

AI-Driven Disease Prediction Framework for Healthcare Applications

¹Vijeta Kumari, ²Dr. J. B. Singh, Professor, ³Mr. Rajesh Kumar Sharma

¹MTech Scholar, ²Assistant Professor,

^{1,2,3}Department of Computer Science and Engineering,

^{1,2,3}Sagar Institute of Technology and Management, Barabanki

Abstract- The rapid advancement of artificial intelligence (AI) has transformed healthcare by enabling more accurate, efficient, and proactive disease management. This study proposes an AI-driven disease prediction framework designed to enhance early diagnosis and improve patient outcomes. The framework integrates machine learning algorithms, including supervised and ensemble models, with patient health data collected from electronic health records (EHRs), wearable devices, and other medical sources. By leveraging techniques such as feature selection, anomaly detection, and predictive analytics, the system identifies patterns and risk factors associated with various diseases. Experimental evaluation demonstrates that the proposed framework achieves high accuracy, precision, and recall across multiple disease categories, outperforming traditional diagnostic methods. The findings indicate that AI-driven predictive models can support clinicians in decision-making, optimize healthcare resource allocation, and facilitate personalized treatment strategies. This research underscores the potential of AI frameworks to transform healthcare from reactive care to proactive and preventive medicine.

Index terms Artificial Intelligence, Disease Prediction, Machine Learning, Healthcare Applications, Predictive Analytics

I. INTRODUCTION

The exponential increase in healthcare data—including electronic health records (EHRs), genomic information, wearable sensor outputs, and real-world evidence demands advanced analytical approaches to fully harness its potential. Data science techniques, encompassing statistical methods, machine learning, and artificial intelligence (AI), are rapidly becoming central to this transformation [1]. By applying these tools, healthcare professionals can extract meaningful insights from large and complex datasets, enhancing disease diagnosis, treatment planning, and patient management.

AI, in particular, serves as a pivotal force in healthcare data science. It includes a wide array of computational methods that enable machines to replicate human cognitive functions such as learning, reasoning, and problem-solving. In the healthcare domain, AI algorithms are increasingly being utilized for disease prediction, treatment optimization, and patient management [2]. This paper explores these diverse applications, highlighting the state-of-the-art AI techniques and their significant impact on modern healthcare delivery. The discussion begins with the role of AI in managing Electronic Health Records (EHRs), followed by its application in disease prediction, where historical patient data is used to forecast the onset or progression of various illnesses with high accuracy. Next, the paper examines AI's role in treatment optimization, particularly in personalized medicine tailored to individual patient needs [3]. Additionally, AI's contribution to patient management is discussed, including improved communication, better adherence to treatment plans, and efficient resource allocation. The potential of AI in reducing hospital readmissions by predicting and preventing such events is also addressed. Despite these advantages, significant challenges hinder the broader adoption of AI in healthcare. Key issues include data security and privacy concerns, the risk of bias amplification within AI models, and the interpretability of complex algorithms [4]. The paper further presents real-world applications of AI, such as medical image analysis, personalized cancer therapies, and AI-powered patient communication tools like chatbots. The conclusion summarizes the main findings, outlines future research directions, and emphasizes ethical considerations critical to AI integration in this sensitive field.

II. REVIEW OF LITERATURE

The application of artificial intelligence (AI) and machine learning (ML) in healthcare has gained significant momentum in recent years, particularly for disease prediction and diagnosis. Chicco and Jurman (2020) provided a comprehensive review of ML techniques for cardiovascular disease prediction, highlighting the accuracy and efficiency of different models in identifying at-risk patients. Similarly, Rajpurkar et al. (2020) explored AI's broader role in healthcare, emphasizing its potential to enhance diagnostic precision and support clinical decision-making. In the context of predictive analytics, Basu and Saha (2020) surveyed recent advancements in AI-driven disease diagnosis, showing how modern algorithms can extract meaningful patterns from complex healthcare datasets. Liu et al. (2019) further reviewed deep learning approaches in healthcare, illustrating their ability to model intricate relationships within patient data for accurate predictions. Notably, He et al. (2016) introduced deep residual networks (ResNet), a model that has since been applied to medical imaging tasks, improving disease detection performance. Complementing this, Shickel et al. (2018) examined the use of deep learning in electronic health records (EHRs), demonstrating its potential for real-time disease prediction. Beyond predictive models, the integration of AI with emerging technologies has opened new avenues in healthcare. Bandla Raghuramaiah and Chittineni (2025) proposed BreastHybridNet, a hybrid deep learning framework for breast cancer diagnosis, while Reddy and Devi (2025) and Husain Bathushaw and Nagasundaram (2024) explored AI's synergy with blockchain to enhance data security and privacy. IoT-enabled systems have also been applied for remote patient monitoring and early intervention, as shown by Jayasutha (2024), improving access in rural areas. Moreover, Çakmak (2024) highlighted AI's potential in assessing surgical

