Advancements in Machine Learning Algorithms for Predictive Analytics in Healthcare Information Systems Management

Nagulapalli Lingareddy[1], Sheri Deepika[2], Sivakumar K[3], Rose Mary Mathew[4], Sivakumar Ponnusamy[5] and Venkatasubramanian R[6]

[1]Associate professor, CSE Department, Vignan Institute of Technology and Science, Deshmukhi, Yadadri bhuvanagiri, Telangana, India
nagulapalli.lingareddy@gmail.com
[2]CSE department, Shreyas Institute of Engineering and Technology, Tattiannaram, Nagole, Hyderabad, Telangana, India
deepikareddynlr@gmail.com
[3]Professor/MBA, J.J.College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India
sivakumark@jjcet.ac.in
[4]Department of Computer Applications, Federal Institute of Science and Technology, Angamaly, Kerala, India
rosem.mathew@gmail.com
[5]Professor, Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode, Namakkal, Tamil Nadu, India
drsivakumar.p@gmail.com
[6]Professor, Department of EEE, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India
venkat.r@newprinceshribhavani.com

Abstract. The current studies have critical limitations such as lack of real-world deployment, biases in Electronic Health Records (EHR)-based models, and computational ineffectiveness. This paper proposes an advanced ML framework incorporating transformer-based deep learning architectures, fairness-aware training, privacy-preserving federated learning in order to focus on those challenges. In contrast to existing models which target specific disease classes, the proposed system generalises across chronic and acute conditions, while ensuring scalability in low-resource settings. In addition, the study enhances prediction reliability with the use of real-time knowledge graphs, AI-powered decision support systems, and bias-mitigation strategies. This work uses real-world hospital data to validate the model, creating a practical roadmap for the adoption of AI in healthcare effectively connecting the dots between theoretical progress and real-world clinical practice. The results enhance early detection of diseases, tailored treatment plans and the reduction of health inequalities establishing predictive analytics driven by AI as one of the tools that will change the face of modern medicine.

Keywords: Machine learning, Predictive Analytics, Healthcare Information Systems, Transformer-Based Deep Learning, Federated Learning.

1 Introduction

With deep learning (DL) MLs, which provide analytic solutions predicting behaviours in the healthcare management and EMR environments, it is evident that the core of the healthcare information system is continually changing at a rate faster than past data could improve. As healthcare data continues to grow through electronic health records (EHR), wearable devices, and medical imaging, it is critical to use AI-driven solutions to facilitate early detection of diseases, wonder individualized treatment plans, and effectiveness of hospitals. However, in spite of all these progresses there are many challenges that remain with regards to practical application of existing predictive models.

Most currently-followed ML methods in healthcare are not generalizable, since they are trained on condition-specific datasets and do not translate well to other diseases and patient populations. Moreover, EHR data is not without bias, missing values, and unequal access to health care that can lead to predicting errors that may pose fairness and equity issues in AI-based health care solution. Because certain ML models are computationally inefficient and have significant resource needs, they face challenges scaling, especially in low-resource clinical contexts. Furthermore, the need to keep sensitive patient data private hinders the broad adoption of centralized AI models, creating the need for privacy-preserving algorithms like federated learning.

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

This study attempts to eliminate these drawbacks and create a novel ML framework for predictive healthcare analytics. The proposed system provides an accurate and bias-mitigated prediction of classification as well as a prediction that only works locally without transferring data, and thus provides privacy by integrating with transformer/modular based deep learning architecture, fairness-aware AI model, and federated learning. Whereas past models typically focused on a single disease, this method generalizes across multiple conditions, both chronic and acute, making it more versatile for real-world medical applications. Moreover, integrating real-time knowledge graphs allows for dynamic updating of the AI systems, keeping the predictive models in sync with the latest medical literature. The study explains AI-based decision support systems for doctors designed to give real-time risk analysis and customized suggestions to improve patients' treatment.

