Stress-GPT: Towards stress quantification with an EEG-based foundation model.

Catherine Lloyd[1,*], Loic Lemoine[1,*], Reiyan Al-Shaikh[1], Kim Tien Ly[2], Hakan Kayan[1], Charith Perera[1], Nhat Pham[1]
[1]Cardiff University, [2]University of Oxford
{lloydc25,lorentelemoinel,al-shaikhr,kayanh,pererac,phamn}@cardiff.ac.uk,ktien@oxfordrobotics.institute

ABSTRACT

Stress has emerged and continues to be a regular obstacle in people's lives. When left ignored and untreated, it can lead to many health complications, including an increased risk of death. In this study, we propose a foundation model approach for stress detection without the need to train the model from scratch. Specifically, we utilise the foundation model "Neuro-GPT", which was trained on a large open dataset (TUH EEG) with 20,000 EEG recordings. We fine-tune the model for stress detection and evaluate it on a 40-subject open stress dataset. The evaluation results with a fine-tuned Neuro-GPT are promising with an average accuracy of 74.4% in quantifying "low-stress" and "high-stress". We also conducted experiments to compare the foundation model approach with traditional machine learning methods and highlight several observations for future research in this direction.

CCS CONCEPTS

• **Applied computing** → *Health informatics*; **Life and medical sciences**; • **Computing methodologies** → *Machine learning*; • **Computer systems organization** → *Embedded and cyber-physical systems.*

KEYWORDS

Biosignals, Foundation Model, Cyber-Physical systems.

*These authors contributed equally to this work.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM MobiCom '24, November 18–22, 2024, Washington D.C., DC, USA
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0489-5/24/11…$15.00
https://doi.org/10.1145/3636534.3698121

ACM Reference Format:
Catherine Lloyd[1,*], Loic Lemoine[1,*], Reiyan Al-Shaikh[1], Kim Tien Ly[2], Hakan Kayan[1], Charith Perera[1], Nhat Pham[1]. 2024. Stress-GPT: Towards stress quantification with an EEG-based foundation model.. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24), November 18–22, 2024, Washington D.C., DC, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3636534.3698121

1 INTRODUCTION

In recent years, stress has emerged as a major public health issue. The fast-paced nature of contemporary life, coupled with increasing demands in both personal and professional spheres, has led to heightened stress levels globally. A 2023 survey of adults in England found that the proportion reporting severe levels of distress increased by 46% between 2020 and 2022 [1]. The WHO [2] briefed a 25% increase in global mental health prevalence in the first year of the COVID-19 pandemic due to multiple stress factors. Prolonged exposure to stress is known to have detrimental and far-reaching effects on our health. These can be both physical, for example affecting the cardiovascular, immune and gastrointestinal systems [3], or mental altering cognitive function and leading to the development of anxiety disorders and depression [4]. Moreover, data from AXA UK and the Centre of Economic and Business Research indicates that work-induced stress alone costs the UK economy £28 billion per year [5]. There is a pressing need for effective monitoring, early detection, and intervention strategies to mitigate its adverse effects on individuals and society.

Key to improving our approaches to stress management is understanding the neural patterns that underpin the stress response. Real-time Electroencephalography (EEG) monitoring is a non-invasive and widely used way of exploring the brain activity associated with stress. EEG time-series data can be converted to reveal the proportion of frequencies that make up the signal. This can provide insight into the mental state of the subject undergoing the stress stimuli and allows the identification of neural patterns. However, due to the inherent variability and complexity of EEG datasets, classifying and interpreting the data has remained a major challenge. Traditional analytical methods such as time-frequency distributions and wavelet transform [6] have been used historically with some success but often struggle to handle the variable and dynamic characteristics of the data effectively.

Until recently, machine learning in physiological sensing systems has been specialised tools, i.e. trained to perform a certain task, such as epileptic seizure detection [7], pain quantification [8], or drowsiness monitoring [9], etc. This approach, however, faces a fundamental challenge. Large amounts of data are required to be collected for each targeted application, which is both labour-intensive and time-consuming. Furthermore, reusing previously collected datasets is not possible if the targeted applications or experiment setups are different.