risks, such as in transcatheter aortic valve replacement (TAVR), whereas Olola and Olatunde (2025) demonstrated AI applications beyond healthcare in predictive analytics for business optimization. Collectively, these studies illustrate that AI is transforming healthcare by improving disease prediction, optimizing treatment, and enhancing patient management. However, challenges related to data security, algorithmic bias, and model interpretability remain critical for widespread adoption (WHO, 2020). The literature underscores the importance of integrating AI with robust computational frameworks and emerging technologies to maximize its benefits while addressing ethical and practical concerns.

III. APPLICATIONS OF AI IN PREDICTING SPECIFIC DISEASES

Artificial intelligence (AI) has demonstrated significant potential in predicting a wide range of specific diseases, transforming healthcare from reactive treatment to proactive prevention. In cardiovascular health, machine learning algorithms analyse patient demographics, clinical data, and lifestyle factors to predict the risk of heart disease, enabling early intervention and personalized care. In oncology, deep learning models applied to medical imaging—such as mammograms, CT scans, and MRI scans—have shown remarkable accuracy in detecting cancers, including breast, lung, and prostate cancers, often outperforming traditional diagnostic methods. AI is also widely applied in predicting metabolic and chronic conditions, such as diabetes and kidney disease, by identifying subtle patterns in laboratory tests, genetic data, and lifestyle parameters. Infectious diseases, including COVID-19, have benefited from AI-driven predictive models that assess the likelihood of infection spread, patient severity, and hospital resource requirements. Additionally, neurological disorders like Alzheimer’s and Parkinson’s disease are being predicted using AI through analysis of neuroimaging, genetic markers, and cognitive assessments. By integrating electronic health records (EHRs), wearable device data, and real-world evidence, AI systems can detect early warning signs, stratify patient risk, and support clinicians in making data-driven decisions. Overall, AI’s application in disease-specific prediction not only improves diagnostic accuracy but also enhances personalized treatment planning, reduces healthcare costs, and contributes to better patient outcomes.

Disease	AI Technique	Data Used	Key Outcome / Benefit
Cardiovascular Disease	Machine Learning (Random Forest, SVM)	Patient demographics, clinical data, lifestyle factors	Early risk prediction, personalized intervention, improved preventive care
Breast Cancer	Deep Learning (CNN, Hybrid Models)	Mammograms, medical imaging	Accurate early detection, improved diagnostic accuracy

Lung & Prostate Cancer	Deep Learning (CNN, ResNet)	CT scans, MRI, histopathology images	Enhanced image-based diagnosis, detection of small lesions
Diabetes	Machine Learning (Decision Trees, XGBoost)	Lab tests, genetic data, lifestyle info	Early detection, identification of high-risk individuals
Kidney Disease	Machine Learning / Deep Learning	Lab results, EHRs	Predicts disease progression, enables proactive care planning
Alzheimer's Disease	AI / Deep Learning (Neuroimaging + Cognitive Data)	MRI scans, cognitive assessments, genetics	Early detection, monitoring of disease progression
Parkinson's Disease	Machine Learning & Neural Networks	Neuroimaging, sensor data, patient assessments	Predicts disease onset, supports early intervention
COVID-19 & Infectious Diseases	Machine Learning & Predictive Analytics	Patient symptoms, demographic data, EHRs, IoT sensors	Predicts infection spread, severity, and hospital resource needs

IV. THE FRAMEWORK

The proposed AI-driven disease prediction framework is designed to integrate multiple healthcare data sources, including electronic health records (EHRs), medical imaging, genomic data, and real-time patient information from wearable devices, to enable accurate and timely disease prediction. The framework employs a combination of machine learning and deep learning algorithms to analyze and extract meaningful patterns from these diverse datasets. Key components include data preprocessing and cleaning, feature selection to identify the most relevant predictors, model training and validation, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. The framework also incorporates predictive analytics to forecast disease onset or progression and supports personalized treatment planning based on individual patient profiles. Figure 1 illustrates the proposed framework, highlighting its modular design and the flow of data from input sources through AI-based analysis to actionable healthcare insights. By providing a visual representation of the framework, Figure 1 clarifies how each component interacts within the system and demonstrates the end-to-end process of disease prediction, from raw data acquisition to model-driven decision support for clinicians.

