Dr. B. Suresh, 2026, 14:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Early Disease Prediction Using Machine Learning And Deep Learning: A Comparative Analysis

Dr. B. Suresh[1], K. Muthuchamy[2]

[1&2]Assistant Professor, Department of Computer Science & Information Technology
Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai

Abstract- Early prediction of disease allows clinicians to intervene before irreversible damage occurs, improving survival, reducing treatment costs, and easing pressure on health systems. Machine learning (ML) and deep learning (DL) have emerged as the two dominant computational approaches for this task. This review paper presents a comparative analysis of ML and DL for early disease prediction, drawing on representative peer-reviewed studies across diabetes, cardiovascular disease, cancer, neurodegenerative disease, and critical care. The paper outlines the principal algorithms, data modalities, benchmark datasets, and preprocessing practices, and compares ML and DL on data requirements, accuracy, interpretability, computational cost, generalisability, and deployment readiness. The evidence indicates that DL achieves clinician-level or better performance for image, signal, and sequence data, whereas classical ML and gradient-boosted ensembles remain competitive, and often superior, on structured tabular clinical data. Persistent barriers include data scarcity and quality, class imbalance, dataset shift, bias, limited interpretability, and a shortage of prospective, externally validated studies. The paper concludes with recommendations and future directions, including explainable AI, federated learning, multimodal modelling, and standardised reporting.

Keywords- Early Disease Prediction, Machine Learning, Deep Learning, Comparative Analysis, Medical Imaging, Electronic Health Records, Explainable AI

I. INTRODUCTION

1. Background

Many of the diseases that account for the largest share of global morbidity and mortality, including cardiovascular disease, diabetes, cancer, and dementia, develop silently over years before symptoms appear. Detecting disease, or the high risk of disease, at an early stage widens the window for lifestyle change, screening, and treatment, and is widely regarded as one of the most cost-effective routes to better population health. Traditional risk calculators and rule-based clinical scores, however, rely on a small number of predictors and linear assumptions, and cannot exploit the growing volume of electronic health records (EHRs), medical images, genomic profiles, and wearable-sensor data.

Artificial intelligence (AI) has therefore moved to the centre of predictive medicine [1]–[3]. Within AI, two families of methods are commonly applied: classical machine learning, which learns predictive functions from hand-crafted features, and deep learning, which uses multi-layer neural networks to learn features directly from raw data [5]. Both families have produced notable results, but they differ in the data they require, the way they are validated, and the extent to which clinicians can understand their output. Choosing between them for a given prediction task is a practical question that researchers and health organisations repeatedly face.

2. Motivation, Scope, and Contributions

Although numerous studies report high accuracy for individual diseases, evidence on how ML and DL compare under similar conditions is scattered across clinical, engineering, and statistical literature. This review consolidates that evidence into a single comparative view. Its contributions are: (i) a concise description of the main ML and DL methods used for early prediction; (ii) a summary of data modalities, public datasets, and preprocessing practices; (iii) a disease-wise synthesis of representative studies; (iv) a structured comparison of ML and DL across clinically relevant criteria; and (v) a discussion of challenges and future directions.

© 2026 Dr. B. Suresh, This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

3. Review Methodology

This is a narrative review. Representative peer-reviewed articles and benchmark datasets were selected from major sources, including PubMed, IEEE Xplore, ScienceDirect, Nature, and Google Scholar, using combinations of the terms early disease prediction, machine learning, deep learning, electronic health records, and medical imaging. Priority was given to highly cited studies, studies published in reputed journals, studies with external or multi-site validation, and systematic reviews that directly compare modelling approaches. The review is not a formal systematic review, and the studies discussed are illustrative rather than exhaustive. Reported performance figures are quoted from the original studies and are not directly comparable across studies because datasets, populations, and endpoints differ.

4. Structure of the Paper

Section 2 describes ML and DL methods. Section 3 covers data sources, datasets, and preprocessing. Section 4 reviews disease-specific applications. Section 5 presents the comparative analysis. Section 6 discusses challenges, Section 7 outlines future directions and recommendations, and Section 8 concludes.

