Within-corpus performance on SAM40

Table 2 gives the LOSO results on SAM40 with all 32 channels. When each stressor was contrasted with relaxation, so that the classes were balanced, the best macro-F1 ranged from 68.1% (arithmetic) to 70.3% (Stroop), and the best AUROC reached 0.804. The three stressors were therefore separated from relaxation about equally well. Pooling all stressors created a 1:3 class ratio, and accuracy became misleading: the random forest reached 75.7% accuracy, the proportion of the majority class, with a macro-F1 of only 51.3% because it labelled almost every epoch as stress. Performance varied strongly between participants, with per-participant macro-F1 between 0.29 and 0.94.

Table 2 - Within-corpus LOSO on SAM40 (32 channels, 40 folds). Macro-F1 in % (mean ± SD over participants), AUROC in parentheses. Bold: best macro-F1 per row

Contrast	LogReg	Random forest	RBF-SVM	Riemannian
Relax vs all stressors (1:3)	59.9 ± 10.9 (0.734)	51.3 ± 8.3 (0.740)	**61.2 ± 9.9** (0.724)	58.9 ± 12.3 (0.732)
Relax vs arithmetic	64.2 ± 12.0 (0.756)	**68.1 ± 10.6** (0.754)	64.5 ± 9.9 (0.743)	62.2 ± 12.1 (0.748)
Relax vs Stroop	67.5 ± 10.5 (0.783)	**70.3 ± 9.3** (0.785)	67.6 ± 11.7 (0.781)	69.3 ± 11.9 (0.804)
Relax vs mirror image	63.3 ± 12.4 (0.748)	**69.4 ± 8.9** (0.764)	66.3 ± 11.7 (0.760)	63.5 ± 12.6 (0.750)

Abbreviations: AUROC, area under the receiver operating characteristic curve; LogReg, logistic regression; LOSO, leave-one-subject-out; RBF-SVM, support vector machine with radial basis function kernel; SD, standard deviation

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

Cross-corpus generalisation gap (H1)

Table 3 and Figure 2 compare levels 1 and 2 on the ten shared electrodes. Transfer lowered macro-F1 significantly in 22 of the 24 model–direction–stressor settings ($p<0.01$). The largest drop, 37.6 points, occurred for logistic regression trained on SAM40 arithmetic and tested on EEGMAT, where macro-F1 fell from 0.664 to 0.288. Both non-significant cases involved the random forest in the SAM40→EEGMAT direction, where its within-EEGMAT baseline was already weak (0.565). H1 is supported.

Table 3 - Macro-F1 on the target corpus (10 shared electrodes). SAM40→EEGMAT: cross-corpus values, with the within-EEGMAT LOSO baseline in the first row. EEGMAT→SAM40: cross-corpus / within-SAM40 LOSO for the same contrast

Direction	SAM40 contrast	LogReg	Random forest	RBF-SVM	Riemannian
SAM40→ EEGMAT	*within EEGMAT*	*0.664*	*0.565*	*0.674*	*0.690*
	Arithmetic	0.288	0.573	0.477	0.586
	Stroop	0.511	0.463	0.452	0.553
	Mirror image	0.364	0.552	0.479	0.390
	All (1:3)	0.445	0.282	0.439	0.498
EEGMAT→ SAM40	Arithmetic	0.507 / 0.634	0.333 / 0.662	0.380 / 0.642	0.381 / 0.655
	Stroop	0.383 / 0.671	0.333 / 0.696	0.339 / 0.673	0.400 / 0.711
	Mirror image	0.454 / 0.656	0.333 / 0.686	0.365 / 0.660	0.378 / 0.654
	All (1:3)	0.387 / –	0.200 / –	0.253 / –	0.319 / –

Abbreviations: LogReg, logistic regression; LOSO, leave-one-subject-out; RBF-SVM, support vector machine with radial basis function kernel

