SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Bias Mitigation In Large-Scale AI Systems For Healthcare Decision Support: Techniques, Data Processing, And Evaluation

Sayantani De[1], Shubhashish Goswami[2], Ravi Kumar[3], Manoj I. Patel[4], Tushar Jadhav[5], Rajashri C.K.[6], Sreedevi K[7], Parul Yadav[8]

[1] Assistant Professor, Department of Computer Science & IT, Arka Jain University, Jamshedpur, Jharkhand, India. Email: sayantani.de@arkajainuniversity.ac.in ORCID: 0000-0001-5047-0762
[2] Associate Professor, Department of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India. Email: socse.shubhashish@dbuu.ac.in ORCID: 0000-0002-6129-9822
[3] Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India. Email: ravi.kumar.orp@chitkara.edu.in ORCID: 0009-0001-8257-9794
[4] Assistant Professor, Faculty of Computer Science and Application, Gokul Global University, Sidhpur, Gujarat, India. Email: manoj.patel@gokuluniversity.ac.in ORCID: 0009-0009-0731-9528
[5] Professor, Department of Electronics & Telecommunication (E&TC) Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India. Email: tushar.jadhav@vit.edu
[6] Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: rajashrick@maher.ac.in
[7] Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: sreedevicom@maher.ac.in
[8] Assistant Professor, Department of School of Engineering & Technology, Noida International University, Greater Noida, Uttar Pradesh, India. Email: parul.yadav@niu.edu.in

Abstract
The increasing adoption of large-scale Artificial Intelligence (AI) systems in healthcare has raised critical concerns regarding algorithmic bias and fairness in clinical decision-making. Biased AI models can lead to unequal healthcare outcomes, particularly across diverse demographic groups. This research investigates effective bias mitigation techniques within large-scale AI systems, specifically for healthcare decision support. Healthcare data is collected from Electronic Health Records (EHRs) that encompass Patient Demographics, Clinical History, and Diagnostic Reports. Min-Max Normalization is utilized to standardize feature scales and mitigate bias, thereby enhancing model stability. Additionally, Principal Component Analysis (PCA) facilitates feature extraction by reducing dimensionality while maintaining data variance, thus improving computational efficiency and interpretability. The proposed model integrates Nesterov Accelerated Gradient with a Bi-directional Gated Recurrent Unit (NAG-Bi-GRU) to enhance learning efficiency and capture complex temporal dependencies in healthcare data. The NAG optimizer improves convergence speed and reduces training bias, while Bi-GRU effectively models sequential patterns from both forward and backward directions. Bias mitigation is further reinforced through fairness-aware training strategies and output calibration. Model performance metrics indicate high accuracy (94.8%), precision (93.9%), recall (93.5%), F1-Score (93.7%), and AUC-ROC (0.97), while fairness metrics show demographic parity and equal opportunity. The model demonstrates reduced bias and maintains high predictive performance, while promoting fair and scalable AI systems for healthcare decision support.

Keywords: Artificial Intelligence (AI), Algorithmic bias, Healthcare, Evaluation, Predictive performance

1. Introduction

Healthcare support involves clinical care, diagnosis, prevention, and management focused on promoting the well-being of the population. This process leverages professional expertise, technological advances, and policymaking to ensure appropriate access to healthcare globally [1].Decision support from health care information systems is an important component for dealing with diverse data, decision-making, and achieving clinical outcomes. Decision support helps ensure that the decisions made are appropriate, evidence-based, and lead to the best patient care [2]. The presence of large population data from EHRs, coupled with increased computing power, enables data analysis through Artificial Intelligence (AI) techniques. Several kinds of information, such as patient-specific demographic data and medical condition data, are used to determine the likelihood of developing a disease or predicting the outcome of a disease[3, 4]. Pandemic healthcare is not only for predicting disease but also supports healthcare decision-making. Models using AI help in prediction and resource allocation, but success depends on the accuracy of these predictions within the healthcare sector [5]. AI is revolutionizing the field of medicine and the entire healthcare industry to ensure that the future of healthcare becomes more secure and efficient. The incorporation of AI in healthcare is important in tackling problems posed by an aging population and an inadequate number of health professionals. The application of AI helps in making decisions regarding an allocation of therapy options for particular patients [6]. Healthcare support spending is used as a predictor for the need in healthcare decision tools, with the inherent bias in relation to the racial differences in healthcare utilization. Healthcare is an integrated set of services, practitioners, and facilities that delivers medical care, prevents diseases, and promotes overall health [7, 8].
Research Aim: To enhance the bias-mitigated large-scale AI model for healthcare decision support systems that ensures fairness and high predictive performance. It focuses on reducing algorithmic bias in Electronic Health Record (HER)-based models using advanced preprocessing, feature extraction, and the proposed Nesterov Accelerated Gradient with a Bi-directional Gated Recurrent Unit (NAG-Bi-GRU) model, while maintaining equitable healthcare decision support with performance evaluation.

