

# Methodologies Using Artificial Intelligence to Detect Cognitive Decrements in Aviation Environments

G. Merrill Rice; Steven Linnville; Dallas Snider

- INTRODUCTION:** Despite significant advancements in aerospace engineering and safety protocols over the last decade, U.S. Naval mishap rates have remained essentially unchanged. This paper explores how researchers may leverage current artificial intelligence (AI) technologies to enhance aviation safety.
- METHODS:** A critical review was performed identifying aviation research protocols which have incorporated machine learning (ML) to enhance the accuracy of detecting common aviation hazards leading to cognitive decrements. The review proposes a three-step methodology for creating protocols to identify cognitive decrements in aviators: 1) sensor selection; 2) preprocessing techniques; and 3) ML algorithm development. Natural language processing was utilized to assist with the development of aviation-related denoising and ML algorithm tables.
- RESULTS:** Several psychophysiological biosensors, enhanced by ML modeling, show promise in identifying cognitive deficits secondary to fatigue, hypoxia, and spatial disorientation. The most cited biosensors integrated with ML models include electroencephalographic, electrocardiographic, and eye-tracking devices. The application of preprocessing techniques to biosensor data is a critical methodological step prior to applying ML algorithms for data training and classification. ML algorithms utilized were categorized into supervised, unsupervised, and semi-supervised types, often used in combination for more accurate predictions.
- DISCUSSION:** Current literature suggests that AI, when used in conjunction with various psychophysiological sensors, can predict and potentially mitigate common aeromedical hazards such as fatigue, spatial disorientation, and hypoxia in simulated settings. The miniaturization of preprocessing and ML algorithmic hardware is the next phase of transitioning AI to operational environments for real-time continuous monitoring.
- KEYWORDS:** aviation safety, machine learning, denoising, cognitive deficits, hazard identification.

Rice GM, Linnville S, Snider D. *Methodologies using artificial intelligence to detect cognitive decrements in aviation environments*. *Aerosp Med Hum Perform*. 2025; 96(4):327–338.

Over the last decade, rates of U.S. Naval aviation Class A mishaps for all platforms have remained essentially constant, averaging 1.41 per 100,000 flight hours (Fig. 1).<sup>1</sup> Resultant loss of life and total cost of all naval aviation mishaps over this last decade has been 100 fatalities and over \$8 billion.<sup>2</sup> At the heart of these events lie human factors, which have consistently contributed to upwards of 80% of these mishaps.<sup>3</sup> For the U.S. Navy, among the potentially detectable aeromedical preconditions for these human factors involving Class A, B, and C mishaps, the top three include fatigue, spatial disorientation (SD), and respiratory/physiological events<sup>3</sup> (Fig. 2). The common denominator for these preconditions is a degradation of the aviator's cognitive performance. How might aeromedical researchers leverage the use of artificial intelligence (AI) to

combat the most common detectable human factors that contribute to aviation mishaps in real time?

Generally, there are several ways AI may be broken down categorically into various classifications (Fig. 3). For example, Mukhamediev recently categorized AI into seven various

From the Naval Safety Command, Norfolk, VA, Naval Medical Research Unit Dayton, Wright-Patterson Air Force Base, OH, and University of West Florida, Pensacola, FL.

This manuscript was received for review in July 2024. It was accepted for publication in November 2024.

Address correspondence to: Dr. G. Merrill Rice, 375 A Street, Norfolk, VA 23511, United States; gmerrillrice@gmail.com.

Copyright © by The Authors.

This article is published Open Access under the CC-BY-NC license.

DOI: <https://doi.org/10.3357/AMHP.6555.2025>



## CLASS A AVIATION MISHAPS

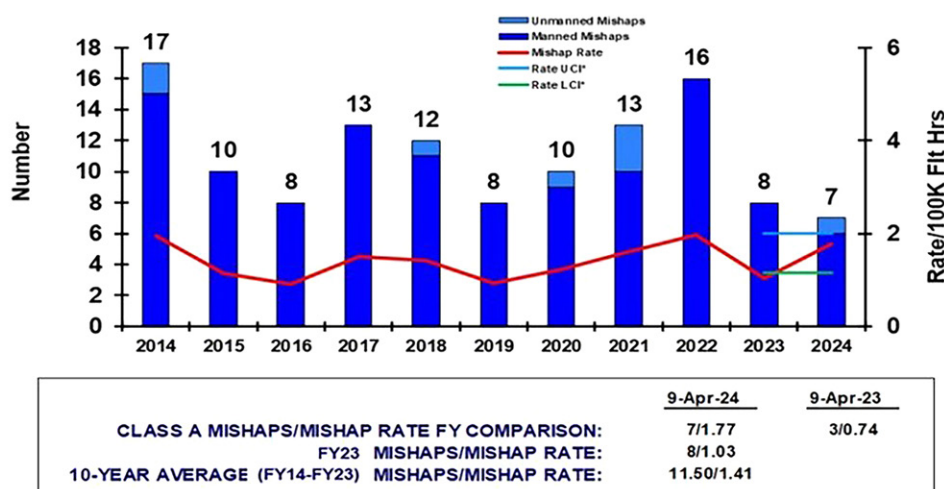


Fig. 1. U.S. Navy Class A mishap rates FY2014Q1 – FY2024Q1.

classifications: machine learning (ML), natural language processing (NLP), planning, robotics, expert systems, speech, and vision recognition.<sup>4</sup> The underlying foundation for each of these categories is the development of ML algorithms and computational models that enable machines to simulate intelligent behavior. Although the development of these algorithms is central to simulating and identifying the aviation hazard or precondition one seeks to model, there are other key steps in protocol development that are essential prior to using AI to mitigate a potential mishap.

The objective of this paper is to provide a critical review of the available research integrating psychophysiological sensors with ML algorithms to detect cognitive decrements in pilots. Psychophysiological sensors, as the name implies, are biosensors that in some way convey the cognitive state of the subject through physiological measurement; for brevity purposes, these will be deemed “biosensors.” Several systematic reviews of the literature on the topic, specifically with regards to the identification of aviation hazards, suggest a three-step methodology by which researchers create the foundations of their protocols

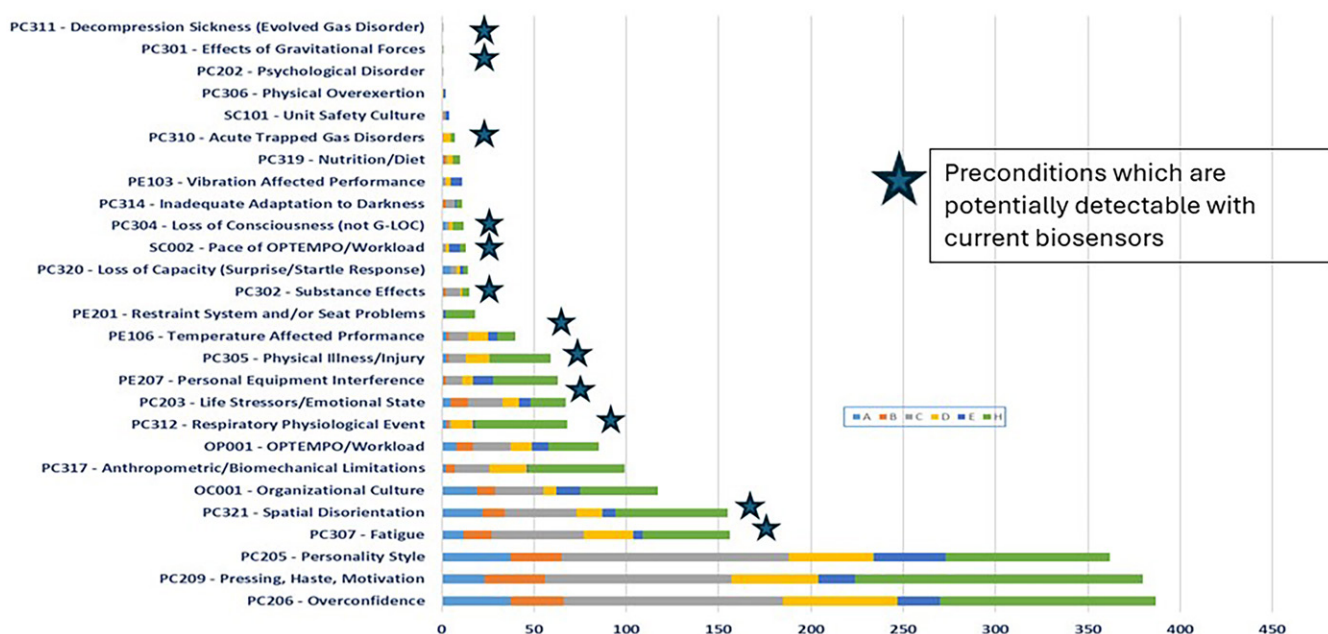
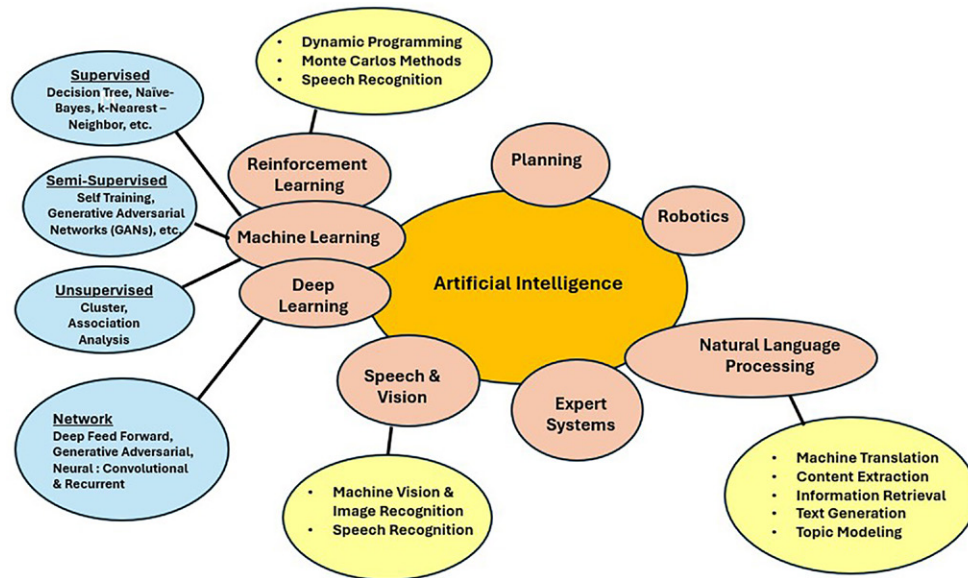


