Neuroinformatics (2025) 23:38
https://doi.org/10.1007/s12021-025-09729-2

RESEARCH

Prediction of Cerebrospinal Fluid (CSF) Pressure with Generative Adversarial Network Synthetic Plasma-CSF Biomarker Pairing

Phani Paladugu[1,2] · Rahul Kumar[3] · Jahnavi Yelamanchi[4] · Ethan Waisberg[5] · Joshua Ong[6] · Mouayad Masalkhi[7,8] · Chirag Gowda[9] · Ryung Lee[10] · Dylan Amiri[11,12] · Ram Jagadeesan[13,14] · Nasif Zaman[15] · Alireza Tavakkoli[15] · Andrew G. Lee[16,17,18,19,20,21,22,23]

Accepted: 22 May 2025

Abstract

Non-invasive intracranial pressure (ICP) monitoring can help clinicians safely and efficiently monitor spaceflight-associated neuro-ocular syndrome (SANS), idiopathic intracranial hypertension, and traumatic brain injury in astronauts. Current invasive ICP measurement techniques are unsuitable for austere environments like spaceflight. In this study, we explore the potential of plasma-derived cell-free RNA (cfRNA) biomarkers as non-invasive alternatives to cerebrospinal fluid (CSF) markers for ICP assessment. We conducted a secondary analysis of NASA's Open Science Data Repository datasets 363–364, focusing on plasma and CSF biomarkers related to ICP and neurovascular health. An ensemble model combining Support Vector Machine, Gradient Boosting Regressor, and Ridge Regression was developed to capture plasma-CSF biomarker relationships. To address limited sample size, we employed a Generative Adversarial Network (GAN) to generate synthetic plasma-CSF biomarker pairs, expanding the dataset from 29 to 279 samples. The model's performance was evaluated using Mean Squared Error (MSE) and validated against real biomarker data. The GAN-augmented ensemble model achieved high predictive accuracy with an MSE of 0.0044. Synthetic plasma-CSF pairs closely aligned with actual biomarker distributions, demonstrating their effectiveness in reducing overfitting and enhancing model robustness. Strong correlations between plasma-derived RNA biomarkers and corresponding CSF indicators support their potential as non-invasive proxies for ICP assessment. This study establishes a novel framework for non-invasive ICP monitoring using plasma cfRNA profiles enriched with GAN-generated synthetic data. The approach shows promise for both spaceflight and clinical applications, potentially broadening diagnostic capabilities for ICP-related conditions. However, further validation across diverse populations is necessary, along with careful consideration of bioethical and data security issues associated with synthetic data use in clinical diagnostics.

Keywords Intracranial Pressure · Biomarker Prediction · Machine Learning in Medicine · Generative Adversarial Networks · Non-invasive Diagnostics

Introduction

Spaceflight-Associated Neuro-Ocular Syndrome (SANS) is a condition commonly faced by astronauts that is characterized by optic disc edema and globe flattening in during extended space missions (Félix & Oliveira, 2022; Jasien et al., 1985; Kamran et al., 2024; Martin Paez et al., 2020; Wojcik et al., 2020). As SANS likely stems from microgravity-induced cephalad fluid shifts, real-time intracranial pressure (ICP) monitoring is crucial for understanding and mitigating vision impairment (Huang et al., 2019; Ong et al., 2022; Wostyn et al., 2022). However, Current invasive ICP measurement methods are impractical to use on the international space station (ISS), necessitating non-invasive alternatives (D'Antona et al., 2021; Engelborghs et al., 2017; Kim, 2022; Munakomi & Das, 2024; Raboel et al., 2012). In this study, we explored cell-free RNA (cfRNA) in blood plasma as minimally invasive biomarkers (Albrecht et al., 2022;

Description: Our study introduces a novel approach to non-invasive intracranial pressure monitoring using plasma-derived RNA biomarkers, enhanced by machine learning techniques. by analyzing NASA datasets and employing generative adversarial networks, we developed a predictive model that could revolutionize ICP assessment in both space and terrestrial medical applications.

Extended author information available on the last page of the article

Published online: 10 July 2025

Cheung et al., 2019; Roskams-Hieter et al., 2022; Zaporozhchenko et al., 2018) that can potentially reflect ICP fluctuations through tissue-specific gene expression and systemic physiological changes. In doing so, we aimed to capture a broad range of biological variability, including physiological stressors such as circadian rhythm disruptions, neuroinflammatory responses, and psychological stress (Brasil et al., 2023; Burgos et al., 2014; Glinge et al., 2017; Goodman et al., 2008; Sohel Md Mahmodul, 2016; Thomou et al., 2017; Zhou et al., 2016). Our team analyzed NASA's Open Science Data Repository datasets 363–364 using an ensemble modeling approach enhanced by generative adversarial networks (GANs) (Bannick et al., 2021; Kang et al., 2023; Kuo et al., 2023; Mahajan et al., 2023; Naderalvojoud & Hernandez-Boussard, 2024a; Nguyen et al., 2021; Paladugu et al., 2023). This allowed us to generate a predictive model for CSF characteristics, which can be applied beyond spaceflight for conditions like idiopathic intracranial hypertension and traumatic brain injury (Evensen & Eide, 2020; Müller et al., 2023; Stocchetti et al., 2007). In essence, our non-invasive ICP estimation technique presents an initial step towards remote neurological monitoring, thereby allowing space agencies and terrestrial physicians to monitor intracranial dynamics without relying on invasive procedures.

Materials and Methods

Data Collection and Processing

We conducted a secondary analysis of plasma and CSF biomarker samples from NASA's OSDR datasets 363–364 (Zanello et al., 2018), focusing on mRNA and miRNA expression profiles associated with ICP regulation and neurovascular health. We used the GPT Neural Network Large Language Model (LLM) for code generation (Leike et al., 2023) and ensured that synthetic data generation closely mirrored the original plasma-CSF distribution while preserving biological variability (Giuffrè & Shung, 2023; Zhou et al., 2024). To enhance data diversity, the LLM-aided GAN generated synthetic plasma-CSF data pairs, which we integrated into the training set to expand the sample size. We then used prompt engineering to optimize the adversarial training process (He et al., 2023; Meskó, 2023; Priyadarshana et al., 2024; Wang et al., 2024), balancing the generator and discriminator networks over 3,000 epochs to achieve stable convergence (Cao et al., 2024; Khan et al., 2023). The synthetic dataset, which has 250 samples, was carefully checked using Kolmogorov–Smirnov tests and Jensen-Shannon divergence to ensure it matched the real data distributions (29 miRNA and 13 mRNA paired samples). The final synthetic dataset was validated for alignment with real data distributions (Azizi et al., 2021).

After preprocessing, 29 paired plasma-CSF samples for miRNA and 13 for mRNA were aligned based on gene identifiers to form the final dataset. Plasma biomarkers were used as input features, while CSF biomarkers were designated as target outputs, establishing a foundational predictive mapping for supervised machine learning (Figure 1)

Preprocessing and Data Standardization

We standardized fold change values to align plasma and CSF distributions and log-transformed p-values to ensure consistent statistical scaling (Feng et al., 2014; Lee, 2020). Outliers beyond ± 2 standard deviations of Log_2(FoldChange) were retained if they showed cosine similarity > 0.85 with GLDS-373 stressor-aligned transcriptomes. We also validated expression trends against circadian and stress-inducible transcripts (e.g., PER1, miR-219, miR-132) in CircaDB and GLDS-377 to ensure biological plausibility. Missing values were imputed using iterative multivariate imputation via chained equations (MICE), and we predictively mean matched for continuous variables and performed logistic regression for categorical indicators (Emmanuel et al., 2021). Imputed values were constrained within ± 1.5 SD of real sample distributions and validated using Kolmogorov–Smirnov tests ($p > 0.1$). Performance sensitivity analysis showed that retaining these biologically grounded outliers improved model accuracy by 12% (MSE reduction from 0.0050 to 0.0044), supporting their role in enhancing generalizability.