II. MACHINE LEARNING AND DEEP LEARNING METHODS

1. Classical Machine Learning

Classical ML methods learn a mapping from a vector of input features, such as age, blood pressure, glucose level, or imaging-derived measurements, to a disease outcome. Commonly used algorithms include logistic regression, support vector machines (SVM) [8], k-nearest neighbours, naïve Bayes, decision trees, random forests [9], and gradient-boosted trees such as XGBoost [10]. These methods are computationally light, work with comparatively small datasets, and produce outputs that are easier to explain. Their performance, however, depends heavily on the quality of feature engineering and feature selection, which require domain expertise [4].

2. Deep Learning

Deep learning uses neural networks with many layers to learn hierarchical representations directly from data [5]. Convolutional neural networks (CNNs) [11] dominate image analysis, including retinal, radiological, dermatological, and histopathological images. Recurrent networks such as LSTM [12] and GRU models capture temporal dependencies in longitudinal EHRs, ECG, and physiological signals. Autoencoders learn compact unsupervised representations of high-dimensional records [13], and Transformer architectures [14] use attention to model long-range relationships in clinical text, sequences, and images. Table 1 summarises the main algorithms.

Table 1 – Principal ML and DL Algorithms Used in Early Disease Prediction

Algorithm	Family	Key characteristics
Logistic / Linear Regression	ML	Statistical baseline; coefficients easily interpreted; limited with non-linear data
Support Vector Machine (SVM)	ML	Maximum-margin classifier; effective in high-dimensional, small-sample settings [8]
Random Forest (RF)	ML	Ensemble of decision trees; robust to noise; provides feature importance [9]
Gradient Boosting (XGBoost)	ML	Sequential tree ensemble; strong performance on tabular clinical data [10]
k-Nearest Neighbours / Naïve Bayes	ML	Simple, fast baselines; sensitive to feature scaling or independence assumptions
Convolutional Neural Network (CNN)	DL	Learns spatial features from images (X-ray, CT, MRI, retinal, histology) [11]
Recurrent Neural Network / LSTM / GRU	DL	Models temporal sequences such as longitudinal EHR, ECG and vital signs [12]

Algorithm	Family	Key characteristics
Autoencoder	DL	Unsupervised representation learning from high-dimensional records [13]
Transformer	DL	Attention-based; models long-range dependencies in text, records and images [14]

III. DATA SOURCES, DATASETS, AND PREPROCESSING

1. Data Modalities

Early prediction models draw on several data modalities: structured EHR data (demographics, diagnoses, laboratory results, medications), medical images (X-ray, CT, MRI, PET, retinal photographs, dermoscopy, histopathology), physiological signals (ECG, EEG, photoplethysmography), genomic and biomarker data, and data from wearables and mobile devices. Structured EHR data is the most common input for ML, whereas images and signals are the natural territory of DL [6]. A growing number of studies combine several modalities in multimodal models.

2. Benchmark Datasets

Table 2 lists widely used public datasets. Small tabular datasets such as the UCI Heart Disease and Pima Indians Diabetes collections remain popular for benchmarking classical ML but are too small and homogeneous to support claims about clinical generalisability. Large resources such as MIMIC-III [15] have enabled work on critical-care prediction, while image repositories have driven rapid progress in DL.

Table 2 – Commonly Used Public Datasets for Disease Prediction

Dataset	Disease area	Data type	Typical use
UCI Heart Disease (Cleveland)	Cardiovascular	Clinical attributes (about 300 records)	Benchmarking classical ML
Pima Indians Diabetes	Diabetes	Clinical measurements (768 records)	Benchmarking classical ML
Wisconsin Breast Cancer (WBCD)	Breast cancer	Cell-nucleus features from biopsy images	Tabular classification
MIMIC-III / MIMIC-IV	Multiple (ICU)	De-identified critical-care EHR [15]	Mortality, sepsis, deterioration
ChestX-ray14	Thoracic diseases	Chest radiographs (over 100,000 images)	CNN-based detection
ADNI	Alzheimer's disease	MRI, PET, cognitive and genetic data	Early neurodegeneration
EyePACS / Messidor-2	Diabetic retinopathy	Retinal fundus photographs	Screening with CNNs
ISIC Archive	Skin cancer	Dermoscopic images	Lesion classification