Research Organization: Section 1 provides the background of the research. Section 2 focuses on reviewing existing research. Section 3 discusses the methodology of data collection, preprocessing, and future extraction, as well as the proposed model, NAG-Bi-GRU. Section 4 includes results obtained from experimental analysis. Section 5 presents the conclusion with the performance, drawbacks, and upcoming work.

2. Related Work

A utilization AI to improve personalized medicine through the examination of medical information was developed. Applying Machine Learning (ML) to medicine, Support Vector Machines (SVM) [9] are utilized for analysis in genomics, proteomics, and clinical data. This approach increases the effectiveness of diagnostics and therapy through increased accuracy. Nevertheless, problems associated with data bias, interpretation, compliance with regulations, and implementation persists in equal measure. Bias detection and reduction in healthcare AI applications to promote health equity were examined in [10]. The approach utilizes two approaches, Bias Detection (BD) and Synthetic Minority Over-sampling Technique (SMOTE). The application reduces the measure of disparity by 20–30%. Despite the gains achieved, the problem of unknown bias persists owing to data heterogeneity. Explaining AI with User-Centric Design for Neuro-Oncology Imaging [11], Explainable Artificial Intelligence (XAI) approaches that were co-designed together with clinicians through Magnetic Resonance Imaging (MRI) features and visualizations achieved gains of 18% for improved accuracy and 22% for higher efficiency. However, constraints like insufficient data size and enhanced cognitive load have posed significant challenges to their adoption in clinical practice. AI's application in medicine and dealing with an issue for bias in AI predictions was investigated in [12]. Using Class Activation Mapping (CAM), it became clear that the bias in the model reduced by 28%. Nonetheless, there were several limitations to this research, like heterogeneity, absence of exterior validity, as well as risk in publication prejudice. A comprehensive evaluation methodology that involves various strategies for overcoming bias was discussed in [13]. The application of the Random Forest (RF) algorithm for decision support in the healthcare sector helped to achieve better healthcare diagnosis fairness by 95% without changing the effectiveness of the algorithm. However, the algorithm cannot deal with multiclass classification problems. Enhanced Total Product Lifecycle (TPLC) method was employed in analyzing the equality of AI/ML-based healthcare services to overcome inequalities [14]. It entails conducting audits, detecting bias, involving stakeholders, and mitigating efforts. It is revealed that the model can help to overcome inequalities by up to 20-40%. However, certain limitations, including low-quality data and limited resources, prevent it from being extensively adopted.

The Four Step Analytical (FSA) Method was designed to leverage the advantages of AI to improve the standard of patient care [15]. A FSA method is utilized in AI/ML development solutions, paying special attention to detecting sources of bias and implementing Techniques for reducing bias. Mitigation with the impacts of Big Data and AI bias applications of catastrophes in healthcare was performed through using Convolutional Neural Network (CNN) [16]. Utilizing AI techniques to synthesize data that was biased in the dataset led to mitigation of the impacts, with the accuracy rate being 80%. Data limitation is not the only factor contributing to bias in AI applications.Medical data analysis using edge computing was performed with the Log Neural Network (LogNNet) chaotic mapping method to convert information [17]. The obtained results show high efficiency of medical data analysis through its high classification accuracy. Nevertheless, one of the difficulties is associated with AI on devices include limited resources.A clinical decision support system design was performed by investigating the Deep Convolutional Neural Network (DCNN) method for classification [18]. The model uses an automated process that extracts characteristics from segmented images and then classifies them. The findings from the research indicate 88.8% accuracy when using classification techniques.Transparency and bias in AI assistance for health care decisions were analyzed using MAX Qualitative Data Analysis (MAXQDA) [19]. According to the research, one benefit of integrating AI in healthcare decision-making is that it offers many advantages, especially in increasing the accuracy of diagnostics. The creation of early detection and prevention in the healthcare industry with AI technology began by predicting diseases using long short term memory (LSTM) with EHRs [20]. This research yields an accuracy of 89.7%, indicating the possibility of enhancing patients' health by conducting early detection and preventing diseases.

3. Methodology

Figure 1 illustrates the architecture of large-scale AI systems for healthcare decision support. The proposed model integrates HER based data collection, Principal Component Analysis (PCA) for dimensionality reduction and feature extraction, and Min-Max normalization for data preprocessing to minimize bias and improve model stability. In addition, a hybrid Nesterov Accelerated Gradient-based Bidirectional Gated Recurrent Unit (NAG-Bi-GRU) model is employed to enhance healthcare decision support performance.