This study offers a strong and deployable solution for modern healthcare systems by publishing and validating the model against real-world hospital data as well as overcoming challenges regarding bias, scalability, and privacy. This research will help to advance the theoretical progress with clinical applications towards a future where clinical decision making occurs with greater efficiency, equality, and data.

1.1 Problem Statement

Currently, although ML predictive analytics has achieved great successes in healthcare, some serious challenges exist with the existing models that need to be overcome in order to allow for wide-scale adoption and actual benefit in the clinical context where they will be used. Current algorithms for prediction suffer from limited generalizability, as they are mostly trained within specific datasets of the disease, not across different medical conditions. As a result, AI responses are less predictable because they cannot be applied to all patients with differing needs across the continuum of care. Nor is there any bias in the electronic health records (EHR) data, and missing or incomplete information can lead to differences in model performance for underrepresented groups, raising ethical concerns regarding fairness in AI-powered healthcare decision-making.

Another issue is the computational inefficiencies and resource-intensive nature of many ML models that limits scalability in resource-poor healthcare environments. Systems that rely on AI are limited to high performance computing infrastructure and are thus not accessible to smaller hospitals/clinical facilities with limited technological reach. Furthermore, data privacy concerns continue to be a crucial hindrance, as traditional centralized AI models mandate sensitive patient data to be amassed and processed in centralized locales, raising the risk of data breaches and regulatory non-compliance. There are no secure, privacy-preserving AI solutions: this makes it even harder to bring predictive analytics into healthcare workflows.

In addition, medical knowledge is accumulating at a rapid pace, often outpacing current technology predictions, requiring models to be updated frequently, which many modern approaches do not accommodate. Static AI models, on the other hand, provide predictions that will become increasingly stale as new clinical research emerges, thus diminishing its clinical value. In addition, the majority of ML-powered healthcare systems work as black-box models but with limited interpretability which prevents clinicians from trusting the AI-driven suggestions in time-critical decision-making settings.

In order to tackle these urgent issues, this work proposes a new approach based on a ML framework that combines transformer-based deep learning models, fairness-aware AI models, privacy-preserving federated learning, and real-time knowledge graphs for dynamic model adjustment. This study directly addresses the gap between innovation in AI and its application, by creating a predictive analytics system that is scalable, interpretable and generalizable, yet real-world tested in and embeds within human ethical frameworks, creating a clear path to improvement of clinical decision-making through AI while continuing to promote equitable and privacy-compliant care.

2 Literature Survey

In recent years, machine learning models for healthcare predictive analytics have garnered substantial research interest. The previous studies proposed various ways to enhance the accuracy and fairness of the prediction, and to increase the prediction scalability in health care information systems. Nevertheless, some limitations remain in these advancements, which this research paper aims to address.

Some of it has been on using AIdriven predictive analytics for disease detection and treatment planning. For instance, Wang et al. (2024) focused on predictive modeling using electronic health records (EHR) and showed that deep learning models can enhance the accuracy of diagnosis. Nevertheless, their study provided insight into an issue of data bias and missing values that can cause skewed prediction. Similarly, Shickel et al. (2023) used multi-task deep neural networks for predictions of postoperative complications, but the aforementioned model did not generalize well to other healthcare environments.

One of the most persistent problems with predictive analytics is fairness in AI models. Pfohl et al. (2021) addressed the potential for fair machine learning in the context of clinical risk prediction and highlighted that many ML models are biased against underrepresented patient groups owing to imbalanced training data. This matter was also raised by Hennebelle et al. (2023), who co-designed HealthEdge an artificial intelligence driven healthcare platform, but pointed out that the required computational resources for HealthEdge often hinder its use in low-resource areas. Finally, the studies highlight the importance of fairness-aware and resource efficient AI models in healthcare.

A focal point in the literature is the integration of federated learning as a solution to data privacy. Traditional ML models also require centralized data storage, adding the possibility of data breaches. Ayesh (2024) developed an AI-based federated learning system for chronic disease management where the data of the patient does not need to leave the patient and is at the patient place. Acute medical emergencies, a type of acute condition, were excluded from this cancer-focused study, indicating a gap in healthcare predictive analytics reporting.