3. Preprocessing and Feature Engineering

Clinical data are noisy, incomplete, and imbalanced. Typical preprocessing steps include handling missing values through imputation, outlier treatment, normalisation, and encoding of categorical variables. Because disease cases are usually rare relative to healthy cases, class imbalance is addressed by resampling methods such as SMOTE [7], cost-sensitive learning, or threshold adjustment. Feature selection techniques (filter, wrapper, and embedded methods) reduce dimensionality and overfitting in ML pipelines, while DL relies on data augmentation, transfer learning from pre-trained networks, and regularisation to cope with limited labelled data. Appropriate evaluation uses sensitivity, specificity, precision, F1-score, area under

the ROC curve (AUC), and calibration, rather than accuracy alone, which can be misleading when the outcome is rare.

4. Evaluation Metrics and Validation Practice

Because early-prediction tasks usually involve rare outcomes, overall accuracy is a poor yardstick: a model that labels every patient as healthy can appear highly accurate while detecting no disease. Studies should therefore report several complementary metrics, summarised in Table 3, and should state the operating threshold at which sensitivity and specificity were measured.

Table 3 – Commonly Used Evaluation Metrics for Disease Prediction Models

Metric	Definition	Clinical relevance
Sensitivity (recall)	TP / (TP + FN)	Proportion of diseased patients detected; critical when missing disease is costly
Specificity	TN / (TN + FP)	Proportion of healthy patients correctly cleared; limits unnecessary work-up
Precision (PPV)	TP / (TP + FP)	Reliability of a positive alert; falls sharply when prevalence is low
F1-score	2 x (Precision x Recall) / (Precision + Recall)	Balances precision and recall on imbalanced data
AUC-ROC	Area under sensitivity vs 1 - specificity curve	Threshold-independent discrimination; can look optimistic for rare outcomes
Calibration	Agreement of predicted and observed risk	Needed when predicted probabilities guide treatment decisions

Equally important is the validation design. Random train-test splits on a single dataset tend to overstate performance, particularly when records from the same patient appear in both partitions. Stronger designs use patient-level splits, k-fold cross-validation for small datasets, temporal validation on later admissions, and external validation on data from different institutions [34]. Calibration plots and decision-curve analysis show whether predictions are reliable and whether acting on them would benefit patients. A typical end-to-end pipeline therefore moves from data collection and preprocessing, through feature engineering or representation learning and model training, to internal validation, external validation, and finally clinical deployment with continuous monitoring.

IV. DISEASE-WISE APPLICATIONS

1. Diabetes And Diabetic Complications

Kavakiotis et al. [25] reviewed ML and data mining in diabetes research and found SVMs, ensemble methods, and neural networks to be widely used for prediction, diagnosis, and complication assessment, with data quality and small cohorts as recurring limitations. For diabetic retinopathy screening, Gulshan et al. [16] trained a deep CNN on retinal fundus photographs and reported a sensitivity of 97.5% and specificity of 93.4% on the EyePACS-1 validation set, demonstrating that DL can support large-scale screening where specialist access is limited.

2. Cardiovascular Disease

For cardiovascular risk, Weng et al. [24] compared random forests, logistic regression, gradient boosting, and neural networks on routine primary-care data from the United Kingdom and found that the neural network achieved the highest AUC (0.764), exceeding the ACC/AHA guideline algorithm (0.728). In the DL domain, Choi et al. [22] applied a recurrent network to longitudinal EHRs to predict heart failure onset and reported an AUC of 0.883 compared with 0.834 for logistic regression, and Attia et al. [21] showed that a CNN applied to the 12-lead ECG could identify left-ventricular systolic dysfunction with an AUC of 0.93. These studies show that DL adds most value when temporal or signal structure is available.