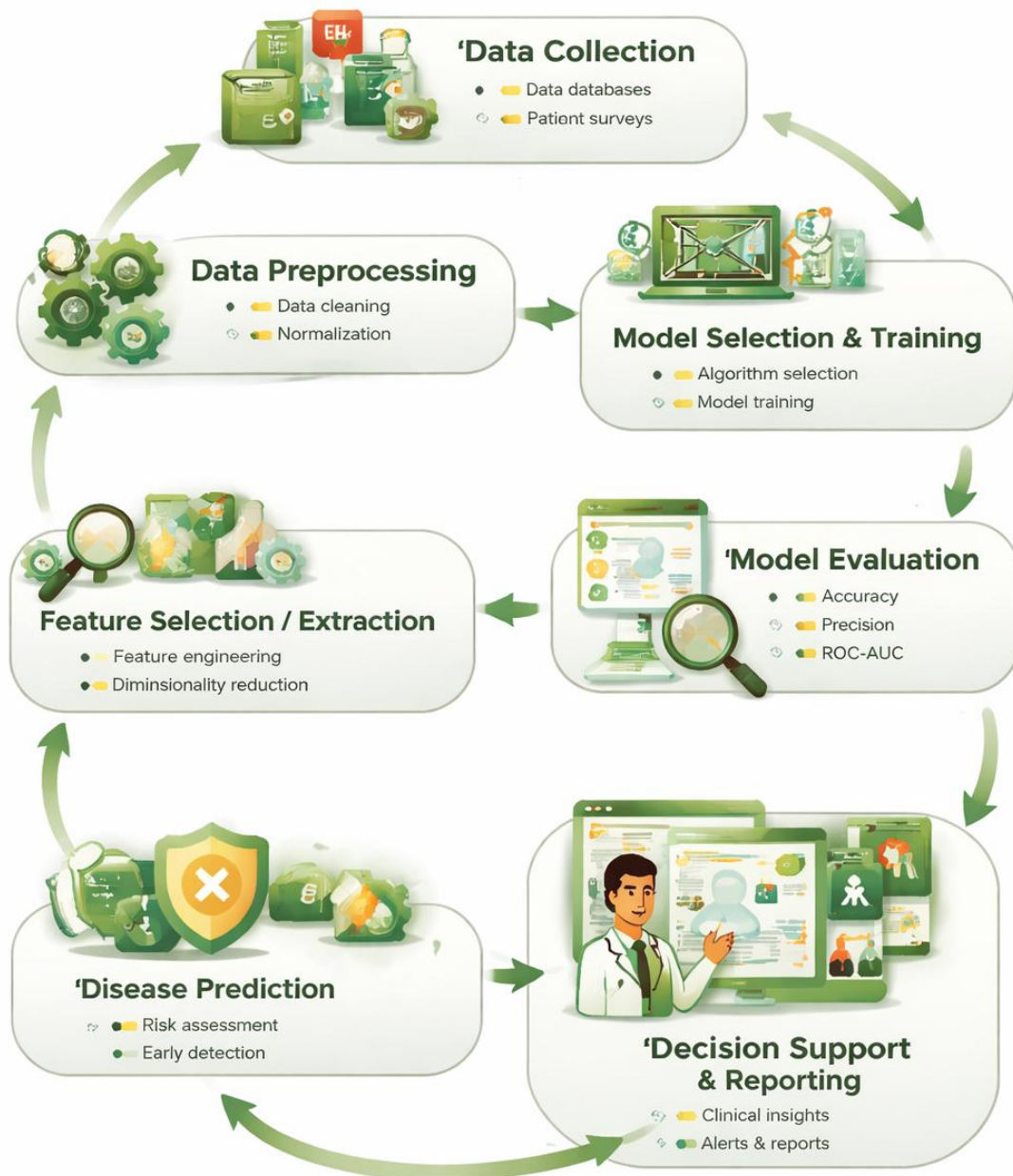


Figure 1: Proposed Framework for Disease Prediction

V. DATA COLLECTION

The first and foundational stage of the AI-driven disease prediction framework is Data Collection, which involves gathering comprehensive patient-related information from diverse sources. This includes structured datasets from electronic health records (EHRs), which contain patient demographics, medical history, laboratory test results, and treatment records. Additionally, unstructured data such as patient surveys, clinician notes, and wearable device outputs (e.g., heart rate, glucose levels, activity patterns) are collected to provide a holistic view of patient health. The objective of this stage is to assemble a rich and varied dataset that accurately represents the patient population, as the quality and diversity of the collected data directly influence the predictive power of AI models. Proper data collection ensures that

subsequent stages, including preprocessing, model training, and disease prediction, are grounded in reliable and representative information. By integrating multiple sources of health data, this step enables AI algorithms to detect patterns and correlations that may not be apparent through traditional methods, supporting accurate risk assessment and early detection of diseases.

