scientific data

OPEN

DATA DESCRIPTOR

A finer-grained high altitude EEG dataset for hypoxia levels assessment

Yingjun Si[1,2], Yu Zhang[3] ✉, Xi Zhang[1,2] ✉, Sicong Liu[3], Honghao Zhang[4] & Hui Yang[1,2] ✉

The study reports on a high-altitude EEG dataset comprising 64-channel EEG signals from 23 subjects, aiming at achieving a finer-grained assessment of hypoxia levels. Four hypoxia levels were induced by creating a gradient of oxygen partial pressure through changes in altitude and external hypoxia stimulation. The dataset was collected in a hypoxic chamber that simulates altitude changes, allowing for a refined classification of different hypoxia levels based on ranges of oxygen saturation. The total recorded EEG data amounts to approximately 10.25 hours. Validation results indicate that the four hypoxia levels can be effectively recognized using EEG signals. Compared to binary classification, our fine-grained dataset allows for more precise detection of hypoxia levels. This dataset is anticipated to have significant research and practical value in developing accurate methods for identifying hypoxia levels. As a valuable and standardized resource, it will enable extensive analysis and comparison for researchers in the field of high-altitude hypoxia.

Background & Summary

Plateaus are common geographical features, encompassing nearly 45% of the Earth's land surface[1]. Each year, in addition to the indigenous population, numerous individuals from low-altitude regions migrate to higher elevations for occupational or educational purposes[2,3]. The reduced atmospheric pressure at these elevations can induce harmful hypoxic physiological responses[4], potentially leading to altitude-related conditions such as pulmonary and cerebral edema[5,6]. If timely intervention is not provided to meet the body's oxygen requirements, hypoxia at high altitudes can pose significant health risks[7]. Therefore, delivering effective oxygen supplementation in a timely manner is crucial[8]. However, when the oxygen supply is limited, it is necessary to accurately determine the actual oxygen demand to extend the availability of oxygen. Thus, precise monitoring and assessment of hypoxia levels are essential.

Although many physiological signals are associated with the hypoxia levels of the human body, there is no consensus on which signal is the most representative. The human respiratory and circulatory systems can indirectly reflect the degree of hypoxia through these physiological indicators. Respiration and oxygen saturation are commonly used to assess the hypoxia levels of the human body[9]. Although they are recognized parameters, their measurement at a specific body site does not represent the systemic hypoxic condition of the entire body[10]. In fact, no single indicator can characterize hypoxia levels of the whole body. As is well known, despite constituting only 2% of the body's weight, the brain consumes about 20% of the oxygen supply to maintain its normal function[11]. This implies that the brain is highly sensitive to hypoxia[6,12]. Without prompt intervention to meet cerebral oxygen requirements, hypoxia at high altitudes can pose a significant health risk, causing irreversible cerebral damage[13]. In the field of brain science, there are numerous established techniques for monitoring the state of the brain, such as magnetoencephalography (MEG), cerebral blood flow, and electroencephalography (EEG)[14]. Among them, EEG has emerged as a potent tool for investigating cerebral function in high-altitude environments due to its high temporal resolution, cost-effectiveness, and portability[15].

EEG datasets for plateau hypoxia identification remain scarce, undoubtedly limiting the potential for related research. This scarcity is due to the significant time and financial costs required for high-altitude experiments.

[1]School of Life Sciences, Northwestern Polytechnical University, Xi'an, 710129, China. [2]Engineering Research Center of Chinese Ministry of Education for Biological Diagnosis, Treatment and Protection Technology and Equipment, Northwestern Polytechnical University, Xi'an, 710129, China. [3]School of Computer Science, Northwestern Polytechnical University, Xi'an, 710129, China. [4]School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, 710072, China. ✉e-mail: zhangyu@nwpu.edu.cn; oca-john@hotmail.com; kittyyh@nwpu.edu.cn

Dataset	Scenario	Participants	Data category	Sampling rate	Duration of a single record	Total duration
FACED[16]	Office	Healthy young adults	EEG	250 Hz or 1000 Hz	30 s	28.7 H
O'Toole's dataset[17]	Hospital	Neonates Hypoxic-ischaemic encephalopathy	EEG	256 Hz or 200 Hz	1 H	169 H
Harespod[18,19]	Hypobaric oxygen chamber	Healthy young adults	Respiratory waveform, SpO2	100 Hz	50 Min	12 H
FGHA-EEG[43]	Hypobaric oxygen chamber	Healthy young adults	EEG	1000 Hz	5 Min	10.25 H

Table 1. Comparison in terms of some characteristics of similar datasets with ours.

Even when high-quality datasets are obtained, researchers often lack the incentive to make their results publicly available. Consequently, it is challenging to find even those studies with limited relevance. The FACED dataset is a large-scale, fine-grained affective computing EEG dataset collected in an office environment[16]. Despite sharing a focus on the importance of EEG signals, its applicability to plateau hypoxia tasks is limited due to differences in specific scenarios. O'Toole's dataset comprises neonatal EEG recordings of infants diagnosed with hypoxic-ischemic encephalopathy[17], facilitating researchers in grading the severity of infantile illnesses. However, since the data originates from low-altitude regions, it cannot be utilized for high-altitude research. The Harespod dataset[18,19], created by our team, is a publicly available dataset specifically developed for studying hypoxia in high-altitude scenarios. Its strength lies in the collection of continuous, synchronized multi-physiological signals, but it does not include more representative physiological signals such as EEG. In conclusion, current publicly available datasets exhibit certain deficiencies in supporting the fine-grained identification of plateau hypoxia states.

To address these issues, we propose an open-access high-altitude EEG dataset for finer-grained assessment of hypoxia levels. This dataset comprises EEG data aimed at studying hypoxia levels through the brain's response to oxygen depletion signals. EEG can capture millisecond-level changes in brain electrical activity[20,21], allowing researchers to explore richer information. Theoretically, EEG is the optimal physiological indicator for assessing hypoxia levels. Although prior research in EEG-based hypoxia detection at high altitudes has achieved certain advancements[22–26], most studies have focused on binary classification, simplifying EEG data into two levels: normal and hypoxic. This approach results in insufficient granularity and reduced resolution, limiting effective intervention in high-altitude hypoxia scenarios. Given the need for precise health monitoring at high altitudes, binary classification may not accurately reflect individual hypoxia levels in these settings. In our experiment, we simulated altitudes ranging from 2 *km* to 4 *km*, exposing participants to corresponding hypoxic conditions and recording their EEG signals. Researchers used the recorded oxygen saturation levels as a basis for classification, supporting the division of hypoxia into four distinct levels. A comparison of some characteristics of similar datasets with ours is summarized in Table 1. While public datasets are expected to be finer-grained, our dataset meets the requirements for real-world applications in classifying hypoxia in high-altitude environments. It is anticipated that this dataset will aid in identifying fine-grained hypoxia levels, contributing to the development of new neurophysiological biomarkers for the diagnosis and treatment of hypoxia. Furthermore, this dataset will be crucial for the future development of fine-grained hypoxia classification algorithms based on EEG and corresponding hypoxia monitoring systems, ultimately reducing reliance on experts and providing more detailed guidance for subsequent treatments, such as oxygen therapy.