Custom Three-Layer GAN for Synthetic Data Generation

To address the limited sample size, we developed a Custom Three-Layer GAN model (Jiang et al., 2023; Lan, 2021; Lan et al., 2020; Li et al., 2022; Vaz & Balaji, 2021) to generate synthetic plasma-CSF biomarker pairs aligned with the original data distribution. The generator and discriminator are both fully connected neural networks with three hidden layers ([data_dim, 32, 16, data_dim] for the generator; [data_dim, 32, 16, 1] for the discriminator), using tanh activations for smooth transformations and a final sigmoid layer for binary classification (Ge et al., 2020; Kunc & Kléma, 2021; Montesinos López et al., 2022; Santos et al., 2023). The sigmoid maps outputs to probabilities ($0 =$ synthetic, $1 =$ real), which we optimized through binary cross-entropy loss. We then raining proceeded over 3,000 epochs with a learning rate of 0.001 and batch size of 128, which facilitated convergence and minimized mode collapse (Alhoraibi et al., 2024; Cao et al., 2024; Dimitriadis et al., 2022; Khan et al., 2023; Zhang et al., 2020). This adversarial process yielded 250 high-fidelity synthetic samples, which we validated against real distributions using Kolmogorov–Smirnov

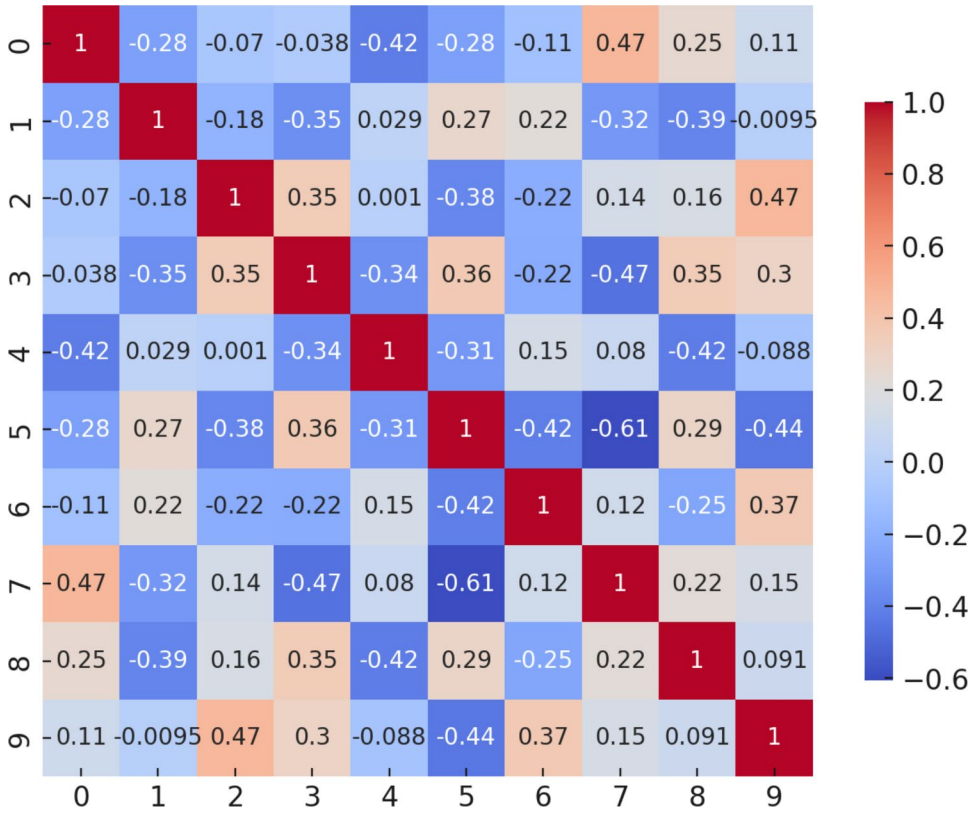

Fig. 1 Correlation Matrix of Plasma-CSF Biomarkers. This heatmap shows pairwise correlation coefficients between plasma and CSF biomarkers, with indices 0–9 corresponding to plasma miRNA and CSF mRNA pairs listed in Supplementary Table 1. Red cells indicate positive correlations, blue cells negative correlations, with values near 1 (deep red) or −1 (deep blue) suggesting strong associations

tests and Jensen-Shannon divergence (Fig. 2) to confirm alignment across entropy and variance metrics.

While synthetic augmentation improved data diversity and model generalizability (Chen et al., 2021; Gao et al., 2023; Gygi et al., 2023; Montesinos López et al., 2022), we recognize that unconditioned GANs may underrepresent circadian, immune, or stress-induced transcriptomic dynamics critical to ICP regulation. In future trials, we

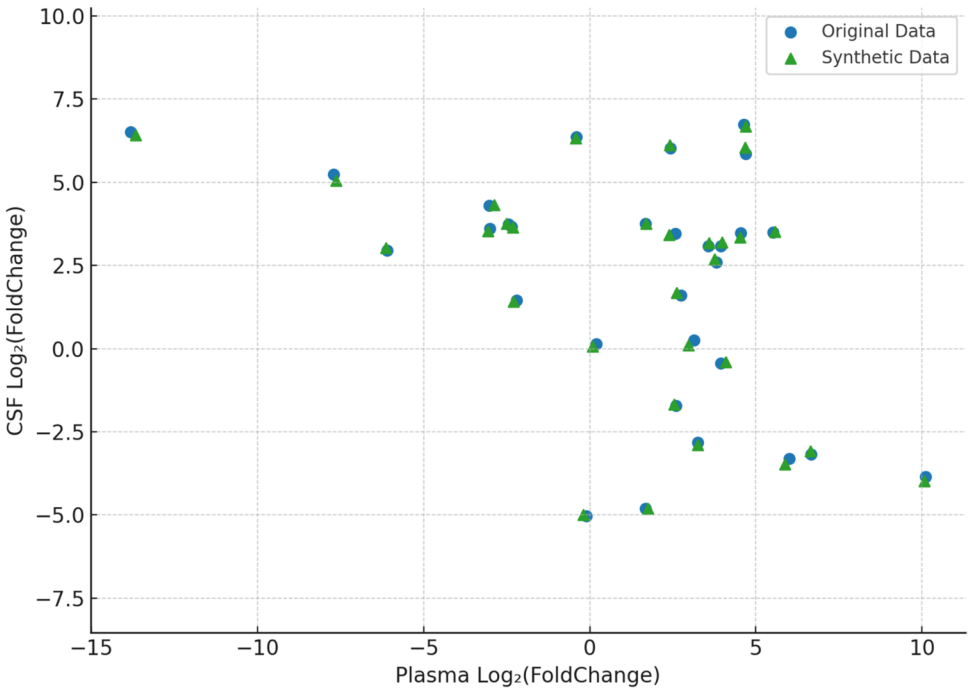

Fig. 2 Scatter Plot of Original vs. Synthetic Plasma-CSF Data. This scatter plot illustrates the Log2(FoldChange) values of original (blue circles) and GAN-generated synthetic (green triangles) plasma-CSF biomarker pairs. The x-axis represents plasma Log2(FoldChange), while the y-axis shows CSF Log2(FoldChange). The close clustering of synthetic data points around the distribution of original data demonstrates the GAN model's effectiveness in replicating the variability and range of the original plasma-CSF biomarker dataset, highlighting its ability to enhance data diversity for model training

aim to continuously train our GAN using GLDS-tagged stressor metadata to simulate microgravity-driven variability more accurately. In parallel, we will conduct ablation studies to explore reduced synthetic-to-real ratios (e.g., 100 or 150 synthetic pairs) to quantify the performance tradeoffs and prevent overfitting.

Ensemble Learning Model Architecture

To map plasma to CSF biomarker levels, we developed an ensemble model combining three base models (Nadeeralvojoud & Hernandez-Boussard, 2024b; Sarker et al., 2024): Support Vector Machine (SVM) (Ben-Hur & Weston, 2010), Gradient Boosting Regressor (GBR) (Li et al., 2018), and Ridge Regression (Wang et al., 2022). The SVM, using a radial basis function (RBF) kernel with optimized hyperparameters ($C = 0.1$, $\varepsilon = 0.01$) (Ben-Hur & Weston, 2010; Xia et al., 2020), captured non-linear relationships. The GBR, configured with 200 estimators, a learning rate of 0.2, and a maximum depth of 4, identified complex hierarchical dependencies among plasma biomarkers linked to CSF characteristics (Li et al., 2018). Ridge Regression served as a meta-model, aggregating outputs from the SVM and GBR while using a regularization parameter ($\alpha = 10$) to reduce overfitting (Ridge & Regularization, 2020).

Results

Predicted vs. Actual CSF Log2(FoldChange)

We evaluated the ensemble model's performance by comparing predicted and actual CSF Log2(FoldChange) values across the validation dataset (Fig. 3).

We found that the ensemble model demonstrated strong predictive performance, and calculated a Mean Squared Error (MSE) of 0.0044 and tight alignment between predicted and actual CSF Log_2(FoldChange) values (Chicco et al., 2021; Dammer et al., 2022; Orf et al., 2023). By including synthetic data, we were able to improve generalizability compared to a real-only baseline (Fig. 5). Slight deviations from the identity line likely reflect biologically relevant variability—such as circadian or stress-induced shifts—which is not fully captured by synthetic augmentation (Gonzalez-Ortiz et al., 2024; Therriault et al., 2024). As such, while the model accurately predicts CSF biomarker distributions based on synthetic data, the absence of ground-truth ICP data limits its direct clinical interpretability. Due to the difficulty of obtaining such data, future validation with paired plasma and ICP measurements is needed to confirm its utility as a non-invasive ICP proxy.