This paper explores the foundation model approach in stress detection using EEG signals. The developed method aims to provide an objective and accurate model to quantify stress levels by leveraging a pre-trained foundation model on a large and publicly available EEG dataset, thus, eliminating the need for training the model from scratch. **We make the following contributions.**

(1) We explored the application of Neuro-GPT [10], a foundation model trained on a large open dataset (TUH EEG) [11] with 20,000 EEG recordings and fine-tuned it for stress detection.
(2) We conducted the evaluation on a publicly available dataset of 40 subject EEG recordings [12] exposed to stress stimuli. The results are promising with an average accuracy of 74.4%. We also conducted experiments to compare the foundation model approach with traditional machine learning methods and highlight several observations for future research in this direction.

2 METHODOLOGY

In this study, we employ the SAM 40 public stress EEG dataset [8] to fine-tune the Neuro-GPT foundation model.

2.1 Data Pre-processing

Dataset Overview. The SAM 40 EEG dataset is a collection of electroencephalogram data obtained from 40 subjects, who have undergone cognitive tasks known to induce stress using a 32-electrode Emotiv Epoc Flex gel kit. The tasks were the identification of symmetry in mirror images, arithmetic equations, Stroop colour-word test and a period of relaxation. Each task lasted 25 seconds and was repeated thrice, producing a total of 480 samples. The subjects were also asked to self-report their stress level for each stress-inducing task on a 1 to 10 scale. The pre-processed data, provided by Ghosh et al. 2022 [8], was utilised in this study. This data was pre-processed with band-pass filtering from 0.5-45 Hz, along with artefact removal via a Savitzky-Golay filter and wavelet thresholding.

Signal analysis. Short-time Fourier transforms were generated for each channel across the 4 conditions, for each subject and trial. This was accomplished through spectrograms on Matlab with the parameters set to a sampling frequency of 128 Hz, a window length of 512, an overlap length of 500, and an nFFT of 512.

To allow for further analysis, topographic maps were generated from 0-60 Hz in 10 Hz intervals for each subject, to visualise the spatial distribution of each frequency band. This was completed with EEGLAB GUI. An example set of heat maps is shown in Fig. 1 for a subject undergoing an arithmetic test and relaxation. There is a clear drop in intensity at around 45 Hz across the stress and relaxed conditions. The stress signal appears to show "bursts" of gamma activity as opposed to the continuous gamma activity shown in the relaxed data above 45 Hz.

Across the 40 subjects and the 4 test conditions, the most active area of brain activity is in the frontal area of the brain. This is observed at 10 Hz intervals from 0 to 60 Hz, reflecting this area's importance in the activity of frequency bands corresponding to mental states (alpha, beta and gamma). Therefore in future analysis, it is noted that electrodes in this area (Fz, Fp1, F7, F3, F4, F8 and Fp2) are of heightened relevance. Moreover, a similar pattern is observed in all test conditions when power spectral density is plotted against frequency. This reflects the initial spectrogram observations of a shift at around 45 Hz in intensity.

Lastly, for further analysis, we extracted statistical features from selected subjects to determine where there are perceivable differences between stress samples. These features include mean, variance, skewness, and kurtosis.

Samples Selection. As the data was recorded with 32 electrodes, it was important to identify the relevant electrodes for stress detection. We analysed the channels to determine this and to ensure the most appropriate channels were selected if necessary. Furthermore, the dataset's subject rating system raises a possible conflict between manual and automated classification. An arithmetic sample can be rated 1 with another arithmetic sample rated 10 while still corresponding to the same class. With the ratings on the opposite ends of the rating spectrum, this implies that a class may have conflicting representations. For our purpose of stress detection, this would confuse the classifier, as samples that may not be quantifiable as stress would be included in a stress class.

2.2 Fine-tuning Stress

Algorithm Architecture. Neuro-GPT leverages both an EEG encoder and a GPT model. The learning algorithm can be used without either feature, allowing for multiple types of fine-tuning strategies. We will leverage two strategies outlined in Neuro-GPT: "Encoder+GPT" and "Encoder-Only".