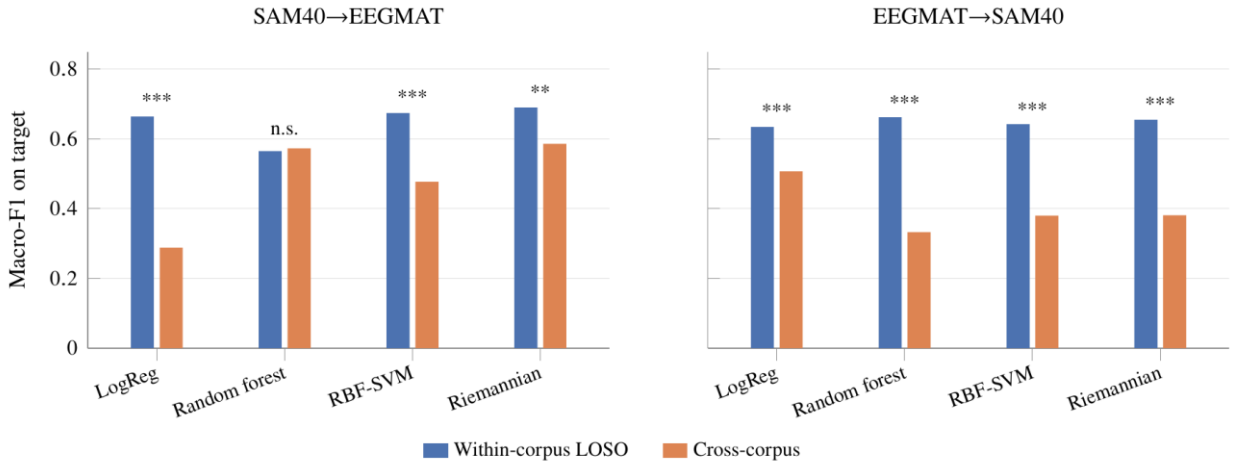


*Figure 2 - Generalisation gap for the shared stressor (SAM40 arithmetic vs relaxation, EEGMAT serial subtraction vs rest). Blue: LOSO within the target corpus; orange: model trained on the other corpus. Paired Wilcoxon test over target participants: ** $p<0.01$, *** $p<0.001$, n.s. not significant*

Asymmetric and inverted transfer

The two directions behaved differently (Figure 3). Models trained on SAM40 kept part of their ability to rank EEGMAT epochs, with AUROC between 0.64 and 0.69 for the random forest and the Riemannian classifier. Models trained on EEGMAT did not. On SAM40 their AUROC was below 0.5 in 9 of the 12 single-stressor settings and between 0.50 and 0.57 in the other three, and with all stressors pooled it was below 0.5 for every model (0.39–0.49). A classifier trained on EEGMAT therefore ranks SAM40 relaxation epochs as more stress-like than SAM40 task epochs. The EEGMAT-trained random forest, which saw three rest epochs for every task epoch, predicted rest for almost every SAM40 epoch (macro-F1 0.333 in all three contrasts). An AUROC consistently below chance cannot be explained by noise: the feature pattern that separates task from rest in one protocol is partly reversed in the other.

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

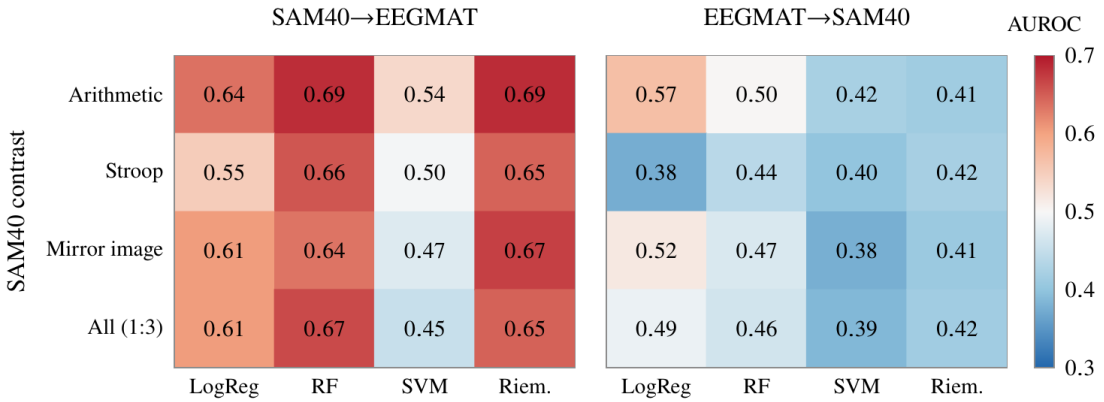


Figure 3 – Cross-corpus AUROC for every SAM40 contrast and classifier. Red cells are above chance (0.5); blue cells are below chance, i.e. the decision direction learned on the source corpus is reversed on the target

Stressor type (H2)

In the SAM40→EEGMAT direction the SAM40 arithmetic contrast, which uses the same kind of stressor as EEGMAT, gave the highest target AUROC for all four classifiers (Figure 3, first row of the left panel). The margin over the next-best stressor was small, between 0.01 and 0.04, and macro-F1 did not follow the same order because the decision threshold shifts between corpora. We therefore regard the result as consistent with H2 rather than as proof of it.

Unsupervised alignment (H3)

Table 4 and Figure 4 compare transfer with and without alignment. Two patterns are clear. First, re-centring helped within a corpus but not across corpora. It raised within-EEGMAT LOSO macro-F1 from 0.690 to 0.728 (AUROC 0.800) and within-SAM40 arithmetic macro-F1 from 0.655 to 0.668, yet it lowered cross-corpus AUROC in all four settings and left the EEGMAT→SAM40 AUROC below chance (0.375 and 0.377). Second, CORAL helped in the harder direction. For EEGMAT→SAM40 it improved the AUROC of every band-power model by 0.04–0.19, lifting all of them above 0.5 (0.57–0.62), with significant macro-F1 gains for the SVM and the random forest ($p<0.001$). In the SAM40→EEGMAT direction its effect was mixed: logistic regression gained 0.22 in macro-F1, largely through a better-placed threshold, while the random forest lost ground. H3 is only partly supported: alignment narrows the gap in some settings but closes it In none.