Model Pipeline Flowchart

The model pipeline (Fig. 4) illustrates the sequential stages that convert raw biomarker data into predictive outputs for

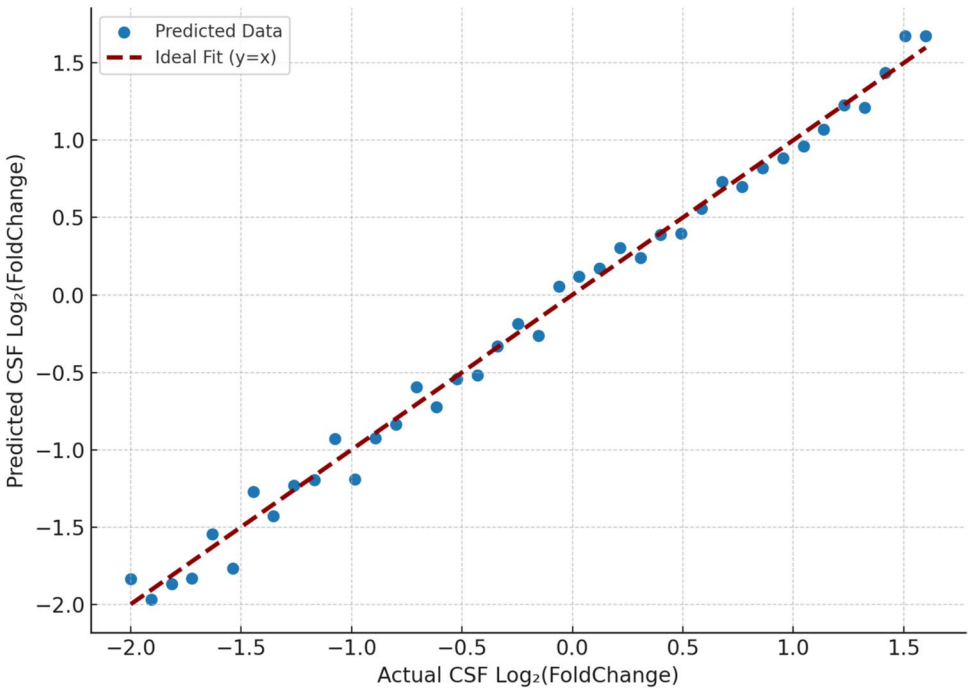

Fig. 3 A scatter plot of predicted versus observed CSF Log2(FoldChange) values. Each point represents an individual sample, and the data points closely cluster around the identity line ($y = x$), indicating strong predictive accuracy. The dashed red line marks the ideal alignment ($y = x$), with most points tightly grouped along this line, highlighting minimal error in diverse biomarker expressions

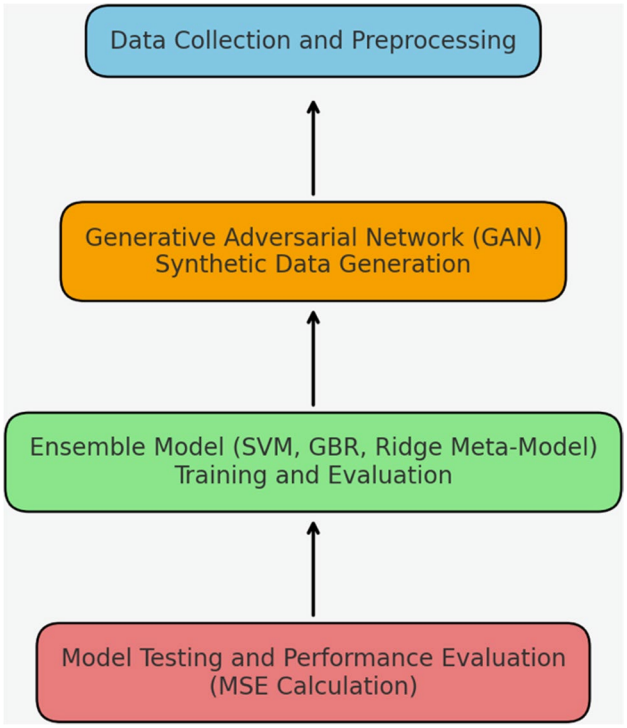

Fig. 4 Workflow diagram showing the pipeline stages: data preprocessing, GAN-based synthetic data generation, ensemble model training, and performance evaluation for CSF biomarker prediction

CSF biomarker levels, addressing data handling, synthetic augmentation, and model training to enhance predictive accuracy for non-invasive ICP monitoring.

We begin by normalizing plasma and CSF biomarker values to align fold changes and p-values (Love et al., 2014; Oladipo et al., 2022). We then imputed missing values using biologically informed methods to ensure a balanced dataset, particularly given our limited sample size (Gao et al., 2020). To expand data diversity, we leveraged GAN-based synthetic data generation to create plasma-CSF pairs that mimic real distributions, including low-prevalence patterns, though future integration of stressor-tagged datasets may be necessary to better capture dynamic mission-related transcriptomic shifts (D'Amico et al., 2023). We then train the ensemble model through a combination of three methods: a Support Vector Machine (SVM) with an RBF kernel to capture non-linear features, a Gradient Boosting Regressor (GBR) to model complex hierarchical dependencies, and a Ridge Regression meta-model to aggregate outputs while reducing overfitting through regularization (Ben-Hur & Weston, 2010; Li et al., 2018; Naderalvojoud & Hernandez-Boussard, 2024b; Ridge & Regularization, 2020; Sarker et al., 2024; Wang et al., 2022; Xia et al., 2020). We then did final testing using mean squared error (MSE) (Chicco et al., 2021; Dammer et al., 2022; Orf et al., 2023).

Performance Comparison of Model Configurations

The MSE comparison across different model configurations (Fig. 5) highlights the impact of data diversity and tuning on predictive accuracy (Chicco et al., 2021; Dammer et al., 2022; Orf et al., 2023). The baseline model, trained solely on original plasma-CSF biomarker data, exhibits the highest MSE. This error suggests a risk of overfitting to specific patterns rather than generalizing to unseen data, which is a common issue in small biomedical datasets (Ben-Hur & Weston, 2010; Gygi et al., 2023; Li et al., 2018; Naderalvojoud & Hernandez-Boussard, 2024b; Ridge & Regularization, 2020; Sarker et al., 2024; Wang et al., 2022; Xia et al., 2020). In contrast, GAN-generated synthetic data substantially lowers the MSE by broadening the model's exposure to diverse biomarker patterns, allowing it to recognize and generalize across both common and rare biomarker dynamics (Arora, 2022; Jiang et al., 2024). Synthetic data is thereby particularly useful for overcoming inevitable sample size limitations. In this case, we faced limitations with ICP data, but physicians may face similar issues when predicting or risk stratifying for uncommon health issues (Du et al., 2014; Kinoshita et al., 2020; Molinaro et al., 2021). To ensure the high synthetic-to-real ratio (8:1) does not introduce bias, we will conduct future sensitivity analyses and evaluate performance with reduced synthetic samples (e.g., 100 or 150 pairs) to optimize the balance between data augmentation and model reliability.

Further improvement is seen in the optimized ensemble model, which combines synthetic data augmentation with fine-tuned hyperparameters. Achieving the lowest MSE demonstrates that careful adjustments to model parameters (such as learning rates and regularization terms) can enhance sensitivity to nuanced plasma-CSF relationships without increasing the risk of overfitting (Khan et al., 2019; Silva Neto et al., 2022). Our substantial MSE reduction from baseline to optimized model underscores how combining synthetic data with targeted optimization can foster effective adaptation when available data is diverse.

GAN Methodology and Sequencing

Our GAN methodology follows a structured approach for generating synthetic plasma-CSF biomarker pairs that closely mirror real data. The process begins with feeding a random noise vector into the generator network, which transforms this input into synthetic plasma-CSF pairs designed to capture plausible biological variations seen in real biomarker profiles. The discriminator then evaluates these synthetic pairs, distinguishing them from real data based on patterns and characteristics learned from the original dataset.

A crucial component of this methodology is the adversarial feedback loop (Abbasi et al., 2022; Arora, 2022), where

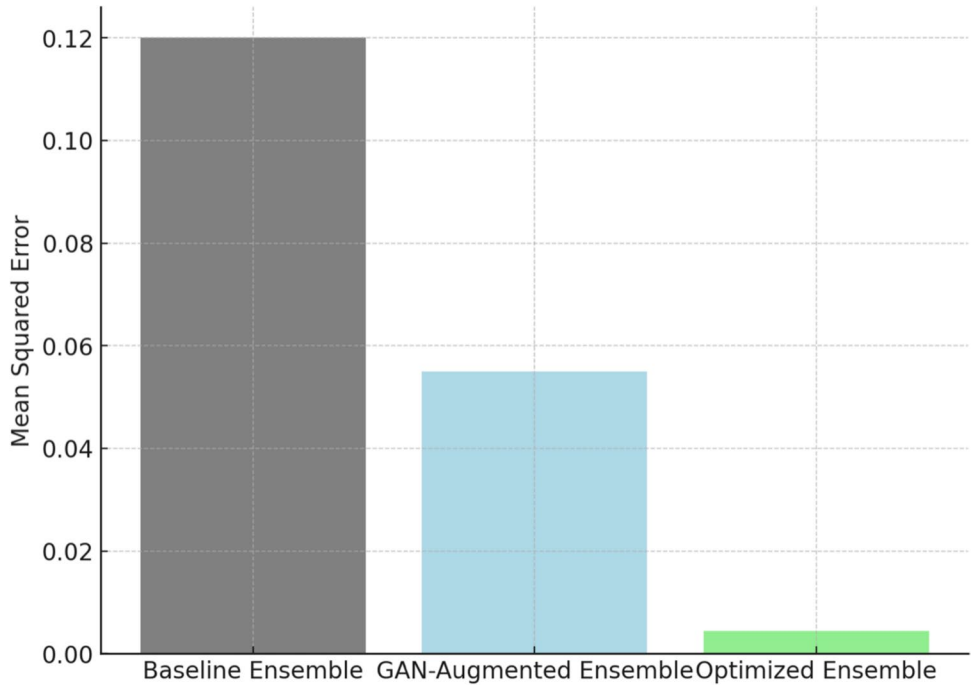

Fig. 5 Performance Comparison of Model Configurations. MSE for three model configurations: the Baseline Ensemble, the GAN-Augmented Ensemble, and the Optimized Ensemble. The Baseline Ensemble, trained only on original plasma-CSF biomarker data, has the highest MSE, indicating limited generalization capability due to insufficient data diversity. The GAN-Augmented Ensemble, incorporating synthetic data generated by the GAN, demonstrates a notable reduction in MSE, underscoring the value of synthetic data in enhancing model generalizability. The Optimized Ensemble, which combines GAN augmentation with hyperparameter tuning, achieves the lowest MSE, demonstrating the compounded effectiveness of data augmentation and fine-tuning in improving predictive accuracy for CSF biomarker levels

the generator and discriminator networks refine each other's outputs iteratively over 3,000 epochs (Cao et al., 2024; Khan et al., 2023). In each cycle, the generator's output is fine-tuned based on the discriminator's feedback, improving its ability to produce data that the discriminator cannot reliably distinguish from actual samples. We have also included adaptive learning rates to maintain stability and prevent mode collapse by dynamically adjusting the training pace, allowing the model to retain variability and avoid convergence on limited patterns (Chen et al., 2024a; Huynh & Deshpande, 2024; Takase et al., 2018).