Methods

Experimental environment setting. We simulated high-altitude environments in our indoor laboratory using a hypoxic chamber. The components of the chamber and the equipment required for the experiment are detailed in Fig. 1. The hypoxic chamber includes a glass hood (DYC 5000 A, China) designed to withstand reduced pressure, two suction pumps (THOMAS, VTE 8, 6 m^3/h, Germany), a vacuum gauge (YiChuan, YNZ-100BF, $-$ 0.1-0 *MPa*, China) to monitor pressure changes, and two inlet valves (BTAOO, 60 L/min, China) for air renewal within the hood. Temperature control was maintained by an air conditioner set to a constant 24 °C. The temperature was kept between 24 and 26 °C, while humidity was consistently maintained between 53% and 58%. Participants could enter and exit the chamber using a pulley mechanism located at its summit. In Fig. 1, the blue arrow indicates a hose directing airflow from the base inlet to the chamber's summit, while the yellow arrow shows the pumps discharging air from the lower side, ensuring a cycle of fresh air within the chamber.

The human body may struggle to adapt to abrupt or significant changes in altitude[27]. A gradient altitude adjustment scheme was designed to provide the body with ample time for adaptation to each altitude change, setting five distinct altitude levels. Symptoms of altitude sickness typically appear around 2 *km*[28–31], with intolerable physiological reactions observed at 4 *km*[32,33]. Higher altitudes pose increased safety risks, which is why our experiment limited altitude exposure to the range of 2 to 4 *km*. Considering the subjects' tolerance, the total duration of the experiment should not be excessively long. And a sufficient number of experimental groups is necessary to obtain comprehensive data on the variability of stress responses. Specifically, a 5 *min* EEG recording with eyes open in a resting state was conducted before the hypoxic chamber was sealed. In the altitude range of 2 to 4 *km*, a 5 *min* stabilization period was established for every additional 0.5 *km* of elevation, and data was collected during this process. The altitude simulation scheme is illustrated in Fig. 2, where red triangles indicate the start and end times of data collection for the entire experiment. The solid blue line in the graph represents the time period during which data was recorded, and the blue dashed line indicates the total duration of the experiment. Each experiment consisted of approximately 30 *min* of data.

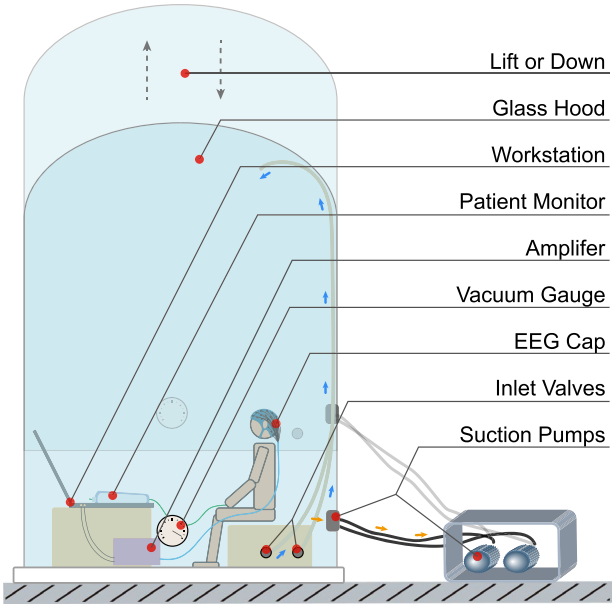

Fig. 1 A schematic diagram depicting the hypobaric chamber for high-altitude simulation, the physiological signal acquisition device, and the airflow direction.

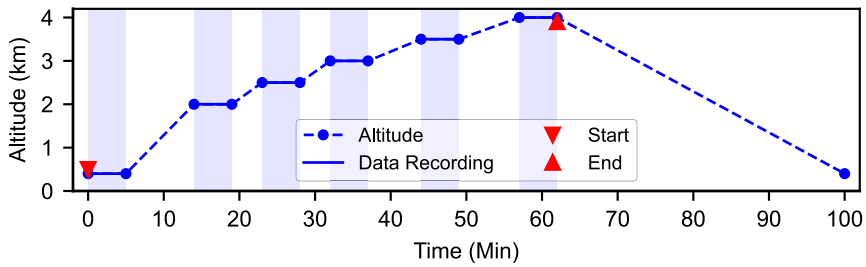

Fig. 2 Altitude gradient adjustment scheme. The blue dashed lines indicate the altitude adjustment protocol during the experiment. The blue solid lines represent periods of stable altitude, during which EEG data was collected. Red triangles mark the start and end times of the data collection process.

The proportion of oxygen in the air remains constant regardless of altitude. In our experiment, the hypoxic chamber was treated as a semi-open system, with gas exchange facilitated by suction pumps and inlet valves, minimally affecting the oxygen ratio. To maintain a specific simulated altitude, we regulated the speed of airflow in and out, thereby controlling the absolute concentration of oxygen. Before the experiment, it was essential to verify the chamber parameters to ensure that the negative pressure reading (vacuum level) corresponded to the desired simulated altitude.

Once the hypoxic chamber achieved a specific altitude, we adjusted the differential speed of air entering and exiting by modifying the number of pumps and the turns of the inlet valves. These adjustments were made to alter the pressure conditions within the chamber. During the experiment, minor fluctuations in altitude, as indicated by the vacuum gauge, typically occurred over a period of 1 to 3 *min*. If a specific set of parameters could maintain the simulated altitude with the desired precision for a continuous duration of 10 *min*, this parameter configuration was deemed satisfactory and documented for future reference. Table 2 presents the gradient elevation regulation scheme and the optimal parameter combinations used to maintain altitude in the hypoxic chamber.

Experiment procedure. After completing the exploratory tests, a standardized and reliable experimental procedure was developed to simulate high-altitude environments. During the experimental preparation phase, it is essential to ensure that both the hypoxic chamber and signal acquisition equipment are functioning correctly. Fresh air exchange was facilitated by keeping an inlet valve open, with the initially closed valve adjusted as needed throughout the experiment. Participants were required to remain awake during EEG recording, minimizing head and body movements, as well as activities such as blinking and swallowing. To ensure data quality, we segmented data collection at specific altitudes and time intervals, which was necessary due to the difficulty participants faced in maintaining prolonged adherence to the experimental protocol.

The specific protocol of the experiment is as follows:

Altitude (*km*)	Air pressure (*kPa*)	Vacuum level (*MPa*)	No. of pumps	Turns of inlet valve
2.0	78.9	0.0178	1	6
2.5	72.84	0.0238	1	4
3.0	67.24	0.0295	1	2
3.5	61.64	0.035	1	0.5
4.0	56.04	0.041	2	8

Table 2. The control variables for maintaining altitude stability in a hypoxic chamber include the number of pumps and the turns of the intake valve. The number of pumps determines the rate at which gas is removed from the chamber, while the turns of the intake valve regulate the rate at which external air enters the chamber.