3. Cancer

In oncology, DL has demonstrated expert-level performance across imaging modalities. Esteva et al.

[17] trained a CNN for skin lesion classification that performed on par with 21 board-certified dermatologists. McKinney et al. [18] reported that an AI system for breast cancer screening achieved fewer false positives and false negatives than radiologists on United Kingdom and United States data. Ardila et al. [19] developed a three-dimensional CNN for lung cancer screening on low-dose CT, reporting an AUC of 94.4% on National Lung Screening Trial data. Classical ML on biopsy-derived features, such as the Wisconsin Breast Cancer dataset, also yields high accuracy, but on small and curated data that limit generalisation.

4. Neurodegenerative Disease

Early identification of Alzheimer's disease is challenging because pathology precedes symptoms by many years.

Ding et al. [20] developed a DL model using FDG-PET brain scans that achieved an AUC of 0.98 and detected disease on average about six years before the final clinical diagnosis in the tested cohort. Such results are promising but derive from limited independent test sets, highlighting the need for larger prospective validation.

5. Critical Care and Hospital Settings

In hospital settings, Rajkomar et al. [23] applied deep learning to EHR data from more than two hundred thousand hospitalised adults and reported AUROC values of 0.93 to 0.94 for in-hospital mortality, outperforming traditional clinical predictors.

Miotto et al. [13] showed that an unsupervised autoencoder representation (Deep Patient) learned from a large EHR database could improve prediction across many diseases. At the same time, external validation of a widely deployed proprietary sepsis model reported much poorer discrimination than the vendor had claimed and a failure to identify most sepsis cases [29], a cautionary example that high development performance does not guarantee clinical utility. Table 4 summarises the representative studies.

Table 4 – Representative Studies on Early Disease Prediction Using ML and DL

Study	Disease / task	Data	Method	Reported performance
Gulshan et al. [16]	Diabetic retinopathy	Retinal images	Deep CNN	Sensitivity 97.5%, specificity 93.4% (EyePACS-1)
Esteva et al. [17]	Skin cancer	Clinical images	Deep CNN	Performance on par with 21 dermatologists
McKinney et al. [18]	Breast cancer	Screening mammograms	Deep learning ensemble	Fewer false positives and false negatives than radiologists (US and UK data)
Ardila et al. [19]	Lung cancer	Low-dose CT	3D CNN	AUC 94.4% on NLST data
Ding et al. [20]	Alzheimer's disease	FDG-PET brain scans	CNN (Inception-V3)	AUC 0.98; detection about 6 years before diagnosis
Attia et al. [21]	LV systolic dysfunction	12-lead ECG	CNN	AUC 0.93
Choi et al. [22]	Heart failure onset	Longitudinal EHR	GRU (RNN)	AUC 0.883 vs 0.834 for logistic regression

Study	Disease / task	Data	Method	Reported performance
Rajkomar et al. [23]	In-hospital mortality	EHR (FHIR format)	LSTM / attention models	AUROC 0.93–0.94
Weng et al. [24]	Cardiovascular risk	Primary-care records (UK)	RF, GB, neural network	Neural network AUC 0.764 vs 0.728 for ACC/AHA guideline
Kavakiotis et al. [25]	Diabetes	Clinical / genetic data	Review of ML methods	SVM and ensembles widely used; data quality a key issue

V. COMPARATIVE ANALYSIS: MACHINE LEARNING VERSUS DEEP LEARNING

1. Comparison Across Key Criteria

Table 5 compares ML and DL on criteria that matter for clinical adoption. In general, ML suits structured data and modest sample sizes, while DL excels when raw high-dimensional data, such as images and signals, are available in large quantity.