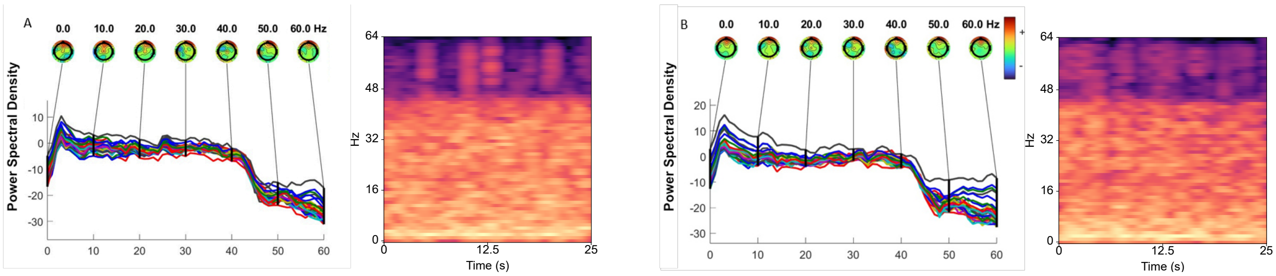

Figure 1: Power spectral density over frequency plots, accompanied by topographic brain maps at intervals of frequency every 10 Hz. This data is from Subject 1 trial 1, arithmetic (A) and relaxed (B).

The former uses the complete Neuro-GPT model with the latter using the model after the GPT architecture is removed.

Briefly, fixed-length chunks of EEG data are passed through an EEG encoder, which extracts "spatiotemporal features" and creates learning embeddings. These learning embeddings make use of a "casual masking" technique, masking tokens in a sequence. These sequences are fed into the GPT model, where the model "learns to predict" the masked tokens. The reconstruction loss is then computed using the original and the predicted token.

Data Preparation. We configured the Neuro-GPT release with their pre-trained model for it to be fine-tuned on the SAM dataset. Originally, Neuro-GPT used the BCI Competition IV Motor Imagery dataset [13] for fine-tuning, which required its own adjustments to be compatible with the pre-trained model. The data preparation in our fine-tuning had to match their configuration as closely as possible to produce more reliable results.

The raw SAM 40 EEG data arrays were extracted and processed. This included up-sampling the data to 250 Hz, applying a bandpass filter at 0.5 Hz and 100 Hz, and applying a notch filter at 50 Hz to eliminate the mains hum. SAM 40 was recorded with 32 electrodes, compared to the 22 electrodes used for both TUH and BCI. Out of these 32 electrodes, only 16 were a match with the Neuro-GPT electrodes. This resulted in 6 missing electrodes. Our samples were reduced to the 16 matching channels with an added selection of 6 channels. Finally, as NeuroGPT is configured to accept zipped numpy arrays as input, we exported the samples in the same format.

The SAM 40 dataset consisted of three "stress" activity classes: "Arithmetic", "Mirror", and "Stroop" with each stress event trial given a stress rating from 1-10. We leveraged this system to produce a more consistent two-class stress system: "Low-stress" and "High-stress". A threshold of 6 is applied to the stress ratings, with all samples below a rating of 6, alongside all relax samples, being classed as low-stress. All samples with a rating of 6 and above are classed as high-stress. From our analysis of this dataset, it was observed that there was confusion between classes and stress at different ratings. Therefore, we deemed the binary classification system appropriate for our use case.

Sample Handling. Originally, Neuro-GPT attempts to handle samples that consist of multiple trials that are hard-coded in their trial extraction system. As our samples consist of only one 25-second sample, the trial handling processes were adjusted to handle each sample as its own individual trial. Due to misalignment between the Neuro-GPT datasets, a 22x22 matrix multiplication is required to line up their electrodes. This aspect of the fine-tuning was adjusted to be compatible with our 16 matching electrodes. The remaining electrode spots were filled up by the next set of electrodes in the SAM-40 dataset.

Fine-Tuning. Prior to fine-tuning the model with our new data, the Neuro-GPT findings were replicated to ensure consistency and reliability. We then proceeded to fine-tune the pre-trained model with our new data. The number of classes was reduced from the original 4 to 2, with 0 representing "Low stress" and 1 representing "High stress". We fine-tuned Neuro-GPT with both "Encoder+GPT" and "Encoder-only" strategies. Other training parameters, such as batch size, were kept to the original defaults provided by Neuro-GPT. For "Encoder+GPT", the number of chunks was increased to 12 for one set of results and 1 for another. The chunk length of 2 remained for "Encoder-only" due to model compatibility. The chunk size was kept at 500. Neuro-GPT uses cross-fold validation. Our new data was split into 9 folds. Each fine-tuning method was conducted nine times, once for each fold. This produced three sets of nine results and models.