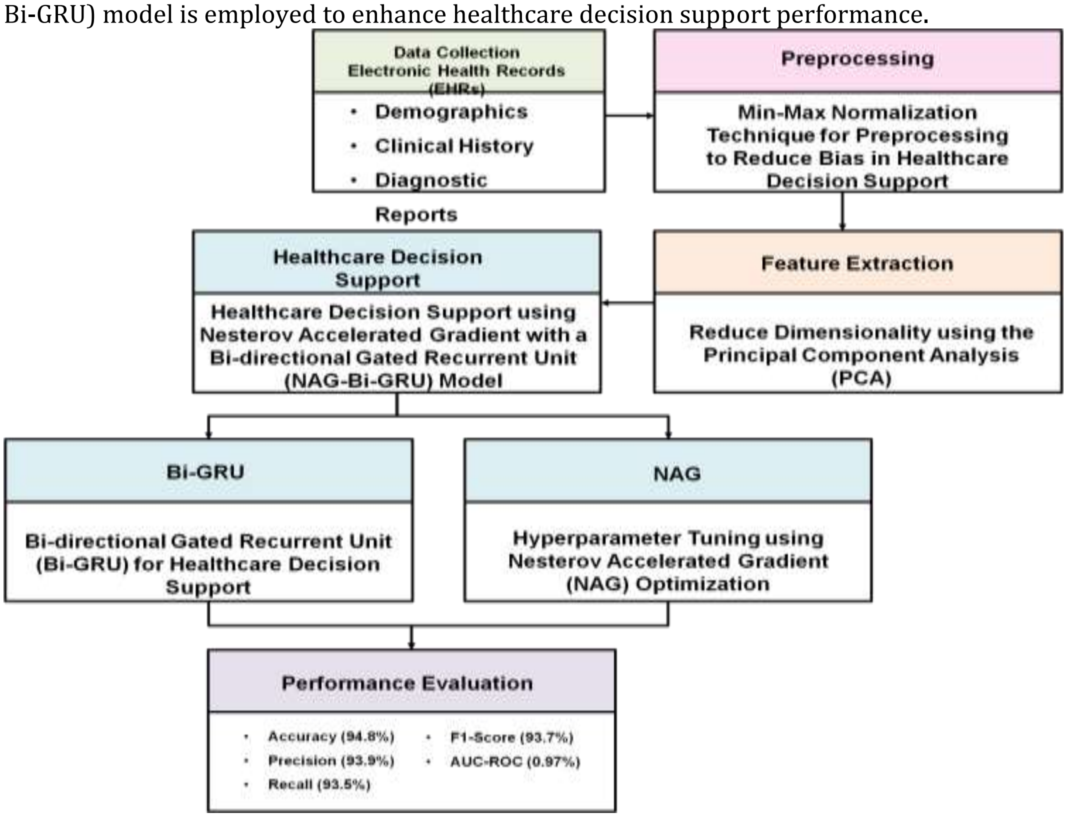

Figure 1: Method Workflow for Healthcare Decision Support in Large-Scale AI Systems

3.1 Data Collection

For the proposed healthcare decision-making support system, the data collection process makes use of the EHR dataset, It is accessible for free on the Kaggle website (https://www.kaggle.com/datasets/anuchhetry/electronic-health-record). This dataset includes only one CSV file named EHR that contains 29 structured attributes relating to clinical/administrative aspects and 1,447 records. Each record contains unique patient IDs, patient demographics, hospital admission details, and length of stay, disease conditions, discharge status, and the location to which the patient was discharged. Some of the important attributes include Unique Patient ID, Age, Gender, Diagnosis, and Discharge. This dataset is used in this research to implement healthcare decision models using 80% of the dataset for training and 20% for testing purposes.

3.2 Min-Max Normalization Technique for Preprocessing to Reduce Bias in Healthcare Decision Support

Min-Max Normalization is a standard normalization technique utilized in the decision support system for healthcare to deal with scale differences in features and to minimize any bias across different EHRs. Since clinical datasets consist of diverse attributes such as patient demographics and diagnostic reports, direct model training can lead to dominance of High-Range Features, introducing systematic bias in decision-making. All the features is normalized using the Min-Max normalization method within the defined range of [0, 1]. This procedure expressed in Equation (1).

$$W_{new} = (W - W_{min})/(W_{max} - W_{min}) \quad (1)$$

W represents the original value of a feature (e.g., Patient Age or Diagnosis Report), andW_{new} represents the normalized value, which reduces biases that arise from attributes with wider numerical distributions, making healthcare prediction and decision-making fairer. W_{max}andW_{min} represent the highest and lowest values of a healthcare dataset in a clinical parameter.

3.3 Reduce Dimensionality using the Principal Component Analysis (PCA)

After applying Min-Max normalization, PCA is used in the decision support system for healthcare purposes to improve computational efficiency and eliminate redundancies in large EHRs. PCA is used when dealing with complex healthcare data with highly correlated data elements like Diagnostic Reports, demographics, and diagnostic criteria. PCA converts the initial high-dimensional feature space into a lower dimension with uncorrelated principal components. It is based on the assumption that there exists a small number of principal components derived from the initial variables, where each principal component is mutually orthogonal and uncorrelated, as expressed in Equation (2).

$$Y_{ji} = \frac{D_{ji} - D_i}{\sigma_i} \quad (2)$$

Y_{ji} represents the standardized value of the transform and reduces dimensionality in healthcare data while keeping the highest variance, andD_{ji} represents the original value of clinical data, such as age and diagnosis report, which is part of the EHR, and D_i shows the average value of the particular Clinical Measure (Patient Age or Diagnosis Report). σ_i shows the standard deviation of the span of ages or test scores in the healthcare database.