Fig. 2. HFACS 8.0, leading aeromedical preconditions associated with Class A, B, and C Naval mishaps 2013-FY2024.



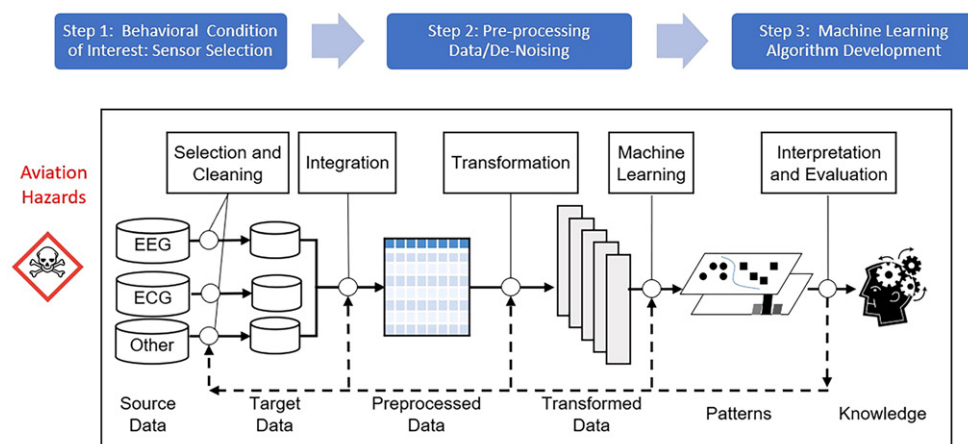
**Fig. 3.** Subcategories of artificial intelligence (AI).

to identify cognitive decrements for aviators.<sup>5–7</sup> These steps may be broken down to include: hazard identification/sensor selection; preprocessing or denoising techniques; and development of ML algorithms. Within the following paragraph, to illustrate this methodology for using AI to identify cognitive decrements in the cockpit, we will provide a synopsis of a recently published aeromedical protocol whose objective was to identify a precondition (hypoxia) that could conceivably result in cognitive decline and potentially a mishap.<sup>8,9</sup>

Realizing the need for real-time sensors to detect cognitive performance decrements in the cockpit, Rice et al., evaluated dry electroencephalograms' (EEG) ability to detect hypoxia.<sup>8</sup> As compared to wet EEG, dry EEG, as the name implies, does not require extensive preparation of the subject to connect and does not require transducer gel to improve signal transduction. Both advantages lend themselves to transitioning this technology to an operational environment. Their research suggested that a reduction in overall dry-EEG power could identify hypoxia in lieu of aviators not

recognizing their own meaningful decreases in oxygen saturation and cognitive performance.<sup>8</sup> Snider et al. advanced this work further by reducing the variance of the data sets through the preprocessing technique of principal component analysis (PCA) and then applying various ML algorithms, such as decision tree (DT), neural network (NN), and naïve Bayes (NB).<sup>9</sup> By doing so, these researchers increased the sensitivity and specificity of dry-EEG technology to detect hypoxia to greater than 97%.

Utilizing the framework described in the previous aeromedical protocol example, which is consistent with the methodology of most protocols found within available systematic reviews on aviation safety and ML,<sup>5–7</sup> we find three common steps in the aforementioned process of using AI to prevent aviation mishaps. These are: 1) biosensor selection; 2) preprocessing of data/denoising; and 3) ML algorithm development. Below in **Fig. 4**, a schematic process of AI's use to identify hazards which may affect pilots' cognitive performance and subsequently result in mishap is illustrated.



**Fig. 4.** Methodology steps of integrating artificial intelligence (AI) with psychophysiological sensors.

**Table I.** Denoising Techniques for Time Series Data.

TECHNIQUE	DESCRIPTION
Notch Filtering <sup>10</sup>	Removes specific frequency components, e.g., 50 or 60 Hz power line noise.
Band-Pass Filtering <sup>11</sup>	Retains frequencies within a specific range (e.g., 0.5–50 Hz) & removes frequencies outside it.
Adaptive Filtering <sup>12</sup>	Uses adaptive algorithm to remove noise by adjusting its parameters based on signal characteristics.
Spatial Filtering <sup>13</sup>	CSP & Surface Laplacian enhance the signal-to-noise ratio by highlighting specific spatial signals while reducing noise.
CSP <sup>14</sup>	CSP identifies patterns to distinguish between conditions or classes, and commonly used in BCI to understand which parts of the brain are active under different conditions.
SL <sup>15</sup>	SL enhances local signal differences in the brain to reduce noise and highlight important information, helping to pinpoint EEG signal origins.
WT <sup>16</sup>	WT decomposes EEG signals into frequency components at different resolutions, enabling noise removal—effective for eliminating transient artifacts.
EMD <sup>17</sup>	EMD decomposes EEG signals into different oscillatory modes known as IMFs, allowing noise removal by analyzing and reconstructing the signal from selected IMFs.
BSS <sup>18</sup>	BSS uses denoising techniques like CCA and FastICA to separate EEG signals into independent sources, isolating and removing noise.
CCA <sup>19,20</sup>	CCA identifies relationships between two variable sets, helping to detect individual sources in a mixture.
FastICA <sup>21</sup>	FastICA efficiently separates mixed signals into independent sources by identifying statistically independent signals.
Regression <sup>22</sup>	Regression methods can model and subtract noise, such as using EOG signals to remove eye movement artifacts from EEG signals.
ICA <sup>23</sup>	ICA separates a multivariate signal into independent components, effectively isolating artifacts like eye blinks, muscle noise, and other interferences.
PCA <sup>9,24</sup>	PCA reduces EEG data dimensionality by transforming it into uncorrelated principal components, aiding in noise identification and removal.
ML Approaches <sup>25</sup>	Recent advances use ML models, including neural networks, to identify and remove noise from EEG signals, trained to filter out specific artifacts.

BCIs: brain computer interfaces; BSS: blind source separation; CCA: canonical correlation analysis; CSP: common spatial patterns; EMD: empirical mode decomposition; EOG: electrooculogram; FastICA: fast independent component analysis; ICA: independent component analysis; IMFs: intrinsic mode functions; ML: machine learning; PCA: principal component analysis; SL: surface Laplacian; WT: wavelet transform

Acknowledging that most aeromedical researchers may not be routinely exposed to denoising techniques, such as PCA, or ML algorithms, such as DT or NB, we have developed quick reference tables to orientate the reader as to their purpose when these terms arise (**Tables I and II**). The overarching goal of this review is that aeromedical practitioners may use this paper as a blueprint for future research involving AI to identify and mitigate cognitive performance decrements in the cockpit.