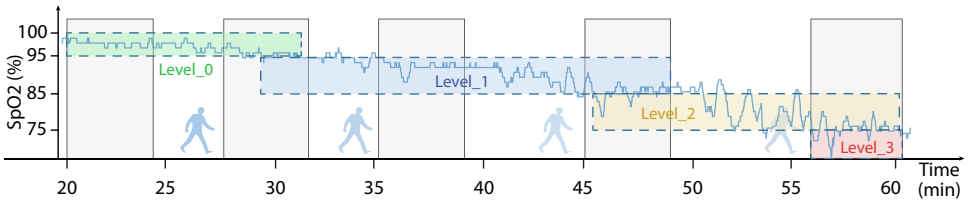

Fig. 3 The scheme for labeling the collected EEG data based on oxygen saturation levels.

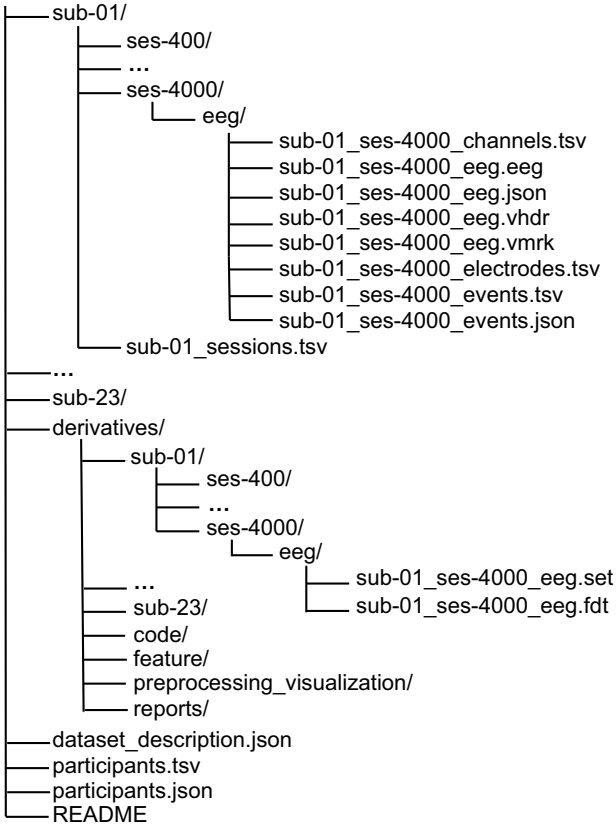

Fig. 4 The structure of the EEG dataset in BIDS format.

- Without adjusting for altitude, we conducted a control experiment. Researchers and participants engaged in a 5 *min* EEG recording with eyes open in a resting state. Oxygen saturation was continuously recorded throughout the entire experiment. Subsequently, a researcher outside the oxygen chamber closed the glass hood and reduced the air pressure by activating a suction pump. The simulated altitude was confirmed in real-time by the researcher outside the oxygen chamber, who monitored the negative pressure gauge and adjusted equipment parameters as needed.

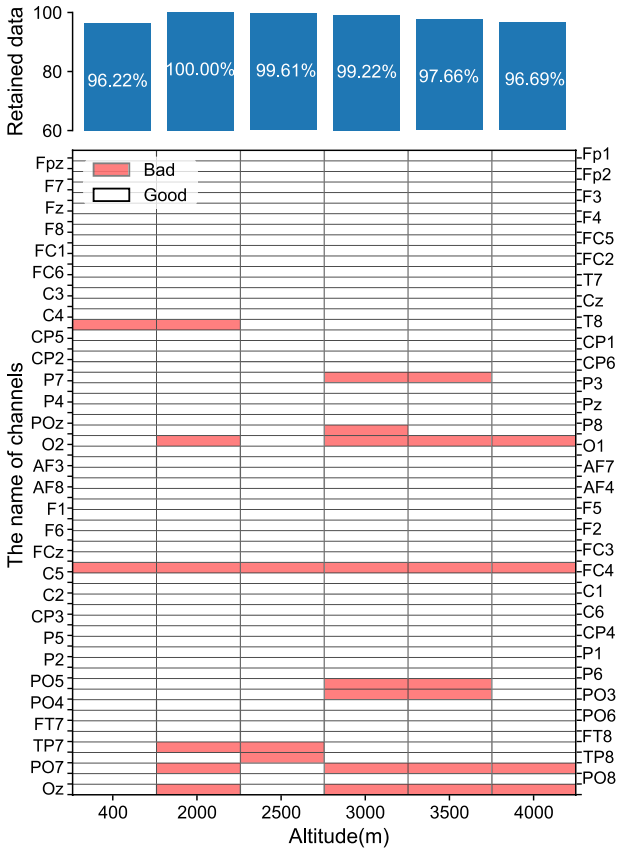

Fig. 5 The EEG data quality distribution across channels and altitudes for subject 09. The quality distribution chart illustrates the quality of EEG data collected from subject Sub09 during the experiment. The horizontal axis represents altitude groups, while the vertical axis lists all the EEG electrodes. The red blocks denote bad channels, whereas the white blocks indicate good ones.

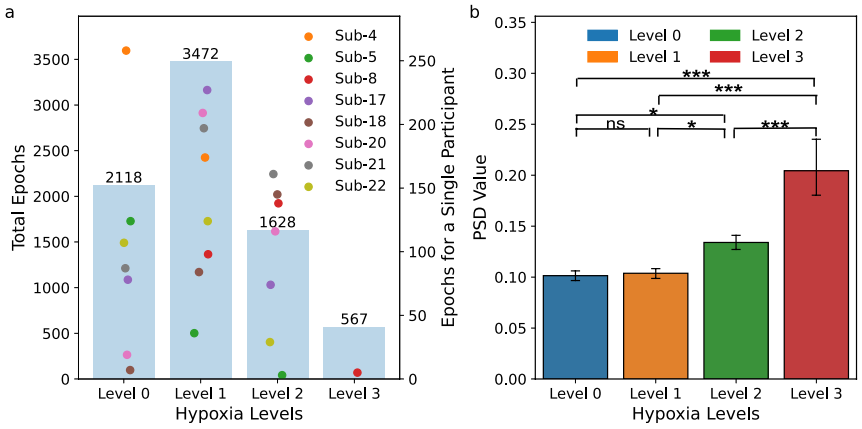

Fig. 6 The distribution of epoch counts and PSD across different levels of hypoxia. (**a**) The bar chart presents the distribution of epoch counts at various hypoxia levels in the preprocessed data. The scatter plot illustrates the distribution of epochs for each volunteer across different hypoxia levels. (**b**) represents the post-hoc Tukey HSD Test results for PSD across four hypoxia levels. Perform post-hoc multiple comparisons Tukey HSD test on the average PSD under different levels. The level of significance is indicated by asterisks: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