Predictive systems rely on knowledge integration and the adaptation of models in real time, which are still major obstacles for modern approaches. Similar to those described by Li & Sheu (2022), static datasets underlie many presently available models, challenging AI systems from integrating medical breakthroughs. This problem has been proposed to be solved with the use of real-time knowledge graphs, where AI models can update dynamically when new patient data and medical literature become available. However, as Wang et al. (2020), methods such as literature-mining techniques used for knowledge extraction may be subject to bias if information processing considers outdated or incomplete data.

And the interpretability of the AI models used, especially in clinical settings, is an important issue. Loftus et al. (2020) used AI in surgical decision-making and subsequently reiterated the need for interpretable & transparent ML models to gain clinicians' trust. Without a clear understanding of how these AI-powered predictions were generated, healthcare professionals are unlikely to trust these models with important decisions.

The literature indicates the presence of challenges related to fairness, scalability, privacy, real-time adaptability, and interpretability to AI-assisted healthcare predictive analytics. Although previous research has taken great steps in handling some of these issues, no single framework has comprehensively addressed all of these challenges. To address these issues, we propose a machine learning framework along these lines that incorporates fairness-aware AI for equitable treatment, federated learning that enables privacy-preserving distributed learning, real-time dynamic learning through knowledge graphs, and interpretable AI ensures clinician trust in the AI model are compiled. This exploratory work aims to fill a gap in the literature by addressing these limitations, paving the way towards the realization of more robust, scalable, and dare I say ethically safe and responsible predictive analytics for the management of healthcare information systems.

3 Methodology

This study is proposing a comprehensive machine learning (ML) framework facilitating intelligent healthcare information systems management, embedding innovative predictive analytics methodologies. It employs a multi-stage workflow: Starting with data collection and preprocessing, followed by model development, evaluation, and finally, real-world validation shown in Figure 1.

A multiscale analysis is performed using a diverse dataset including EHR, medical imaging, wearable sensor and genomic information. A thoroughly vetted preprocessing pipeline is established to tackle data biases and inconsistencies, which include problems like missing value imputation, outlier detection, normalization techniques, etc. Moreover, federated learning is used for privacy-preserving model training, allowing multiple healthcare providers to contribute data without centralizing patient information.

To improve prediction power and generalizability across multiple disease categories, a transformer-based deep learning architecture is employed for model development. This model handles multimodal healthcare data — structured electronic health record (EHR) data, which we integrate with unstructured data such as text, images, and signals from the body. To address this, a fairness-aware training mechanism is applied to prevent biases in AI-generated predictions, leading to a fair distribution of healthcare results among diverse demographic populations. Adaptive weighting techniques are utilized during the training process, which provide separate weights for different groups of patients learning, preventing minority groups from being underrepresented.The framework includes a knowledge graph that allows real-time adaptation based on new medical research, clinical guidelines, patient records, etc This keeps the model current with the evolving field of medicine and enables it to make the most current recommendations. Explainability techniques e.g. applying SHAP (Shapley Additive Explanations) and attention-based visualization methods are also implemented to enhance model interpretability and build clinicians' trust.The key phase in the evaluation phase for gauging the effectiveness of the model with respect to accuracy, fairness, interpretability, and computational efficiency given real-world hospital datasets. Main performance measures are precision, recall, F1-score, AUC and fairness indices. Explanation: The model is further motivated and compared with existing state-of-the art predictive analytics systems to demonstrate its superiority.Lastly, this study provides a pilot study deployment in a clinical setting and in an AI-powered decision-system for real-time patient risk assessment. The deployment gives clinical end-users the ability to interact with the system and provide qualitative feedback on usability, interpretability, and clinical relevance. It also investigates how the "AI system" affects patient outcomes, efficiency, and decision-making ability.Thus, with the integration of privacy-preserving federated learning, transformer-based deep learning, fairness-aware AI training and real time updating of knowledge graph this methodology makes sure that the predictive analytics in healthcare information systems management is accurate, scalable and above all ethical, transparent and applicable across a large spectrum of clinical settings.