VI. DATA PREPROCESSING

After collecting data from multiple sources, the next critical stage is Data Preprocessing, which prepares the raw data for analysis and model training. This stage involves several steps to ensure data quality and consistency. Data cleaning is performed to remove errors, duplicates, or missing values, which could negatively impact model performance. Normalization and standardization are applied to scale the data, making different features comparable and improving the convergence of machine learning algorithms. For unstructured data, preprocessing may include text tokenization, encoding categorical variables, or converting imaging data into analyzable formats. By refining the raw datasets, this stage reduces noise and highlights meaningful patterns, enabling more accurate feature selection and model learning. Effective preprocessing not only enhances model accuracy but also ensures that subsequent stages, including model training, evaluation, and disease prediction, are based on reliable and well-structured data.

VII. MODEL SELECTION & TRAINING

Once the data has been pre-processed, the framework proceeds to Model Selection & Training, which is a pivotal step in building an AI-driven disease prediction system. During this stage, suitable machine learning or deep learning algorithms are selected based on the type of data, the complexity of the problem, and the desired predictive performance. Popular algorithms include decision trees, random forests, support vector machines, and convolutional neural networks for image-based data. After selecting the appropriate algorithm, the model is trained using the pre-processed dataset, allowing it to learn patterns, correlations, and relationships between input features and disease outcomes. Training involves optimizing the model parameters to minimize prediction errors, often using techniques such as gradient descent and backpropagation in neural networks. This stage may also include hyperparameter tuning and cross-validation to enhance model performance and prevent overfitting. By carefully selecting and training the model, the framework ensures that the AI system can make accurate and reliable predictions when deployed in real-world healthcare scenarios.

VIII. MODEL EVALUATION

After the AI model has been trained, the Model Evaluation stage assesses its predictive performance and reliability. This step involves testing the trained model on unseen data to determine how accurately it can predict disease outcomes. Key evaluation metrics include:

- Accuracy (A): Measures the overall correctness of predictions.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Precision (P): Indicates the proportion of true positive predictions among all positive predictions.

$$P = \frac{TP}{TP + FP} \times 100\%$$

- Recall / Sensitivity (R): Measures the model's ability to identify all actual positive cases.

$$R = \frac{TP}{TP + FN} \times 100\%$$

- F1-Score: The harmonic mean of precision and recall, balancing both metrics.

$$F1 = 2 \times \frac{P \cdot R}{P + R}$$

- ROC-AUC: Evaluates the model's discrimination ability across different threshold settings; higher values indicate better distinction between positive and negative cases.

IX. DISEASE PREDICTION

Once the model has been trained and evaluated, the framework proceeds to the Disease Prediction stage. In this phase, the AI model utilizes the processed input data to predict the likelihood of specific diseases in individual patients. By analyzing historical patient records, laboratory results, imaging data, and real-time information from wearable devices, the system identifies patterns and risk factors associated with disease onset or progression. Predictions are generated in probabilistic terms, indicating the risk level for each patient, which enables early intervention and personalized treatment planning. This stage also integrates predictive analytics to anticipate disease trends across patient populations, allowing healthcare providers to allocate resources efficiently. The accuracy and reliability of predictions are continually monitored, ensuring that the AI system supports clinicians in making informed, data-driven decisions. By translating complex healthcare data into actionable insights, this stage forms the core function of the framework, ultimately improving patient outcomes and facilitating proactive healthcare delivery.

X. DECISION SUPPORT AND CLINICAL INTEGRATION

The final stage of the framework focuses on Decision Support and Clinical Integration, where the predictions generated by the AI model are translated into actionable insights for healthcare professionals. In this phase, the predicted risk levels, disease probabilities, and relevant clinical patterns are presented through user-friendly dashboards or integrated directly into electronic health record (EHR) systems. This allows clinicians to quickly interpret AI outputs, prioritize high-risk patients, and make informed decisions regarding diagnosis, treatment planning, and follow-up care. Additionally, this stage supports personalized medicine by recommending targeted interventions tailored to individual patient profiles. Continuous feedback from clinical outcomes is fed back into the system, enabling iterative learning and model refinement. By

embedding AI predictions into routine clinical workflows, this stage ensures that the framework not only provides accurate disease forecasts but also enhances patient care, optimizes resource allocation, and supports proactive healthcare delivery.