Comparison with traditional models. We implemented the state-of-the-arm algorithms with our low-stress and high-stress classes to provide more established comparisons to our LLM fine-tuning. With three comparison methods (2D-CNN, SVM, and XGBoost), we produce three further binary classifiers. Unlike the fine-tuned model that uses raw EEG signals as the input, we have to conduct additional pre-processing steps, i.e., extracting Mel spectrograms, for our comparison classifiers to work.

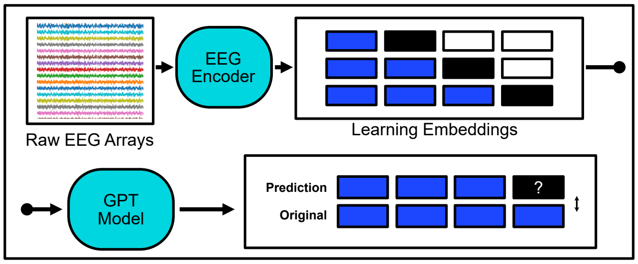

Figure 2: Neuro-GPT EEG encoder and GPT mask prediction pipeline.

3 STRESS QUANTIFICATION RESULTS WITH NEURO-GPT

3.1 Evaluation Metrics.

In this study, we use the last evaluation accuracy as the accuracy for a given fold. We collect the accuracy from all nine folds in each method and compute the average accuracy for a given method. The highest fold accuracy and lowest fold accuracy are also taken to provide a best and worst scenario from our fine-tuning.

3.2 Classification Results.

The initial replication with "Encoder+GPT" and "Encoder-only" using the pre-trained model met the provided results from Neuro-GPT, falling within their accuracy ranges. Our binary classification method produced an average accuracy of 71.3% when fine-tuned with both the encoder and GPT. The best fold saw an accuracy of 86.8% and 62.3% on the worst fold. An accuracy of 74.4% was produced when fine-tuned with only the encoder, featuring an accuracy of 96.2% on the best and 67.9% on the worst fold.

Both methods show reliable classification of the low-stress and high-stress separated samples. However, evaluation losses increase after some time. The learning curves produced demonstrate that the model may benefit from more samples in the dataset, specifically more distinctly class-separated samples.

As seen in table 1, the "Encoder-only" method produces better results than its GPT-included counterpart, with better average, best, and worst folds. "Encoder+GPT"'s best fold of 86.8% accuracy was outperformed by its counterpart with an accuracy of 96.2%. This may be due to a lower number of chunks, as when one chunk of used, "Encoder+GPT" produced an accuracy of 92.5% on the same fold. While this may be true, our results follow the same trend as Neuro-GPT's fine-tuning where the "Encoder-Only" strategy outperformed the rest.

Table 1: Fine-tuned Neuro-GPT's performance

Method	**Avg**	**Min** [fold]	**Max** [fold]
Encoder-only	0.744	0.679 [8]	0.962 [7]
Encoder+GPT (12 chunks)	0.713	0.623 [6]	0.868 [7]
Encoder+GPT (1 chunk)	0.738	0.623 [8]	0.925 [7]

3.3 Comparison with Traditional Machine Learning Methods.

We conducted the experiments to compare the classification performance of our fine-tuned Neuro-GPT with three traditional machine learning algorithms, 2D-CNN, SVM, and XGBoost. For the traditional models to work, we extracted the Mel spectrograms from the raw EEG signals and use them as the inputs. The 2D-CNN model achieved an overall accuracy of 85% while SVM and XGBoost showed similar overall accuracy at 86% and 93%, respectively. The best fold accuracies of our fine-tuned model resemble reliable classifiers and are able to outperform the SVM and 2D-CNN classifiers. However, the three algorithms yield better accuracy more consistently than the fine-tuned GPT model. It is important to note that the traditional methods were trained specifically on the target dataset and require Mel spectrograms to work while the Neuro-GPT only fine-tunes from a generic foundation model and only requires raw EEG data. Thus, it gives us some insights that the foundation model approach can present a more scalable solution with relatively comparable performance to traditional methods. The further inclusion of domain knowledge through hand-engineered features could potentially enhance the robustness of the fine-tuned foundation model.