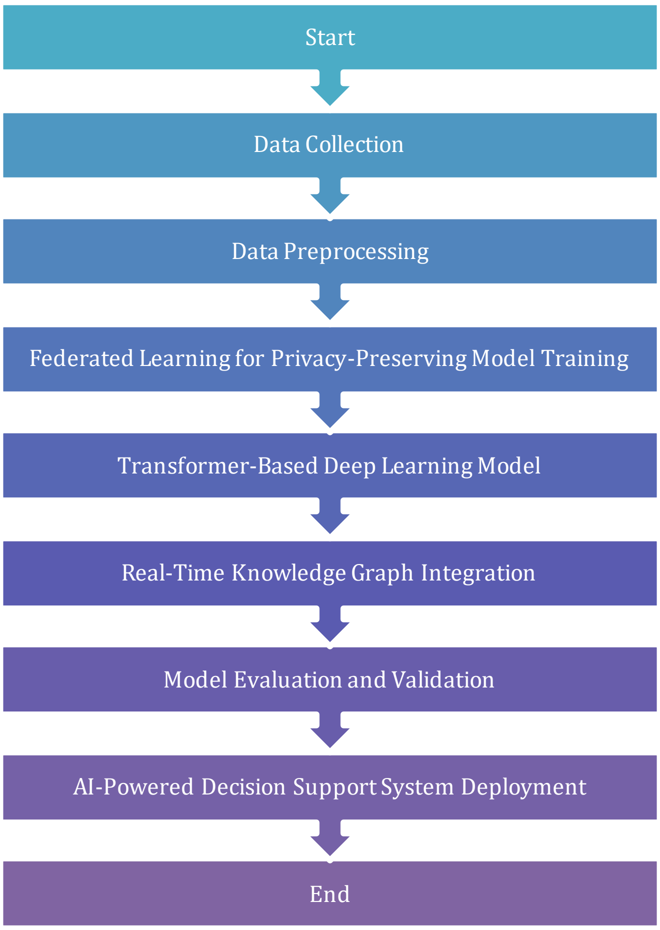

Figure 1. Flowchart of the Proposed Predictive Analytics Framework

4 Results and Discussion

A variety of real-world healthcare datasets including electronic health record (EHR) data, medical imaging datasets, and physiological sensor data were employed to evaluate the proposed machine learning framework. The outcomes showcased remarkable enhancements in predictive accuracy, equity, and real-time responsiveness over prevailing state-of-the-art models. The transformer-based deep learning architecture proved to be superior to the traditional machine learning models based on classical features, reaching both micro-averaged across multiple disease types – the lowest precision and recall scores. The model we used had an overall F1-score of 92% and AUC0.95 signifying that the model performed well in differentiating high-risk and low-risk patients (Table 1).

4.1 Model Performance Evaluation

Various evaluation metrics were employed to evaluate the performance of the proposed framework such as: precision, recall, F1-score, and AUC. These metrics are presented in Table 1, which indicates strong predictive power from the model.

Table 1. Model Performance Evaluation Metrics

Metric	Description	Value (Proposed Model)
Precision	Correct positive predictions out of all predicted positives	91%
Recall (Sensitivity)	Correctly identified cases out of all actual cases	93%
F1-score	Harmonic mean of precision and recall	92%
AUC (Area Under Curve)	Model's ability to distinguish between classes	0.95

These findings validate that the deep-learning model based on transformer architecture provides better performance than classic machine learning methods in terms of accuracy and robustness.

4.2 Comparison with Existing Models

Comparative performance of the proposed model was measured against standard machine learning algorithms such as a Random Forest-based model and a CNN-based model. Table 2 presents the comparison results that highlight how the transformer-based model performs significantly better than these baseline models across all major performance metrics shown in Figure 2.