## METHODS

For each of the above methodological steps, biosensor selection, preprocessing techniques, and ML algorithm development, we performed a critical review of the literature,

identifying aviation-applicable citations that would provide the reader a basic conceptual understanding of how current researchers are integrating ML models with psychophysiological sensors to identify cognitive deficits.

Specifically, within “Preprocessing and Denoising Techniques” in the development of Table I, we utilized the NLP program Chat GPT v. 3.5 (San Francisco, CA, United States), with the query “Denoising techniques for EEG” as a starting point, and cross-referenced this list with published aviation-applicable references and protocols (see **Supplement A**, which can be found in the online version of this article). ChatGPT is an AI tool that responds to user questions and can handle a variety of tasks, making it more flexible than traditional AI systems that are designed for specific functions like face recognition or playing chess. In some ways, it mimics human thinking,

**Table II.** Description of the Three Major Types of Machine Learning.

TYPE	LEARNING	PREDICTION	ALGORITHM EXAMPLES
Supervised <sup>26–28</sup> : Trains with labeled input data	Algorithm goes through the examples and adjusts its internal rules to improve predictions.	After training, predictions are made on previously unseen data without labels.	LR, Log R, DT, NN, NB, LSTM, k-NN, LDA, CNN, GB, RF
Unsupervised <sup>29–31</sup> : Uses unlabeled input data during training	Algorithm organizes the input data into clusters based on common features, without knowing what those clusters are.	After learning the clusters, the algorithm places new data into the defined clusters discovered.	Clustering (e.g., k-means, hierarchical clustering), Dimensionality Reduction (e.g., PCA), Association Rule Learning
Semi-supervised <sup>32–35</sup> : Uses both labeled and unlabeled data during training	Algorithm first uses the labeled data and then makes guesses about the unlabeled data. As it processes the unlabeled data, it improves its learning by finding more patterns in both labeled and unlabeled examples.	Once trained, the algorithm makes predictions on new, unseen data using the patterns learned during training to make more accurate predictions, even though most of its training data didn't have labels.	Self-Training (1 model), Co-Training (2 models), Semi-Supervised Clustering (modified k-means with labeled/unlabeled data), Transductive Support Vector Machine (boundary developed between labeled/unlabeled data), Labeled propagation (building a similarity network using labeled data to identify patterns in unlabeled data)

LR: linear regression; LogR: logistic regression; DT: decision tree; RF: random forest; NB: naïve Bayes; NN: neural network; GB: gradient boosting; kNN: k-nearest neighbor; LDA: linear discriminant analysis; CNN: convoluted neural network; LSTM: long short-term memory



known as artificial general intelligence. The information that is generated from ChatGPT is not the final product and often requires editing and cross-referencing. The supplement provided demonstrates the output provided to the user from ChatGPT and the editing and verification required to present this data in a scientifically valid format.

## RESULTS

### Biosensor Selection

There have been numerous biosensors utilized to assess cognitive states of pilots over the last decade.<sup>5,36</sup> For example, EEG, electrocardiograms (ECG), galvanic stimulation recorder (GSR), near-infrared spectroscopy (NIRS), electrooculograms (EOG), electromyogram (EMG), and eye-tracking (ET) have all been used directly to monitor cognitive states or indirectly as surrogate markers of current or impending cognitive deficits.<sup>36</sup> A preponderance of the recent research involving ML and aviation has relied upon noninvasive dry or wet EEG because of its temporal ability to monitor cognitive states directly in real time.<sup>5,37,38</sup> Subsequently, this review will focus on EEG as the primary biosensor used in combination with ML, and to a lesser extent, other sensors such as ECG and ET will also be discussed.

As we will be focusing much of the discussion on the interpretation of EEG data, it is appropriate to provide a brief primer as to how EEG data is typically characterized. Classically, neuroscientists describe the various frequencies of brainwaves from highest to lowest as gamma ( $\gamma$ ), beta ( $\beta$ ), alpha ( $\alpha$ ), theta ( $\theta$ ), and delta ( $\delta$ ).<sup>39,40</sup> The frequencies have been correlated with various levels of cognitive functioning:  $\gamma$  (38–100 Hz) associated with high levels of cognitive processing;  $\beta$  (16–38 Hz) associated with alertness and concentration;  $\alpha$  (8–16 Hz) associated with relaxation and calmness;  $\theta$  (3–8 Hz) associated with meditation and presleep states; and  $\delta$  (1–3 Hz) associated with deep sleep and cognitive disorders.<sup>40,41</sup> These ranges vary slightly upon which references you read; however, in general,

they tend to be consistent at identifying predominant cognitive states. A systematic review of the current research involving EEG indices to access cognitive human performance suggests that the power of these individual frequencies, and to a lesser degree, the amplitudes of event-related potentials (ERPs: i.e., stimuli-induced, millisecond “snapshots” of EEG), are the primary features of EEGs extracted to identify performance decrements.<sup>37</sup> These spatiotemporal changes in EEG frequency, power, and amplitude can be exploited and introduced within ML models to accurately predict in real time the mental states of those monitored. Moreover, there are several types of EEG sensors that have been regularly used for research purposes which are cited in the literature. A nonexhaustive list of these products was recently compiled using Google Scholar by Liu in his most recent review of cognitive neuroscience and robotics and updated for this paper (Table III).<sup>38</sup>

One of the first papers to enhance the interpretation of biosensor data with ML algorithms was performed by Harrivel et al.<sup>42</sup> Noting that most commercial aviation accidents were due to a loss of flight crew airplane state of awareness, her team evaluated attention-related human performance limiting states (AHPLS) of 24 commercial pilots with multimodal psychophysiological sensing. Extracting features from five different biological sensing modalities [EEG, heart rate variability (HRV), ECG, respiration, and GSR], they identified unique indices of attention for each modality and subsequently trained ML algorithms to accurately identify AHPLS. Specifically for EEG, they extracted power spectral density (PSD) estimates for the various brain wave frequencies 1–40 Hz and their corresponding EEG channels. PSD, as the name implies, refers to the distribution of power into individual frequency components of the signal, and has been used extensively in neuroscience research to better classify epileptic seizures.<sup>43</sup> The selected features of each sensor modality in Harrivel’s study were then trained on four ML models: 1) random forest (RF); 2) gradient boosting (GB); 3) Nu-Support Vector Machine (Nu-SVM); and 4) polynomial kernels. They noted that using the combination of EEG, respiration, and GSR features provided them the best

**Table III.** Summary of Popular EEG Headsets (Statistics 7/15/2024).

MANUFACTURER (SENSOR HEADSET)	ELECTRODES	SAMPLING RATE	CITATION REF.
Compumedics Neuroscan	32/64/128/256	Up to 20 KHz	21,000 +
Electrical Geodesics Inc	32/64/128/256	1 KHz	15,800+
BioSemi	Up to 256	Up to 16 KHz	14,500+
Emotiv	2/5/14	Up to 8 KHz	13,600+
NeuroSky	Up to 256	1 KHz	7000+
Neuroelectrics	8/20/32	Up to 500 Hz	5160+
Brain Products	Up to 256	Up to 5 KHz	4840+
OpenBCI	Up to 19	Up to 2 KHz	3040+
ANT Neuro	32/64/128/256	Up to 16 KHz	2550+
Advance Brain Monitoring	9/28	Up to 256 Hz	2260+
Muse	Up to 256	Up to 100 KHz	2030+
Neurolink	Up to 3072	Up to 20 KHz	1400+
Cognionics	Up to 128	500 Hz	730+
Wearable Sensing	Up to 64	Up to 16 KHz	640+
mBrainTrain	Up to 32	Up to 1 KHz	370 +
g.tech	8/16/32/64	250/500 Hz	320+

accuracy at determining AHPLS in their study population.<sup>42</sup> Although groundbreaking with regards to augmenting biosensor data with ML, specifications about EEG features extracted were limited to the mention of PSD and wavelet decomposition without any mention of which EEG bandwidths were of particular importance or any preprocessing techniques being employed.