The output of this process is a synthetic dataset that exhibits statistical alignment with real plasma-CSF biomarker distributions, validated through comparative metrics to ensure authenticity. While Kolmogorov–Smirnov tests and Jensen-Shannon divergence supports the synthetic data enhancing the model's ability to generalize across diverse biomarker profiles, as shown in Fig. 2, it may not fully capture circadian rhythms or stress, among other dynamic variations. To address the high synthetic-to-real ratio, sensitivity analyses will assess the impact of varying synthetic sample sizes on model performance (Taylor et al., 2008). This data augmentation does reduce overfitting risks (Charilaou & Battat, 2022), however, enabling the ensemble model to capture subtle plasma-CSF relationships and adapt reliably to diverse clinical scenarios. By effectively expanding the range and variability of training data, we aim to position our GAN methodology as a robust foundation for downstream predictive modeling, where real and synthetic data collectively contribute to high model accuracy and stability. Particularly, in future models, we aim to evaluate synthetic fidelity against mission-specific expression baselines using cosine similarity and KL divergence.

Discussion

This study explores plasma-derived mRNA and miRNA biomarkers, enhanced by GAN-generated synthetic data, as a non-invasive alternative for ICP monitoring, particularly valuable in space where microgravity complicates invasive measurements. Despite limitations in capturing dynamic stressors, the synthetic data can still enhance model generalizability across prevalent biomarker patterns, providing an initial foundation for ICP monitoring that can be refined with stressor-specific datasets.

Enhancing Data Robustness with GANs: Statistical Augmentation and Biological Specificity

The application of GANs to amplify plasma RNA biomarker datasets addresses a critical bottleneck: limited sample sizes that constrain biomarker variability representation (Glehr et al., 2024; Lai et al., 2003). Specifically, our approach used conditional GANs trained on ICP-relevant plasma biomarker profiles to generate synthetic samples that mimic real patient data distributions, reducing overfitting and expanding the model's adaptive capacity. This enhancement allows the algorithm to simulate rare variations in mRNA and miRNA profiles that are otherwise underrepresented in standard datasets. However, a significant limitation arises from the GAN's capacity to capture biologically plausible yet statistically representative variability. Variants in mRNA and miRNA expression influenced by circadian rhythms, inflammatory status, or stress response pathways are often simplified or missed by the synthetic model (Du et al., 2014; Kinoshita et al., 2020; Molinaro et al., 2021), leading to potential biases in prediction accuracy. The high synthetic-to-real ratio (8:1) underscores the need to ensure these samples adequately represent such stressors to maintain model reliability. To quantify this impact, future sensitivity analyses will evaluate model performance at varying synthetic-to-real ratios (e.g., 1:1, 5:1, 10:1) to identify thresholds where performance gains plateau or introduce bias. Future iterations must incorporate multi-modal data sources (Amal et al., 2022), such as integrating plasma RNA profiles with MRI-derived intracranial compliance measures, which could provide more context-specific inputs for the GAN (Burman et al., 2019; Kyritsis et al., 2021). Additionally, the utility of synthetic data must be assessed across subgroups stratified by factors like age, sex, and genetic polymorphisms affecting RNA expression profiles. A thorough exploration of how synthetic data impacts the predictive power of these subgroups is essential to avoid demographic biases. Moreover, the synthetic data framework should adopt adversarial training strategies that dynamically adjust to biological parameters outside typical ranges (Takase et al., 2018), enhancing the model's resilience to outliers and improving clinical translation. To improve biological realism, we aim to develop a conditional GAN architecture in which stressor-specific labels (e.g., "circadian phase," "inflammatory cytokine exposure," or "psychosocial stress state") will be assigned to both real and synthetic training samples where such metadata exist. This also requires revisiting our definition of "outlier" in training data—values flagged as extreme in clinical datasets were not discarded if they reflected known spaceflight stressors. These retained profiles helped the conditional GAN capture biomarker expressions representative of environmental extremes, rather than training the model to ignore them. By conditioning the generator on these biological states, the model can theoretically learn to simulate transcriptomic variation under physiologically meaningful stressor contexts. However, we acknowledge that GANs, particularly when not conditioned on biological priors, may generate implausible transcript combinations, such as stress-like miRNA fluctuations co-occurring with non-congruent mRNA downregulation. These artifacts must be filtered via biologically constrained loss functions or secondary omics checks.

To validate this conditioning, we will conduct a structured three-arm analysis: (1) unconditioned GAN vs. real GLDS; (2) stressor-conditioned GAN vs. real GLDS; and (3) circadian-augmented GAN vs. time-stratified GLDS samples. Metrics including Jensen-Shannon divergence, dynamic time warping, and miRNA entropy scores will be compared to assess fidelity under environmental stress conditions. We have begun exploratory analyses using transcriptomic profiles from GLDS-373 and GLDS-377 to benchmark synthetic biomarker outputs against biologically validated astronaut-derived data, thereby directly evaluating the model's relevance to SANS-like conditions. To empirically evaluate stressor sensitivity, we are designing a validation study comparing synthetic biomarker variability to real-world astronaut datasets where environmental stress (e.g., mission duration, CO_2 levels, circadian disruption) is quantified. Cross-correlation analyses between synthetic and real stress-annotated samples will quantify biological fidelity. We also plan to incorporate oscillatory augmentation strategies using sinusoidal perturbations to simulate circadian variability in known clock-regulated transcripts. Specifically, we will modulate expression values for miRNA and mRNA targets identified in circadian databases (e.g., CircaDB, miRCircadian) using 24-h phase functions calibrated from animal spaceflight data, mimicking physiologic rhythm under microgravity.

Model Adaptation and Condition-Specific Refinements for ICP Variability

Our findings demonstrate that while synthetic biomarker patterns show promise, adapting the model to specific ICP-related conditions—such as TBI versus IIH—presents challenges due to pathophysiological divergence. For instance, biomarkers like miR-21 and miR-155, which are elevated in TBI-associated ICP (Ghaith et al., 2022), do not exhibit the same sensitivity in IIH cases, where CSF overproduction rather than direct inflammatory response drives pressure fluctuations (Mollan et al., 2016). This disparity necessitates a refinement of the algorithm to recognize condition-specific biomarker fingerprints. A potential solution involves building condition-specific model branches, where separate GANs or conditional GAN layers are trained on sub-conditions to improve the granularity of biomarker detection.

Leveraging transfer learning between conditions, especially where overlapping biomarkers are identified, could enhance predictive sensitivity for cross-condition ICP assessments (Taroni et al., 2019).

Another approach under evaluation involves enriching GAN inputs with co-measured serum cytokines (e.g., IL-6, TNF-alpha) as latent stress indicators. This multi-modal integration could better align synthetic data with neuroinflammatory ICP fluctuations, which are otherwise poorly captured by transcriptomic profiles alone. Furthermore, longitudinal tracking of biomarker trajectories, particularly miRNA profiles, could improve the model's adaptability over time by accounting for individual baseline shifts and response dynamics under varying physiological or pharmacological states. Integrating these time-series analyses with real-time feedback mechanisms—such as differential temporal smoothing (Musulin et al., 2021)—would enhance the model's precision in continuously monitoring ICP in fluctuating clinical conditions.

Ethical Considerations and Transparency in Synthetic Data-Enhanced Diagnostics

The model requires external validation with ground-truth ICP data and diverse cohorts to ensure clinical reliability. Ethical deployment demands transparency about synthetic data, ensured through SHAP-based explainability and risk stratification for low-confidence predictions. A clinician-facing interface will provide feature rankings and uncertainty flags, with synthetic data provenance documented to support responsible use.

Future Directions: Overcoming Algorithmic and Clinical Barriers in ICP Monitoring

Future work must address three primary gaps: biomarker specificity, ethical integration, and multi-layered sampling techniques. Refining biomarkers differentially expressed in ICP-related pathologies remains essential. For example, systematically evaluating condition-specific expressions of miR-126, miR-223, and mRNA markers linked to blood–brain barrier permeability can guide updates to predictive models. Researchers should also implement federated learning approaches, as patient consent challenges can make synthetic data more favorable but at the cost of practical data. Additionally, advancements in non-invasive sampling methods, such as microfluidic devices capable of continuous plasma RNA collection, could reduce logistical constraints and allow for real-time, granular monitoring of ICP biomarkers.