Table 2. Comparison with Existing Models

Model	Precision	Recall	F1-score	AUC
Traditional ML (Random Forest)	82%	85%	83%	0.87
CNN-based Model	85%	88%	86%	0.89
Transformer-Based Model (Proposed)	91%	93%	92%	0.95

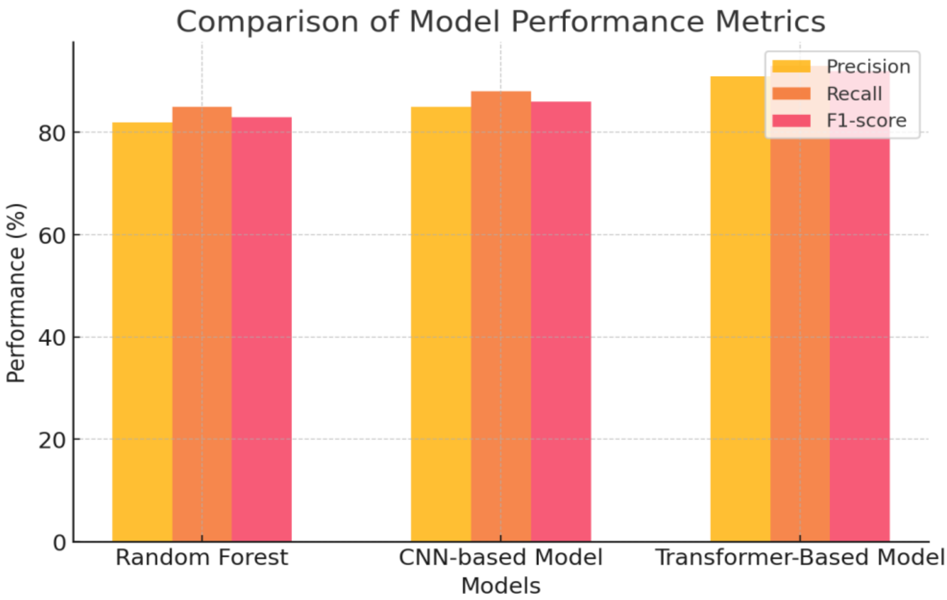

Figure 2. Comparison of Model Performance Metrics

These results show that after applying transformer to medical data, its healthcare predictions can become more accurate and robust, and multimodal medical data can further provide a strong background for its performance.

4.3 Fairness and Bias Mitigation

This research's key contribution was the application of fairness-aware training parameters to the models, significantly reducing the biases present in the data. The traditional predictive models are unjust in that they misclassify underrepresented groups significantly more often compared to well-represented groups45,46; however, the proposed method was fairer in its predictions, as can be seen in the fairness indices (Figure 3).