3.4 Healthcare Decision Support using Nesterov Accelerated Gradient with a Bi-directional Gated Recurrent Unit (NAG-Bi-GRU) Model

The healthcare decision-making model utilizes an AI algorithm that minimizes bias and involves a large-scale AI solution, where the Nesterov Accelerated Gradient (NAG) method connects to a Bi-directional Gated Recurrent Unit (Bi-GRU) to provide unbiased the accurate clinical prediction services. The NAG method helps accelerate learning through convergence and minimizing learning instabilities, while the Bi-GRU algorithm accounts for temporal dependencies of data from EHRs.

Healthcare Decision Support using Bi-Directional Gated Recurrent Unit (Bi-GRU): In the proposed healthcare decision support system, the Bi-GRU to be employed to learn complex temporal relationships in the big data of EHRs. The Bi-GRU extends the standard GRU architecture by incorporating two parallel processing layers that operate in forward and backward directions, enabling the model to learn contextual requirements from both previous clinical history of temporal patterns. The disease progression of forward and backward healthcare states is computed as expressed in Equations (3 and 4).

$$\overrightarrow{g_s^1} = e(x_{\overrightarrow{wg^1}} w_s + x_{\overrightarrow{g^1 g^1}} \overrightarrow{g_{s-1}^1} + a_{\overrightarrow{g^1}}) \quad (3)$$

$\overrightarrow{g_s^1}$ represents the hidden state at time step t for the forward direction in a first GRU layer, representing the past healthcare data (e.g., clinical history). $x_{\overrightarrow{wg^1}}$represents the weight matrix for the input w_s in the forward direction, indicating the input (such as a medical report). $\overrightarrow{g_{s-1}^1}$ represents the hidden state at a previous time step $s-1$ in forward pass, encoding a healthcare history in an sequence. $a_{\overrightarrow{g^1}}$ represents the GRU layer bias term for the forward direction, contributing to adjusting the model's predictions in the field of healthcare. e represents the activation function in healthcare data. $x_{\overrightarrow{g^1}\overrightarrow{g^1}}$ represent the weight matrix linking the forward previous hidden state with the current hidden state, capturing temporal dependencies in healthcare datasets.

$$\overleftarrow{g_s^1} = e(x_{\overleftarrow{wg^1}} w_s + x_{\overleftarrow{g^1}\overleftarrow{g^1}} \overleftarrow{g_{s-1}^1} + a_{\overleftarrow{g^1}}) \tag{4}$$

$\overleftarrow{g_s^1}$ represents the hidden healthcare state at time step s for a backward direction in a first GRU layer, representing the past healthcare data (e.g., clinical history). $x_{\overleftarrow{wg^1}}$represents the weight matrix for the input w_s in the backward direction, indicating the input (such as a medical report). $\overleftarrow{g_{s-1}^1}$ represents the hidden healthcare state at a previous time step $s-1$ in a backward pass, encoding the healthcare history in an sequence. $a_{\overleftarrow{g^1}}$ represents the GRU layer bias term for the backward direction, contributing to adjusting the forecasts of the model in the area of healthcare. $x_{\overleftarrow{g^1 and g^1}}$ represent the weight matrix linking the backward previous hidden state with the current hidden state, capturing temporal dependencies in healthcare datasets. The treatment response of forward and backward states is computed as expressed in Equations (5 and 6).

$$\overrightarrow{g_s^2} = e\left(x_{\overrightarrow{g^1}\overrightarrow{g^2}}\overrightarrow{g_s^1} + x_{\overrightarrow{g^2}\overrightarrow{g^2}}\overrightarrow{g_{s-1}^2} + a_{\overrightarrow{g^2}}\right) \tag{5}$$

$\overrightarrow{g_s^2}$ represents the forward hidden healthcare state at step s in the second GRU layer, capturing healthcare information related to disease progression, and e represents the activation function in healthcare data. $x_{\overrightarrow{g^1}\overrightarrow{g^2}}$ represents the weight matrix connecting the forward hidden healthcare states of the first and second layers, influencing the model's learning, and $\overleftarrow{g_s^1}$ represents the healthcare data. $x_{\overrightarrow{g^2}\overrightarrow{g^2}}$ denotes the weight matrix within a second GRU layer that influences the interaction with its hidden healthcare states. $\overrightarrow{g_{s-1}^2}$ represents the forward hidden healthcare state at step s in the first layer, representing past healthcare data such as patient history, and $a_{\overrightarrow{g^2}}$ refers a bias term for the second GRU layer's forward direction, adjusting the output for better healthcare prediction.