- When simulating an altitude close to 2.0 *km*, participants were instructed to prepare for data collection and minimize physical activity. Upon reaching 2.0 *km*, we recorded 5 *min* of eyes-open resting-state EEG. The EEG data was promptly annotated with oxygen saturation levels using the *eego*^{TM} software acquisition system. As indicated in Table 2, to maintain the simulated altitude for 5 *min*, the system employed a single pump, and

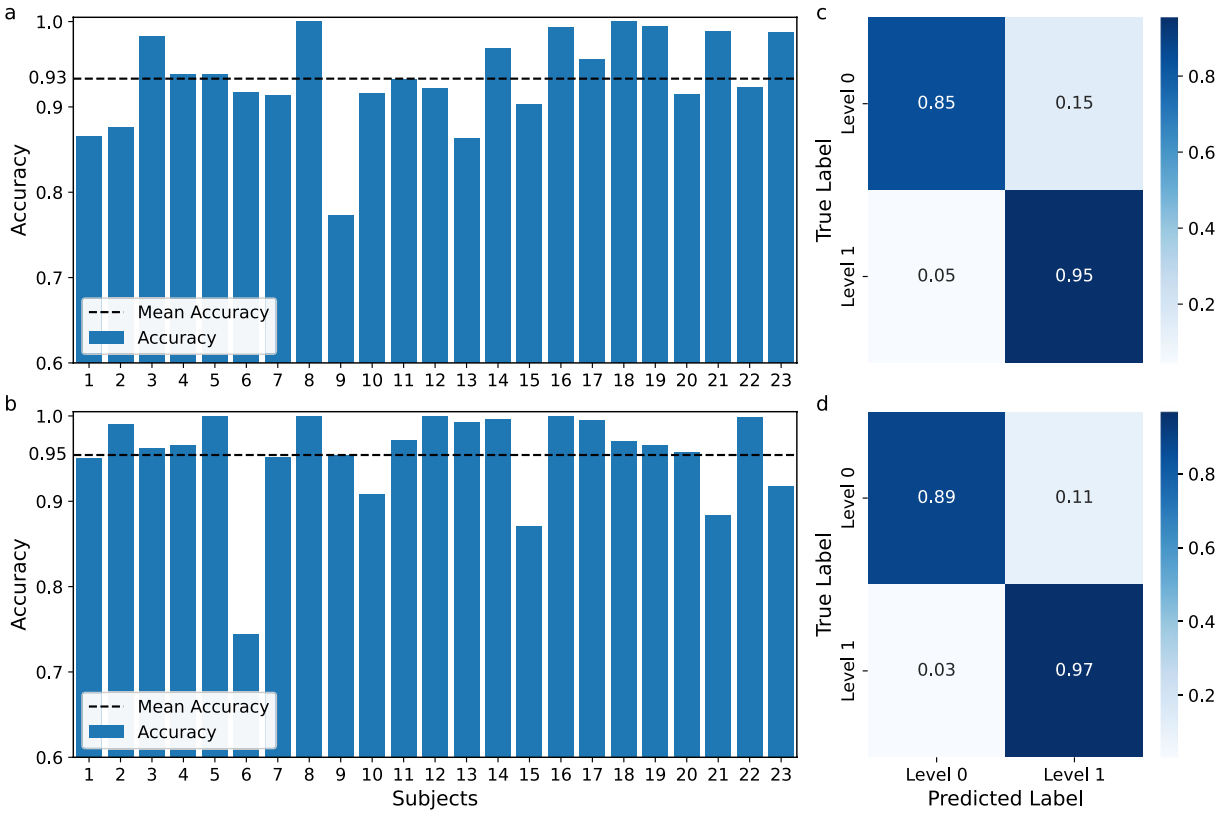

Fig. 7 The binary classification test results of SVM and EEGNet. (**a**) and (**b**) illustrate the classification accuracies for a binary-class problem using SVM and EEGNet. Each subject's classification performance is represented by the blue bars, while the black dashed line indicates the mean classification accuracy across all subjects. (**c**) and (**d**) display the mean confusion matrices from a 10-fold cross-validation for the binary-class hypoxia levels classification, using SVM and EEGNet.

the intake valve was rotated six turns for adjustment. After the stabilization phase, the second air pump was engaged for tandem operation, incrementally increasing the simulated altitude.

- Upon reaching 2.5 *km*, 3.0 *km*, and 3.5 *km*, labels were marked in the 5 *min* EEG data separately. The external researcher repeated the aforementioned steps using the parameter combinations from Table 2 to sustain the designated altitudes and collect physiological data at each specific altitude.
- Based on the details in Table 2, two pumps were operational while the intake valve was turned eight times to maintain the altitude for 5 *min*. After this period, one pump was deactivated to facilitate a decrease in simulated altitude, marking the end of the data collection phase. As the negative pressure approached zero, the researcher inside the chamber was allowed to open the inlet valve, which had been closed during the experiment, to accelerate the admission of air.

Subjects. A total of 23 subjects (10 males and 13 females) were recruited for the study. All subjects were born in plain areas and had no prior experience living in high-altitude regions. Their ages ranged from 23 to 32 years (mean = 26.52, std = 2.29). The subjects were all healthy, right-handed individuals with no history of neurological or psychiatric disorders, no other illnesses, and no medication use. They were instructed to abstain from coffee and alcohol for at least 24 *h* before the experiment and to ensure a normal sleep pattern. Prior to the commencement of the experiment, all subjects signed an informed consent form, which included consent for both participation in the experiment and the open sharing of data for scientific research purposes. The experimental procedures adhered to the guidelines of the Helsinki Declaration and were approved by the Northwestern Polytechnical University Medical and Laboratory Animal Ethics Committee (202302035). The study was conducted between November 20, 2022, and March 28, 2023. Subjects were anonymized and indexed as Sub01 ~ Sub23. Two subjects (Sub21 and Sub22) experienced physical discomfort at an altitude of 4 *km*, leading to the termination of their experiments. However, data from the completed experimental trials were retained for analysis. The data from certain trials exhibited a number of bad leads approaching or exceeding 50%, or contained a substantial amount of low-quality segments, significantly compromising the usability of the data. Consequently, the data from these trials will be excluded from the dataset.

