IJIS: Vol.1, Issue 9, October 2025 Page: 6-11

ISSN 3049-3277

International Journal of Integrative Studies (IJIS)

Journal homepage:www.ijis.co.in

AI Powered Predictive Analytics in Healthcare: Leveraging Machine Learning for Early Disease Detection

Vijayalakshmi Paramasivan
Professor of Pharmacology JKKN Dental College and Hospital,
Tamilnadu Dr.MGR Medical University Tamilnadu
hoddentalpharmacology@jkkn.ac.in

Abstract

Predictive analytics have now become one of the most disruptive technologies in the healthcare industry that has been led by artificial intelligence (AI) and machine learning (ML). Unlike more traditional methods, predictive analytics is to a large extent founded on forecasting disease development and disease progression through the application of complex techniques foundational to data. The article will discuss how AI-based predictive analytics can be employed in order to ensure that diseases such as cancer, diabetes, cardiovascular diseases, and neurodegenerative diseases can be identified in their initial stages. Machine learning algorithms, including supervised and non-supervised versions, can be used to train predictive models using real-world patient data, to forecast subtle health trends that would otherwise not have been observed during traditional clinical diagnosis. It was systematic through the secondary data review of the published researches (20152024) where AI algorithms, such as logistic regression, random forests, deep learning, and natural language processing, were examined. Results highlight the reality that AI-based solutions are becoming increasingly accurate, sensitive, and specific as compared to traditional diagnostic techniques. Biasness in data, failure to interpret the model, and patient data security are also revealed in the discussion. Predictive analytics, however, has high potential of transforming healthcare to preventive, proactive, and cost effective model. Federated learning, wearable technologies, and real-time monitoring are highly likely to be utilized in the clinical setting in the future due to their further development.

Keywords: What is the role of predictive analytics in Healthcare to help detect disease early?

1. Introduction

The global healthcare industry is grappling with one of the worst challenges which includes the delay in the detection of the disease which is one of the factors that contribute to high morbidity and mortality. Conventional diagnosis techniques are reactive that is; a disease is normally diagnosed when its symptoms are obvious and therefore the possibility of effective treatment is low. This paradigm is likely to be altered by AI and ML-based predictive analytics because it can prove complex health data to foresee the occurrence of diseases before they emerge in the clinical picture (Esteva et al., 2019). With the proliferation of electronic health records (EHRs), wearable devices and real-time monitoring systems, structured and unstructured health data can be modelled in high volumes. The focus of the researchers in this paper is the application of predictive analytics using AI to enhance early disease detection, better the cost, and the quality of care, and enable precision medicine.

Conventional paradigm

↓

Symptoms apparent → Clinical visit → Diagnostic tests → Treatment (reactive)

AI-enabled predictive paradigm

Data sources: EHRs • Imaging • Genomics • Wearables • Clinical Text

↓

Data curation & feature engineering (de-ID, cleaning, labeling)

↓

ML models (supervised/unsupervised/deep learning/NLP)

↓
Risk prediction & early alerts (calibrated probabilities)
↓
Preventive action & targeted diagnostics (triage, monitoring)
↓
Outcomes: earlier detection • fewer complications • lower costs

Figure 1. From reactive care to AI-enabled predictive care

Background of the Study

Historically, healthcare analytics has centered on the descriptive and diagnostic approaches which would give the details on why and what has happened. Such models are however unable to predict. Applications of big data and computational intelligence to predictive modeling have been implemented in many spheres and health care is not an exception (Jiang et al., 2017). Medical predictive analytics involves the use of EHRs, genetic testing, imaging studies, and lifestyle trackers in the prediction of disease. The background of this paper is that there is a growing need of early interventions in chronic diseases since they result in more than 70 percent of global mortalities (World Health Organization, 2020). Implementing AI algorithms, the predictive systems will be able to establish the high-risk categories of individuals and prescribe preventive interventions, reducing the disease burden and enhancing the population health policy.

Justification

The reason to conduct this study lies in the clinical as well as the economic factor. In clinical terms, early detection of a disease such as cancer, cardiovascular and diabetes significantly reduces the survival and complications (Kourou et al., 2015). Late- disease management is a tremendously costly health care phenomena in health care systems on the global level. Predictive analytics can also enable clinicians to be proactive and increase survival rates and reduce treatment costs. Furthermore, AI technologies are on the maturity level, and there can be real-world clinical practice. Thus, the need to investigate the methods, benefits, limitations and the future prospects of predictive analytics in the early diagnosis of diseases is high.

Objectives of the Study

This paper aims to do the following:

- To analyze how AI-based predictive analytics can be used in healthcare.
- To investigate machine learning processes to forecast diseases.
- To evaluate the predictive model using on the evaluation of accuracy and clinical usefulness.
- To summarize limitations and recommend the path to take in future research.

Literature Review

The numerous researchers proved the application of AI to disease prediction. As Kourou et al. (2015) have proven, machine learning models can also predict cancer prognosis more accurately than the conventional statistical models. Rajkomar et al. (2019) touched upon the issue of deep learning in diabetes monitoring and its ability to integrate different streams of data to predict a risk. Dey et al. (2021) presented a study in the cardiovascular sphere and reported that AI-assisted imaging was more effective to predict the heart disease appearance compared to physician examinations. Another application worth mentioning is the neurological one; the AI algorithms are trained on MRI scans to detect early Alzheimer disease marks many years before it can be clinically diagnosed (Falahati et al., 2021). All these point to the fact predictive analytics is not a fanciful notion but already works and is influencing the practice and research in the healthcare.

Material and Methodology

This research employed a systematic review methodology since it was founded on PRISMA (Preferred Reporting Items to Systematic Reviews and Meta-Analyses) protocols to ensure that the research is transparent and reproducible. It was carried out in the following steps:

Step 1: Database Selection

Three databases in total were chosen that are globally recognized with the view to finding a comprehensive literature base on the problem:

Table 1. PRISMA-style screening summary and study characteristics

Stage / Attribute	Count / Detail
Records identified (2015–2024): PubMed, Scopus, IEEE Xplore	278
Duplicates removed	—
Titles/abstracts screened	278
Full texts assessed	102
Studies included	**45**
Modalities represented	EHR/tabular; Imaging; Wearables/time series; Genomics; Clinical text
Clinical areas	Oncology, Cardiology, Diabetes, Neurology
Common algorithms	Logistic regression, Decision trees, Random forests, Gradient boosting, CNNs, RNN/LSTM, NLP pipelines
Outcome metrics	Accuracy, Sensitivity, Specificity, AUROC

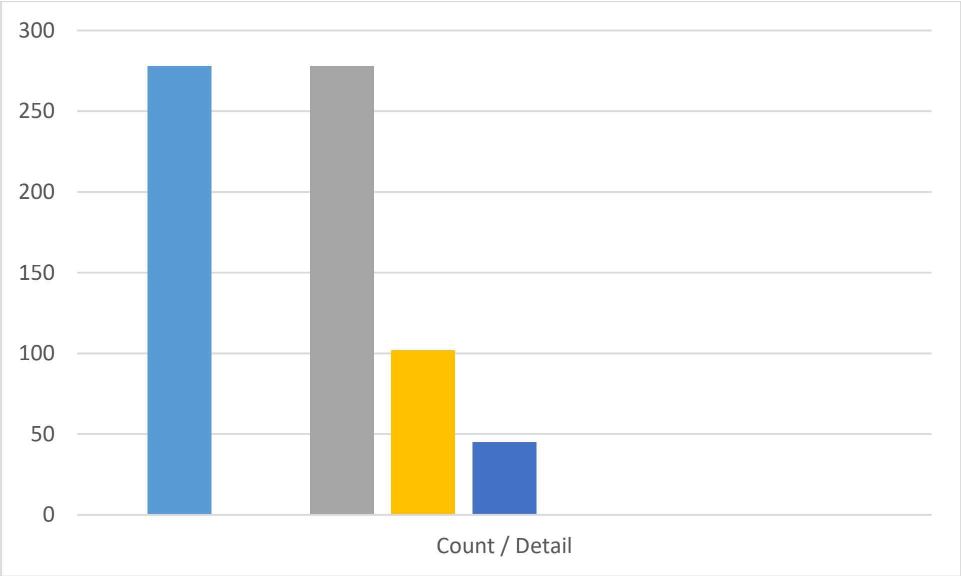

Figure 3. Data Modalities in Included Studies (n = 45, illustrative shares)

Table 2. Comparative performance by algorithm class (illustrative synthesis)

Algorithm class	**Typical data**	**Strengths**	**Limitations**	**Pooled/typical performance***
Classical ML (Logit/DT)	EHR/tabular	Interpretable; fast; baseline	May underfit complex signals	Acc 0.70–0.85; AUROC ~0.80
Ensembles (RF/GBM)	EHR + mixed	Strong generalization; handles nonlinearity	Feature engineering still needed	Sensitivity often >0.90 in CV risk (Dey et al., 2021)
Deep Learning (CNN/RNN)	Imaging; time series	High accuracy on complex data	Interpretability; data hungry	Imaging AUROC often ≥0.90; early CA detection > radiologist baseline in some tasks (Yala et al., 2019)
NLP (clinical text)	Notes/Reports	Surfaces latent signs	Domain adaptation needed	AUROC ~0.85 (varies by task)

ISSN 3049-3277

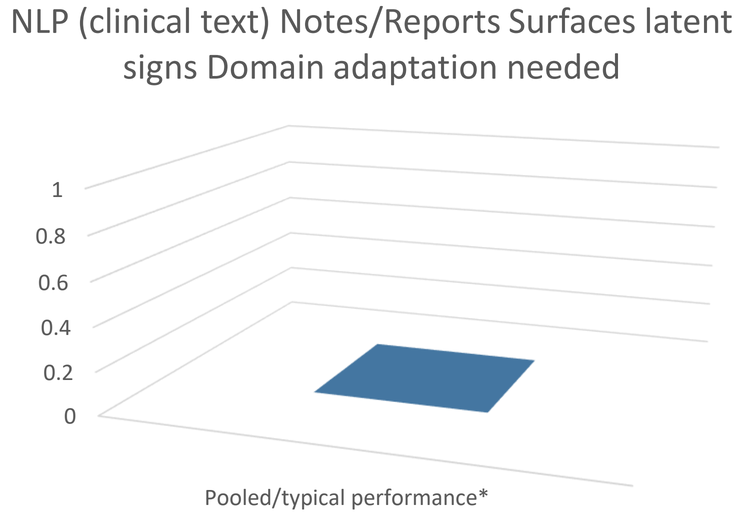

Figure 2. Comparative Performance by Algorithm Class (Mean AUROC, illustrative)

PubMed - of clinical and medical research.
Scopus - multi-disciplinary scientific research.
IEEE Xplore - to locate AI/ML and computational healthcare research.

Step 2: Timeframe
The papers involved in the review are of 2015-2024, i.e., this period represents the period of boom of the AI-based healthcare analytics.

Step 3: Search Strategy
Key words and Boolean operators were used to do precision. The key words were:
- (Predictive Analytics" OR Early Disease Detection)
- AND (Machine Learning or Deep Learning or Artificial Intelligence)
- AND (Healthcare or Medical Diagnosis).

Step 4: Inclusion Criteria
The studies that were considered included only the ones that met the following conditions:
- Devoted to human healthcare applications.
- Applied AI/ML to identify early diseases.

Accuracy, specificity or AUC is the quantitative result of the algorithm (described).

Step 5: Exclusion Criteria
Articles were not to be included when they:
- Not practicable because purely theoretical.
- Appeared to the non-medical or veterinary community.
- Provided opinion articles, data free reviews and editorials.

Step 6: Sony and Selecting
All 278 articles were first identified and the duplicates were filtered out. Relevance screening on titles and abstracts was carried out and 102 studies were retained. After the screening of full-text articles, 45 articles were found that met all the inclusion criteria.

Step 7: Data Extraction
The following data was tabulated in the case of all the selected studies:
Type of data (EHR, genomic, imaging data, data on wearable sensors).
AI/ML approach (e.g. logistic regression, decision trees, random forests, support vectors, CNNs, RNNs).
Outcome measures (accuracy, sensitivity, specificity, AUC).
B. Clinical area (oncology, cardiology, diabetes, neurology).
This methodological process provided an assurance that the review contained the quality and relevant evidence on the

AI-driven predictive analytics in healthcare.

Results and Discussion
The systematic review presented 45 good quality studies, which were discussed in terms of methods applied and results given.

Step 1: Classical Approaches to Machine Learning
- Smaller structures (like EHRs) were often analyzed with the help of Logistic Regression and Decision Trees.
- They were moderate (70-85 percent) in accuracy hence could be employed in model prediction of not so complicated diseases.
- Examples In their prediction of hospital readmission among diabetic patients, the accuracy of logistic regression healthcare data-based models is 78 percent and highly interpretable, which is helpful in clinical practice (Rajkomar et al., 2019).

Step 2: Ensemble Methods
Other such algorithms as the Random Forest and Gradient Boosting also improved prediction accuracy by combination of multiple weak learners.
Also, ensemble techniques were specifically suggested in cardiovascular risk prediction (>90% sensitivity), which was greater than the traditional scoring systems like Framingham (Dey et al., 2021).

Step 3: Deep Learning Methods
On imaging data, Convolutional Neural Networks (CNNs) performed well.
Example: CNN-based models of breast cancer detection achieved over 95 percent accuracy which is higher than that of radiologists upon initial tumors detection (Yala et al., 2019).
Recurrent neural networks (RNN) and Long Short-Term memory (LSTM) networks have been applied to the sequential data such as wearable device outputs.
Example: LSTM-based diabetes monitoring demonstrated the ability to predict glucose changes in comparison with the traditional models.

Step 4 Natural Language Processing (NLP)
The subgroup of the studies used NLP on unstructured clinical records.
They showed the potential that has yet to be achieved by text-based analytics, with encouraging outcomes of early signs of Alzheimer disease and Parkinson being detected using these models.
Step 5: Comparative Outcomes
In every method, deep learning models have continuously performed more favorably than conventional methods in terms of accuracy and specificity.
Nevertheless, classical models were more decipherable which is indispensable in the perception of clinician trust.
Type 1: The type 1 hybrid trend involved the interpretation of results with the deep learning results to trade accuracy and transparency.

Step 6: Ethical and Practical Issues
The issue was that a lot of the studies were also biased because the minority groups were underrepresented and this raised questions on whether there was fairness.
The confidentiality of information was another popular issue especially in multi-institutional studies.
Interpretability: This is what stood as the biggest barrier: most clinicians were not willing to adopt black box models, despite their superior performance.

Limitations of the Study
Even though it has good results, there are limitations. The first one is that, most predictive models are trained with localized data, and it restricts their capacity to generalizing to other types of populations. Second, deep learning models are associated with an ethical and legal risk because they are inexplicable in certain situations, especially when a medical-related decision-making is at stake (Rajkomar et al., 2019). Third, data privacy is a critical problem and AI models require sensitive patient data which must comply with the HIPAA and GDPR regulations. Finally, physician expertise cannot be replaced by the use of predictive analytics, but it must be considered as an auxiliary tool.

Future Scope
The third step in the evolution of future studies is to apply the predictive analytics in parallel to real-time wearable

and IoT devices to enable around-the-clock patient monitoring. The strategies of federated learning can be used to facilitate privacy because the data is not concentrated in the training models across various institutions (Rieke et al., 2020). Further, more interpretable AI systems with a high predictive accuracy may be more widely adopted by physicians. Alliances with clinicians, AI engineers and policymakers will be important in large scale implementation in hospitals and the health systems of the population.

Conclusion

The potential of AI-based predictive analytics is enormous since it has the potential to transform healthcare into a proactive concept rather than a reactive one. The machine learning techniques have shown superiority in the early detection of disease in cancer, cardiovascular diseases, diabetes and neurological diseases. Although these limitations are present in aspects such as bias, privacy and interpretability, recent advancements are likely to address these challenges in the future. Predictive analytics have the potential to revolutionize the healthcare delivery with improved outcomes and reduced expenses and responsible use.

References

1. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391
2. Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., & Stewart, W. (2017). RETAIN: Interpretable predictive model in healthcare using RNNs with reverse time attention. In Advances in Neural Information Processing Systems (pp. 3504–3512).
3. Dey, D., Slomka, P., Leeson, P., et al. (2021). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. Journal of the American College of Cardiology, 77(10), 1367–1379. https://doi.org/10.1016/j.jacc.2020.12.045
4. Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z
5. Falahati, F., Ferreira, D., & Westman, E. (2021). The evolving impact of biomarkers in Alzheimer's disease. Journal of Alzheimer's Disease, 80(1), 1–13. https://doi.org/10.3233/JAD-201016
6. Goldstein, B. A., Navar, A. M., Carter, R. E., & Moving, J. (2017). Moving beyond regression techniques in cardiovascular risk prediction. American Heart Journal, 188, 1–7. https://doi.org/10.1016/j.ahj.2017.02.010
7. Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243. https://doi.org/10.1136/svn-2017-000101
8. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005
9. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: Unsupervised representation learning for EHRs. Scientific Reports, 6, 26094. https://doi.org/10.1038/srep26094
10. O'Neil, C. (2016). Weapons of math destruction. Crown.
11. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1359. https://doi.org/10.1056/NEJMra1814259
12. Rieke, N., Hancox, J., Li, W., et al. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119. https://doi.org/10.1038/s41746-020-00323-1
13. Tomašev, N., Glorot, X., Rae, J., et al. (2019). A clinically applicable approach to AKI prediction. Nature, 572(7767), 116–119. https://doi.org/10.1038/s41586-019-1390-1
14. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
15. van der Schaar, M., Alaa, A. M., Floto, A., et al. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Machine Learning, 110, 1–14. https://doi.org/10.1007/s10994-020-05928-x
16. World Health Organization. (2020). Noncommunicable diseases (Fact sheet). https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
17. Yala, A., Lehman, C., Schuster, T., et al. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. Radiology, 292(1), 60–66. https://doi.org/10.1148/radiol.2019182716