Table 4 – Cross-corpus macro-F1 / AUROC with and without unsupervised alignment. ΔF1: aligned minus unaligned macro-F1, with the paired Wilcoxon p-value in parentheses

	SAM40→EEGMAT		EEGMAT→SAM40	
Method	F1 / AUROC	ΔF1 (*p*)	F1 / AUROC	ΔF1 (*p*)
SAM40 contrast: arithmetic				
Riemannian	0.586 / 0.686		0.381 / 0.414	
+ RA	0.551 / 0.633	−0.034 (0.12)	0.396 / 0.375	+0.015 (0.67)
LogReg	0.288 / 0.637		0.507 / 0.569	
+ CORAL	0.509 / 0.631	**+0.221** (<0.001)	0.501 / 0.612	−0.006 (0.77)
RBF-SVM	0.477 / 0.538		0.380 / 0.422	
+ CORAL	0.511 / 0.569	+0.034 (0.31)	0.482 / 0.602	**+0.102** (<0.001)
Random forest	0.573 / 0.685		0.333 / 0.503	
+ CORAL	0.521 / 0.591	−0.051 (0.12)	0.389 / 0.622	**+0.056** (<0.001)
SAM40 contrast: all stressors (1:3)				
Riemannian	0.498 / 0.648		0.319 / 0.415	
+ RA	0.476 / 0.582	−0.022 (0.38)	0.366 / 0.377	+0.047 (0.006)
LogReg	0.445 / 0.606		0.387 / 0.488	

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

Table 4 (Continuation) – Cross-corpus macro-F1 / AUROC with and without unsupervised alignment. ΔF1: aligned minus unaligned macro-F1, with the paired Wilcoxon p-value in parentheses

	SAM40→EEGMAT		EEGMAT→SAM40	
Method	F1 / AUROC	ΔF1 (*p*)	F1 / AUROC	ΔF1 (*p*)
RBF-SVM	0.439 / 0.453		0.253 / 0.386	
+ CORAL	0.422 / 0.517	−0.017 (0.64)	0.367 / 0.571	+0.114 (<0.001)
Random forest	0.282 / 0.665		0.200 / 0.462	
+ CORAL	0.213 / 0.551	−0.069 (0.003)	0.265 / 0.583	+0.064 (<0.001)

Abbreviations: AUROC, area under the receiver operating characteristic curve; CORAL, correlation alignment; F1, macro-averaged F1 score; LogReg, logistic regression; RA, Riemannian re-centring; RBF-SVM, support vector machine with radial basis function kernel

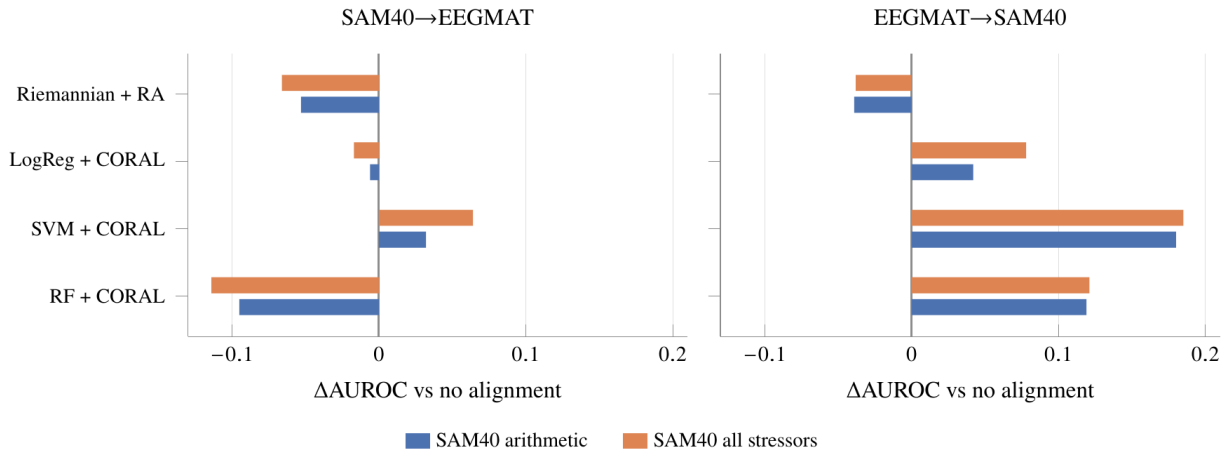


Figure 4 - Change in cross-corpus AUROC after unsupervised alignment, relative to the same classifier without alignment. Positive values mean better transfer

What the classifiers rely on (H4)

Figure 5 shows random-forest importances for the relaxation versus mirror-image contrast. The delta band carried the largest share of importance, followed by alpha, and the two frontopolar electrodes, Fp2 and Fp1, ranked highest among channels. Slow activity at frontopolar sites is the usual signature of blinks and eye movements, so part of the within-corpus performance could come from task-related eye behaviour rather than from a cortical response to stress. The prominence of alpha agrees with the well-known suppression of alpha activity under cognitive load [24].