To ensure participant safety in our experimental protocols, the following criteria were established for the premature termination of the study: (1) Researchers or participants perceived real-time simulation environments as posing safety risks. (2) Rapid increase in heart rate. (3) Oxygen saturation fell below 60%[34–36]. These guidelines

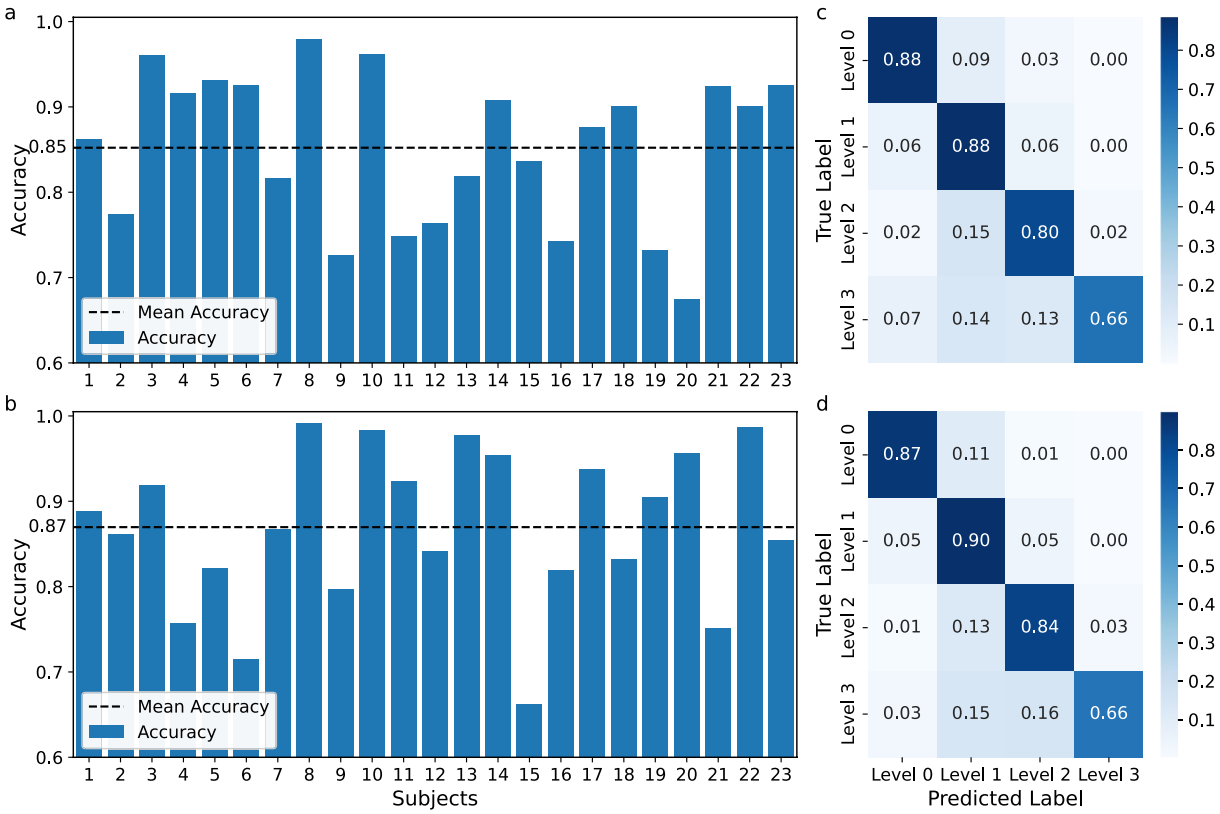

Fig. 8 The 4-category classification test results of SVM and EEGNet. (**a**) and (**b**) illustrate the classification accuracies for a 4-category problem using SVM and EEGNet. Each subject's classification performance is represented by the blue bars, while the black dashed line indicates the mean classification accuracy across all subjects. (**c**) and (**d**) display the mean confusion matrices from a 10-fold cross-validation for the 4-category hypoxia levels classification, using SVM and EEGNet.

are designed to promptly identify and mitigate any potentially dangerous physiological responses during the research activities. Additionally, participants were allowed to request the termination of the trial at any time. As a safety measure, oxygen bottles with ample supplies were available within the chamber in case of an emergency.

Data acquisition. The experimental EEG signals were acquired using a sophisticated portable EEG system (*eego*^{TM} mylab, ANT Neuro, Netherlands). A total of 64 Ag/AgCl wet electrodes, applied with conductive gel, were positioned according to the International 10–20 system standards. Throughout the experiment, the impedance remained below 10 $k\Omega$. The experiment was conducted with a sampling rate of 1000 *Hz*. The EEG data were systematically recorded with the reference electrode at CPz and the ground electrode at AFz. Notably, the EEG amplifier defaulted to these specified reference and ground electrodes. The sensor-collected signals were amplified and transmitted to a workstation for data storage.

Oxygen saturation is the primary physiological parameter used to measure participants' hypoxia levels in high-altitude scenarios, and the calculated oxygen saturation needs to be recorded. Data were collected using photoplethysmography with a finger clip oximetry device (BerryMed, BSJ09001C, China) provided with the patient monitor (BerryMed, JHY-40, China). During the experiment, the finger clip sensor was worn on the middle or index finger of the participant's left hand. Data acquisition for the oximetry waveform was executed at a sampling rate of 100 *Hz*, creating a waveform representing oxygen saturation levels with amplitudes varying between 0 and 100. These waveforms were documented at precise 0.01 *s* intervals. Concurrently, calculations of oxygen saturation and heart rate were performed every second, with their values ranging from 0 to 100.

EEG and SpO_2 data were collected simultaneously. Given the absence of other non-invasive gold standards for detecting hypoxia, oxygen saturation and heart rate serve as auxiliary measures for assessing hypoxia. SpO_2 was used to determine the hypoxia levels of the subjects, and the EEG data for all subjects were manually marked with hypoxia levels in the acquisition system. The selection of hypoxia level markers was based on the following considerations: the generally accepted normal range for SpO_2 is between 95% and 100%[37,38]. Cyanosis is typically identifiable at approximately 85% oxygen saturation[39], and SpO_2 below 75% is considered indicative of severe hypoxia[7,40]. When the termination criteria are met, personnel inside the hypoxic chamber promptly use portable oxygen bottles to provide oxygen to the subject. In Fig. 3, the horizontal axis represents time, while the shaded gray area indicates the period during which experimental data was collected. The vertical axis denotes oxygen saturation levels, with the blue line illustrating the recorded values of oxygen saturation concurrent with the EEG signal acquisition. Based on the previously described method of categorizing oxygen saturation levels,

the collected EEG signals are labeled into four levels. These four levels are marked with dashed line matrices in green, blue, yellow, and red, respectively.

Data preprocessing. The dataset was collected in a hypoxic chamber to simulate potential practical application scenarios. To validate the dataset, we conducted a preprocessing procedure for further analysis using the DISCOVER-EEG pipeline[41]. The sampling rates were first standardized to 500 *Hz*. Subsequent steps included applying a high-pass filter with a passband from 0.25 to 0.75 *Hz* to filter the data and reject bad channels. The EEG was then re-referenced to the average reference. Both ICA and interpolation were used to mitigate the effects of artifacts and bad channels. Most parameters were set to the default values of the DISCOVER-EEG pipeline, though some require special attention. The "ChannelCriterion" was set to 0.8 to exclude channels of lower quality. The "ICLabel probability threshold" was set to 0.8, removing components with high probabilities of being classified as "muscle" or "eye movement." The "FlatLineCriterion" was set to 5 seconds, which filters out channels with prolonged signal absence, possibly due to electrode malfunction. The "LineNoiseCriterion" was set to 4, meaning channels with excessive line noise were flagged as abnormal. After preprocessing, the data were segmented into epochs for further analysis. Deviating from the DISCOVER-EEG pipeline, the M1 and M2 channels were removed post-filtering. The 'linked-mastoids' reference was avoided due to potential issues with impedance changes during recording, which can be problematic[42]. To obtain clean EEG data, a FIR bandpass filter spanning 1 to 47 *Hz* was employed. Certain trials exhibited bad leads close to or exceeding 50%, or contained a substantial amount of low-quality segments, significantly impacting the usability of the data. Consequently, the data from these trials will be excluded from the dataset. For more information on preprocessing critical details, please refer to the "discover-eeg-master" in the "code" folder.