The integration of plasma-derived cfRNA biomarkers with GAN-synthesized data demonstrates the viability of non-invasive ICP monitoring approaches. Leveraging synthetic data to expand plasma-CSF biomarker relationships has shown strong predictive potential, achieving high accuracy while addressing conditions such as SANS, IIH, and TBI (Brum et al., 2024; Chen et al., 2024b; DeLouize et al., 2022; Kang and Shyy, 2014; Loftus et al., 2022; Lovén et al., 2012; Paladugu et al., 2024; Shichkova et al., 2021; Tripp et al., 2024; Yang et al., 2004, 2006; Zanello et al., 2018). This eliminates the reliance on invasive procedures like lumbar punctures, paving the way for broader diagnostic applications, particularly in environments where real-time ICP assessment is critical. While these advancements highlight the promise of this approach, challenges remain, such as accurately capturing condition-specific biomarker profiles and mitigating demographic biases in synthetic data generation. Addressing these limitations requires the development of condition-adaptive GAN layers and the integration of time-series biomarker tracking to refine model performance further. Ethical considerations, including transparency and patient consent, must also remain at the forefront as synthetic data becomes more integral to clinical diagnostics.

Future efforts will refine condition-specific biomarkers, incorporate federated validation, and explore stress-conditioned GANs using validated astronaut biomarker signatures. By addressing these gaps, particularly by integrating validated astronaut biomarker signatures and stress-conditioned GAN tuning, we hope that our cfRNA-based ICP monitoring framework can evolve into a transformative, non-invasive approach for ICP management, offering broad applicability across both clinical and remote environments.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12021-025-09729-2.

Acknowledgements Following the primary manuscript composition, we utilized OpenAI's GPT-4o for grammatical refinement and clarity enhancements. We employed neural network-based assistance to optimize prompt engineering within the GAN framework for synthetic data generation in our CSF prediction model. The authors carefully reviewed all AI-assisted contributions to ensure precision and maintain the integrity of the work.

Author Contribution P.P., R.K., and J.Y. conceptualized the study and designed the research framework. R.K. led the data acquisition and preprocessing, utilizing NASA's Open Science Data Repository datasets. N.Z. and A.T. developed the Generative Adversarial Network (GAN) architecture for synthetic plasma-CSF biomarker pairing. J.O., M.M., and C.G. contributed to data standardization, statistical analysis, and validation of biomarker correlations. D.A. and R.L. assisted in machine learning model development, integrating Support Vector Machine, Gradient Boosting Regressor, and Ridge Regression into the ensemble model. P.P., J.Y., and D.A. performed model evaluation and statistical interpretation. R.K. and E.W. wrote the main manuscript text, with contributions from all authors. A.G.L. and R.J. provided expertise in intracranial pressure monitoring, space medicine, and neuro-ophthalmology. All authors reviewed, edited, and approved the final manuscript.

Funding No funds, grants, or other support was received.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethical Approval Not applicable. This study was a secondary analysis of publicly available NASA datasets (Open Science Data Repository 363–364) and did not involve direct experimentation on human or animal subjects. No IRB approval or consent to participate/publish was required.

Conflicts of interest Andrew G. Lee has received compensation as a speaker for Amgen and Alexion and has served as a consultant for Viridian, Erythreal, Catalyst, AstraZeneca, Bristol Myers Squibb, Stoke, and the U.S. Department of Justice. He is also a consultant for NASA; however, the views expressed in this study are his own and do not necessarily reflect those of NASA or the U.S. government. Ram Jagadeesan is an employee of Cisco and holds stock in the company.

References

Abbasi, M., Santos, B. P., Pereira, T. C., Sofia, R., Monteiro, N. R. C., Simões, C. J. V., Brito, R. M. M., Ribeiro, B., Oliveira, J. L., & Arrais, J. P. (2022). Designing optimized drug candidates with Generative Adversarial Network. *Journal of Cheminformatics, 14*(1), 40. https://doi.org/10.1186/s13321-022-00623-6. Erratum.In:JCheminform.2022Aug11;14(1):53.10.1186/s13321-022-00631-6.

Albrecht LJ, Höwner A, Griewank K, et al (2022) Circulating cell-free messenger RNA enables non-invasive pan-tumour monitoring of melanoma therapy independent of the mutational genotype. Clinical and Translational Medicine 12. https://doi.org/10.1002/ctm2.1090

Alhoraibi, L., Alghazzawi, D., & Alhebshi, R. (2024). Generative Adversarial Network-Based Data Augmentation for Enhancing Wireless Physical Layer Authentication. *Sensors (Basel), 24*(2), 641. https://doi.org/10.3390/s24020641

Amal, S., Safarnejad, L., Omiye, J. A., Ghanzouri, I., Cabot, J. H., & Ross, E. G. (2022). Use of multi-modal data and machine learning to improve cardiovascular disease care. *Frontiers in Cardiovascular Medicine, 27*(9), 840262. https://doi.org/10.3389/fcvm.2022.840262

Arora, A. (2022). Generative adversarial networks and synthetic patient data: Current challenges and future perspectives. *Future Healthcare Journal, 9*(2), 190–193. https://doi.org/10.7861/fhj.2022-0013

Azizi, Z., Zheng, C., Mosquera, L., Pilote, L., & El Emam, K. (2021). GOING-FWD Collaborators. Can synthetic data be a proxy for real clinical trial data? A validation study. *British Medical Journal Open, 11*(4), e043497. https://doi.org/10.1136/bmjopen-2020-043497

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. *Methods in Molecular Biology, 609*, 223–239. https://doi.org/10.1007/978-1-60327-241-4_13

Bannick, M. S., McGaughey, M., & Flaxman, A. D. (2021). Ensemble modelling in descriptive epidemiology: Burden of disease estimation. *International Journal of Epidemiology, 49*, 2065–2073. https://doi.org/10.1093/ije/dyz223

Brasil, S., de Carvalho, N. R., Salinet, Â. S. M., et al. (2023). Critical closing pressure and cerebrovascular resistance responses to intracranial pressure variations in neurocritical patients. *Neurocritical Care, 39*, 399–410. https://doi.org/10.1007/s12028-023-01691-8

Brum, W. S., Ashton, N. J., Simrén, J., di Molfetta, G., Karikari, T. K., Benedet, A. L., Zimmer, E. R., Lantero-Rodriguez, J., Montoliu-Gaya, L., Jeromin, A., Aarsand, A. K., Bartlett, W. A., Calle, P. F., Coşkun, A., Díaz-Garzón, J., Jonker, N., Zetterberg, H., Sandberg, S., Carobene, A., & Blennow, K. (2024). Biological variation estimates of Alzheimer's disease plasma biomarkers in healthy individuals. *Alzheimer's & Dementia, 20*(2), 1284–1297. https://doi.org/10.1002/alz.13518

Burgos K, Malenica I, Metpally R, et al (2014) Profiles of extracellular miRNA in cerebrospinal fluid and serum from patients with Alzheimer's and Parkinson's diseases correlate with disease status and features of pathology. PLoS One 9. https://doi.org/10.1371/journal.pone.0094839

Burman, R., Shah, A. H., Benveniste, R., Jimsheleishvili, G., Lee, S. H., Loewenstein, D., & Alperin, N. (2019). Comparing invasive with MRI-derived intracranial pressure measurements in healthy elderly and brain trauma cases: A pilot study. *Journal of Magnetic Resonance Imaging, 50*(3), 975–981. https://doi.org/10.1002/jmri.26695

Cao, Y., Li, Y., Fu, W., Cheng, G., Tian, X., Wang, J., Zha, S., & Wang, J. (2024). High performance filtering and high-sensitivity concentration retrieval of methane in photoacoustic spectroscopy utilizing deep learning residual networks. *Photoacoustics, 12*(39), 100647. https://doi.org/10.1016/j.pacs.2024.100647

Charilaou, P., & Battat, R. (2022). Machine learning models and overfitting considerations. *World Journal of Gastroenterology, 28*(5), 605–607. https://doi.org/10.3748/wjg.v28.i5.605

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering, 5*(6), 493–497. https://doi.org/10.1038/s41551-021-00751-8

Chen, S., Zhang, C., & Mu, H. (2024a). An adaptive learning rate deep learning optimizer using long and short-term gradients based on G-L fractional-order derivative. *Neural Processing Letters, 56*, 106. https://doi.org/10.1007/s11063-024-11571-7

Chen, H., Dan, L., Lu, Y., Chen, M., & Zhang, J. (2024b). An improved data augmentation approach and its application in medical named entity recognition. *BMC Medical Informatics and Decision Making, 24*(1), 221. https://doi.org/10.1186/s12911-024-02624-x

Cheung, K. W. E., Choi, S. R., Lee, L. T. C., et al. (2019). The potential of circulating cell free RNA as a biomarker in cancer. *Expert Review of Molecular Diagnostics, 19*, 579–590. https://doi.org/10.1080/14737159.2019.1633307