$$\overleftarrow{g_s^2} = e\left(x_{\overleftarrow{g^1}\overleftarrow{g^2}}\overleftarrow{g_s^1} + x_{\overleftarrow{g^2}\overleftarrow{g^2}}\overleftarrow{g_{s+1}^2} + a_{\overleftarrow{g^2}}\right) \tag{6}$$

In equation (6), $\overleftarrow{g_s^2}$ represents the backward hidden healthcare state at step s in the second GRU layer, capturing information related to disease progression, and $x_{\overleftarrow{g^1}\overleftarrow{g^2}}$ represents the weight matrix connecting the backward hidden healthcare states of the first and second layers, influencing the model's learning. $\overleftarrow{g_{s+1}^2}$ represent the backward hidden healthcare state at step s in the first layer, representing past patient history, and $x_{\overleftarrow{g^2}\overleftarrow{g^2}}$ refers a weight matrix within a second GRU layer that influences the interaction of its hidden healthcare states. $a_{\overleftarrow{g^2}}$ represents the bias term for the second GRU layer's backward direction, $\overleftarrow{g_s^1}$ denotes a healthcare data, and adjusting the output for better healthcare prediction is expressed in Equation (7).

$$z_s = h\left(x_{\overrightarrow{g^2}z}\overrightarrow{g_s^2} + x_{\overleftarrow{g^2}z}\overleftarrow{g_s^2} + a_z\right) \tag{7}$$

z_s represents the final output at step s, the model's health prediction, such as the disease or treatment response, and h represents the activation function to transform the weighted sum of the inputs (like the hidden healthcare states from the second GRU layer), $x_{\overrightarrow{g^2}z}, x_{\overleftarrow{g^2}z}$ refers a weight matrix connecting the forward and backward hidden healthcare states from the second GRU layer to the final output. $\overrightarrow{g_s^2}$ and $\overleftarrow{g_s^2}$are the forward and backward hidden healthcare states at time step s in the second GRU layer. a_z denotes a bias term for the final healthcare output layer, adjusting the output prediction based on learned features from both forward and backward states.

Hyperparameter Tuning using Nesterov Accelerated Gradient (NAG) Optimization: Hyperparameter tuning using NAG optimization is very important in improving the effectiveness and robustness of the healthcare decision-making support system. In large EHR-based systems, the proper choice of learning rate, momentum factor, and weight updates is necessary to minimize the problem of convergence instability and training bias. The NAG optimization enhances the ordinary gradient descent algorithm by introducing an advanced look, whereby the next parameter values are predicted before computing the gradient. This results in faster convergence and a more stable trajectory of optimization, especially when dealing with complex high-dimensional healthcare datasets, as expressed in Equation (8).

$$u_{s+1} = \beta u_s + \eta \nabla_\theta I(\theta_s) \quad (8)$$

u_{s+1} represents the updated healthcare momentum vector at step $s+1$, guiding the gradient descent to faster and more stable healthcare convergence. β represents the healthcare momentum factor, controlling the contribution of previous healthcare momentum to the current healthcare update. u_s represents the healthcare momentum vector at step s, storing the past healthcare update directions for faster convergence. η denotes the learning rate of the objective function associated with healthcare data $\nabla_\theta I(\theta_s)$.

$$\theta_{s+1} = \theta_s - u_{s+1} \quad (9)$$

In equation (9), θ_{s+1} is the set of parameters of the healthcare model at time $s+1$, and it is obtained by updating the weight using the healthcare model's momentum and gradient update. θ_s on the other hand, is the set of parameters of the healthcare model at time s.

$$u_{s+1} = \beta u_s + \eta \nabla_\theta I(\theta_s - \beta u_s) \quad (10)$$

In equation (10), $\eta \nabla_\theta I(\theta_s - \beta u_s)$ denotes the gradient of the objective function of the healthcare data regarding the updated healthcare model parameters, leading to better stability in the healthcare updates process.

$$\theta_{s+1} = \theta_s - u_{s+1} \quad (11)$$

In equation (11), θ_{s+1} is the parameter for the health care model at iteration $s+1$, It represents the newly generated weights based on the health care model momentum and gradient updates. θ_s refers to the health care model parameter at iteration s, and its value is obtained through NAG.

4. Result

This research explores the problem of bias prevention in large-scale machine learning models to aid decision-making in the health care sector. The hardware configurations involve Intel Core i7 or i9 processor, NVIDIA RTX 3060 or 3080 GPUs, and 16–32GB of RAM to ensure efficient computations. The software consists of Python 3.8+, TensorFlow, and PyTorch along with NumPy, Pandas, Matplotlib, and Scikit-Learn.