Evolving the methodology of identifying fatigue states with ML, Masse introduced the identification of inattention deafness in the form of alarm omission and alarm detection.<sup>44</sup> Their methodology employed preprocessing techniques of bandpass filtering and independent component analysis (ICA) to reject eye and muscle artifacts. Additionally, PSD was obtained for each brain frequency bandwidth for corresponding time-frequency analysis of omission “hits” or omission errors. They identified that the PSD of  $\gamma$  and mid-wavelength  $\beta$  frequencies tended to be higher for subjects who did not omit auditory alarm alerts. Quantifying these power spectrum differences between frequencies for those who omitted alarm alerts and those who did not, they developed ML models that could accurately identify those individuals who would have omission errors at 74.6% across their study population, with a maximum accuracy of 90.4% for one individual.<sup>44</sup>

Lee *et al.*, in addition to quantifying the various PSD for each EEG bandwidth, utilized the amplitude of the individual frequency bandwidths from aviators as input for his ML models.<sup>45</sup> Combining both the spatial-temporal features of EEG data within layers of convoluted neural network (CNN) ML models and stacking them in front of a long short-term memory (LSTM) ML model, they were able to accurately identify multiple abnormal mental states, such as high workload, low workload, low distraction, high distraction, high fatigue, and low fatigue. Their abnormal mental states were objectively quantified by: complexity of flight operations for workload; counting the number of words in the ATC message while maintaining the predefined conditions of the aircraft for distraction; and for fatigue, the subjects fed the Karolinska sleepiness scale (KSS) as input, which is estimated as the significant index of subjective drowsiness level. Their hybrid ML model “Mentalnet” achieved a 68.8% accuracy, demonstrating the utility of combining various layers of ML models.

Concerning actual in-flight acquisition of EEG data, Caldwell demonstrated the ability to obtain EEG data from sleep-deprived helicopter pilots sensitive enough to detect fatigue.<sup>46</sup> Taheri-Gorji *et al.*, recently contributed to the field by establishing that EEG feature extraction may be enhanced by ML algorithms during actual flight.<sup>47</sup> His team evaluated 16 pilots with a 20-channel, dry-EEG device during training flights in either a Piper Archer or a Cessna 172S. They characterized pilot workload by the complexity of flight operations. For example, straight-level flight would be considered low workload, whereas precision approach would be considered high workload. EEG features extracted were either the aforementioned PSD or the log energy entropy of each bandwidth. Log energy entropy describes the amount of information carried by a signal or how much randomness is in a signal.<sup>48</sup> They trained their ML models on over 200 EEG features of  $\delta$ ,  $\theta$ ,  $\alpha$ , and  $\beta$

wavelengths from various EEG channels, ultimately achieving a 93% accuracy at determining low, medium, and high workload states.

Thus far, the vast amount of discussion has been on the identification of cognitive workload and fatigue with ML models, using predominately EEG data; but, what about the other two leading causes of abnormal mental states that are associated with aviation mishaps, SD, and hypoxia? Recently, researchers have noted EEG features which have identified vection illusions and unperceived somatosensory illusions.<sup>49,50</sup> Specifically, Hoa noted statistically significant increases in  $\alpha$  waves of the right frontal cortex for subjects who experienced unrecognized vection.<sup>49</sup> Sciortino noted widespread power spectral decreases in  $\alpha$  and  $\beta$  for subjects exposed to a perceptual illusion in which participants experienced a fake model hand as being part of their own body, a.k.a., the rubber hand illusion.<sup>50</sup> In the foreseeable future, researchers could develop analogous machine training models described previously to evaluate changes in workload and fatigue states to accurately predict SD, using the EEG indices noted by these studies, in actual flight.

Regarding hypoxia, our initial aeromedical protocol example developed by Rice *et al.*, later further evaluated by Snider *et al.*, established the utility of EEG indices, specifically decreasing PSD of  $\beta$  waves and  $\theta$  waves, and the ability of DT, NB, and NN ML to predict hypoxia over 97%.<sup>8,9</sup> Liu, demonstrated similar utility of support vector machine (SVM) ML models in differentiating sustained attention performance of adults exposed to high altitude compared to healthy norms by utilizing EEG indices of ERPs, achieving an accuracy of 92.54%.<sup>51</sup>

In summary, regarding the use of EEGs as biosensors to detect changes in mental states, the literature suggests that feature extraction of high-frequency  $\beta$  PSD typically decreases in fatigued states, and that  $\alpha$ - and  $\theta$ -wave PSD have been shown to increase. Although fewer studies exist regarding the utilization of ML to identify the aviation hazards of SD and hypoxia, both conditions have demonstrated unique changes in PSD and, more recently, ERP and gravity frequency of PSD transition. Taken together, these EEG feature extractions, in combination with current ML models, hold promise of high accuracy in their ability to identify cognitive decrements in aviation environments.

There is a substantial body of scientific evidence that features of ECG, specifically HRV, may be useful in identifying aspects of mental workload and cognitive states in a nonaviation environment.<sup>52,53</sup> HRV is a physiological phenomenon characterized by fluctuations in the time intervals between consecutive heartbeats, and it reflects the influence on the sinus node of the two limbs of the autonomic nervous system (ANS)—sympathetic (SNS) and parasympathetic (PNS).<sup>54,55</sup>

HRV indices have been identified within flight simulators to index cognitive workload states.<sup>56–58</sup> Specifically, the ratio of low-frequency (LF) HRV (LF: 0.04–0.15 Hz; index of SNS) and high-frequency (HF) HRV (HF: 0.15–0.4 Hz; index of PNS) has been observed to increase due to the predominance of SNS during stressful events.<sup>57</sup> Capitalizing on this observation with

regard to HRV, Qin applied both unsupervised ML, in the form of Toeplitz Inverse Covariance Based Clustering (TICC), and supervised ML, in the form of SVM models, to these ECG features and achieved a 91.8% accuracy at identifying mental fatigue produced by prolonged flight missions.<sup>59</sup> Indeed, in some cases, HRV has performed as well or better than EEG as a biosensor in predicting pilots' cognitive workload during take-off, cruise, and landing phases when combined with common ML algorithms such as SVM or K-Nearest Neighbors (k-NN).<sup>60</sup>

Less is known with regard to HRV's ability to specifically identify SD and hypoxia in aviation environments. Lower HRV has been demonstrated in numerous pilot communities who have undergone SD training.<sup>61</sup> As suggested previously, lower HF HRV may be seen in a variety of stressful conditions. So, although this type of biosensor index is not unique to SD, it could be used in conjunction with other biosensors to infer abnormal spatial perception. As for hypoxia, HRV has been shown to decrease with mild normobaric hypoxia at 10,000 ft (3048 m), equivalent to slightly above commercial aviation cabin altitudes. This decrease in HRV was enhanced when combined with higher cognitive workload states, thus suggesting a synergistic response to two stressful conditions.<sup>62</sup> Similarly, Castro-Herrera recently exposed 44 aviators to acute severe hypobaric hypoxia at a simulated altitude of 25,000 ft (7620 m), resulting in decreases in both HF and LF HRV upon arriving at terminal altitude.<sup>63</sup> Other researchers have questioned the validity of LF HRV and LF/HF ratio in determining cardiac sympatho-vagal balance and subsequent physiological states.<sup>64–66</sup> From the current literature, data suggest the ANS response to hypoxia is complex and may not be useful to specifically characterize acute hypoxia, especially with higher heart rates, regardless of applying ML models to the data.

Eye-tracking (ET) applications for aviation to identify various cognitive states have evolved greatly over the last decade, so much so that a total of three systematic reviews have been performed on the topic.<sup>67–69</sup> To summarize, these reviews evaluated the literature on the subjects involved (civilian pilots, military pilots, air traffic control), type of visual equipment used, eye metrics extracted, and aviation hazard they were attempting to identify (fatigue, SD, hypoxia, cognitive workload). All reviews concluded that ET has the potential to be effective in terms of preventing errors or injuries by detecting, for example, fatigue or performance decrements.

As of this writing, there are four main methods used to measure eye movements. These methods are electro-oculography (EOG), scleral contact lens/search coil, photo-oculography (POG) and video-oculography (VOG).<sup>68,70</sup> In ET aviation research, distinct eye metrics have been identified and related to different cognitive, emotional, and physiological states, which can be used to gain a wider understanding of the human mind.<sup>69</sup> These eye metrics are fixation, saccadic movements, pupillary response, and eye blink rate.