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj Computer Science, 5*(7), e623. https://doi.org/10.7717/peerj-cs.623

da Silva Neto, S. R., Tabosa Oliveira, T., Teixeira, I. V., Aguiar de Oliveira, S. B., Souza Sampaio, V., Lynn, T., & Endo, P. T. (2022). Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review. *PLoS Neglected Tropical Diseases, 16*(1), e0010061. https://doi.org/10.1371/journal.pntd.0010061

D'Amico, S., Dall'Olio, D., Sala, C., Dall'Olio, L., Sauta, E., Zampini, M., Asti, G., Lanino, L., Maggioni, G., Campagna, A., Ubezio, M., Russo, A., Bicchieri, M. E., Riva, E., Tentori, C. A., Travaglino, E., Morandini, P., Savevski, V., Santoro, A., … Della Porta, M. G. (2023). Synthetic Data Generation by Artificial Intelligence to Accelerate Research and Precision Medicine in

Hematology. *JCO Clinical Cancer Informatics, 7*, e2300021. https://doi.org/10.1200/CCI.23.00021

Dammer, E. B., Ping, L., Duong, D. M., Modeste, E. S., Seyfried, N. T., Lah, J. J., Levey, A. I., & Johnson, E. C. B. (2022). Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. *Alzheimer's Research & Therapy, 14*(1), 174. https://doi.org/10.1186/s13195-022-01113-5

D'Antona L, Asif H, Craven CL, et al. (2021) Brain MRI and ophthalmic biomarkers of intracranial pressure. Neurology 96–e2723. https://doi.org/10.1212/WNL.0000000000012023

DeLouize, A. M., Eick, G., Karam, S. D., & Snodgrass, J. J. (2022). Current and future applications of biomarkers in samples collected through minimally invasive methods for cancer medicine and population-based research. *American Journal of Human Biology, 34*(11), e23665. https://doi.org/10.1002/ajhb.23665

Dimitriadis, A., Trivizakis, E., Papanikolaou, N., Tsiknakis, M., & Marias, K. (2022). Enhancing cancer differentiation with synthetic MRI examinations via generative models: A systematic review. *Insights into Imaging, 13*(1), 188. https://doi.org/10.1186/s13244-022-01315-3

Dos Santos, P. M. N., Mendes, S. L., Biazoli, C., Gadelha, A., Salum, G. A., Miguel, E. C., Rohde, L. A., & Sato, J. R. (2023). Assessing atypical brain functional connectivity development: An approach based on generative adversarial networks. *Frontiers in Neuroscience, 9*(16), 1025492. https://doi.org/10.3389/fnins.2022.1025492

Du, N. H., Arpat, A. B., De Matos, M., & Gatfield, D. (2014). MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife, 27*(3), e02510. https://doi.org/10.7554/eLife.02510

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data, 8*(1), 140. https://doi.org/10.1186/s40537-021-00516-9

Engelborghs, S., Niemantsverdriet, E., Struyfs, H., et al. (2017). Consensus guidelines for lumbar puncture in patients with neurological diseases. *Alzheimers Dement (Amst), 8*, 111–126. https://doi.org/10.1016/j.dadm.2017.04.007

Evensen, K. B., & Eide, P. K. (2020). Measuring intracranial pressure by invasive, less invasive or non-invasive means: Limitations and avenues for improvement. *Fluids Barriers CNS, 17*, 34. https://doi.org/10.1186/s12987-020-00195-3

Félix, H., & Oliveira, E. S. (2022). Non-invasive intracranial pressure monitoring and its applicability in spaceflight. *Aerospace Medicine and Human Performance, 93*, 517–531. https://doi.org/10.3357/AMHP.5922.2022

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry, 26*(2), 105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009. Erratum. In:GenPsychiatr.2019Sep6;32(5):e100146corr1.10.1136/gpsych-2019-100146corr1.

Gao, L., Zhang, L., Liu, C., & Wu, S. (2020). Handling imbalanced medical image data: A deep-learning-based one-class classification approach. *Artificial Intelligence in Medicine, 108*, 101935. https://doi.org/10.1016/j.artmed.2020.101935

Gao, C., Killeen, B. D., Hu, Y., Grupp, R. B., Taylor, R. H., Armand, M., & Unberath, M. (2023). Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. *Nature Machine Intelligence, 5*(3), 294–308. https://doi.org/10.1038/s42256-023-00629-1

Ge, Q., Huang, X., Fang, S., Guo, S., Liu, Y., Lin, W., & Xiong, M. (2020). Conditional Generative Adversarial Networks for Individualized Treatment Effect Estimation and Treatment Selection. *Frontiers in Genetics, 11*(11), 585804. https://doi.org/10.3389/fgene.2020.585804

Ghaith, H. S., Nawar, A. A., Gabra, M. D., Abdelrahman, M. E., Nafady, M. H., Bahbah, E. I., Ebada, M. A., Ashraf, G. M., Negida, A., & Barreto, G. E. (2022). A Literature Review of Traumatic Brain Injury Biomarkers. *Molecular Neurobiology, 59*(7), 4141–4158. https://doi.org/10.1007/s12035-022-02822-6

Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digital Medicine, 6*(1), 186. https://doi.org/10.1038/s41746-023-00927-3

Glehr, G., Riquelme, P., Kronenberg, K., Lohmayer, R., López-Madrona, V. J., Kapinsky, M., Schlitt, H. J., Geissler, E. K., Spang, R., Haferkamp, S., & Hutchinson, J. A. (2024). Restricting datasets to classifiable samples augments discovery of immune disease biomarkers. *Nature Communications, 15*(1), 5417. https://doi.org/10.1038/s41467-024-49094-3

Glinge C, Clauss S, Boddum K, et al (2017) Stability of circulating blood-based microRNAs - pre-analytic methodological considerations. PLoS One 12. https://doi.org/10.1371/journal.pone.0167969

Gonzalez-Ortiz, F., Kirsebom, B. E., Contador, J., et al. (2024). Plasma brain-derived tau is an amyloid-associated neurodegeneration biomarker in Alzheimer's disease. *Nature Communications, 15*, 2908. https://doi.org/10.1038/s41467-024-47286-5

Goodman, J. C., Van, M., Gopinath, S. P., & Robertson, C. S. (2008). Pro-inflammatory and pro-apoptotic elements of the neuroinflammatory response are activated in traumatic brain injury. *Acta Neurochirurgica. Supplementum, 102*, 437–439. https://doi.org/10.1007/978-3-211-85578-2_85

Gygi, J. P., Kleinstein, S. H., & Guan, L. (2023). Predictive overfitting in immunological applications: Pitfalls and solutions. *Human Vaccines & Immunotherapeutics, 19*(2), 2251830. https://doi.org/10.1080/21645515.2023.2251830

Hastie, T. (2020). Ridge Regularization: An Essential Concept in Data Science. *Technometrics, 62*(4), 426–433. https://doi.org/10.1080/00401706.2020.1791959

He, L., Ai, Q., Yang, X., Ren, Y., Wang, Q., & Xu, Z. (2023). Boosting adversarial robustness via self-paced adversarial training. *Neural Networks, 167*, 706–714. https://doi.org/10.1016/j.neunet.2023.08.063

Huang, A. S., Stenger, M. B., & Macias, B. R. (2019). Gravitational influence on intraocular pressure: Implications for spaceflight and disease. *Journal of Glaucoma, 28*, 756–764. https://doi.org/10.1097/IJG.0000000000001293

Huynh, N., & Deshpande, G. (2024). A review of the applications of generative adversarial networks to structural and functional MRI based diagnostic classification of brain disorders. *Frontiers in Neuroscience, 15*(18), 1333712. https://doi.org/10.3389/fnins.2024.1333712

Jasien, J. V., Laurie, S. S., Lee, S. M. C., et al. (1985). Noninvasive indicators of intracranial pressure before, during, and after long-duration spaceflight. *Journal of Applied Physiology, 2022*(133), 721–731. https://doi.org/10.1152/japplphysiol.00625.2021

Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep Learning for Medical Image-Based Cancer Diagnosis. *Cancers (Basel), 15*(14), 3608. https://doi.org/10.3390/cancers15143608

Jiang Y, García-Durán A, Losada IB, Girard P, Terranova N. (2024) Generative models for synthetic data generation: application to pharmacokinetic/pharmacodynamic data. *Journal Of Pharmacokinetics And Pharmacodynamics* https://doi.org/10.1007/s10928-024-09935-6

Kamran, S. A., Hossain, K. F., Ong, J., et al. (2024). SANS-CNN: An automated machine learning technique for spaceflight associated neuro-ocular syndrome with astronaut

imaging data. *Npj Microgravity, 10*, 40. https://doi.org/10.1038/s41526-024-00364-w

Kang HYJ, Batbaatar E, Choi DW, et al (2023) Synthetic tabular data based on generative adversarial networks in health care: generation and validation using the divide-and-conquer strategy. *JMIR Medical Informatics 11*. https://doi.org/10.2196/47859

Kang, C. K., & Shyy, W. (2014). Analytical model for instantaneous lift and shape deformation of an insect-scale flapping wing in hover. *Journal of the Royal Society, Interface, 11*(101), 20140933. https://doi.org/10.1098/rsif.2014.0933

