

## Mitigating Bias in Healthcare AI With the Gerchberg-Saxton Algorithm



---

Machine learning has significantly advanced various industries, particularly in healthcare, where deep learning is revolutionising drug discovery, early diagnosis, medical imaging, and personalised treatments. However, these advancements come with challenges, notably issues of interpretability, data availability, and inherent biases. Bias in machine learning refers to systematic inaccuracies disadvantaging certain groups, while fairness is the absence of such prejudice. In healthcare, biases often stem from non-representative training data, leading to models that do not generalise well across different populations.

Bias can arise in several ways: from biased datasets influencing algorithm outcomes (data-to-algorithm bias), from the algorithm itself (algorithm-to-user bias), or from user-generated data reflecting inherent biases (user-to-data bias). Addressing these biases is critical, especially in healthcare, where biased models can lead to misdiagnoses and incorrect treatments with severe consequences.

[A recent study published in the Journal of Healthcare Informatics Research](#) focuses on mitigating data-to-algorithm bias using a method based on the Gerchberg-Saxton (GS) algorithm, which operates in the frequency domain. By distributing information more uniformly among features used in model training, this method aims to reduce racial bias in deep learning classification. The approach ensures more consistent model accuracy across different racial groups, thereby improving equity in healthcare by enhancing model performance for underrepresented populations.

### Exploring Innovative Methods to Advance Fairness in AI

The rapid expansion of Artificial Intelligence (AI) necessitates addressing fairness and bias to ensure equitable decision-making and minimise potential harm. Various methods have been proposed, each with unique strengths and challenges.

Data augmentation generates synthetic samples to improve the representation of under-represented groups in training datasets. Although successful to some extent, it can struggle to represent specific populations accurately and may introduce new biases if not carefully implemented .

Adversarial learning involves training predictor and adversary models simultaneously to enhance prediction accuracy while reducing bias. This method has shown promise, especially with novel fairness prediction schemes, but requires significant computational resources and further exploration for complex datasets .

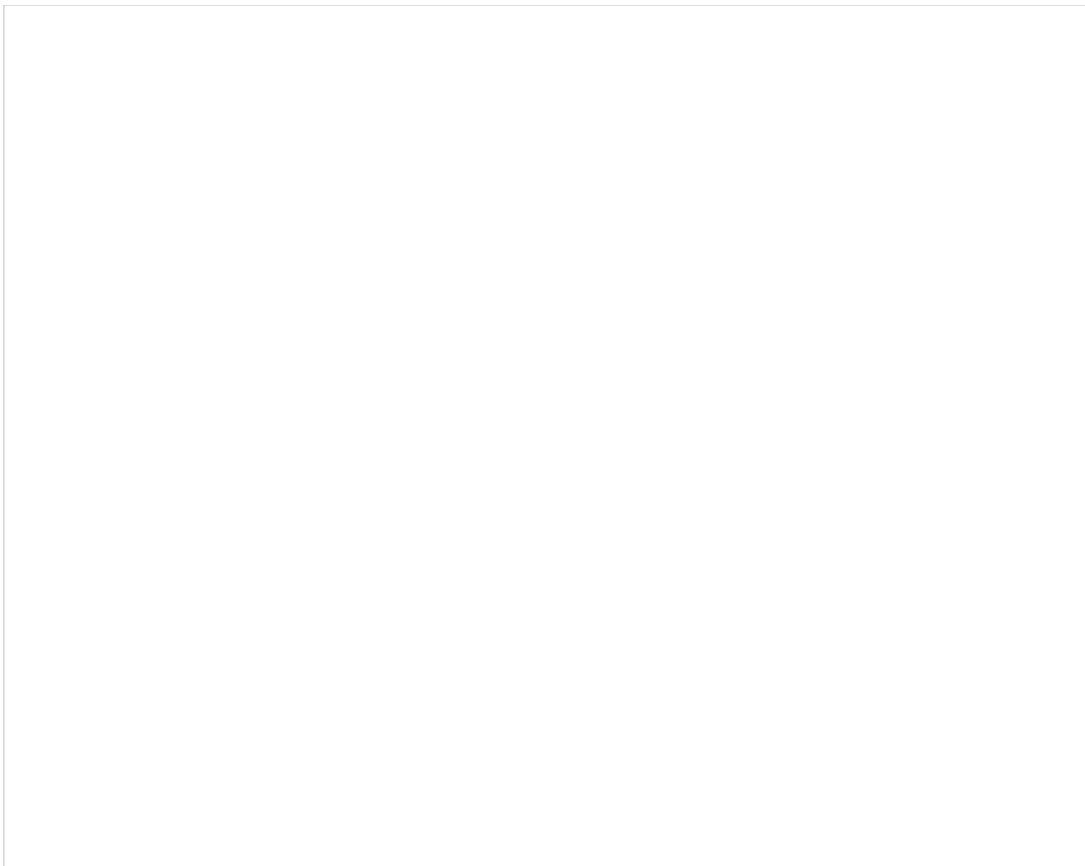


Image Source: [Journal of Healthcare Informatics Research](#)

Fairness constraints within learning processes are another area of research, but there is ongoing debate about the most appropriate constraints, indicating this as a future area for innovation. Deep reinforcement learning has also been explored for fairness, but results are mixed and often not generalisable.