Data Records

The values 400, 2000, 2500, etc., mentioned in the article represent altitudes in meters. These values were used as grouping labels during data recording to distinguish signals from various altitudes. The segmented data is available in BIDS format and has been organized into the FGHA-EEG Dataset[43] (Finer-Grained High Altitude EEG Dataset). The structure of the dataset is illustrated in Fig. 4. In the main directory, each of the 23 participants has their own data folder, while all derived data is stored in one folder. Most participants provided EEG data recorded at six different altitudes, along with electrode information, channel details, and event markers. Notably, as in the Subjects section, the final set of data for two participants is missing due to altitude sickness. And data from some participants at certain altitudes were excluded due to quality issues. Additionally, four important text files offer detailed experimental information. The derived folder contains intermediate data generated during the processing of the raw data. Each participant's subfolder contains preprocessed data. The results, including preprocessing visualizations, feature data, reports, and related code, are stored in separate folders. The following are introductions to four key files: "dataset_description.json" provides an overview of the dataset, including details about the authors, keywords, abstract, data collection date and location, license, dataset type, and BIDS version. "participants.tsv" contains participant-related data such as sex and age. "participants.json" explains the columns present in the "participants.tsv" file. Lastly, the "README" offers general information about the dataset, including contact details.

Technical Validation

Basic dataset information. Since the dataset was recorded in a discomforting hypoxic environment, it contains some artifacts. To ensure data quality, we identified bad channels and time segments during the preprocessing phase. These steps are illustrated using Sub09 as an example. Figure 5 shows experimental data quality distribution across channels and altitudes. Additionally, the proportion of acceptable quality data segments across all altitude groups is calculated. This figure demonstrates that the obtained dataset is free from artifacts, ensuring its quality is sufficient for subsequent data analysis.

To quantify the data volume for each hypoxia level and facilitate subsequent analysis, we segmented the data into epochs using hypoxia level markers. Each epoch spans 6 seconds, with a 50% overlap to ensure continuity and comprehensive coverage between consecutive segments. The dimensions of each epoch are 61 by 3000, with 61 denoting electrode channels and 3000 indicating sampling points. The dataset comprises 7785 segments. Not all individuals experienced all four hypoxia levels. For example, hypoxia level 3 did not occur in the data of Sub05, Sub17, Sub18, and Sub20 ~ Sub22, as shown in Fig. 6(a). This is due to higher hypoxia tolerance in some individuals[44,45], which prevented their oxygen saturation levels from falling below 75% during the experiment. Alternatively, some data may have been removed due to quality issues. The proportions of data volumes for hypoxia levels 0, 1, 2, and 3 are approximately 3.73:6.12:2.87:1.00, as shown in Fig. 6(a). This indicates an imbalance in sample sizes among the different categories.

Power spectral density of four levels. To investigate potential differences in power spectral density (PSD) across four hypoxia levels, we extracted the PSDs of 61 channels from preprocessed data. Specifically, we calculated the PSD of each data segment at every level using Welch's method and averaged the results to obtain the mean PSD for each hypoxia level. We then performed a variance analysis and pairwise comparisons of the average PSDs using ANOVA and post-hoc Tukey HSD tests. This allowed us to assess overall differences across hypoxia levels and specific differences between pairs. Results were considered statistically significant for p values < 0.05. The ANOVA indicated significant differences in average PSDs among the levels ($F(3, 61240) = 38.368$, $p < 0.001$), suggesting that different hypoxia levels can be reliably identified using our dataset. The post-hoc Tukey HSD test revealed no significant difference between levels 0 and 1 ($p = 0.996$), but significant differences were observed among other level pairs, as shown in Fig. 6(b). This confirms the validity and effectiveness of hypoxia level division in our dataset.

Classification analysis. Classification analysis was conducted to validate the dataset's effectiveness. We aimed to determine if the dataset could support finer-grained recognition by classifying 4 hypoxia levels on a 6 *s* time scale. A 10-fold cross-validation procedure was applied to the preprocessed data. Figure 6(a) shows that Sub05 had fewer than 10 epochs at hypoxia level 2, affecting data partitioning in intra-subject recognition. To ensure a similar distribution between training and test data, categories with fewer than 10 epochs were excluded. For each subject, the dataset was divided based on class proportions, with 90% of the EEG data used for training and 10% for testing.

We conducted two types of validation: binary classification and 4-category classification recognition. First, the classical method based on PSD features combined with Support Vector Machine (SVM) was used for hypoxia recognition. This method extracts PSD features from the raw EEG data and then applies the SVM algorithm for binary and 4-category classification. The procedure was repeated 10 times, and classification performance was assessed by averaging accuracies across 10-fold. The binary classification accuracy for hypoxia recognition was 92.59 $\pm$ 0.20% (mean $\pm$ standard error, with all subsequent classification accuracies reported similarly). Additionally, we employed a deep learning algorithm, EEGNet, for recognizing hypoxia levels. A binary classification accuracy of 94.82 $\pm$ 0.12% was achieved with a 10-fold procedure. The binary classification accuracies of our dataset outperformed the performance of a previous study[26]. For the 4-category classification, the SVM and EEGNet models achieved recognition accuracies of 84.89 $\pm$ 0.29% and 86.16 $\pm$ 0.15%, respectively.

Figure 7(a),(b), illustrate the binary classification accuracies for each participant using the SVM and EEGNet models. The EEG data from 23 subjects yielded an average accuracy of 93% using the SVM algorithm, while the EEGNet model achieved an average accuracy of 95%. These figures highlight the variability in accuracy obtained from data across different individuals. Figure 7(c),(d) present the confusion matrices for binary hypoxia level recognition using the SVM and EEGNet models. Figure 8(a),(b) show that the average accuracy for the 4-category classification across 23 participants is 85% for the SVM algorithm and 87% for the EEGNet model. Figure 8(c),(d) display the confusion matrices for the 4-category hypoxia levels recognition using the SVM and EEGNet models. As expected, the performance in 4-category classification is lower than in binary classification, which aligns with previous research findings[16]. These results confirm the feasibility of using EEG signals to differentiate between four hypoxia levels. Analysis of the average confusion matrices reveals that EEGNet outperforms SVM in recognizing imbalanced data. Both models encounter difficulties distinguishing hypoxia level 3 from the other levels, likely due to the smaller data volume for hypoxia level 3, which may prevent the models from adequately learning its characteristics.