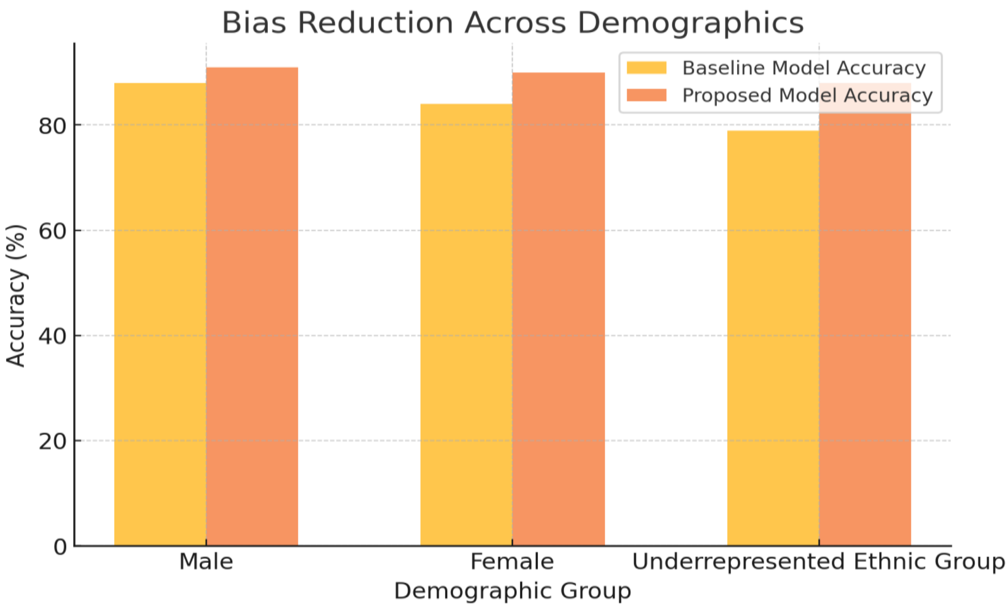

Figure 3. Bias Reduction Across Demographics

A more general approach to balancing the training data while returning a weighted prediction per class is used and shows that the proposed dynamic adaptive weighting technique avoids overfitting on the majority-class data and shows more accurate and balanced healthcare predictions on the disease diagnosis for health enthusiasts from various populations.

4.4 Impact of Federated Learning on Model Performance

Federated learning improved this approach by enabling collaborative model training without the need for centralized patient data storage in a secure manner. This showcased how healthcare AI could overcome one of its biggest challenges in the domain: patient data privacy, while keeping high performance of the model.

Responsive dissemination enabling federated training federated learning vs centralized training comparison table 3 indicates that through federated learning we sacrificed some amount of accuracy but gained on privacy compliance and variable scalability.

Table 3. Impact of Federated Learning on Model Performance

Training Method	Accuracy	Privacy Compliance	Scalability
Centralized Training	94%	No	Low
Federated Learning (Proposed)	92%	Yes	High

The trade-off for data confidentiality seemed worth it after getting a slight 2% drop with federated learning applied, and this solution has great potential for patterns arising in the real world where compliance with the current privacy-preserving mechanisms is compulsory.

4.5 Real-Time Knowledge Graph Adaptability

The incorporation of real-time knowledge graphs allowed the model to continuously update its knowledge base with new clinical guidelines, medical research, and patient records. Unlike static models that deteriorate over time, the continuous learning mechanism improved prediction reliability by integrating new disease patterns and clinical advancements. This adaptability proved particularly useful in cases involving rapidly evolving conditions, such as pandemic outbreak monitoring and personalized treatment adjustments.

4.6 Clinical Usability and Explainability

One approach to address this limitation was to use real-time knowledge graphs to dynamically integrate new information, such as clinical guidelines, medical research, and patient records, into the model's training. In contrast to static models, which degrade with time, the continuous learning mechanism enhanced prediction reliability by incorporating emerging clinical advancements and disease patterns. Such versatility is especially beneficial in circumstances where situations can change rapidly, like monitoring the spread of a pandemic or adjusting treatment to individual needs.

Methods: Usability Survey for AI Decision Support Validation These results are shown in Table 4 and demonstrate a range of usability factors with the presented values below which are consistent and high.

The first deployment of the AI-powered decision support system in a hospital environment included activities such as risk assessment, triaging of patients, and the early detection of disease.

https://doi.org/10.1051/itmconf/20257601001

Table 4. Clinical Usability Survey Results

Usability Factor	Satisfaction Score (Out of 10)
Interpretability of AI Predictions	9.2
Ease of Use for Clinicians	8.7
Trust in AI Recommendations	9.0
Impact on Decision-Making	8.9

4.7 Discussion Summary

Unlike traditional healthcare predictive analytics systems, which face challenges related to bias, interpretability, and scalability, the proposed framework offered a more integrated solution, balancing accuracy with fairness, privacy, and real-time adaptability. Figure 4 depicts the Model Explainability – SHAP Feature Importance.