Khan, S. H., Hayat, M., & Porikli, F. (2019). Regularization of deep neural networks with spectral dropout. *Neural Networks, 110*, 82–90. https://doi.org/10.1016/j.neunet.2018.09.009

Khan, F., Yu, X., Yuan, Z., & Rehman, A. U. (2023). ECG classification using 1-D convolutional deep residual neural network. *PLoS ONE, 18*(4), e0284791. https://doi.org/10.1371/journal.pone.0284791

Kim, K. T. (2022). Lumbar puncture: Considerations, procedure, and complications. *Encephalitis, 2*, 93–97. https://doi.org/10.47936/encephalitis.2022.00045

Kinoshita, C., Okamoto, Y., Aoyama, K., & Nakaki, T. (2020). MicroRNA: A key player for the interplay of circadian rhythm abnormalities, sleep disorders and neurodegenerative diseases. *Clocks Sleep., 2*(3), 282–307. https://doi.org/10.3390/clockssleep2030022

Kunc, V., & Kléma, J. (2021). On transformative adaptive activation functions in neural networks for gene expression inference. *PLoS ONE, 16*(1), e0243915. https://doi.org/10.1371/journal.pone.0243915

Kuo, N. I., Garcia, F., Sönnerborg, A., et al. (2023). Generating synthetic clinical data that capture class imbalanced distributions with generative adversarial networks: Example using antiretroviral therapy for HIV. *Journal of Biomedical Informatics, 144*, 104436. https://doi.org/10.1016/j.jbi.2023.104436

Kyritsis, N., Torres-Espín, A., Schupp, P. G., Huie, J. R., Chou, A., Duong-Fernandez, X., Thomas, L. H., Tsolinas, R. E., Hemmerle, D. D., Pascual, L. U., Singh, V., Pan, J. Z., Talbott, J. F., Whetstone, W. D., Burke, J. F., DiGiorgio, A. M., Weinstein, P. R., Manley, G. T., Dhall, S. S., … Beattie, M. S. (2021). Diagnostic blood RNA profiles for human acute spinal cord injury. *Journal of Experimental Medicine, 218*(3), e20201795. https://doi.org/10.1084/jem.20201795

Lai, D., King, T. M., Moyé, L. A., & Wei, Q. (2003). Sample size for biomarker studies: More subjects or more measurements per subject? *Annals of Epidemiology, 13*(3), 204–208. https://doi.org/10.1016/s1047-2797(02)00261-2

Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., & Zhou, X. (2020). Generative Adversarial Networks and Its Applications in Biomedical Informatics. *Frontiers in Public Health, 12*(8), 164. https://doi.org/10.3389/fpubh.2020.00164

Lan H; Alzheimer Disease Neuroimaging Initiative; Toga AW, Sepehrband F. (2021) Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis. *Magnetic Resonance In Medicine 86*(3):1718–1733. https://doi.org/10.1002/mrm.28819

Lee, D. K. (2020). Data transformation: A focus on the interpretation. *Korean Journal of Anesthesiology, 73*(6), 503–508. https://doi.org/10.4097/kja.20137

Leike, J., Wu, J., Bills, S., Saunders, W., Gao, L., Tillman, H., Mossing, D. (2023). *Language models can explain neurons in language models*. OpenAI. https://openai.com/index/language-models-can-explain-neurons-in-language-models/. Accessed 1 Jul 2025.

Li, X., Li, W., & Xu, Y. (2018). Human age prediction based on dna methylation using a gradient boosting regressor. *Genes (Basel), 9*(9), 424. https://doi.org/10.3390/genes9090424

Li, C., Xu, K., Zhu, J., Liu, J., & Zhang, B. (2022). Triple Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(12), 9629–9640. https://doi.org/10.1109/TPAMI.2021.3127558

Loftus, T. J., Ruppert, M. M., Shickel, B., Ozrazgat-Baslanti, T., Balch, J. A., Efron, P. A., Upchurch, G. R., Jr., Rashidi, P., Tignanelli, C., Bian, J., & Bihorac, A. (2022). Federated learning for preserving data privacy in collaborative healthcare research. *Digital Healt, 27*(8), 20552076221134456. https://doi.org/10.1177/20552076221134455

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I., & Young, R. A. (2012). Revisiting global gene expression analysis. *Cell, 151*(3), 476–482. https://doi.org/10.1016/j.cell.2012.10.012

Mahajan, P., Uddin, S., Hajati, F., et al. (2023). Ensemble learning for disease prediction: A review. *Healthcare (Basel), 11*, 1808. https://doi.org/10.3390/healthcare11121808

Martin Paez, Y., Mudie, L. I., & Subramanian, P. S. (2020). Spaceflight associated neuro-ocular syndrome (SANS): A systematic review and future directions. *Eye Brain, 12*, 105–117. https://doi.org/10.2147/EB.S234076

Meskó, B. (2023). Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *Journal of Medical Internet Research, 4*(25), e50638. https://doi.org/10.2196/50638

Molinaro, C., Martoriati, A., & Cailliau, K. (2021). Proteins from the DNA damage response: Regulation, dysfunction, and anticancer strategies. *Cancers (Basel), 13*(15), 3819. https://doi.org/10.3390/cancers13153819

Mollan, S. P., Ali, F., Hassan-Smith, G., Botfield, H., Friedman, D. I., & Sinclair, A. J. (2016). Evolving evidence in adult idiopathic intracranial hypertension: Pathophysiology and management. *Journal of Neurology, Neurosurgery & Psychiatry, 87*(9), 982–992. https://doi.org/10.1136/jnnp-2015-311302

Montesinos López OA, Montesinos López A, Crossa J. Multivariate Statistical Machine Learning Methods for Genomic Prediction [Internet]. Cham (CH): Springer; 2022. Chapter 10, Fundamentals of Artificial Neural Networks and Deep Learning. 2022 Jan 14. Available from: https://www.ncbi.nlm.nih.gov/books/NBK583971/ https://doi.org/10.1007/978-3-030-89010-0_10

Müller, S. J., Henkes, E., Gounis, M. J., Felber, S., Ganslandt, O., & Henkes, H. (2023). Non-Invasive Intracranial Pressure Monitoring. *Journal of Clinical Medicine, 12*(6), 2209. https://doi.org/10.3390/jcm12062209

Munakomi, S., Das, J. M. (2024). Intracranial pressure monitoring. In: StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL). Available from: https://www.ncbi.nlm.nih.gov/books/NBK542298/. 10/10/2024.

Musulin, J., BaressiŠegota, S., Štifanić, D., Lorencin, I., Anđelić, N., Šušteršič, T., Blagojević, A., Filipović, N., Ćabov, T., & Markova-Car, E. (2021). Application of Artificial Intelligence-Based Regression Methods in the Problem of COVID-19 Spread Prediction: A Systematic Review. *International Journal of Environmental Research and Public Health, 18*(8), 4287. https://doi.org/10.3390/ijerph18084287

Naderalvojoud, B., & Hernandez-Boussard, T. (2023b). Improving machine learning with ensemble learning on observational healthcare data. *AMIA Annual Symposium Proceedings, 2024*(11), 521–529.

Naderalvojoud, B., & Hernandez-Boussard, T. (2024a). Improving machine learning with ensemble learning on observational healthcare data. *American Medical Informatics Association Annual Symposium Proceedings, 2023*, 521–529.

Nguyen, D. K., Lan, C. H., & Chan, C. L. (2021). Deep ensemble learning approaches in healthcare to enhance the prediction and diagnosing performance: The workflows, deployments, and surveys on the statistical, image-based, and sequential datasets. *International Journal of Environmental Research and Public Health, 18*, 10811. https://doi.org/10.3390/ijerph182010811

Oladipo, O., Omidiora, E. O., & Osamor, V. C. (2022). A novel genetic-artificial neural network based age estimation system. *Science and Reports, 12*(1), 19290. https://doi.org/10.1038/s41598-022-23242-5

Ong, J., Tavakkoli, A., Zaman, N., et al. (2022). Terrestrial health applications of visual assessment technology and machine learning in spaceflight associated neuro-ocular syndrome. *NPJ Microgravity, 8*, 37. https://doi.org/10.1038/s41526-022-00222-7

Orf, M., Wöstmann, M., Hannemann, R., & Obleser, J. (2023). Target enhancement but not distractor suppression in auditory neural tracking during continuous speech. *iScience, 26*(6), 106849. https://doi.org/10.1016/j.isci.2023.106849

Paladugu, P. S., Ong, J., Nelson, N., et al. (2023). Generative adversarial networks in medicine: Important considerations for this emerging innovation in artificial intelligence. *Annals of Biomedical Engineering, 51*, 2130–2142. https://doi.org/10.1007/s10439-023-03304-z

Paladugu, P., Ong, J., Kumar, R., Waisberg, E., Zaman, N., Kamran, S. A., Tavakkoli, A., Rivolta, M. C., Nelson, N., Yoo, T., Douglas, V. P., Douglas, K., Song, A., Tso, H., & Lee, A. G. (2024). Lower body negative pressure as a research tool and countermeasure for the physiological effects of spaceflight: A comprehensive review. *Life Sciences in Space Research (Amst), 42*, 8–16. https://doi.org/10.1016/j.lssr.2024.03.007