4 DISCUSSION

SAM-40 Dataset Observations. It was observed that the SAM 40 dataset lacked distinction between the task samples and the relax samples when analysed via Mel spectrograms. This was seen across the key electrodes. This determines that Mel spectrograms may not be suitable for stress detection with this dataset due to its setup and methodology. A self-reported stress rating system can be too subjective, as different people experience stress differently resulting in what can be considered inconsistent ratings, as shown in Fig. 3. For example, a person may provide a stress level of 7, while another person better prepared for stress may provide a lower rating while experiencing the same amount of stress.

We observed that the largest range (of low stress to high stress) ratings in a given patient were 7. Through the statistical feature extraction, perceivable differences are noticed in the subject's variance. However, as this is in the maximum range, other subjects may produce less variation in this feature as their self-reported level ranges are lower.

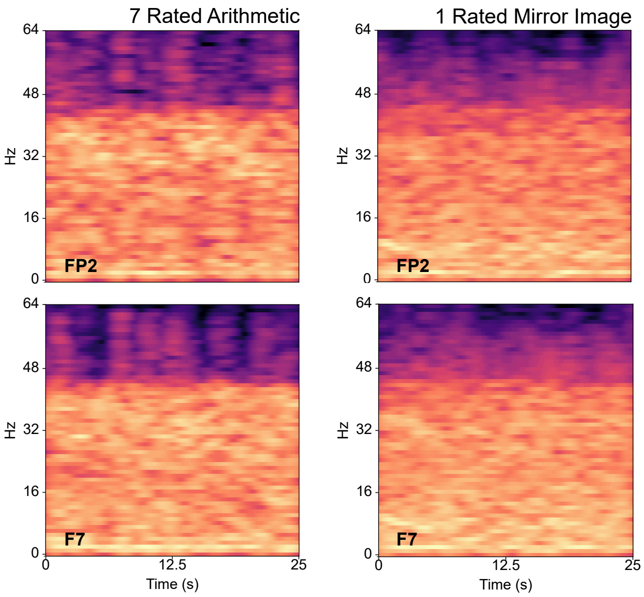

Figure 3: A comparison between a self-rated stress level 7 (high-stress) sample and a stress level 1 (low-stress) sample in the SAM-40 dataset.

Furthermore, the tasks and relax samples were not produced in isolation, and this may produce conflicts and confusion between the samples. It is stated in SAM 40 that a subject's next trial would start after their stress rating for the prior trial was given. The last task of a trial would be an arithmetic task, with the start of a trial being a relaxed event. On average, arithmetic is the most stressful task based on all subject's self-reporting. This raises an open question on whether relaxed samples in following trials are affected by being conducted after the most stressful activity.

Lastly, the mismatch in electrodes between Neuro-GPT and SAM 40 may have harmed the results. With only 16 similar electrodes, SAM 40's setup may not be completely compatible with Neuro-GPT. With complete channel compatibility, it can be argued that there may be an increase in the model's performance. To resolve this, Neuro-GPT can be rebuilt to target the electrode configuration used in SAM 40. Nevertheless, based on the evaluation results featuring the adapted 16+6 channel system, fine-tuning can be considered a step in the right direction.

5 RELATED WORK

Stress Quantification with Physiological Sensors. Physiological biosensors are enhancing stress monitoring in clinical and personal health contexts by providing real-time feedback [14]. Despite progress, challenges persist. Hou et al. [15] developed an EEG-based stress recognition algorithm with promising accuracy but calls for larger datasets. Samson and Koh [16] note improvements with wearable sensors like CortiWatch and SKINTRONICS for real-time cortisol detection while emphasizing the need for better integration. Kocielnik et al. [17] introduced a combined sensor wristband and questionnaire framework, effectively linking stress data to activities and behaviour. Yoon, Sim and Cho [18] presented a flexible stress monitoring patch, with enhanced sensitivity and a lifespan of 9 days. Jovanov et al. [19] detailed the WISE system, which uses HRV for long-term stress monitoring.