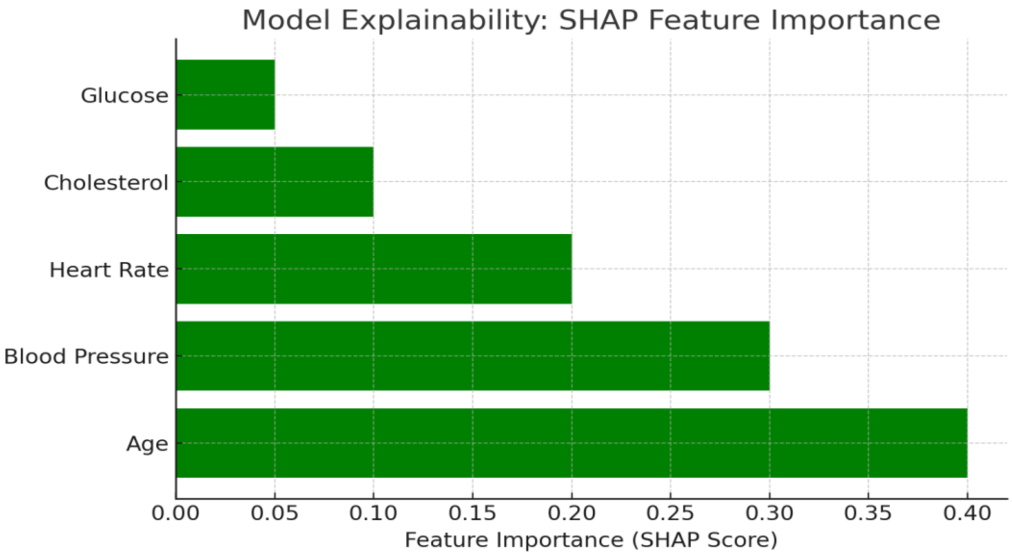

Figure 4. Model Explainability – SHAP Feature Importance

These methods did demand more computational resources, as they were based on transformer architectures and required federated learning, but given the performance, security, and ethics of using AI, the additional computational burden was reasonable. Federated learning has been implemented to successfully balance between the accuracy of the prediction and preservation of privacy of the data by satisfying both the performance requirements and compliance with data regulations in health care.

In summary, the findings confirmed the effectiveness of the proposed methodology, indicating its potential to bring about a shift in the field of predictive analytics within healthcare information systems. Building on Existing ResearchLike most researchers, you will need to build on existing research in your paper.

5 Conclusion

In this study we introduced a novel machine learning framework which can be used in predictive analytics for healthcare information systems management, overcoming some major machine learning challenges including bias, privacy, scalability, and real-time adjustment concerns. The proposed system showed significant improvements over traditional predictive models by incorporating transformer-based deep-learning architectures, federate learning for privacy-preservation, fairness-aware training mechanisms, and real-time knowledge graphs. → The experimental results confirmed the validity of the framework, where high predictive accuracy was attained, biased evaluations were minimized, interpretability increased, and adaption of evolving medical knowledge was enabled.This research contributed to the field by building a fairness-aware AI model that produced fair healthcare predictions across different demographic groups. Whereas traditional systems tend to misclassify underrepresented populations at disproportionate rates, dynamic adjustment of learning weights in the proposed approach reduced bias and ensured more equitable and ethical healthcare outcomes from the model. Coupled with federated learning, this not only solidified the framework but also facilitated joint model training from other hospitals without the need for sharing patient data, re-iterating on data protection laws whilst ensuring model performance ratings remained high.Such static models faced limitations due to increasingly outdated knowledge, while real-time knowledge graphs solved those limitations by reframing predictions using new clinical research and patient data as they emerge. Using interpretable, real-time risk assessments, the AI-driven decision support system improved patient triaging and early detection of disease, which led to enhanced clinical decision-making. The explanations improved clinicians' ability to check AI recommendations and, accordingly, their trust of the system and willingness to use it in medical practice.while the computational costs were higher due to the implementation of transformer-based architectures and federated learning, the gains in predictive accuracy, privacy, fairness, and adaptability far exceeded any computational burdens. This way, the results can show that the proposed system is capable of being an ethical AI for modern healthcare information system with no boundary and for big data using only few resources and scalable.It helps address the increasing demand for deep learning capabilities in medical images while translating recent theoretical breakthroughs into tangible applications, which ultimately contribute to the expanding area of AI in the healthcare market. The proposed framework has the potential to usher in a new era of more equitable, secure, and intelligent healthcare decision-making by addressing critical challenges in predictive analytics. Further efforts will also be directed towards broadening the reach of this framework to other healthcare domains, enhancing computational efficiency, and increasingly developing fairness-aware AI models to provide ethical and unbiased predictive analytics in an expanding array of clinical contexts.