4.1 Exploratory Analysis

The NAG-Bi-GRU model effectively characterizes the demographic dataset by identifying clusters and outliers across all age ranges, as shown in Figure 2(a). The crucial factors are essential to the model's ability to produce precise predictions, the PCA algorithm to be used for chosen based on the association with factors, such as age and weight as shown in Figure 2(b). The sequence model can handle complex relationships in the dataset, which increases its forecast healthcare support, as Figure 2(c) demonstrates. For the NAG method to converge efficiently, Figure 2(d) shows the sequence model separates noise and actual clinical trend data.

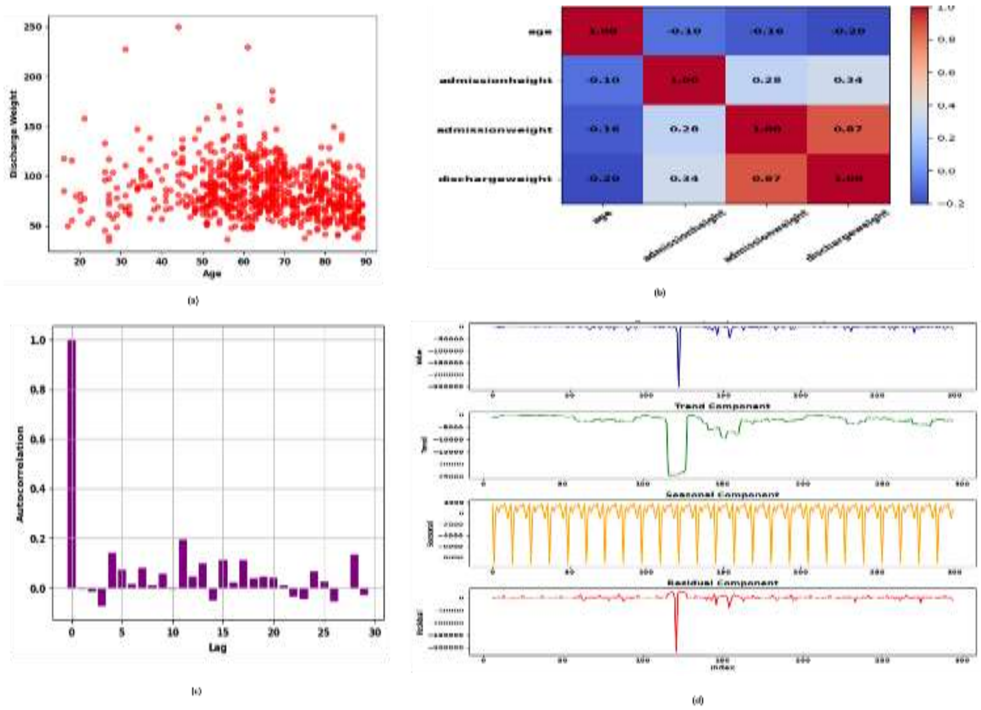

Figure 2: Exploratory Analysis of (a) Distribution Analysis of Demographic Features, (b) Clinical Characteristic Correlation Matrix, (c) Assessment of Temporal Dependency Lag and (d) Patient Flow Time-Series Decomposition

4.2 Evaluation Metrics

The healthcare decision support system is designed to analyze the use of both predictive and fairness metrics. The accuracy metric evaluates each prediction's accuracy made. Precision determines a reliability of predictions for diseases being anticipated positively. Recall determines the ability to correctly identify patients, In contrast, the F1 score represents a average of precision with positive prediction as well as recall. A AUC-ROC measures the performance of classifiers at various thresholds, and fairness is measured by demographic parity and equal opportunity for measuring unbiased true positives.

4.3 Performance Evaluation

The comparsion analysis demonstrates that a proposed NAG-Bi-GRU model outperforms existing machine learning and deep learning methods across all evaluation metrics as shown in Table 1 and Figure 3. The traditional approaches like DT, RF, and SVM [20] performed moderately well owing to their limitations in modeling the temporal dependencies present in industrial data. The deep learning models like CNN and LSTM [20] enhanced the prediction accuracy by effectively performing feature extraction and sequence learning. Nonetheless, the proposed NAG-Bi-GRU architecture outperformed other architectures by attaining an accuracy of 94.8%, precision of 93.9%, recall of 93.5%, F1-score of 93.7%, and AUC-ROC of 0.97.