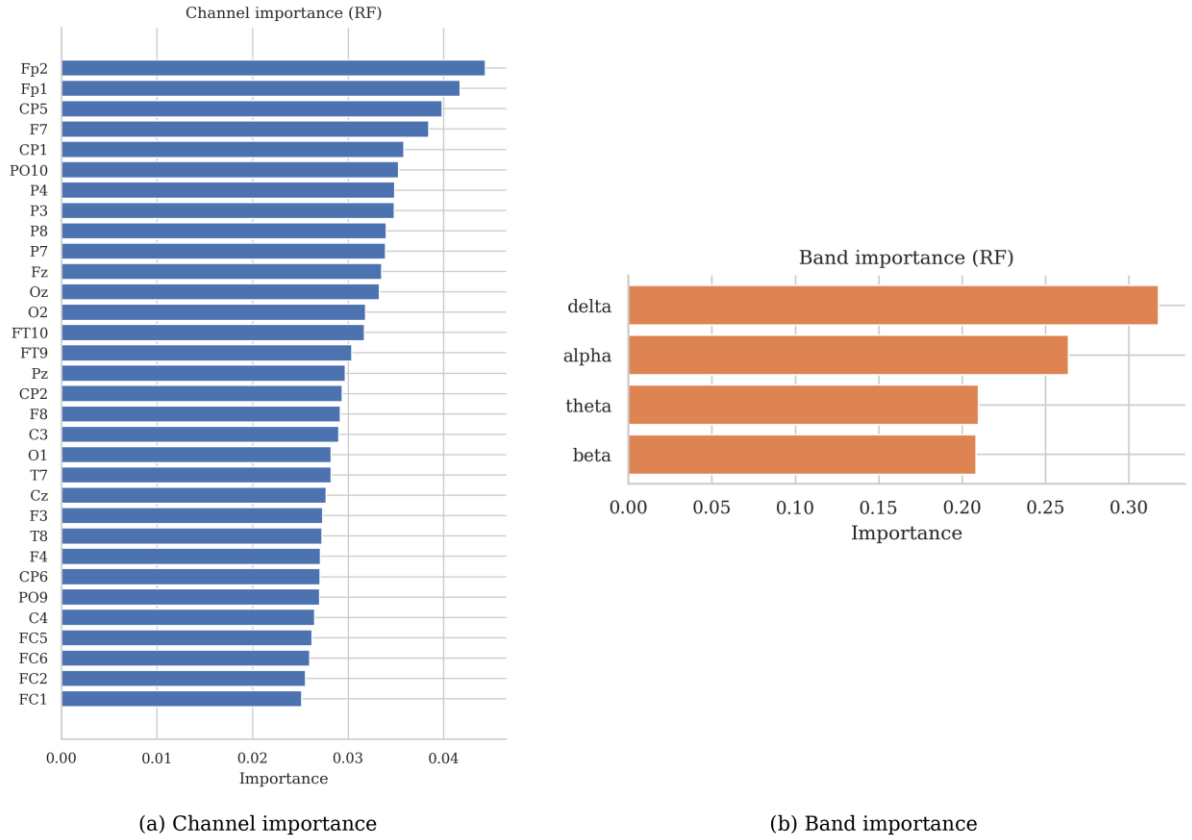
The ICA control (Table 5) shows that ocular activity is not the main source of discriminative information. After removing eye-related components, macro-F1 of the band-power classifiers fell by only 1.1–3.0 points and AUROC by 0.03–0.05. The Riemannian classifier lost more (8.9 points). This is probably in part a methodological effect: removing components lowers the rank of the epoch covariance matrices, and the tangent-space mapping is sensitive to that.