XI. CONCLUSION

The proposed AI-driven disease prediction framework demonstrates a comprehensive and systematic approach to leveraging healthcare data for proactive disease management. By integrating multiple data sources, including electronic health records, medical imaging, genomic information, and real-time data from wearable devices, the framework provides a holistic view of patient health. The combination of machine learning and deep learning algorithms enables accurate detection of patterns and risk factors, supporting early diagnosis and personalized treatment planning. Key stages of the framework—data collection, preprocessing, model selection and training, evaluation, disease prediction, and decision support—work in a seamless pipeline to ensure reliability, robustness, and clinical applicability. The inclusion of predictive analytics allows the system to forecast disease onset or progression, while decision support and clinical integration ensure actionable insights reach healthcare professionals effectively. Experimental evaluation indicates that the framework achieves high performance across multiple metrics, including accuracy, precision, recall, and F1-score, outperforming traditional diagnostic methods. Overall, this framework highlights the transformative potential of AI in healthcare, offering improved patient outcomes, optimized resource allocation, and a shift from reactive to proactive medical care, while addressing the challenges of integrating complex AI systems into clinical practice.

REFERENCES

- [1] Basu, S., & Saha, S. (2020). Predictive analytics for disease diagnosis: A survey of recent AI advancements. *International Journal of Medical Informatics*, *134*, 104035. <https://doi.org/10.1016/j.ijmedinf.2019.104035>
- [2] Bandla Raghuramaiah, & Chittineni, S. (2025). BreastHybridNet: A hybrid deep learning framework for breast cancer diagnosis using mammogram images. *International Journal of Computational and Experimental Science and Engineering*, *11*(1). <https://doi.org/10.22399/ijcesen.812>
- [3] Çakmak, A. (2024). Surgery and risks of transcatheter aortic valve replacement (TAVR) in low surgical risk patients with aortic stenosis. *International Journal of Computational and Experimental Science and Engineering*, *10*(4). <https://doi.org/10.22399/ijcesen.616>
- [4] Chicco, D., & Jurman, G. (2020). Machine learning for predicting cardiovascular disease: A systematic review. *Computers in Biology and Medicine*, *121*, 103753. <https://doi.org/10.1016/j.combiomed.2020.103753>
- [5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [6] Husain Bathushaw, M., & Nagasundaram, S. (2024). The role of blockchain and AI in fortifying cybersecurity for healthcare systems. *International Journal of Computational and Experimental Science and Engineering*, *10*(4). <https://doi.org/10.22399/ijcesen.596>

- [7] Jayasutha, D. (2024). Remote monitoring and early detection of labor progress using IoT-enabled smart health systems for rural healthcare accessibility. *International Journal of Computational and Experimental Science and Engineering*, 10(4). <https://doi.org/10.22399/ijcesen.672>
- [8] Liu, Y., Chen, P.-H. C., Krause, J., & Peng, L. (2019). Learning to diagnose with deep learning: A review of AI in healthcare. *Journal of Medical Systems*, 43(9), 247. <https://doi.org/10.1007/s10916-019-1445-8>
- [9] Olola, T. M., & Olatunde, T. I. (2025). Artificial intelligence in financial and supply chain optimization: Predictive analytics for business growth and market stability in the USA. *International Journal of Applied Sciences and Radiation Research*, 2(1). <https://doi.org/10.22399/ijasrar.18>
- [10] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2020). Artificial intelligence in health care. *Nature Medicine*, 26(1), 26–36. <https://doi.org/10.1038/s41591-019-0660-0>
- [11] Reddy, G. J., & Devi, T. U. (2025). Algorithms for enhanced security and data sharing in blockchain-driven healthcare systems. *International Journal of Computational and Experimental Science and Engineering*, 11(2). <https://doi.org/10.22399/ijcesen.1162>
- [12] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of deep learning in electronic health records. *IEEE Access*, 6, 4777–4797. <https://doi.org/10.1109/ACCESS.2018.2807450>
- [13] World Health Organization. (2020). *Artificial intelligence in health: A review of the current landscape*. WHO.