Usage Notes

The code based on DISCOVER-EEG[41] (https://github.com/crisglav/discover-eeg) was modified for the preprocessing of this dataset. This dataset provides EEG recordings at different altitudes along with detailed hypoxia level markers, enabling researchers to view and manipulate raw data for specific research purposes. Utilizing the EEG dataset, which includes altitude and hypoxia level labels, allows for the extraction of additional features such as phase, PSD, and brain functional connectivity. This dataset holds significant potential for various applications in cognitive neuroscience and high-altitude hypoxia recognition research, including: (1) mining neuroelectrophysiological markers for the diagnosis and treatment of hypoxia, (2) developing EEG-based fine-grained identification algorithms for hypoxia, (3) understanding the underlying mechanisms of hypoxia's effects on brain function.

Although our dataset was collected from simulated high-altitude environments, the findings can serve as a reference for assessing human hypoxia levels in other oxygen-limited environments, such as aviation and deep-sea exploration. However, it is important to be aware of certain limitations. The participants are healthy young adults, so the findings may not be applicable to other populations, such as children and the elderly. To ensure participant safety, the altitude in this dataset was limited to 4 *km* and below for only 5 *min*. Consequently, these findings may not apply to higher altitudes. Additionally, individual differences during the 5 *min* data collection period may result in some participants not exhibiting the expected hypoxia levels. The dataset was collected by segmenting data at specific altitudes and time intervals rather than through continuous recording during altitude ascent, thereby limiting its use for research on consecutive altitude increases.

Code availability

For preprocessing, we utilized the DISCOVER-EEG pipeline. Data processing and visualization were conducted using Python 3.10, compatible with both Linux and Windows systems. The Python libraries employed for data processing and analysis included NumPy (1.26.4), h5py (3.10.0), Statsmodels (0.14.1), Pandas (2.2.2), and MNE (1.6.1). For machine learning and deep learning, we used Scikit-learn (1.2.2) and TensorFlow (2.15.0) for data classification. Data visualization was performed with Matplotlib (3.7.5). All custom code is available in the GitHub repository https://github.com/siyingjun007/FGHA-EEG-Dataset-Code.

Received: 19 June 2024; Accepted: 7 November 2024;
Published online: 18 December 2024

References

1. Encyclopedia britannica, plateau summary https://www.britannica.com/summary/plateau-landform (2003).
2. Wu, T. & Kayser, B. High altitude adaptation in tibetans. *High Altitude Medicine & Biology* **7**, 193–208, https://doi.org/10.1089/ham.2006.7.193 (2006).
3. Tremblay, J. C. & Ainslie, P. N. Global and country-level estimates of human population at high altitude. *Proceedings of the National Academy of Sciences* **118**, e2102463118, https://doi.org/10.1073/pnas.2102463118 (2021).
4. Richalet, J.-P., Hermand, E. & Lhuissier, F. J. Cardiovascular physiology and pathophysiology at high altitude. *Nature Reviews Cardiology* **21**, 75–88, https://doi.org/10.1038/s41569-023-00924-9 (2024).
5. Li, Y. & Liu, Y. Oxygen enrichment and its application to life support systems for workers in high-altitude areas. *International journal of occupational and environmental health* **20**, 207–214, https://doi.org/10.1179/2049396714Y.0000000068 (2014).
6. Zhang, X. & Zhang, J. The human brain in a high altitude natural environment: A review. *Frontiers in Human Neuroscience* **16**, 915995, https://doi.org/10.3389/fnhum.2022.915995 (2022).