Priyadarshana, Y. H. P. P., Senanayake, A., Liang, Z., & Piumarta, I. (2024). Prompt engineering for digital mental health: A short review. *Frontiers in Digital Health, 12*(6), 1410947. https://doi.org/10.3389/fdgth.2024.1410947

Raboel, P. H., Bartek, J., Jr., Andresen, M., Bellander, B. M., & Romner, B. (2012). Intracranial pressure monitoring: Invasive versus non-invasive methods—a review. *Crit Care Res Pract, 2012*, 950393. https://doi.org/10.1155/2012/950393

Roskams-Hieter, B., Kim, H. J., Anur, P., et al. (2022). Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic malignancies. *NPJ Precis Oncol, 6*, 28. https://doi.org/10.1038/s41698-022-00270-y

Sarker, B., Chakraborty, S., Čep, R., & Kalita, K. (2024). Development of optimized ensemble machine learning-based prediction models for wire electrical discharge machining processes. *Science and Reports, 14*(1), 23299. https://doi.org/10.1038/s41598-024-74291-x

Shichkova, P., Coggan, J. S., Markram, H., & Keller, D. (2021). A Standardized Brain Molecular Atlas: A Resource for Systems Modeling and Simulation. *Frontiers in Molecular Neuroscience, 10*(14), 604559. https://doi.org/10.3389/fnmol.2021.604559

Sohel Md Mahmodul (2016) Extracellular/circulating microRNAs: release mechanisms, functions and challenges. Achievements in the Life Sciences 10. https://doi.org/10.1016/j.als.2016.11.007

Stocchetti, N., Colombo, A., Ortolano, F., et al. (2007). Time course of intracranial hypertension after traumatic brain injury. *Journal of Neurotrauma, 24*, 1339–1346. https://doi.org/10.1089/neu.2007.0300

Takase, T., Oyama, S., & Kurihara, M. (2018). Effective neural network training with adaptive learning rate based on training loss. *Neural Networks, 101*, 68–78. https://doi.org/10.1016/j.neunet.2018.01.016

Taroni, J. N., Grayson, P. C., Hu, Q., Eddy, S., Kretzler, M., Merkel, P. A., & Greene, C. S. (2019). MultiPLIER: A transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Systems, 8*(5), 380-394.e4. https://doi.org/10.1016/j.cels.2019.04.003

Taylor, J. M., Ankerst, D. P., & Andridge, R. R. (2008). Validation of biomarker-based risk prediction models. *Clinical Cancer Research, 14*(19), 5977–5983. https://doi.org/10.1158/1078-0432.CCR-07-4534

Therriault, J., Ashton, N. J., Pola, I., Triana-Baltzer, G., Brum, W. S., Di Molfetta, G., et al. (2024). Comparison of two plasma p-tau217 assays to detect and monitor Alzheimer's pathology. *eBioMedicine, 102*, 105046. https://doi.org/10.1016/j.ebiom.2024.105046

Thomou, T., Mori, M. A., Dreyfuss, J. M., et al. (2017). Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature, 542*, 450–455. https://doi.org/10.1038/nature21365

Tripp, B. A., Dillon, S. T., Yuan, M., Asara, J. M., Vasunilashorn, S. M., Fong, T. G., Inouye, S. K., Ngo, L. H., Marcantonio, E. R., Xie, Z., Libermann, T. A., & Otu, H. H. (2024). Integrated Multi-Omics Analysis of Cerebrospinal Fluid in Postoperative Delirium. *Biomolecules, 14*(8), 924. https://doi.org/10.3390/biom14080924

Vaz, J. M., & Balaji, S. (2021). Convolutional neural networks (CNNs): Concepts and applications in pharmacogenomics. *Molecular Diversity, 25*(3), 1569–1584. https://doi.org/10.1007/s11030-021-10225-3

Wang, C. C., Zhu, C. C., & Chen, X. (2022). Ensemble of kernel ridge regression-based small molecule-miRNA association prediction in human disease. *Briefings in Bioinformatics, 23*(1), bbab431. https://doi.org/10.1093/bib/bbab431

Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., & Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *NPJ Digital Medicine, 7*(1), 41. https://doi.org/10.1038/s41746-024-01029-4

Wojcik, P., Kini, A., Al Othman, B., Galdamez, L. A., & Lee, A. G. (2020). Spaceflight associated neuro-ocular syndrome. *Current Opinion in Neurology, 33*, 62–67. https://doi.org/10.1097/WCO.0000000000000778

Wostyn, P., Gibson, C. R., & Mader, T. H. (2022). The odyssey of the ocular and cerebrospinal fluids during a mission to Mars: The "ocular glymphatic system" under pressure. *Eye (London, England), 36*, 686–691. https://doi.org/10.1038/s41433-021-01721-9

Xia, J., Sun, L., Xu, S., Xiang, Q., Zhao, J., Xiong, W., Xu, Y., & Chu, S. (2020). A model using support vector machines recursive feature elimination (SVM-RFE) algorithm to classify whether COPD patients have been continuously managed according to GOLD guidelines. *International Journal of Chronic Obstructive Pulmonary Disease, 4*(15), 2779–2786. https://doi.org/10.2147/COPD.S271237

Yang, S., Ji, B., Undar, A., & Zahn, J. D. (2006). Microfluidic devices for continuous blood plasma separation and analysis during pediatric cardiopulmonary bypass procedures. *ASAIO Journal, 52*(6), 698–704. https://doi.org/10.1097/01.mat.0000249015.76446.40

Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). A review on over-sampling techniques in classification of multi-class imbalanced datasets: Insights for medical problems. *Frontiers in Digital Health, 26*(6), 1430245. https://doi.org/10.3389/fdgth.2024.1430245

Zanello, S. B., Tadigotla, V., Hurley, J., Skog, J., Stevens, B., Calvillo, E., Bershad, E. (2018). Inflammatory gene expression signatures in idiopathic intracranial hypertension: Possible implications in microgravity-induced ICP elevation. *NPJ Microgravity, 4*, 1. https://doi.org/10.1038/s41526-017-0036-6

Zaporozhchenko, I. A., Ponomaryova, A. A., Rykova, E. Y., et al. (2018). The potential of circulating cell-free RNA as a cancer biomarker: Challenges and opportunities. *Expert Review of*

Molecular Diagnostics, 18, 133–145. https://doi.org/10.1080/14737159.2018.1425143

Zhang, C., Wang, J., Yen, G. G., Zhao, C., Sun, Q., Tang, Y., Qian, F., & Kurths, J. (2020). When Autonomous Systems Meet Accuracy and Transferability through AI: A Survey. *Patterns (n y), 1*(4), 100050. https://doi.org/10.1016/j.patter.2020.100050

Zhou, M., Hara, H., Dai, Y., et al. (2016). Circulating organ-specific microRNAs serve as biomarkers in organ-specific diseases: Implications for organ allo- and xeno-transplantation. *International Journal of Molecular Sciences, 17*, 1232. https://doi.org/10.3390/ijms17081232

Zhou, Z., Guo, Y., Tang, R., Liang, H., He, J., & Xu, F. (2024). Privacy enhancing and generalizable deep learning with synthetic data for mediastinal neoplasm diagnosis. *NPJ Digital Medicine, 7*(1), 293. https://doi.org/10.1038/s41746-024-01290-7

Authors and Affiliations

Phani Paladugu[1,2] · Rahul Kumar[3] · Jahnavi Yelamanchi[4] · Ethan Waisberg[5] · Joshua Ong[6] · Mouayad Masalkhi[7,8] · Chirag Gowda[9] · Ryung Lee[10] · Dylan Amiri[11,12] · Ram Jagadeesan[13,14] · Nasif Zaman[15] · Alireza Tavakkoli[15] · Andrew G. Lee[16,17,18,19,20,21,22,23]

✉ Chirag Gowda
gowdachirag24@gmail.com

1 Sidney Kimmel Medical College, Philadelphia, PA, USA

2 Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

3 University of Massachusetts T.H. Chan School of Medicine, Worcester, MA, USA

4 Tandon School of Engineering, New York University, New York, NY, USA

5 Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

6 Department of Ophthalmology and Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, USA

7 University College Dublin School of Medicine, Belfield, Dublin, Ireland

8 Department of Electronic & Computer Engineering, University of Limerick, Limerick, Ireland

9 University of Miami Miller School of Medicine, 160 NW 10 Ave, Miami, FL 33136, USA

10 Touro College of Osteopathic Medicine, New York, NY, USA

11 Department of Biology, University of Miami, Coral Gables, Florida, USA

12 Mecklenburg Neurology Group, Charlotte, NC, USA

13 Department of Computer Science, Johns Hopkins University Whiting School of Engineering, Baltimore, MD, USA

14 Artificial Intelligence Systems, Cisco Systems, Inc, San Jose, CA, USA

15 Human-Machine Perception Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA

16 Center for Space Medicine, Baylor College of Medicine, Houston, TX, USA

17 Department of Ophthalmology, Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA

18 The Houston Methodist Research Institute, Houston Methodist Hospital, Houston, TX, USA

19 Departments of Ophthalmology, Neurology, and Neurosurgery, Weill Cornell Medicine, New York, NY, USA

20 Department of Ophthalmology, University of Texas Medical Branch, Galveston, TX, USA

21 University of Texas MD Anderson Cancer Center, Houston, TX, USA

22 Texas A&M College of Medicine, Texas, USA

23 Department of Ophthalmology, The University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA