Performance and Reliability Landscape of AI Models in Healthcare Diagnostics

Dr. Suman Jayakumar[1], *Dr.Shashi Kant Gupta*[2]
[1] Post-Doctoral Research Scholar, Lincoln University College, Malaysia
[2] Adjunct Professor, Lincoln University College, Malaysia
Email ID: jayakumarsuman@gmail.com, shashigupta@lincoln.edu.my

Abstract: Healthcare diagnostic has been significantly transformed by Artificial Intelligence (AI) that leads to personalized treatment, precise prediction of disease, and early detection. There is also greater deal of work done on privacy preservation using AI as well as some emerging studies towards trustworthiness too in clinical studies. Existing review work are noted to be quite fragmented making it quite challenging to extract a holistic inference towards influence of AI on healthcare diagnostic. Therefore, the proposed study introduces a comprehensive, compact, and highly systematic review of AI approaches exercised towards healthcare diagnostic using Machine Learning (ML), Deep Learning (DL). Adopting PRISMA methodology, the proposed study presents varied scale of methodologies used in clinical study. The study contributes to highlighting emerging trends, unsolved challenges, frequently used approach, etc.

Keywords: Healthcare Diagnostics, Artificial Intelligence, Machine Learning, Deep Learning, Privacy, Trust

Introduction

The aim of this study is towards systematic review of state-of-the-art AI methodologies used in healthcare diagnostics, along with reviewing privacy preservation approaches and trustworthy frameworks. The contribution of this manuscript are as follows: i) The paper presents a granular assessment of AI offering a comprehensive mapping in healthcare diagnostics, ii) The paper studies emerging privacy-preserving approaches along with analysis of trustworthy AI dimensions, iii) The paper also presents a quantitative analysis of latest research trends along with studying trade-offs, iv) Finally, the paper identifies research gap and unsolved problems.

Research Method

The proposed study adopts Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA). The initial process is to frame-up Literature Search Strategy where articles from databases e.g. IEEE, PubMed, Web of Science, and Scopus is looked upon based on AI approaches, healthcare applications, and outcomes measures. The inclusion criteria consist of i) articles using ML, DL, privacy related attributes, trust-based methods towards healthcare diagnosis, ii) articles published between 2020-2025. The exclusion criteria consist of i) any articles towards prognosis without diagnostic outcomes, ii) articles without any quantitative assessment, and iii) non-English publication. The third stage is called Study Screening and Selection where 145 articles have been identified in initial search followed by removal of duplicates resulting in 85 articles.

Learning Outcomes

Table 1 exhibits increasing quantity of publications while conventional ML models follows next. Study implementation using encrypted inference and FL are in nascent stage and growing but they are very less in number of publication in contrast to DL and ML approaches. It is also found that there are very less number of work carried out towards trustworthiness in AI and it remains quite a less explored area although it is emerging.

Table 1 Research Trends (2020-2025)

AI Approach	Scopus	Web of Science	PubMed	IEEE
ML Methods [1][2]	1,799	1,598	1,449	1,189
DL Methods [3][4]	2,682	2,423	2,389	1,901
Privacy-Preserving AI [5][6]	675	591	531	411
Trustworthiness in AI [7][8]	478	401	389	331

Most Frequently Addressed Problems

Based on 60 number of studies considered in this manuscript, it is noted that most frequently studied problems is related to diagnosis of cardiovascular disease. There are numerous papers that has already addressed diagnosis related to myocardial infarction, heart failure, coronary artery disease using mainly ML models (Ensemble, XGBoost, RF). The other most frequently addressed research problems are related to diagnosis of lungs and thoracic disease, diabetes, and cancer. All these are mainly researched upon using DL mostly and in some cased ML too but less in contrast to DL approaches.

Less Frequently Addressed Problems

There are few implementation emphasizing on small-cohort disease or rare disorders. Majority of the dataset is focused on adult populations while AI diagnostic considering pediatric is very much limited. There is a very restricted representation of cognitive disorders while few studies have reported real-time validation on clinical deployment. Generalizability is another ignored perspective in almost every research work.

Most Frequently Adopted Approaches

RF and CNN are the most frequently adopted classical ML and DL model used in healthcare diagnostics. Apart from RF, LR and different variants of gradient boosting (AdaBoost, LightGBM, XGBoost) have been reported in various predictive healthcare analysis. CNN adoption is more inclined towards imaging-centric diagnostics (histopathology, ultrasound, X-ray, CT, MRI). Further, CNN is also increasingly integrated with ViT for performance gains. Adoption of FL is quite less in contrast to CNN and transformers in healthcare analytics; however, it is one of the preferred one for training privacy preservation approaches. They are also seen to be fused with DP methods. XAI is just a beginning and less concrete modelling is witnessed now attempting towards interpretability and gaining clinical trust.

Less Frequently Adopted Approaches

Not all approaches are predominantly used; there are certain approaches which have been used but their frequencies of implementation are found quite lower. There are various approaches like specialized boosting methods (LightGBM, XGBoost), hybrid architectures of privacy and trust, blockchain for privacy, GAN are found to be less frequently adopted towards solving issues pertaining to privacy and trustworthiness. Apart from this, the landscape of emerging transformer models has not been assessed outside of core imaging dataset.

Unsolved Problems

The unsolved problems till date are related to external validation and generalizability mainly. Other unsolved problem is related to adoption of small and imbalanced dataset where prediction of rare disease is still underdeveloped. Trade-off existing between performance and clinical interpretability as well as between utility and privacy remains unsolved too. There is also a potential deployment and integration barriers while mitigation approaches towards fairness and bias in research work is yet to be seen.

Conclusion

A multi-dimensional review and systematic study towards AI utilization in healthcare diagnostic has been studied in this paper. The prime emphasis was to actually understand the taxonomies of various enlisted AI models used for solving different problems which are mainly related to diagnosis of specific disease ranges. The paper analyzes clinical context and disease specific ranges while it also discrete determines the specific AI model with optimal diagnostic performance. Various emerging privacy preserving methods e.g. GenAI integration, HE, DP, FL has been studied along with their implications and limitation. The future work can be carried out towards developing a hybrid AI model for addressing the identified research gap or unsolved problems in this study. Hybrid AI model can be developed in future that extends privacy-preserving AI techniques to complex data for real-world applicability. Further, emerging transformer-centric models can be investigated for optimizing predictive efficient complying with highest standard of privacy preservation and trustworthiness.

References

1. Y. Rimal, N. Sharma, S. Paudel, A. Alsadoon, M. P. Koirala, and S. Gill, “Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy,” Sci. Rep., vol. 15, no. 1, p. 13444, 2025, doi: 10.1038/s41598-025-93675-1
2. N. Nasution, M. A. Hasan, and F. Bakri Nasution, “Predicting Heart Disease using machine learning: An evaluation of Logistic Regression, Random Forest, SVM, and KNN models on the UCI Heart Disease dataset,” IT J. Res. Dev., vol. 9, no. 2, pp. 140–150, 2025, doi: 10.25299/itjrd.2025.17941
3. J. Zhou et al., “A deep learning model for early diagnosis of alzheimer’s disease combined with 3D CNN and video Swin transformer,” Sci. Rep., vol. 15, no. 1, p. 23311, 2025, doi: 10.1038/s41598-025-05568-y
4. P. G. Anderson et al., “Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays,” Sci. Rep., vol. 14, no. 1, p. 25151, 2024, doi: 10.1038/s41598-024-76608-2
5. S. Shukla, S. Rajkumar, A. Sinha, M. Esha, K. Elango, and V. Sampath, “Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity,” Sci. Rep., vol. 15, no. 1, p. 13061, 2025, doi: 10.1038/s41598-025-95858-2

6. M. Ferdowsi, M. M. Hasan, and W. Habib, “Responsible AI for cardiovascular disease detection: Towards a privacy-preserving and interpretable model,” Comput. Methods Programs Biomed., vol. 254, no. 108289, p. 108289, 2024, doi: 10.1016/j.cmpb.2024.108289.
7. A. Chauhan, D. Sarkar, G. S. Verma, H. Rastogi, K. Adapa, and M. Duggal, “Evaluating trustworthiness in AI-Based diabetic retinopathy screening: addressing transparency, consent, and privacy challenges,” BMC Med. Ethics, vol. 26, no. 1, p. 140, 2025, doi: 10.1186/s12910-025-01265-7
8. S. C. Nouis, V. Uren, and S. Jariwala, “Evaluating accountability, transparency, and bias in AI-assisted healthcare decision- making: a qualitative study of healthcare professionals’ perspectives in the UK,” BMC Med. Ethics, vol. 26, no. 1, p. 89, 2025, doi: 10.1186/s12910-025-01243-z