7. Chen, X. *et al*. Mechanism, prevention and treatment of cognitive impairment caused by high altitude exposure. *Frontiers in Physiology* **14**, 1191058, https://doi.org/10.3389/fphys.2023.1191058 (2023).
8. Burtscher, J., Mallet, R. T., Pialoux, V., Millet, G. P. & Burtscher, M. Adaptive responses to hypoxia and/or hyperoxia in humans. *Antioxidants & redox signaling* **37**, 887–912, https://doi.org/10.1089/ars.2021.0280 (2022).
9. He, Y. *et al*. High arterial oxygen saturation in the acclimatized lowlanders living at high altitude. *Phenomics* **3**, 329–332, https://doi.org/10.1007/s43657-023-00117-x (2023).
10. Longmore, S. K. *et al*. A comparison of reflective photoplethysmography for detection of heart rate, blood oxygen saturation, and respiration rate at various anatomical locations. *Sensors* **19**, 1874, https://doi.org/10.3390/s19081874 (2019).
11. Siwicka-Gieroba, D., Robba, C., Gołacki, J., Badenes, R. & Dabrowski, W. Cerebral oxygen delivery and consumption in brain-injured patients. *Journal of Personalized Medicine* **12**, 1763, https://doi.org/10.3390/jpm12111763 (2022).
12. Aboouf, M. A., Thiersch, M., Soliz, J., Gassmann, M. & Schneider Gasser, E. M. The brain at high altitude: from molecular signaling to cognitive performance. *International Journal of Molecular Sciences* **24**, 10179, https://doi.org/10.3390/ijms241210179 (2023).
13. Virués-Ortega, J., Garrido, E., Javierre, C. & Kloezeman, K. C. Human behaviour and development under high-altitude conditions. *Developmental science* **9**, 400–410, https://doi.org/10.1111/j.1467-7687.2006.00505.x (2006).
14. Aboalayon, K. A. I., Faezipour, M., Almuhammadi, W. S. & Moslehpour, S. Sleep stage classification using eeg signal analysis: a comprehensive survey and new investigation. *Entropy* **18**, 272, https://doi.org/10.3390/e18090272 (2016).
15. Zhang, H. *et al*. The applied principles of eeg analysis methods in neuroscience and clinical neurology. *Military Medical Research* **10**, 67, https://doi.org/10.1186/s40779-023-00502-7 (2023).
16. Chen, J. *et al*. A large finer-grained affective computing eeg dataset. *Scientific Data* **10**, 740, https://doi.org/10.1038/s41597-023-02650-w (2023).
17. O'toole, J. M. *et al*. Neonatal eeg graded for severity of background abnormalities in hypoxic-ischaemic encephalopathy. *Scientific Data* **10**, 129, https://doi.org/10.1038/s41597-023-02002-8 (2023).
18. Zhang, X. *et al*. A high altitude respiration and spo2 dataset for assessing the human response to hypoxia. *Scientific Data* **11**, 248, https://doi.org/10.1038/s41597-024-03065-x (2024).
19. Zhang, X., Zhang, Y., Si, Y. & Yang, H. Respiration and spo2 dataset for assessing the human response to hypoxia at high altitude. *Figshare* https://doi.org/10.6084/m9.figshare.c.6623344.v1 (2024).
20. Zhang, Y. *et al*. 40 hz light flicker alters human brain electroencephalography microstates and complexity implicated in brain diseases. *Frontiers in Neuroscience* **15**, 777183, https://doi.org/10.3389/fnins.2021.777183 (2021).
21. Subha, D. P., Joseph, P. K., Acharya U, R. & Lim, C. M. Eeg signal analysis: a survey. *Journal of medical systems* **34**, 195–212, https://doi.org/10.1007/s10916-008-9231-z (2010).
22. Hu, M., Li, J., Li, G. & Freeman, W. J. Analysis of early hypoxia eeg based on a novel chaotic neural network. In *Neural Information Processing: 13th International Conference, ICONIP 2006, Hong Kong, China, October 3-6, 2006. Proceedings, Part I 13*, 11–18, https://doi.org/10.1007/11893028_2 (Springer, 2006).
23. Hu, M., Li, J., Li, G., Tang, X. & Ding, Q. Classification of normal and hypoxia eeg based on approximate entropy and welch power-spectral-density. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 3218–3222, https://doi.org/10.1109/IJCNN.2006.247307 (IEEE, 2006).
24. Zhang, J., Li, G., Hu, M., Li, J. & Luo, Z. Recognition of hypoxia eeg with a preset confidence level based on eeg analysis. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 3005–3008, https://doi.org/10.1109/IJCNN.2008.4634221 (IEEE, 2008).
25. Zhang, T., Wang, Y. & Li, G. Effect of intermittent hypoxic training on hypoxia tolerance based on single-channel eeg. *Neuroscience letters* **617**, 39–45, https://doi.org/10.1016/j.neulet.2016.01.063 (2016).
26. Liu, H. *et al*. Machine learning based on event-related eeg of sustained attention differentiates adults with chronic high-altitude exposure from healthy controls. *Brain Sciences* **12**, 1677, https://doi.org/10.3390/brainsci12121677 (2022).
27. West, J. B. High-altitude medicine. *American journal of respiratory and critical care medicine* **186**, 1229–1237, https://doi.org/10.1164/rccm.201207-1323CI (2012).
28. Bärtsch, P. & Saltin, B. General introduction to altitude adaptation and mountain sickness. *Scandinavian journal of medicine & science in sports* **18**, 1–10, https://doi.org/10.1111/j.1600-0838.2008.00827.x (2008).
29. Sharma, P., Pandey, P., Kumari, P. & Sharma, N. K. Introduction to high altitude and hypoxia. In *High Altitude Sickness–Solutions from Genomics, Proteomics and Antioxidant Interventions*, 1-17, https://doi.org/10.1007/978-981-19-1008-1_1 (Springer, 2022).
30. Schoene, R. B. Illnesses at high altitude. *Chest* **134**, 402–416, https://doi.org/10.1378/chest.07-0561 (2008).
31. Ladrido-Ignacio, L. *et al*. Randomised, controlled trial of ginkgo biloba and acetazolamide for prevention of acute mountain sickness: the prevention of high altitude illness trial (phait). *bmj* **38043**, 7C, https://doi.org/10.1136/bmj.38043.501690.7C.
32. Xie, Y. *et al*. Association between arterial blood gas variation and intraocular pressure in healthy subjects exposed to acute short-term hypobaric hypoxia. *Translational Vision Science & Technology* **8**, 22–22, https://doi.org/10.1167/tvst.8.6.22 (2019).
33. Netzer, N. C. *et al*. Spo2 and heart rate during a real hike at altitude are significantly different than at its simulation in normobaric hypoxia. *Frontiers in physiology* **8**, 81, https://doi.org/10.3389/fphys.2017.00081 (2017).
34. Bustamante-Sánchez, A. *et al*. Effects of hypoxia on selected psychophysiological stress responses of military aircrew. *BioMed Research International* **2021**, https://doi.org/10.1155/2021/6633851 (2021).
35. Rice, G. M. *et al*. Gender differences in dry-eeg manifestations during acute and insidious normobaric hypoxia. *Aerospace Medicine and Human Performance* **90**, 369–377, https://doi.org/10.3357/AMHP.5227.2019 (2019).
36. Malle, C. *et al*. Working memory impairment in pilots exposed to acute hypobaric hypoxia. *Aviation, space, and environmental medicine* **84**, 773–779, https://doi.org/10.3357/asem.3482.2013 (2013).
37. Gong, G. *et al*. Study of an oxygen supply and oxygen saturation monitoring system for radiation therapy associated with the active breathing coordinator. *Scientific reports* **8**, 1254, https://doi.org/10.1038/s41598-018-19576-8 (2018).
38. Kane, B., Decalmer, S. & O'Driscoll, B. R. Emergency oxygen therapy: from guideline to implementation. *Breathe* **9**, 246–253, https://doi.org/10.1183/20734735.025212 (2013).
39. Langley, R. & Cunningham, S. How should oxygen supplementation be guided by pulse oximetry in children: do we know the level? *Frontiers in pediatrics* **4**, 138, https://doi.org/10.3389/fped.2016.00138 (2017).
40. Wong, N. Y., van Waart, H., Sleigh, J. W., Mitchell, S. J. & Vrijdag, X. C. A systematic review of electroencephalography in acute cerebral hypoxia: clinical and diving implications. *Diving and Hyperbaric Medicine* **53**, 268, https://doi.org/10.28920/dhm53.3.268-280 (2023).
41. Gil Ávila, C. *et al*. Discover-eeg: an open, fully automated eeg pipeline for biomarker discovery in clinical neuroscience. *Scientific Data* **10**, 613, https://doi.org/10.1038/s41597-023-02525-0 (2023).
42. Jobert, M. *et al*. Guidelines for the recording and evaluation of pharmaco-sleep studies in man: the international pharmaco-eeg society (ipeg). *Neuropsychobiology* **67**, 127–167, https://doi.org/10.1159/000343449 (2013).
43. Si, Y., Zhang, X., Zhang, Y. & Yang, H. A finer-grained high altitude eeg dataset for hypoxia levels assessment. *Figshare* https://doi.org/10.6084/m9.figshare.xxxxx (2025).
44. Chapman, R. F. The individual response to training and competition at altitude. *British Journal of Sports Medicine* **47**, i40–i44, https://doi.org/10.1136/bjsports-2013-092837 (2013).
45. Burykh, E. The problem of assessing individual sensitivity and tolerance to hypoxia in animals and humans. *Journal of Evolutionary Biochemistry and Physiology* **55**, 339–347, https://doi.org/10.1134/S0022093019050016 (2019).

Acknowledgements
This study received financial support from the National Natural Science Foundation of China (Grant Nos. 62172336 and 62032018), as well as the Natural Science Foundation of Ningbo (Grant No. 202003N4057).

Author contributions
The experiment was conceived by H.Y., Y.Z., X.Z. and Y.S.. Specific experiments were carried out by Y.S., X.Z., and H.Z., while Y.S. recording the timestamps of key altitudes and collating the raw data files. Y.S. processed and analyzed the data, organized the dataset and codes into an open access format, and drafted the manuscript. Y.S. collaborated with X.Z. on data visualization. X.Z. was primarily responsible for revising the manuscript, while Y.Z., H.Y. and S.L. provided significant feedback on manuscript revisions and formatting standardization. All authors reviewed the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Y.Z., X.Z. or H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.