Table 5 - Effect of removing ocular components with ICA (SAM40, relaxation vs all stressors, 32 channels, LOSO)

	Macro-F1 (%)		AUROC	
Classifier	without ICA	with ICA	without ICA	with ICA
LogReg	59.9	58.7	0.734	0.708
Random forest	51.3	48.3	0.740	0.707
RBF-SVM	61.2	59.3	0.724	0.674
Riemannian	58.9	50.0	0.732	0.690

Abbreviations: AUROC, area under the receiver operating characteristic curve; ICA, independent component analysis; LogReg, logistic regression; LOSO, leave-one-subject-out; RBF-SVM, support vector machine with radial basis function kernel.

https://doi.org/10.52889/1684-9280-2026-77-5-jto066



(a) Channel importance

(b) Band importance

Figure 5 - Random-forest feature importance for relaxation vs mirror image on SAM40 (32 channels, all epochs). Importances are summed over bands for each channel (a) and over channels for each band (b)

4. Discussion

Subject-independent performance is modest. With LOSO validation, stress versus relaxation on SAM40 reached a macro-F1 of about 0.70 at best. This is far below the accuracies reported for the same dataset, including 99.67% [6]. Those figures were not obtained with LOSO and cannot be compared directly, but the size of the difference fits the inflation known to arise when data from the same participant or recording enter both training and test sets [7, 8]. We saw the same effect for fatigue in our earlier work [10].

Task labels are not state labels. Cross-corpus transfer is the more relevant test for a practical detector, and here performance fell significantly despite shared electrodes, identical preprocessing and per-participant normalisation. The reversed transfer from EEGMAT to SAM40 is the most informative result. In both datasets a label of 1 means that the participant was doing a task, not that the participant felt stressed. The task periods differ between protocols in visual demand, eye movements, possible speech-related muscle activity and duration, and the control periods differ too (a short relaxation block in SAM40, a long resting recording in EEGMAT). A classifier can therefore learn a signature of "task versus this particular control condition" whose sign changes from one dataset to the other. Self-reported stress, which SAM40 provides as rating scales, would be a natural way to check labels in future work.

Why label-free alignment falls short. Re-centring removes differences in the marginal distribution of the signals, that is, each participant's and each corpus's offset in covariance space, and it did improve within-corpus results. It did not repair the reversed transfer. This suggests that the dominant cross-corpus shift is conditional: the relation between EEG features and the label itself differs between protocols, and no method that ignores target labels can correct such a shift. The partial success of CORAL on band power indicates that some of the reversal came from differences in feature scale and covariance, for instance device-dependent spectral tilt, which can be aligned. The rest calls for a small amount of labelled target data or for labels that describe the participant's state.

Implications for post-trauma rehabilitation and biofeedback. A relaxation trainer driven by EEG has to work for new users and, often, new hardware. Our results suggest that such systems will need per-user calibration or adaptation, and that candidate models should be checked for dependence on ocular and other artefact-dominated features before deployment.

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

Limitations. Only two corpora were available, and both use cognitive stressors; an emotional stressor dataset such as DASPS [25] and a multitasking dataset such as STEW [26] are needed to test H2 fully. We evaluated compact classical models; deep networks and adversarial adaptation [19] may behave differently. Restricting the montage to ten shared electrodes discards spatial information available within each corpus. Epochs from one participant are not independent, which is why all statistics were computed per participant.

5. Conclusion

Under a unified, subject-independent protocol, EEG classifiers of acute stress reach moderate performance within a dataset and lose much of it on a new dataset, in one direction to the point of reversing their decisions. Label-free alignment narrows the gap in some settings but does not close it. Cross-corpus evaluation exposes protocol-specific shortcuts that within-dataset benchmarks hide, and it should accompany any claim that EEG can detect stress in practice. For post-trauma rehabilitation and biofeedback, this means that EEG stress detectors will need per-patient calibration and validation in trauma patients before clinical use.

Data availability. SAM40 is available on Figshare (article 14562090) and EEGMAT on PhysioNet (eegmat, version 1.0.0). The analysis code is available from the corresponding author upon reasonable request.

Conflict of interest. The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Funding. This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (program-targeted funding, grant No. BR27198099).

Author contributions. Conceptualization, A.M., B.A. and M.M.; methodology, A.M., B.A. and M.M.; software, A.M.; validation, A.M., A.K. and M.Z.; formal analysis, A.K.; investigation, A.M., B.A. and M.M.; resources, A.K.; data curation, A.M.; writing—original draft preparation, A.M., B.A. and M.M.; writing—review and editing, B.A., M.Z., A.K., M.M. and A.K.; visualization, A.M.; supervision, B.A., M.M. and A.K.; project administration, B.A., M.Z., A.K. and A.K.; funding acquisition, M.Z., and A.K. All authors have read and agreed to the published version of the manuscript.

Generative AI statement. The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI tools were used to assist with language editing and improving clarity. All scientific content, analyses, interpretations, and final manuscript decisions were reviewed and approved by the authors.