Fixation refers to when the eye remains still, meaning the pupil is stationary for approximately 180–300 ms.<sup>71</sup> Per Mengtao's aforementioned review, a majority of aviation

research in the last decade has involved some measurement of fixation.<sup>68</sup> Ziv noted that experienced pilots tended to fixate more on multiple instruments as compared to novice aviators, who often focused on fewer.<sup>70</sup> Moreover, pilot's situational awareness (SA) performance and expertise level can be inferred from the distribution of fixations and fixation duration on relevant areas of interest.<sup>72</sup> Fudali-Czyz observed the effective dwells (dwell times exceeding 600 ms) to the stimulation area can reflect if pilots have incurred SD.<sup>73</sup>

Saccades are rapid eye movements that occur when a person shifts between fixations.<sup>71</sup> Saccades last around 10–100 ms, during which time visual information transfer is suppressed; therefore, it is generally concluded that saccades are not directly related to cognitive processing. However, the literature suggest that saccade velocity may be related to lethargy, stress, and fatigue.<sup>74–76</sup> Scannella noted better utility with the measurement of saccades in detecting cognitive workload as compared to cardiac metrics such as HRV during actual flight.<sup>77</sup> Regarding aviation specific preconditions that may result in mishaps, decreases in saccadic drift and velocity have been found to be associated with both hypoxia and fatigue.<sup>78,79</sup>

Utilizing the biometric indices of fixation and saccadic movements, researchers have incorporated them within ML models to identify pilots' attention distribution and accurately predict their SA.<sup>80</sup> Specifically, Jiang monitored the flight deviations of cadets during specified assigned headings, altitude, and airspeeds. They found that they could accurately determine the SA of these cadets by extracting their main visual area of interest during the flight and applying CNN and LTSM ML to these ET indices. Although not cited as frequently as fixation or saccades, blink rate and pupil diameter have also been incorporated in multimodal ML algorithms to detect fatigue<sup>81</sup> and have the potential to be informative in identifying hypoxia<sup>82</sup> and cognitive workload.<sup>83,84</sup>

In summary, several biosensors have shown individual promise in identifying cognitive deficits during both simulated flight environments and actual flight. Their ability to detect cognitive deficits has been enhanced with recent advancements in ML computer modeling. The most cited biosensors that have been integrated with ML models are EEG, ECG, and ET devices. Not explicitly covered in this section are nonpsychophysiological sensors, such as sensors which monitor deviation in flight control inputs. An important example of such research is Wang's study evaluating joystick deviation as a predictor of space module docking crashes when combined with semi-supervised ML algorithms.<sup>85</sup> Future integration of these engineering sensors with biosensors into multinodal machine-human interfaces is on the foreseeable horizon. On a final note, Mengtao concluded the real-time application of this technology, as with all sensors of these types, is still rare due to preprocessing times of raw data.<sup>68</sup> This topic will be explored in the next section of this review.

### Preprocessing & Denoising Techniques

A variety of time series biodata (e.g., EEG, ECG, GSR, blood oxygen saturation, and eye movements) can be recorded and

mathematically interpreted through ML as a biofeedback system for pilots. However, analyzing these time series data is challenging due to electronic or biological noise.

The following discussion proposes a framework that aims to: 1) clean and interpret these physiological measures, which could impact cognitive performance during common aeromedical hazards (fatigue, hypoxia, and SD); and 2) transition this technology to operational environments. The goal is to use any multidimensional biodata for AI input.

Preprocessing techniques start from the hardware used for data collection to mathematical algorithms that separate noise from signals. ML techniques (supervised, unsupervised, and semi-supervised learning discussed in the next section) are employed to denoise signals. Table I, provided within the introduction of this paper, highlights just a few of these denoising methods. In general, the purpose of these methods focuses on achieving clarity of primary signals, with different denoising techniques serving various purposes.

The importance of preprocessing biosensor data prior to incorporating it within ML models cannot be overemphasized. As an example, initial efforts by Snider to apply ML to Rice's dry-EEG data on aviators exposed to hypoxia without denoising techniques resulted in an accuracy of their ML models of only 67%.<sup>86</sup> However, upon applying these denoising techniques to their ML models, they achieved an accuracy of over 97%.<sup>9</sup>

Since EEG's discovery in 1929 by Hans Berger, noise has complicated its interpretation, despite noise filtering.<sup>87</sup> Medical specialists use EEG extensively but still face challenges with subtle abnormalities or complex conditions,

requiring collaboration to interpret. Like the human mind, machines can now aid in real-time EEG interpretation. The development of an artificial general intelligence model that adapts denoising techniques in real time for specific biosensors could enhance ML accuracy for instantaneous biofeedback. Thus, preprocessed biosensor data can then be used by various ML algorithms, each with unique benefits, as detailed in the next section.

### Machine Learning Algorithm Development

ML algorithms form the foundation of AI, enabling decision-making.<sup>88,89</sup> Successful ML development requires data preprocessing, as described in the previous section. Following this, exploratory data analysis and ML algorithm selection are critical steps. ML algorithms are broadly classified into supervised, unsupervised, and semi-supervised categories, which are discussed below and summarized in Table II. The challenge lies in selecting the appropriate algorithm and parameters, depending on data characteristics and analysis objectives. This section will explore these ML categories and their applications in monitoring pilots' cognitive and physiological states. **Table IV** summarizes the various aviation studies cited in this paper that have incorporated ML models to enhance the interpretation of biosensor data by category and the aviation hazard they were attempting to identify.

Supervised learning uses labeled data during training to map input data to output labels accurately.<sup>26,27</sup> The training dataset, comprising precategorized observations, helps the model learn input-output relationships. By analyzing labeled examples, the model identifies patterns, enabling it to

**Table IV.** Selected Aviation Related Studies Utilizing Machine Learning in this Review.

LEARNING MODEL	AUTHOR	BIOSENSOR	TESTING ENVIRONMENT	AEROMEDICAL HAZARD	ML ALGORITHM
Supervised	Snider et al. <sup>9</sup>	Dry EEG	Simulated Flight	Hypoxia	DT, RF, NB, NN
	Harrivel et al. <sup>42</sup>	Dry EEG, ECG, GSR, fNIRS	Simulated Flight	Attention	RF, GB, SVM, polynomial kernels
	Massé et al. <sup>44</sup>	Dry EEG	Simulated Flight	Omission Errors, Attention, Fatigue	k-NN, LDA, RF
	Lee et al. <sup>45</sup>	Dry EEG	Simulated Flight	Fatigue, Workload, Distraction	CNN, LSTM
	Taheri-Gorgi et al. <sup>47</sup>	Dry EEG	Actual Flight	Workload	SVM, RF, LASSO
	Liu et al. <sup>51</sup>	Dry EEG	Chronic Altitude Exposure	Hypoxia	DT, SVM
	Mohanavelu et al. <sup>60</sup>	Dry EEG, HRV	Simulated Flight	Workload	LDA, SVM, k-NN
	Jiang et al. <sup>80</sup>	ET	Simulated Flight	Situational Awareness	CNN, LSTM
Unsupervised	Li et al. <sup>31</sup>	ECG, EMG, GSR, RR	Simulated Flight	Fatigue, Workload	MFT
	Wu et al. <sup>90</sup>	Wet EEG, ET	Simulated Flight	Fatigue	DANN
Semi-supervised	Memarzadeh et al. <sup>34</sup>	Instrument Data	Actual Flight Data	Mishap Prediction	MFT
	Zhang et al. <sup>35</sup>	Instrument Data	Incident reports	Incident risk prediction	SVM, DNN
	Qin et al. <sup>59</sup>	ECG, ET	Long-Duration Flight Mission	Fatigue	TICC/SVM
	Wang et al. <sup>85</sup>	Instrument Data	Simulated Spacecraft Docking	Mishap Prediction	MLP, CNN, GRU, LSTM

DT: decision tree; RF: random forest; NB: naïve-Bayes; NN: neural network; GB: gradient boosting; kNN: k-nearest neighbor; LDA: linear discriminant analysis; CNN: convoluted neural network; LSTM: long short-term memory; LASSO: least absolute shrinkage and selection operator; TICC: Toeplitz inverse covariance based clustering; DANN: deep autoencoder neural network; MFT: multimodality fusion technology; MLP: multilayer perceptron neural network; GRU: gated recurrent unit; DNN: deep neural network



generalize to new data. Examples of supervised learning include: Boolean classification, which predicts binary outcomes such as whether an email is spam; nominal classification, which assigns inputs to predefined categories, such as classifying images as “dog,” “cat,” or “bird;” and regression, which predicts continuous values, such as house prices.