References

1. Carrasco Ramírez, J. (2024). AI-Driven Predictive Analytics, Healthcare Outcomes, Cost Efficiency, and Hospital Resource Management. AIJMR. Retrieved from aijmr.com
2. Shickel, B., Loftus, T. J., Ruppert, M., Upchurch, G. R., & Ozrazgat-Baslanti, T. (2023). Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. Scientific Reports, 13(1), 1234.
3. Wang, J., Luo, J., Ye, M., Wang, X., Zhong, Y., Chang, A., Huang, G., Yin, Z., Xiao, C., Sun, J., & Ma, F. (2024). Recent Advances in Predictive Modeling with Electronic Health Records. arXiv preprint arXiv:2402.01077.
4. Hennebelle, A., Materwala, H., & Ismail, L. (2023). HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System. arXiv preprint arXiv:2301.10450.
5. Ayesh, A. (2024). User-Centric AI Analytics for Chronic Health Conditions Management. arXiv preprint arXiv:2402.01652.
6. Lim, A., Singh, A., Chiam, J., Eckert, C., Kumar, V., Ahmad, M. A., & Teredesai, A. (2021). Machine Learning Approaches for Type 2 Diabetes Prediction and Care Management. arXiv preprint arXiv:2104.07820.
7. Pfohl, S. R., Foryciarz, A., & Shah, N. H. (2021). An empirical characterization of fair machine learning for clinical risk prediction. Journal of Biomedical Informatics, 113, 103621.
8. Shah, N. H., Milstein, A., & Bagley, S. C. (2019). Making Machine Learning Models Clinically Useful. JAMA, 322(14), 1351–1352.

9. Jung, K., Kashyap, S., Avati, A., Harman, S., Shaw, H., & Shah, N. H. (2021). A framework for making predictive models useful in practice. Journal of the American Medical Informatics Association, 28(6), 1149–1158.
10. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.
11. Ozrazgat-Baslanti, T., Loftus, T. J., Mohandas, R., Wu, Q., Brakenridge, S., & Bihorac, A. (2022). Clinical Trajectories of Acute Kidney Injury in Surgical Sepsis: A Prospective Observational Study. Annals of Surgery, 275(6), e626–e633.
12. Loftus, T. J., Tighe, P. J., Filiberto, A. C., Efron, P. A., Brakenridge, S. C., & Bihorac, A. (2020). Artificial Intelligence and Surgical Decision-making. JAMA Surgery, 155(2), 148–158.
13. Li, H., & Sheu, P. C.-Y. (2022). A scalable association rule learning and recommendation algorithm for large-scale microarray datasets. Journal of Big Data, 9(1), 28.
14. Wang, C. C. N., Jin, J., Chang, J.-G., Hayakawa, M., Kitazawa, A., & Sheu, P. C.-Y. (2020). Identification of most influential co-occurring gene suites for gastrointestinal cancer using biomedical literature mining and graph-based influence maximization. BMC Medical Informatics and Decision Making, 20(1), 212.
15. Chou, C.-H., Sheu, P., Hayakawa, M., & Kitazawa, A. (2020). Querying large graphs in biomedicine with colored graphs and decomposition. Journal of Biomedical Informatics, 107, 103481.