Recent advancements in wearable biosensors have led to significant improvements, including enhanced machine learning for EEG and sleep apnea analysis, novel systems for joint health and cardiovascular monitoring [20], and flexible bio-chips that address noise and accuracy issues in EMG, ECG, PPG, and EEG [21]. Emerging stress-monitoring technologies now integrate sensors in smartwatches and contact lenses for non-invasive tracking of indicators like heart rate and cortisol levels [22], while multi-modal systems combine PPG, EEG, eye-gaze, body motion capture, and GSR sensors for precise real-time data synchronization [23]. While the previous works implement traditional machine learning methods, we explore an alternative direction and utilise the Neuro-GPT foundation model to enhance EEG-based stress detection, exploring challenges and limitations with fine-tuning on a 40-subject dataset.

LLMs in Neuroscience Applications. Large Language Models (LLMs) have demonstrated remarkable potential in neuroscience. GPT-2 has been used [24] to generate synthetic EEG and EMG signals, effectively augmenting real datasets and improving classification accuracy, with Random Forest accuracy and real-time gesture recognition increasing by over 20%. Neuroformer [25], a generative transformer model, excels in predicting neuronal circuit activity and inferring neural connectivity, significantly outperforming traditional models like GLMs, and showing effective multimodal integration. Additionally, GPT-3.5 and GPT-4 have improved the automation of computational neuroscience literature curation for the ModelDB repository, achieving high accuracy in identifying relevant papers and enhancing metadata extraction [26]. Event Stream GPT (ESGPT) extends GPTs to continuous-time sequences of complex events such as electronic health record data, demonstrating significant performance improvements over existing tools like TemporAI and highlighting its potential to enhance research efficiency in non-NLP domains [27]. We follow this trend and attempt to identify a scalable approach to stress quantisation by leveraging an LLM foundation model.

6 CONCLUSION

In this paper, we presented solutions to stress focused EEG collection and detection through frontal electrodes and various machine learning methods. We focused on fine-tuning an LLM foundation model with an open dataset, configuring our own modified binary class system. Through this, we made positive steps towards a reliable EEG-based stress classifier, with an average accuracy of 74.4%. Within the study, we outline several challenges and limitations that were encountered. The produced results and methods can lay the foundations for future work, which may involve larger compatible datasets and more optimised classes and features.

ACKNOWLEDGMENTS

This research was partially supported by the Academy of Medical Science through the Networking Grant (NGR1-1600) and Thales MPhil studentship.

REFERENCES

[1] Ru Jia, Kieran Ayling, Carol Coupland, Trudie Chalder, Adam Massey, Urs Nater, Elizabeth Broadbent, Norina Gasteiger, Wei Gao, Clemens Kirschbaum, et al. Increases in stress hormone levels in a uk population during the covid-19 pandemic: A prospective cohort study. *Psychoneuroendocrinology*, 148:105992, 2023.

[2] World Health Organization. Covid-19 pandemic triggers 25 *World Health Organization*, Mar 2022.

[3] Peter J Gianaros and Tor D Wager. Brain-body pathways linking psychological stress and physical health. *Current directions in psychological science*, 24(4):313–321, 2015.

[4] Joan I Cohen. Stress and mental health: a biobehavioral perspective. *Issues in mental health nursing*, 21(2):185–202, 2000.

[5] AXA. Mind your health in the workplace. 2023.

[6] Amjed S Al-Fahoum and Ausilah A Al-Fraihat. Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices*, 2014(1):730218, 2014.

[7] Athar A Ein Shoka, Mohamed M Dessouky, Ayman El-Sayed, and Ezz El-Din Hemdan. Eeg seizure detection: concepts, techniques, challenges, and future trends. *Multimedia Tools and Applications*, 82(27):42021–42051, 2023.

[8] Hoang Truong, Nam Bui, Zohreh Raghebi, Marta Ceko, Nhat Pham, Phuc Nguyen, Anh Nguyen, Taeho Kim, Katrina Siegfried, Evan Stene, et al. Painometry: wearable and objective quantification system for acute postoperative pain. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 419–433, 2020.