Common algorithms in supervised learning include: linear regression, which predicts continuous values by assuming a linear relationship; logistic regression, which is used for binary classification, thereby modeling the probability of category membership; DTs, which are applicable for classification and regression by splitting data into subsets; and NNs, which model complex relationships using interconnected nodes. An example of an aviation protocol that has utilized supervised ML models is the study by Snider et al., who extracted EEG indices such as PSD values and applied them to DT and NB algorithms to accurately identify hypoxia.<sup>9</sup> Likewise, Masse applied the supervised algorithms of RF and SVM to identify EEG indices predictive of cognitive fatigue.<sup>44</sup>

Unsupervised learning works with unlabeled data, finding patterns and structures without specific guidance.<sup>29</sup> The algorithm independently explores the data to uncover its inherent structure, aiming to discover hidden patterns or groupings by analyzing it to find natural clusters or underlying organizations. This method adapts based on the data's properties, providing insights that might be missed with predefined labels. Key tasks for unsupervised learning include: clustering, which involves grouping similar items based on features, with algorithms like k-means and hierarchical clustering partitioning data into clusters; anomaly detection, which identifies outliers or anomalies by understanding normal data patterns; and data visualization, which simplifies data with techniques like PCA to make it easier to visualize and interpret.

Common unsupervised learning algorithms include clustering algorithms, such as k-means and hierarchical clustering, and dimensionality reduction algorithms, such as PCA and association rule learning. Examples of these types of ML being utilized in recent aviation protocols include recognizing pilots' fatigue status using a deep contractive autoencoder network by Wu et al.,<sup>90</sup> as well as Li's protocol predicting unsafe pilot operations utilizing k-means clustering unsupervised ML.<sup>31</sup>

Semi-supervised learning combines labeled and unlabeled data during training, using a small amount of labeled data with a large amount of unlabeled data.<sup>32</sup> This approach is beneficial when labeled data is scarce or costly and unlabeled data is abundant, such as Xu's evaluation of ML algorithms' ability to interpret wearable-ECG data.<sup>33</sup> The algorithm first learns from labeled data to understand input-output relationships, then utilizes unlabeled data to identify patterns and structures, improving performance by exploiting the information in unlabeled data.

Common semi-supervised learning approaches include: self-training, where the algorithm trains on labeled data, then uses its predictions to label unlabeled data and iteratively refines its predictions; cotraining, where multiple classifiers train on different feature subsets, label the unlabeled data, and train each other; label propagation, where labels propagate from labeled to

unlabeled data based on similarity; and active learning, where a human expert assigns class labels to “kickstart,” augment, or reinforce the learning process. Semi-supervised ML models have recently become a focus of interest in aviation to detect anomalies and predict incident risk.<sup>34,35</sup>

Conceptualizing the methodological steps of sensor selection, preprocessing, and ML algorithm development into operational aviation environments, we can envision how this technology may optimize a pilot's performance in next-generation aircraft. The studies we have summarized suggest EEG, ECG, and ET indices of a pilot may be used to identify cognitive decrements. When simultaneously combined with preprocessing techniques and ML algorithms, they hold the potential of mitigating cognitive decrements and enhancing human performance in real time.

## DISCUSSION

In this paper, we presented a three-step methodology by which AI may be applied to data obtained from biosensors to identify cognitive decrements in aviators. This included sensor selection, preprocessing, and ML algorithm development. Intentionally, data integration was not mentioned specifically. To accurately link the cognitive decline you are trying to detect with your biosensor, effective data integration is essential. For example, in Rice's study, the frequency sampling rate of the cognitive performance task being monitored required the data to be time-matched with the same frequency sampling rate for both biosensors, EEG and oxygen saturation, in order to correlate precisely the independent variable under investigation.<sup>8</sup> So, if one biosensor has a sampling rate of 200 Hz and another biosensor has a sampling rate of 250 Hz, the lowest sampling rate must be utilized and integrated to appropriately correlate with one another. Various computer platforms such as LabVIEW® (Austin, TX, United States), MATLAB® (Natick, MA, United States), and Python® (Fredericksburg, VA, United States) have been used to accomplish this integration within simulated and actual flight.<sup>44,45,47</sup>

There is less published research on more difficult preconditions, such as motivation, overconfidence, and personality style, that could contribute to mishaps but do not yet lend themselves readily to “real-time” AI identification. As such, these preconditions were not the focus of our methodologies. However, these more challenging to detect human behavioral preconditions have been the focus of ML investigations in both the systematic analysis of aviation mishap reports and human factor classification in recent publications.<sup>91–93</sup>

This review purposefully incorporated NLP into sections of this paper to demonstrate the utility of these ML models and conceptualize the capabilities of this technology for aerospace environments. Most meta-analysis and systematic reviews incorporate some form of NLP within their methodology to perform searches for reference inclusion/exclusion in their papers. Some researchers have noted the potential for bias when utilizing NLP models such as Chat GPT.<sup>94</sup> This bias has been

shown to be a product predominately of opinion-generated references. We attempted to exclude this condition when developing our tables by ensuring references cited were peer-reviewed, aviation-related, and incorporated established ML algorithms.

From an educational perspective, we have introduced to the audience several new methodologies of AI for which it is certainly difficult to conceptualize the importance in an operational environment. Specifically, the methodological steps of preprocessing and ML algorithm development are not routinely encountered by aeromedical professionals, so presenting their importance within referenced tables explicitly demonstrates the operational impact they may have. Moreover, utilizing NLP as a bridge to crystallize these concepts from a laboratory to an operational environment is and will become an ever-present methodological tool researchers will use to make scientific gains. Exposing readers to appropriate referencing of such tools is of value to future papers and not so much a novelty but rather our current state of science.

In Alreshidi's systematic analysis of ML and aviation safety, only 10% of the 80 papers included in their review had obtained data during actual flight.<sup>5</sup> None of these studies provided biosensor data feedback in real time to their pilots during flight. Future direction for this research will need to focus on the integration of both biosensors and ML computation into aircraft display systems and/or the helmet visor to provide meaningful real-time data to detect or prevent undesirable cognitive states. Miniaturization of the data preprocessing hardware and maturation of ML algorithmic selection will be the next phase of the transition to operational environments for real-time continuous monitoring.

This paper is not a traditional, systematic review of the literature regarding ML and aviation safety. More appropriately, it should be characterized as a critical review with the primary objective of serving as a guidepost for future aeromedical investigators interested in utilizing AI to enhance their research protocols. Reviews of this type emphasize the conceptual importance of the available literature as compared to systematic and meta-analytic reviews, which have a more structured methodology. As such, the paper, as Grant eloquently stated in his analysis of 14 scientific review types, "should serve as a starting point and not an endpoint."<sup>95</sup>

## ACKNOWLEDGMENTS

The views expressed are those of the authors and do not reflect the official guidance or position of the U.S. Government, the Department of Defense (DoD), or the U.S. Navy. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government. The appearance of external hyperlinks does not constitute endorsement by the DoD of the linked websites, or the information, products, or services contained therein. The DoD does not exercise any editorial, security, or other control over the information you may find at these locations. The imagery in this document is property of the U.S. Navy.

The authors would like to acknowledge the Naval Safety Command for providing mishap data regarding rates of Class A mishaps, mishap fatalities and cost, and preconditions which lead to mishaps over the last decade. Specifically,

we would like to acknowledge the professional contributions of CAPT Michael Hoskins and CAPT Jonathan Erpenbach for their careful review and editing of this manuscript.

*Financial Disclosure Statement:* The authors have no competing interests to declare.

*Authors and Affiliations:* G. Merrill Rice, D.O., M.P.H., Naval Safety Command, Norfolk, VA; Steven Linnville, Ph.D., Naval Medical Research Unit Dayton, Wright-Patterson Air Force Base, OH; and Dallas Snider, Ph.D., University of West Florida, Pensacola, FL, United States.

## REFERENCES

1. Department of the Navy. Naval Safety Command: US Naval Aviation Class A Mishap Rate 2014-2023, Risk Management Database. [Accessed May 31, 2024]. Available from <https://navalsafetycommand.navy.mil/>.
2. Department of the Navy. Naval Safety Command: Total Aviation Mishap Fatalities and Mishap Cost 2014-2023, Risk Management Database. [Accessed May 31, 2024]. Available from <https://navalsafetycommand.navy.mil/>.
3. Department of the Navy. Naval Safety Command: Preconditions Associated with Naval All Class Mishaps 2014-2023, Risk Management Database. [Accessed May 31, 2024]. Available from <https://navalsafetycommand.navy.mil/>.
4. Mukhamediev RI, Popova Y, Kuchin Y, Zaitseva E, Kalimoldayev A, et al. Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges. *Mathematics*. 2022; 10(15):2552.
5. Alreshidi I, Moulitsas I, Jenkins KW. Advancing aviation safety through machine learning and psychophysiological data: a systematic review. *IEEE Access*. 2024; 12:5132–5150.
6. Saeidi M, Karwowski W, Farahani FV, Fiok K, Taiar R, et al. Neural decoding of EEG signals with machine learning: a systematic review. *Brain Sci*. 2021; 11(11):1525.
7. Lim JZ, Mountstephens J, Teo J. Eye-tracking feature extraction for biometric machine learning. *Front Neurobot*. 2022; 15:796895.
8. Rice GM, Snider D, Drollinger S, Greil C, Bogni F, et al. Dry-EEG manifestations of acute and insidious hypoxia during simulated flight. *Aerospace Med Hum Perform*. 2019; 90(2):92–100.
9. Snider D, Linnville SE, Phillips JB, Rice GM. Predicting hypoxic hypoxia using machine learning and wearable sensors. *Biomed Signal Process Control*. 2022; 71 (Part A):103110.
10. Widrow B, Glover JR, McCool JM, Kaunitz J, Williams CS, et al. Adaptive noise cancelling: principles and applications. *Proceedings of the IEEE*. 1975; 63(12):1692–1716.
11. Oppenheim AV, Schaffer RW. *Discrete-time signal processing*. 3<sup>rd</sup> ed. London (UK): Pearson Education; 2009:576–577.
12. Widrow B, Stearns SD. *Adaptive signal processing*. Hoboken (NJ): Prentice-Hall; 1985:16–18.
13. Wu D, King J-T, Chuang C-H, Lin C-T, Jung T-P. Spatial filtering for EEG-based regression problems in brain-computer interface (BCI). *IEEE Trans Fuzzy Syst*. 2018; 26(2):771–781.
14. Lotte F, Guan C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans Biomed Eng*. 2011; 58(2):355–362.
15. Babiloni F, Cincotta F, Carducci F, Rossini PM, Babiloni C. Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging-based head models. *Clin Neurophysiol*. 2001; 112(5):724–727.
16. Samar VJ, Bopadikar A, Rao R, Swartz K. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain Lang*. 1999; 66(1):7–60.
17. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A Math Phys Sci*. 1971; 1998(454):903–995.

18. Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Netw.* 2000; 13(4-5):411–430.
19. Roy RN, Bonnet S, Charbonnier S, Jallon P, Campagne A. A comparison of ERP spatial filtering methods for optimal mental workload estimation. *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC);* August 25–29, 2015; Milan, Italy. Piscataway (NJ): IEEE; 2015:7254–7257.
20. Zhuang X, Yang Z, Cordes D. A technical review of canonical correlation analysis for neuroscience applications. *Hum Brain Mapp.* 2020; 41(13):3807–3833.
21. Spurek P, Tabor J, Struski L, Smieja M. Fast independent component analysis algorithm with a simple closed-form solution. *Knowl Base Syst.* 2018; 161:26–34.
22. Croft RJ, Barry RJ. Removal of ocular artifact from the EEG: a review. *Neurophysiol Clin.* 2000; 30(1):5–19.
23. Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. In: Touretzky D, Mozer M, Hasselmo M, editors. *Advances in neural information processing systems 8.* Cambridge (MA): MIT Press; 1996:145–151.
24. Karamacoska D, Barry RJ, Steiner GZ. Using principal components analysis to examine resting state EEG in relation to task performance. *Psychophysiology.* 2019; 56(5):e13327.
25. Bzdok D, Krzywinski M, Altman N. Machine learning: a primer. *Nat Methods.* 2017; 14(12):1119–1120.
26. Köhl N, Hirt R, Baier L, Schmitz B, Satzger G. How to conduct rigorous supervised machine learning in information systems research: the supervised machine learning report card. *Comm Assoc Inform Syst.* 2021; 48(1):46.
27. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. Supervised machine learning algorithms: classification and comparison. *Int J Comput Trends Tech.* 2017; 48(3):128–138.
28. Ackley JL, Puranik TG, Mavris D. A supervised learning approach for safety event precursor identification in commercial aviation [Webinar]. Reston (VA): AIAA Aviation 2020 Forum; June 15–19, 2020.
29. Talaie Khoei T, Ould Slimane H, Kaabouch N. Deep learning: systematic review, models, challenges, and research directions. *Neural Comput Appl.* 2023; 35(31):23103–23124.
30. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry MW, Mohamed A, Yap BW, editors. *Supervised and unsupervised learning for data science.* Cham (Switzerland): Springer Cham; 2020:3–21.
31. Li X, Qian Y, Chen H, Zheng L, Wang Q, Shang J. An unsupervised learning approach for analyzing unsafe pilot operations based on flight data. *Appl Sci.* 2022; 12(24):12789.
32. van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn.* 2020; 109(2):373–440.
33. Xu H, Yan W, Lan K, Ma C, Wu D, et al. Assessing electrocardiogram and respiratory signal quality of a wearable device (sensecho): semisupervised machine learning-based validation study. *JMIR Mhealth Uhealth.* 2021; 9(8):e25415.
34. Memarzadeh M, Romli FI, Marquez RA, Guzman JJ. Semi-supervised active learning for anomaly detection in aviation. *JAIS.* 2023; 20(4):181–194.
35. Zhang X, Mahadevan S. Ensemble machine learning models for aviation incident risk prediction. *Decis Support Syst.* 2019; 116:48–63.
36. Byrom B, McCarthy M, Schueler P, Muehlhausen W. Brain monitoring devices in neuroscience clinical research: the potential of remote monitoring using sensors, wearables, and mobile devices. *Clin Pharmacol Ther.* 2018; 104(1):59–71.
37. Ismail LE, Karwowski W. Applications of EEG indices for the quantification of human cognitive performance: a systematic review and bibliometric analysis. *PLoS One.* 2020; 15(12):e0242857.
38. Liu S, Wang L, Gao RX. Cognitive neuroscience and robotics: advancements and future research directions. *Robot Comput-Integr Manuf.* 2024; 85:102610.
39. Tivadar RI, Murray MM. A primer on electroencephalography and event-related potentials for organizational neuroscience. *Organ Res Methods.* 2019; 22(1):69–94.
40. Borck C. *Brainwaves: a cultural history of electroencephalography.* Oxfordshire (UK): Routledge; 2018.
41. Stone JL, Hughes JR. Early history of electroencephalography and establishment of the American Clinical Neurophysiology Society. *J Clin Neurophysiol.* 2013; 30(1):28–44.
42. Harriavel AR, Elson JL, Feickert C, Frey C. Psychophysiological sensing and state classification for attention management in commercial aviation. Paper presented at AIAA Infotech @ Aerospace; January 4–8, 2016; San Diego, CA.
43. Shoka A, Ahmed M, Attia M, Elsaid W. Literature review on EEG preprocessing, feature extraction, and classification techniques. *Menoufia J Electron Eng Res.* 2019; 28(1):292–299.
44. Massé E, Barthelemy O, Fabre L. Classification of electrophysiological signatures with explainable artificial intelligence: the case of alarm detection in flight simulator. *Front Neuroinform.* 2022; 16:904301.
45. Lee DH, Park SJ, Jung KH, Youn JH, Kang SG, Kwak SH. Autonomous system for EEG-based multiple abnormal mental states classification using hybrid deep neural networks under flight environment. *IEEE Trans Syst Man Cybern Syst.* 2023; 53(10):6426–6437.
46. Caldwell JA, Hall KK, Erickson BS. EEG data collected from helicopter pilots in flight are sufficiently sensitive to detect increased fatigue from sleep deprivation. *Int J Aviat Psychol.* 2002; 12(1):19–32.
47. Taheri Gorji H, Khalid M, Altalhi M, Mohammed MA, Ghani A. Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. *Sci Rep.* 2023; 13(1):2507.
48. Aydin S, Saraoğlu HM, Kara S. Log energy entropy-based EEG classification with multilayer neural networks in seizure. *Ann Biomed Eng.* 2009; 37(12):2626–2630.
49. Hao C, Zheng X, Guo H, Shi G, Ma W. EEG analysis of visually induced spatial disorientation. Paper presented at: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); October 19–21, 2019; Suzhou, China.
50. Sciortino P, Kayser C. The rubber hand illusion is accompanied by a distributed reduction of alpha and beta power in the EEG. *PLoS One.* 2022; 17(7):e0271659.
51. Liu H, Li X, Zhang Y, Bai L, Xu J, Zhang L. Machine learning based on event-related EEG of sustained attention differentiates adults with chronic high-altitude exposure from healthy controls. *Brain Sci.* 2022; 12(12):1677.
52. Forte G, Favieri F, Casagrande M. Heart rate variability and cognitive function: a systematic review. *Front Neurosci.* 2019; 13:710.
53. Arakaki X, Arechavala RJ, Choy EH, Bautista J, Bliss B, et al. The connection between heart rate variability (HRV), neurological health, and cognition: a literature review. *Front Neurosci.* 2023; 17:1055445.
54. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur Heart J.* 1996; 17(3):354–381.
55. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health.* 2017; 5:258.
56. Alaimo A, Barthelemy O, De Fabre L, Bartoli J, Egan K, et al. Human heart-related indexes behavior study for aircraft pilots allowable workload level assessment. *IEEE Access.* 2022; 10:16088–16100.
57. Cao X, Zhang Y, Hu W, Yang L, Zhang H. Heart rate variability and performance of commercial airline pilots during flight simulations. *Int J Environ Res Public Health.* 2019; 16(2):237.
58. Mansikka H, Virtanen K, Harris D. Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. *Appl Ergon.* 2016; 56:213–219.
59. Qin H, Chen X, Zhang X, Wang S. Detection of mental fatigue state using heart rate variability and eye metrics during simulated flight. *Hum Factors Ergon Manuf Serv Ind.* 2021; 31(6):637–651.
60. Mohanavelu K, Sreedevi V, Ramesh Babu D, Raju K. Machine learning-based approach for identifying mental workload of pilots. *Biomed Signal Process Control.* 2022; 75:103623.