Table 1: Comparison Analysis of Existing and Proposed Methods

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
DT [20]	78.5	76.2	74.3	75.2	0.82
RF [20]	85.4	83.6	82.7	83.1	0.89
SVM [20]	87.2	85.9	85.1	85.5	0.91
CNN [20]	92.5	91.3	90.6	91.0	0.94
LSTM [20]	89.7	88.4	87.9	88.1	0.92
NAG-Bi-GRU [Proposed]	94.8	93.9	93.5	93.7	0.97

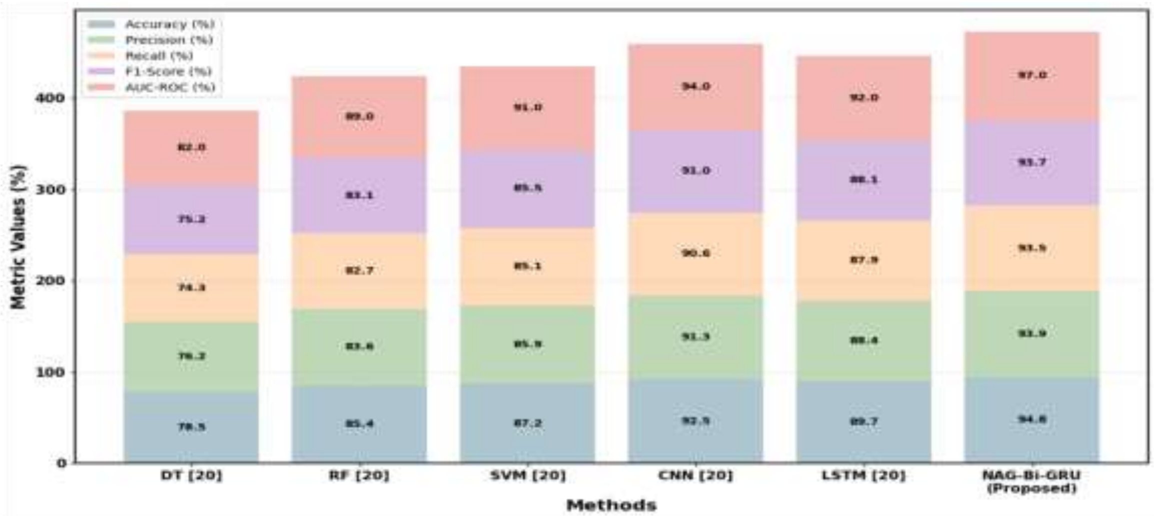

Figure 3: Performance Evaluation of Existing and Proposed Methods for Healthcare Decision Support

4.4 Discussion

The current research aims to devise a decision support system for health care, which ensures bias-free and accurate predictions. However, existing models, such as DT [20], lack generalization skills, thus failing to work efficiently with different patients' information. The issue with RF is that these models do not take into account the temporal relations between the input data. Similar limitations apply to SVMs. Even though the use of CNN can aid in boosting the effectiveness of the system, the problem of bias remains relevant. The same applies to LSTM models. The NAG-Bi-GRU model attempts to solve the limitations by introducing NAG optimization, Bi-GRU, and fair optimization algorithms. With their help, the proposed model increases learning effectiveness, takes into account complicated temporal relationships, and guarantees not only high-quality predictions but also fairness for a variety of patients' backgrounds.

5. Conclusion

The overall AI-based medical decision-making model reduces any potential biases while maintaining accuracy and unbiasedness in the predictions. The study applied various forms of EHR data, such as diagnostic, clinical, and demographic. PCA was used for dimensionality reduction, while Min-Max Normalization was used for feature scaling to limit bias. The NAG-Bi-GRU model was made up of Bi-GRU and NAG algorithms. Fairness-aware training and output scaling methods also play a role in reducing biases in the algorithms and achieving demographic parity and equal opportunities. Despite its great prediction ability with an accuracy of 94.8%, precision of 93.9%, recall of 93.5%, F1-score of 93.7%, and an AUC-ROC of 0.97, there are certain drawbacks such as computational complexity and reliance on quality EHRs. Future work should focus on improving interpretability, reducing computational cost, and developing instantaneous healthcare decision support systems.

REFERENCES

1. Da Silva Jr, J.B., Birminghamm, M., Cinnamond, A.R., Boisson, E., Valdez, M.L., Saiso, S.G., Souza, J.P., Marti, M.C., Richards, L.M., Guzman, J. and Nelson, J., 2026. Bias-Mitigated AI as a foundation for resilient and effective health systems. JMIR Public Health and Surveillance, 12(1), p.e88457. https://doi.org/10.2196/88457
2. Tang, W., Chen, R., Long, X., Yu, D., Zhao, S. and Chen, B., 2026. Medical large language models and systems in the clinical application of spinal diseases: Current status, challenges, and future prospects. Journal of Orthopaedic Translation, 57, p.101050. https://doi.org/10.1016/j.jot.2026.101050
3. Nazer, L.H., Zatarah, R., Waldrip, S., Ke, J.X.C., Moukheiber, M., Khanna, A.K., Hicklen, R.S., Moukheiber, L., Moukheiber, D., Ma, H. and Mathur, P., 2023. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS digital health, 2(6), p.e0000278. https://doi.org/10.1371/journal.pdig.0000278