References

1. McEwen, B. S. (1998). Protective and damaging effects of stress mediators. The New England Journal of Medicine, 338(3), 171–179. https://doi.org/10.1056/NEJM199801153380307

2. Muscatelli, S., Spurr, H., O'Hara, N. N., O'Hara, L. M., Sprague, S. A., & Slobogean, G. P. (2017). Prevalence of depression and posttraumatic stress disorder after acute orthopaedic trauma: A systematic review and meta-analysis. Journal of Orthopaedic Trauma, 31(1), 47–55. https://doi.org/10.1097/BOT.0000000000000664

3. Starr, A. J., Smith, W. R., Frawley, W. H., Borer, D. S., Morgan, S. J., Reinert, C. M., & Mendoza-Welch, M. (2004). Symptoms of posttraumatic stress disorder after orthopaedic trauma. The Journal of Bone and Joint Surgery. American Volume, 86(6), 1115–1121. https://doi.org/10.2106/00004623-200406000-00001

4. van der Kolk, B. A., Hodgdon, H., Gapen, M., Musicaro, R., Suvak, M. K., Hamlin, E., & Spinazzola, J. (2016). A randomized controlled study of neurofeedback for chronic PTSD. PLoS ONE, 11(12), e0166752. https://doi.org/10.1371/journal.pone.0166752

5. Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F., & Al-Nashash, H. (2021). A review on mental stress assessment methods using electroencephalography signals. Sensors, 21(15), 5043. https://doi.org/10.3390/s21155043

6. Golder, K., Haque, M., Islam, M., & Raihan, M. (2025). Enhancing task-specific stress monitoring using artificial intelligence-powered electroencephalography feature selection techniques. In Feature fusion for next-generation artificial intelligence: Sustainable artificial intelligence-powered applications. Springer, 167–179. https://doi.org/10.1007/978-3-031-94386-7_13

7. Li, R., Johansen, J. S., Ahmed, H., Ilyevsky, T. V., Wilbur, R. B., Bharadwaj, H. M., & Siskind, J. M. (2021). The perils and pitfalls of block design for electroencephalography classification experiments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1), 316–333. https://doi.org/10.1109/TPAMI.2020.2973153

8. Brookshire, G., Kasper, J., Blauch, N. M., Wu, Y. C., Glatt, R., Merrill, D. A., Gerrol, S., Yoder, K. J., Quirk, C., & Lucero, C. (2024). Data leakage in deep learning studies of translational electroencephalography. Frontiers in Neuroscience, 18, 1373515. https://doi.org/10.3389/fnins.2024.1373515

9. Saha, S., & Baumert, M. (2020). Intra- and inter-subject variability in electroencephalography-based sensorimotor brain-computer interface: A review. Frontiers in Computational Neuroscience, 13, 87. https://doi.org/10.3389/fncom.2019.00087

10. Mukushev, M., Mengdigali, A., Karibekov, T., Zholdasova, M., Arman, D., & Kustubayeva, A. (2026). Cross-subject electroencephalography-based fatigue classification using machine learning, Riemannian geometry, and compact deep neural networks. Scientific Journal of Astana IT University. https://doi.org/10.37943/25EJPK6829

11. Zheng, W. L., & Lu, B. L. (2015). Investigating critical frequency bands and channels for electroencephalography-based emotion recognition with deep neural networks. IEEE Transactions on Autonomous Mental Development, 7(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497

12. Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & 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

13. Lan, Z., Sourina, O., Wang, L., Scherer, R., & Müller-Putz, G. R. (2019). Domain adaptation techniques for electroencephalography-based emotion recognition: A comparative study on two public datasets. IEEE Transactions on Cognitive and Developmental Systems, 11(1), 85–94. https://doi.org/10.1109/TCDS.2018.2826840

14. Agnelli, F., Blandano, G., Burger, J., D'Amelio, A., Facchi, G., Ghezzi, O., Lanzarotti, R., & Schmid, L. (2026). EEG-based mental stress detection: A comparative and explainable study across tasks and subjects. In Image analysis and processing – ICIAP 2025 workshops. Springer, 17–28. https://doi.org/10.1007/978-3-032-11317-7_2

15. Vos, G., Ebrahimpour, M., van Eijk, L., Sarnyai, Z., & Rahimi Azghadi, M. (2025). A statistical approach for synthetic electroencephalography data generation. arXiv, 16143. https://arxiv.org/abs/2504.16143

16. Zanini, P., Congedo, M., Jutten, C., Said, S., & Berthoumieu, Y. (2018). Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. IEEE Transactions on Biomedical Engineering, 65(5), 1107–1116. https://doi.org/10.1109/TBME.2017.2742541

17. Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. IEEE Transactions on Biomedical Engineering, 59(4), 920–928. https://doi.org/10.1109/TBME.2011.2172210

18. Sun, B., Feng, J., & Saenko, K. (2016). Return of frustratingly easy domain adaptation. Proceedings of the AAAI Conference on Artificial Intelligence, 30(1). https://doi.org/10.1609/aaai.v30i1.10306

19. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. Journal of Machine Learning Research, 17(59), 1–35.