61. Bustamante-Sánchez Á, Clemente-Suárez VJ. Psychophysiological response to disorientation training in different aircraft pilots. *Appl Psychophysiol Biofeedback*. 2020; 45(4):241–247.
62. Temme LA, Wittels HL, Wishon MJ, St. Onge P, McDonald SM, et al. Continuous physiological monitoring of the combined exposure to hypoxia and high cognitive load in military personnel. *Biology*. 2023; 12(11):1398.
63. Castro-Herrera JM, Cedeño-Serna JC, Tuta-Quintero E, Botero-Rosas DA. Heart rate variability as a predictor of hypobaric hypoxia in aircraft pilots. *Rev Latinoam Hipertens*. 2021; 16(4):314–320.
64. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front Physiol*. 2013; 4:26.
65. Goldstein DS, Benthon O, Park MY, Sharabi Y. Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. *Exp Physiol*. 2011; 96(12):1255–1261.
66. Thomas BL, Claassen N, Becker P. Validity of commonly used heart rate variability markers of autonomic nervous system function. *Neuropsychobiology*. 2019; 78(1):14–26.
67. Martínez-Marquez D, Pingali S, Panuwatwanich K, Stewart RA, Mohamed S. Application of eye tracking technology in aviation, maritime, and construction industries: a systematic review. *Sensors*. 2021; 21(13):4289.
68. Mengtao L, Ji Z, Zhang X, Sun G. Leveraging eye-tracking technologies to promote aviation safety- a review of key aspects, challenges, and future perspectives. *Saf Sci*. 2023; 168:106295.
69. Peißl S, Wickens CD, Baruah R. Eye-tracking measures in aviation: a selective literature review. *Int J Aerosp Psychol*. 2018; 28(3-4):98–112.
70. Ziv G. Gaze behavior and visual attention: a review of eye tracking studies in aviation. *Int J Aviat Psychol*. 2016; 26(3-4):75–104.
71. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J. Eye tracking: a comprehensive guide to methods and measures. Oxford (UK): Oxford University Press; 2011.
72. Li W-C, Zhang J, Le Minh T, Cao J, Wang L. Visual scan patterns reflect to human-computer interactions on processing different types of messages in the flight deck. *Int J Ind Ergon*. 2019; 72:54–60.
73. Fudali-Czyż A, Bennett S, Brooks SK, Lubetzky A, Jennings AR, et al. An attentive blank stare under simulator-induced spatial disorientation events. *Hum Factors*. 2024; 66(2):317–335.
74. Jeelani I, Albert A, Han K, Azevedo R. Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. *J Constr Eng Manage*. 2019; 145(1):04018115.
75. Stankovic A, Aitken MR, Clark L. An eye-tracking study of information sampling and decision-making under stress: implications for alarms in aviation emergencies. *Proc Hum Factors Ergon Soc Annu Meet*. 2014; 58(1):125–129.
76. Yan S, Wei Y, Tran CC. Evaluation and prediction of mental workload in user interface of maritime operations using eye response. *Int J Ind Ergon*. 2019; 71:117–127.
77. Scannella S, Lafont V, Stere G, Dewez G, Renard C, et al. Assessment of ocular and physiological metrics to discriminate flight phases in real light aircraft. *Hum Factors*. 2018; 60(7):922–935.
78. Di Stasi LL, McCamy MB, Martinez-Conde S, Gayles E, Hoare C, et al. Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiol Behav*. 2016; 153:91–96.
79. Diaz-Piedra C, Rieiro H, Suárez J, Rios-Tejada F, Catena A, Di Stasi LL. Fatigue in the military: towards a fatigue detection test based on the saccadic velocity. *Physiol Meas*. 2016; 37(9):N62–N75.
80. Jiang G, Liu J, Wang L, Li Z. Transformer network intelligent flight situation awareness assessment based on pilot visual gaze and operation behavior data. *Int J Pattern Recognit Artif Intell*. 2022; 36(5):2259015.
81. Türetkin E, Kökçü F, Atalay V. Real-time eye gaze tracking for human-machine interaction in the cockpit. *Proceedings of AI and Optical Data Sciences III*; March 2, 2022; San Francisco, California. Bellingham (WA): Society of Photo-Optical Instrumentation Engineers (SPIE); 2022.
82. Thropp JE, Scallon JFV, Buza P. PERCLOS as an indicator of slow-onset hypoxia in aviation. *Aerosp Med Hum Perform*. 2018; 89(8):700–707.
83. Othman N, Romli FI. Mental workload evaluation of pilots using pupil dilation. *IREASE*. 2016; 9(3):80–84.
84. Chen S, Epps J. Using task-induced pupil diameter and blink rate to infer cognitive load. *Hum Comput Interact*. 2014; 29(4):390–413.
85. Wang Y, Yu C, Xiao Y, Fan L, Wang S, et al. Crash prediction using deep learning in a disorienting spaceflight analog balancing task. *Front Physiol*. 2022; 13:806357.
86. Snider D, Linnville S, Phillips J, Drollinger S, Banuelos J, Rice G. Using machine learning to build a hypoxia detection model [Abstract]. *Aerosp Med Hum Perform*. 2020; 91(3):240.
87. Caeira MW, Caboclo LO, de Paola L. An appraisal to Hans Berger by the time of his 150th birthday: the human EEG and tales of blood flow, heat and brain waves. *Arq Neuropsiquiatr*. 2023; 81(12):1163–1168.
88. Murphy KP. Machine learning: a probabilistic perspective. Cambridge (MA): MIT Press; 2012.
89. Russell SJ, Norvig P. Artificial intelligence: a modern approach. 4th ed. London (UK): Pearson; 2016.
90. Wu EQ, Zhao W, Li S, Zhang X, Shi J. Pilots' fatigue status recognition using deep contractive autoencoder network. *IEEE Trans Instrum Meas*. 2019; 68(10):3907–3919.
91. Harris D, Li WC. Using neural networks to predict HFACS unsafe acts from the pre-conditions of unsafe acts. *Ergonomics*. 2019; 62(2):181–191.
92. Madeira T, Marques DA, Augusto A, Silva JC. Machine learning and natural language processing for prediction of human factors in aviation incident reports. *Aerosp Sci Technol*. 2021; 8(2):47.
93. Morais C, Ziminski K, Chatzi E, Pimentel M. Identification of human errors and influencing factors: a machine learning approach. *Saf Sci*. 2022; 146:105528.
94. Duncan C, McCulloh I. Unmasking bias in Chat GPT responses. *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*; November 6–9, 2023; Kusadasi, Türkiye. New York (NY): Association for Computing Machinery; 2024.
95. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J*. 2009; 26(2):91–108.