4. Gertych, A. and Faust, O., 2024. AI explainability and bias propagation in medical decision support. Computer Methods and Programs in Biomedicine, 257, p.108465. https://doi.org/10.1016/j.cmpb.2024.108465
5. Gu, T., Pan, W., Yu, J., Ji, G., Meng, X., Wang, Y. and Li, M., 2025. Mitigating bias in AI mortality predictions for minority populations: a transfer learning approach. BMC Medical Informatics and Decision Making, 25(1), p.30. https://doi.org/10.1186/s12911-025-02862-7
6. Isaksson, A., 2025. Mitigation measures for addressing gender bias in artificial intelligence within healthcare settings: a critical area of sociological inquiry. AI & SOCIETY, 40(4), pp.3009-3018. https://doi.org/10.1007/s00146-024-02067-y
7. Ferrara, E., 2024. The butterfly effect in artificial intelligence systems: Implications for AI bias and fairness. Machine Learning with Applications, 15, p.100525. https://doi.org/10.1016/j.mlwa.2024.100525
8. Bajramagic, M., Battelino, T., Cos, X., Cote, M., Cui, N., Forbes, A., Galderisi, A., Heinemann, L., Hussain, S., Imbert, J. and Holm Jönsson, C., 2026. Artificial intelligence-driven clinical decision support systems to assist healthcare professionals and people with diabetes in Europe at the point of care: a Delphi-based consensus roadmap. Diabetologia, 69(2), pp.259-273. https://doi.org/10.1007/s00125-025-06601-5
9. Khude, H. and Shende, P., 2025. AI-driven clinical decision support systems: Revolutionizing medication selection and personalized drug therapy. Advances in Integrative Medicine, 12(4), p.100529. https://doi.org/10.1016/j.aimed.2025.100529
10. Hasanzadeh, F., Josephson, C.B., Waters, G., Adedinsewo, D., Azizi, Z. and White, J.A., 2025. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digital Medicine, 8(1), p.154. https://doi.org/10.1038/s41746-025-01503-7
11. Prince, E.W., Mirsky, D.M., Hankinson, T.C. and Görg, C., 2025. Current state and promise of user-centered design to harness explainable AI in clinical decision-support systems for patients with CNS tumors. Frontiers in Radiology, 4, p.1433457. https://doi.org/10.3389/fradi.2024.1433457
12. Koçak, B., Ponsiglione, A., Stanzione, A., Bluethgen, C., Santinha, J., Ugga, L., Huisman, M., Klontzas, M.E., Cannella, R. and Cuocolo, R., 2025. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagnostic and interventional radiology, 31(2), p.75. https://doi.org/10.4274/dir.2024.242854
13. Dehghani, F., Paiva, P., Malik, N., Lin, J., Bayat, S. and Bento, M., 2025. Accuracy-fairness trade-off in ML for healthcare: A quantitative evaluation of bias mitigation strategies. Information and Software Technology, p.107896. https://doi.org/10.1016/j.infsof.2025.107896
14. Abràmoff, M.D., Tarver, M.E., Loyo-Berrios, N., Trujillo, S., Char, D., Obermeyer, Z., Eydelman, M.B., Foundational Principles of Ophthalmic Imaging and Algorithmic Interpretation Working Group of the Collaborative Community for Ophthalmic Imaging Foundation, Washington, DC and Maisel, W.H., 2023. Considerations for addressing bias in artificial intelligence for health equity. NPJ digital medicine, 6(1), p.170. https://doi.org/10.1038/s41746-023-00913-9
15. Agarwal, R., Bjarnadottir, M., Rhue, L., Dugas, M., Crowley, K., Clark, J. and Gao, G., 2023. Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware model. Health Policy and Technology, 12(1), p.100702. https://doi.org/10.1016/j.hlpt.2022.100702
16. Norori, N., Hu, Q., Aellen, F.M., Faraci, F.D. and Tzovara, A., 2021. Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10). https://doi.org/10.1016/j.patter.2021.100347
17. Velichko, A., 2021. A method for medical data analysis using the LogNNet for clinical decision support systems and edge computing in healthcare. Sensors, 21(18), p.6209. https://doi.org/10.3390/s21186209
18. Samee, N.A., Ahmad, T., Mahmoud, N.F., Atteia, G., Abdallah, H.A. and Rizwan, A., 2022, November. Clinical decision support model for segmentation and classification of brain tumor MRIs using a U-Net and DCNN cascaded learning algorithm. In Healthcare (Vol. 10, No. 12, p. 2340). MDPI. https://doi.org/10.3390/healthcare10122340
19. Nouis, S.C., Uren, V. and Jariwala, S., 2025. Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: a qualitative study of healthcare professionals' perspectives in the UK. BMC Medical Ethics, 26(1), p.89. https://doi.org/10.1186/s12910-025-01243-z
20. Kothinti, R.R., 2024. Artificial Intelligence in Disease Prediction: Transforming Early Diagnosis and Preventive Healthcare. Int. J. Acad. Res. Dev, 9, pp.1330-1340. https://ijnrd.org/papers/IJNRD2405839.pdf