20. Ghosh, R., Deb, N., Sengupta, K., Phukan, A., Choudhury, N., Kashyap, S., Phadikar, S., Saha, R., Das, P., Sinha, N., & Dutta, P. (2022). SAM 40: Dataset of 40 subject electroencephalography recordings to monitor the induced stress while performing Stroop color-word test, arithmetic task, and mirror image recognition task. Data in Brief, 40, 107772. https://doi.org/10.1016/j.dib.2021.107772

21. Zyma, I., Tukaev, S., Seleznov, I., Kiyono, K., Popov, A., Chernykh, M., & Shpenkov, O. (2019). Electroencephalograms during mental arithmetic task performance. Data, 4(1), 14. https://doi.org/10.3390/data4010014

22. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation, 101(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

23. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 7, 267. https://doi.org/10.3389/fnins.2013.00267

24. Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Research Reviews, 29(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

25. Baghdadi, A., Aribi, Y., Fourati, R., Halouani, N., Siarry, P., & Alimi, A. (2021). Psychological stimulation for anxious states detection based on EEG-related features. Journal of Ambient Intelligence and Humanized Computing, 12, 8519–8533. https://doi.org/10.1007/s12652-020-02586-8

26. Lim, W. L., Sourina, O., & Wang, L. P. (2018). STEW: Simultaneous task electroencephalography workload dataset. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(11), 2106–2114. https://doi.org/10.1109/TNSRE.2018.2872924

Жарақаттан кейінгі оңалту мен биологиялық кері байланыс үшін ЭЭГ негізінде жедел стрессті анықтау: Модельдердің жаңа деректер жиынтықтарына шектеулі тасымалдануы

Меңдіғали А. [1*], Мұқушев М.Қ. [2], Әбдікенов Б.Б. [3], Құстубаева А.М. [4], Қамзанова А.Т. [5], Жолдасова М.К. [6]

[1] Зерттеуші, Жасанды интеллект және деректер ғылымы мектебі, Astana IT University, Астана, Қазақстан. E-mail: 242886@astanait.edu.kz

[2] Ассистент-профессор, Креативті индустрия мектебі, Astana IT University, Астана, Қазақстан. E-mail: m.mukusev@astanait.edu.kz

[3] Жасанды интеллект ғылыми-инновациялық орталығының директоры, Astana IT University, Астана, Қазақстан. E-mail: beibit.abdikenov@astanait.edu.kz

[4] Ми институтының директоры, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: almkusto@kaznu.kz

[5] Постдокторант-зерттеуші, Ми институты, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: kamzanova.altyngul@kaznu.kz

[6] Зерттеуші, Ми институты, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: manzur777@gmail.com

Түйіндеме

Жарақаттан кейінгі психологиялық стресс жарақат алған пациенттерде жиі кездеседі және сауығудың нашарлауымен байланысты. ЭЭГ негізіндегі биологиялық кері байланыс оңалту кезінде оны бақылауға және азайтуға көмектесуі мүмкін, бірақ тек стресс детекторы жаңа пациенттер мен жаңа жабдықта жұмыс істеген жағдайда жұмыс жасайды. Жедел стресстің жарияланған ЭЭГ классификаторлары әдетте бір деректер жиынтығында оқытылып, сол жиынтықта тестіленеді, және 90%-дан жоғары дәлдік жиі кездеседі. Біз бұл нәтиженің қаншалықты бөлігі модель бұрын көрмеген қатысушылар үшін, сондай-ақ басқа жабдықпен және стрессті тудырудың басқа хаттамасымен жазылған деректер жиынтығы үшін сақталатынын зерттедік.

Әдістері. Сау еріктілердегі жедел когнитивтік стресс бойынша екі ашық деректер жиынтығы бірыңғай өңдеу тізбегімен өңделді: SAM40 (40 қатысушы; Струп тесті, арифметикалық тапсырма және айналы бейнені тану тапсырмасы релаксацияға қарсы; 32 арналы Emotiv, 128 Гц) және EEGMAT (36 қатысушы; тізбекті азайту тынығу күйіне қарсы; 19 арналы клиникалық жүйе, 500 Гц). Төрт ықшам классификатор (жиілік диапазондарының қуаты бойынша логистикалық регрессия, кездейсоқ орман және RBF-ядролы SVM, сондай-ақ риман геометриясының жанама кеңістігіндегі классификатор) бір қатысушыны шығарып тастау арқылы валидациямен (LOSO), он ортақ электродта екі бағыттағы корпусаралық тасымалдаумен және бақылаусыз теңестірудің екі әдісімен (риман бойынша қайта орталықтандыру және CORAL) бағаланды.

Нәтижелері. SAM40 ішінде бір стрессорды релаксацияға қарсы ажыратуда LOSO бойынша ең жоғары macro-F1 70,3% болды (ең жоғары AUROC 0,804). Корпусаралық тасымалдау модель, бағыт және стрессор комбинацияларының 24-інің 22-сінде macro-F1 мәнін статистикалық маңызды түрде төмендетті (жұптық Уилкоксон критерийі, $p < 0,01$), ең көбі 37,6 пайыздық тармаққа. Тасымалдау асимметриялы болды: EEGMAT деректерінде оқытылған классификаторлар SAM40 деректерінде 12 жағдайдың 9-ында 0,5-тен төмен AUROC көрсетті, яғни үйренілген шешім бағыты керісінше өзгерді.