[9] Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, et al. Wake: a behind-the-ear wearable system for microsleep detection. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 404–418, 2020.

[10] Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: developing a foundation model for eeg. *arXiv preprint arXiv:2311.03764*, 2023.

[11] Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in neuroscience*, 10:196, 2016.

[12] Rajdeep Ghosh, Nabamita Deb, Kaushik Sengupta, Anurag Phukan, Nitin Choudhury, Sreshtha Kashyap, Souvik Phadikar, Ramesh Saha, Pranesh Das, Nidul Sinha, et al. Sam 40: Dataset of 40 subject eeg recordings to monitor the induced-stress while performing stroop color-word test, arithmetic task, and mirror image recognition task. *Data in Brief*, 40:107772, 2022.

[13] Clemens Brunner, Robert Leeb, Gernot M"uller-Putz, Alois Schl"ogl, and Gert Pfurtscheller. Bci competition 2008–graz data set a. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 16:1–6, 2008.

[14] Onur Parlak. Portable and wearable real-time stress monitoring: A critical review. *Sensors and Actuators Reports*, 3:100036, 2021.

[15] Xiyuan Hou, Yisi Liu, Olga Sourina, Yun Rui Eileen Tan, Lipo Wang, and Wolfgang Mueller-Wittig. Eeg based stress monitoring. In *2015 IEEE international conference on systems, man, and cybernetics*, pages 3110–3115. IEEE, 2015.

[16] Cheyenne Samson and Ahyeon Koh. Stress monitoring and recent advancements in wearable biosensors. *Frontiers in bioengineering and biotechnology*, 8:1037, 2020.

[17] Rafal Kocielnik, Natalia Sidorova, Fabrizio Maria Maggi, Martin Ouwerkerk, and Joyce HDM Westerink. Smart technologies for long-term stress monitoring at work. In *proceedings of the 26th IEEE international symposium on computer-based medical systems*, pages 53–58. IEEE, 2013.

[18] Sunghyun Yoon, Jai Kyoung Sim, and Young-Ho Cho. A flexible and wearable human stress monitoring patch. *Scientific reports*, 6(1):23468, 2016.

[19] Emil Jovanov, AO'Donnell Lords, Dejan Raskovic, Paul G Cox, Reza Adhami, and Frank Andrasik. Stress monitoring using a distributed wireless intelligent sensor system. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):49–55, 2003.

[20] Marco Di Rienzo and Ramakrishna Mukkamala. Wearable and nearable biosensors and systems for healthcare, 2021.

[21] Dahyun Nam, Jae Min Cha, and Kiwon Park. Next-generation wearable biosensors developed with flexible bio-chips. *Micromachines*, 12(1):64, 2021.

[22] Ju-Yu Wu, Congo Tak-Shing Ching, Hui-Min David Wang, and Lun-De Liao. Emerging wearable biosensor technologies for stress monitoring and their real-world applications, 2022.

[23] Aashish N Patel, Tzyy-Ping Jung, Terrence J Sejnowski, et al. A wearable multi-modal bio-sensing system towards real-world applications. *IEEE Transactions on Biomedical Engineering*, 66(4):1137–1147, 2018.

[24] Jordan J Bird, Michael Pritchard, Antonio Fratini, Anikó Ekárt, and Diego R Faria. Synthetic biological signals machine-generated by gpt-2 improve the classification of eeg and emg through data augmentation. *IEEE Robotics and Automation Letters*, 6(2):3498–3504, 2021.

[25] Antonis Antoniades, Yiyi Yu, Joseph Canzano, William Wang, and Spencer LaVere Smith. Neuroformer: Multimodal and multitask generative pretraining for brain data. *arXiv preprint arXiv:2311.00136*, 2023.

[26] Ziqing Ji, Siyan Guo, Yujie Qiao, and Robert A McDougal. Automating literature screening and curation with applications to computational neuroscience. *Journal of the American Medical Informatics Association*, 31(7):1463–1470, 2024.

[27] Matthew McDermott, Bret Nestor, Peniel Argaw, and Isaac S Kohane. Event stream gpt: a data pre-processing and modeling library for generative, pre-trained transformers over continuous-time sequences of complex events. *Advances in Neural Information Processing Systems*, 36, 2024.