The paper introduces a model-agnostic method using the Gerchberg-Saxton (GS) algorithm, traditionally used in computer vision and signal processing. This algorithm optimises information distribution, offering a novel approach to tackling bias in AI. This method draws from the established success of the GS algorithm in other domains, presenting a new pathway for AI fairness.

Overall, while significant progress has been made in bias mitigation, the search for universally effective methods continues. This work contributes to the ongoing dialogue by leveraging the GS algorithm in a novel application to improve AI fairness.

### **Mitigating Bias to Safeguard Predictive Model Accuracy**

This study used the Medical Information Mart for Intensive Care (MIMIC-III) version 1.4 database, a widely utilised resource in clinical research for developing machine learning algorithms for applications such as mortality prediction, sepsis detection, and disease phenotyping. The database, maintained by the MIT Laboratory for Computational Physiology, includes de-identified health data from over 40,000 ICU patients at the Beth Israel Deaconess Medical Center between 2001 and 2012.

The study aimed to predict mortality rates within 24 hours of ICU admission and examined bias across different racial and ethnic groups. Using thirty vital sign features, a deep learning classification model was trained on a dataset of 13,980 patients, which was categorised into five groups based on self-reported ethnicity: European American (9814 patients), African American (1690), Eastern Asian American (346), Hispanic American (641), and Others (1489).

The MIMIC-III dataset is known to contain biases that can affect predictive model accuracy. This study aimed to mitigate such biases using the Gerchberg-Saxton (GS) algorithm. The GS algorithm was applied to improve demographic and error rate parity across different groups. The models' performance was evaluated based on accuracy, demographic parity, and error rate parity using true positive, true negative, false positive, and false negative rates. These evaluations were conducted on both the original and GS-applied datasets across three different population cohorts, each tested five times to ensure consistency and reproducibility. The results demonstrated that applying the GS algorithm can help mitigate racial and ethnic biases, leading to more equitable and unbiased mortality predictions.

## Leveraging the Gerchberg-Saxton Algorithm for Equitable Information Distribution

In this study, authors innovatively applied existing tools to mitigate bias across diverse racial and ethnic populations in a deep learning (DL) model using the MIMIC-III database. Known for its inherent selection biases, the MIMIC-III dataset was subjected to the Gerchberg-Saxton (GS) algorithm to improve information distribution.

The GS algorithm, a model-agnostic technique, focuses on transforming data rather than altering the model itself. This approach is crucial because it addresses the dataset's sampling bias, independent of the model type. The GS algorithm's main function is to equalise information distribution across different racial and ethnic groups within the dataset.

However, there are limitations to this method. The GS algorithm is computationally intensive, making it challenging to process large datasets in one go. Data must be segmented and processed in batches, which requires careful attention to ensure equitable representation of all groups in each batch. Additionally, the GS-transformed data might differ structurally from the original, potentially complicating biological interpretations. Despite this, the transformed data can enhance model performance without altering the dataset's content.

## Effective Counters Bias by Equalising Feature Magnitudes

The approach aimed to disperse information from features predominantly informing predictions for European-American groups to all other groups, ensuring equitable information dissemination before partitioning data into training, testing, and validation subsets. Post-GS analysis showed that the algorithm reduced bias and improved prediction accuracy and AUC scores across all racial and ethnic groups. The study highlights two key findings: the GS algorithm effectively counters bias by equalising feature magnitudes, and it improves prediction accuracy and fairness across demographics. This was validated using SHAP calculations and Shannon entropy, which confirmed more uniform feature contributions, leading to more equitable model training.

While further research is needed to fully explore the GS algorithm's potential across different medical data settings and modalities, this study demonstrates a significant step towards bias mitigation in medical DL applications. The GS algorithm's application resulted in more uniform and equitable predictions for various racial and ethnic groups, suggesting its potential to advance efforts in reducing biases in AI models used in healthcare.

Source: [Journal of Healthcare Informatics Research](#)

Image Credit: [iStock](#)

Published on : Tue, 21 May 2024