Table 5 – Comparative Summary of Machine Learning and Deep Learning

Criterion	Classical machine learning	Deep learning
Data type best suited	Structured / tabular data, moderate sample size	Images, signals, text, long sequences, multimodal data
Feature engineering	Manual; domain knowledge needed	Automatic representation learning
Data requirement	Works with hundreds to thousands of samples	Typically needs large, labelled datasets or pre-training
Computational cost	Low to moderate	High (GPUs, long training)
Interpretability	Higher (coefficients, tree importance)	Lower; needs post-hoc explanation (saliency, SHAP)
Risk of overfitting	Moderate	High on small datasets
Typical clinical strength	Risk scores from EHR and lab variables	Image-based screening and early lesion detection
Deployment complexity	Easier to validate and maintain	Greater monitoring, drift and regulatory burden

2. Predictive Performance

The evidence on accuracy is nuanced. For imaging and signal tasks, DL models repeatedly match or exceed specialist performance [16]–[19], and a systematic review of DL versus health-care professionals found broadly equivalent diagnostic performance, while noting that few studies compared both groups on the same data and that reporting was often poor [28]. For structured clinical data, the picture differs. A systematic review of clinical prediction models found no performance benefit of ML over logistic regression when studies at low risk of bias were considered [26], and benchmarking on typical tabular data shows that tree-based ensembles still outperform DL on medium-sized datasets [27]. DL gains appear when longitudinal records are modelled as sequences or when the dataset is very large [22], [23]. Thus, the claim that DL is uniformly superior is not supported; the data modality and sample size largely determine which approach is preferable.

3. Interpretability and Trust

Clinicians need to understand why a model flags a patient. ML models such as logistic regression and

decision trees are comparatively transparent, and random forests and boosted trees provide feature-importance measures. DL models are opaque, and post-hoc methods such as saliency maps, Grad-CAM, and SHAP [31] are used to approximate explanations. Surveys of medical explainable AI caution that such explanations can be unstable or misleading and must be evaluated with clinicians [32]. Interpretability therefore remains a more pressing obstacle for DL than for ML.

4. Computational Cost and Deployment

ML models can be trained on standard hardware and updated easily, which suits resource-constrained hospitals and primary-care settings, including many in low- and middle-income countries. DL models require substantial computing resources, large labelled datasets, and careful monitoring for dataset shift when equipment, populations, or coding practices change. Hybrid pipelines, in which DL extracts features from images and ML combines them with clinical variables, offer a practical compromise.

5. Generalisability and Fairness

Both families suffer when training and deployment populations differ. Obermeyer et al. [30] showed that a widely used health-management algorithm exhibited racial bias because it used healthcare cost as a proxy for health need, illustrating that the problem lies in the label and the data as much as in the algorithm. Models trained at a single site often degrade elsewhere, which is why external and multi-site validation is essential for either approach.

6. Challenges

- Data quality and availability: missing, inconsistent, and unstructured data; small labelled datasets; and limited access owing to privacy regulation.
- Class imbalance and label noise: rare outcomes and diagnoses that are uncertain or delayed.
- Generalisation and dataset shift: performance loss across hospitals, devices, and populations.
- Bias and fairness: under-representation of groups and biased proxy labels can widen health disparities [30].
- Interpretability and clinician trust: black-box models are difficult to audit and justify [32].
- Validation gap: most studies are retrospective, single-site, and not prospectively tested; external validation is uncommon [28].
- Regulation, privacy, and integration: compliance with data-protection laws, regulatory approval, and integration into clinical workflow.

7. Future Directions and Recommendations

Several directions can strengthen the field. First, explainable AI should be built into model development and evaluated with clinicians rather than added afterwards [31], [32]. Second, federated learning allows models to be trained across institutions without sharing patient-level data, addressing privacy while increasing diversity [33]. Third, multimodal models that combine imaging, EHR, genomics, and wearable data, together with large pre-trained and Transformer-based foundation models, may improve early detection beyond single-modality systems [14]. Fourth, researchers should report studies according to TRIPOD guidance [34], include calibration and decision-curve analysis, and compare against strong baselines such as logistic regression and gradient boosting, so that claims of DL superiority are tested fairly [26], [27]. Fifth, prospective multi-centre validation, bias audits, and post-deployment monitoring should become standard before clinical adoption [29], [30]. Finally, for practitioners, a pragmatic rule is to begin with interpretable ML on structured data, and to adopt DL where image, signal, or long-sequence data justify its cost.