Қорытынды. Тапсырма арқылы анықталған стресс белгілері хаттамаға тән ақпаратты қамтиды және ол деректер жиынтығы ауысқанда сақталмайды. ЭЭГ негізіндегі стресс детекторының жарақаттан кейінгі оңалтуда немесе биологиялық кері байланыста қолдануға дайын екендігі туралы кез келген тұжырымның алдында корпусаралық тестілеу жүргізілуі тиіс. Клиникада қолдану үшін әр пайдаланушыға калибрлеу және науқастар тобында валидация қажет болады.

Түйін сөздер: жарақаттан кейінгі стресс, жарақаттан кейінгі оңалту, биологиялық кері байланыс, ЭЭГ, жедел психологиялық стресс, корпусаралық жалпылау.

Детекция острого стресса по ЭЭГ для посттравматической реабилитации и биологической обратной связи: Ограниченная переносимость моделей на новые наборы данных

Мендыгали А. [1*], Мукушев М.К. [2], Абдыкенов Б.Б. [3], Кустубаева А.М. [4], Камзанова А.Т. [5], Жолдасова М.К. [6]

[1] Исследователь, Школа искусственного интеллекта и науки о данных, Astana IT University, Астана, Казахстан. E-mail: 242886@astanait.edu.kz

https://doi.org/10.52889/1684-9280-2026-77-5-jto066

[2] Ассистент-профессор, Школа креативных индустрий, Astana IT University, Астана, Казахстан. E-mail: m.mukusev@astanait.edu.kz
[3] Директор Научно-инновационного центра «Искусственный интеллект», Astana IT University, Астана, Казахстан.
E-mail: beibit.abdikenov@astanait.edu.kz
[4] Директор Института мозга, Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан. E-mail: almkusto@kaznu.kz
[5] Исследователь-постдокторант, Институт мозга, Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан.
E-mail: kamzanova.altyngul@kaznu.kz
[6] Исследователь, Институт мозга, Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан.
E-mail: manzur777@gmail.com

Резюме

Психологический стресс после травмы часто встречается у пациентов с травмами и связан с худшим восстановлением. Биологическая обратная связь на основе ЭЭГ могла бы помочь отслеживать и снижать его в ходе реабилитации, но только если детектор стресса работает на новых пациентах и новом оборудовании. Опубликованные ЭЭГ-классификаторы острого стресса обычно обучают и тестируют на одном наборе данных, и часто сообщается о точности выше 90%. Мы оценили, какая часть этой эффективности сохраняется для участников, которых модель не видела, и для набора данных, записанного на другом оборудовании и по другому протоколу индукции стресса.

Методы. Два открытых набора данных с острым когнитивным стрессом у здоровых добровольцев обработаны единым конвейером: SAM40 (40 участников; тест Струпа, арифметическая задача и задача распознавания зеркальных изображений против релаксации; 32-канальная система Emotiv, 128 Гц) и EEGMAT (36 участников; последовательное вычитание против покоя; 19-канальная клиническая система, 500 Гц). Четыре компактных классификатора (логистическая регрессия, случайный лес и SVM с RBF-ядром на мощности частотных диапазонов, а также классификатор в касательном пространстве римановой геометрии) оценивались методом валидации с исключением одного испытуемого (LOSO), при межкорпусном переносе в обоих направлениях на десяти общих электродах и с двумя методами неконтролируемого выравнивания: римановым перецентрированием и CORAL.

Результаты. Внутри SAM40 лучший macro-F1 при LOSO для одного стрессора против релаксации составил 70,3% (лучший AUROC 0,804). Межкорпусный перенос значимо снижал macro-F1 в 22 из 24 комбинаций модели, направления и стрессора (парный критерий Уилкоксона, $p<0,01$), максимум на 37,6 процентного пункта. Перенос был асимметричным: классификаторы, обученные на EEGMAT, достигали на SAM40 AUROC ниже 0,5 в 9 из 12 случаев, то есть выученное направление решения инвертировалось.

Выводы. Метки стресса, определенные задачей, несут специфичную для протокола информацию, которая не сохраняется при смене набора данных. Межкорпусное тестирование должно предшествовать любым заявлениям о готовности ЭЭГ-детектора стресса к применению в посттравматической реабилитации или в биологической обратной связи. Для клинического использования потребуются калибровка под пользователя и валидация на пациентах.

Ключевые слова: посттравматический стресс, посттравматическая реабилитация, биологическая обратная связь, ЭЭГ, острый психологический стресс, межкорпусная генерализация.

https://doi.org/10.52889/1684-9280-2026-77-5-jto066