VIII. CONCLUSION

This review compared machine learning and deep learning for early disease prediction. Deep learning has achieved expert-level results in imaging- and signal-based screening for conditions such as diabetic retinopathy, skin, breast and lung cancer, and cardiac dysfunction, and shows promise in sequence-based modelling of health records. Classical machine learning, particularly regularised regression and gradient-boosted ensembles, remains highly competitive on structured clinical data, offers lower cost, and is easier to interpret and maintain. Neither family is universally superior; the best choice depends on data modality, sample size, interpretability

requirements, and the resources available for deployment.

Translating these advances into routine care will depend less on further gains in accuracy than on rigorous external and prospective validation, fairness assessment, transparent reporting, explainability, and privacy-preserving collaboration. Future work that addresses these issues will help ensure that early disease prediction systems are not only accurate but also trustworthy, equitable, and useful to clinicians and patients.

REFERENCES

1. E. J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
2. P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in Health and Medicine," Nature Medicine, vol. 28, no. 1, pp. 31–38, 2022.
3. A. Esteva et al., "A Guide to Deep Learning in Healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
4. C. K. Sidey-Gibbons and J. A. Sidey-Gibbons, "Machine Learning in Medicine: A Practical Introduction," BMC Medical Research Methodology, vol. 19, art. 64, 2019.
5. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
6. B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589–1604, 2018.
7. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
8. C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
9. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
10. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
11. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 25, 2012
12. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
13. R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," Scientific Reports, vol. 6, art. 26094, 2016.
14. A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
15. A. E. W. Johnson et al., "MIMIC-III, a Freely Accessible Critical Care Database," Scientific Data, vol. 3, art. 160035, 2016.
16. V. Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
17. A. Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017
18. S. M. McKinney et al., "International Evaluation of an AI System for Breast Cancer Screening," Nature, vol. 577, no. 7788, pp. 89–94, 2020.
19. D. Ardila et al., "End-to-End Lung Cancer Screening with Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography," Nature Medicine, vol. 25, no. 6, pp. 954–961, 2019
20. Y. Ding et al., "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain," Radiology, vol. 290, no. 2, pp. 456–464, 2019
21. Z. I. Attia et al., "Screening for Cardiac Contractile Dysfunction Using an Artificial Intelligence-Enabled Electrocardiogram," Nature Medicine, vol. 25, no. 1, pp. 70–74, 2019.
22. E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using Recurrent Neural Network Models for Early Detection of Heart Failure Onset," Journal of the

American Medical Informatics Association, vol. 24, no. 2, pp. 361–370, 2017.
23. A. Rajkomar et al., "Scalable and Accurate Deep Learning with Electronic Health Records," npj Digital Medicine, vol. 1, art. 18, 2018.
24. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data?" PLoS ONE, vol. 12, no. 4, e0174944, 2017.
25. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," Computational and Structural Biotechnology Journal, vol. 15, pp. 104–116, 2017.
26. E. Christodoulou et al., "A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models," Journal of Clinical Epidemiology, vol. 110, pp. 12–22, 2019.
27. L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data?" in Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track, vol. 35, 2022.
28. X. Liu et al., "A Comparison of Deep Learning Performance Against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis," The Lancet Digital Health, vol. 1, no. 6, pp. e271–e297, 2019.
29. A. Wong et al., "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients," JAMA Internal Medicine, vol. 181, no. 8, pp. 1065–1070, 2021
30. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," Science, vol. 366, no. 6464, pp. 447–453, 2019
31. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017
32. E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021.
33. N. Rieke et al., "The Future of Digital Health with Federated Learning," npj Digital Medicine, vol. 3, art. 119, 2020.
34. S. Durairaj and V. Vetrivel, "The Effect of AI (Artificial Intelligence) in Employee Performance Evaluation on Employee Retention in the Information Technology Sector," in Proc. International Conference on Digital Transformation in Business: Navigating the New Frontiers Beyond Boundaries, 2024. https://doi.org/10.2991/978-94-6463-433-4_8