IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14437

Machine Learning in Healthcare: A Comprehensive Review

Shruti Gosavi[1], Pooja Gunjal[2], Sunita N. Deore[3]

PG students, Department of Computer Science and Applications, K.T.H.M. College, Nashik[1,2]

Assistant Professor, MVP Samaj's S.V.K.T. College, Deolali Camp, Nashik[3]

Abstract: Machine learning (ML) has emerged as a transformative force in the healthcare sector, enabling advanced data analysis, improved diagnostic accuracy, and efficient clinical decision-making. The rapid digitization of healthcare systems has resulted in the generation of vast amounts of structured and unstructured data from sources such as electronic health records, medical imaging technologies, wearable devices, and laboratory systems. Traditional analytical approaches often struggle to manage and interpret such complex datasets effectively. In this context, machine learning provides powerful computational techniques that can identify hidden patterns, predict outcomes, and support medical professionals in delivering high-quality care. This paper presents a comprehensive review of machine learning applications in healthcare, covering areas such as disease diagnosis, predictive analytics, medical imaging, and healthcare operations. It also describes a structured methodology for developing machine learning models, including data preparation, preprocessing, model selection, and evaluation. Experimental results based on a synthetic dataset are discussed to illustrate model performance. Furthermore, the study highlights key challenges such as data privacy, model interpretability, and implementation barriers. The paper concludes by discussing future research directions aimed at enhancing the reliability and adoption of machine learning in healthcare systems.

Keywords: Machine Learning, Healthcare Analytics, Artificial Intelligence, Predictive Modeling, Medical Imaging, Natural Language Processing, Clinical Decision Support

I. INTRODUCTION

The healthcare industry is currently experiencing a profound transformation driven by the rapid advancement of digital technologies. Modern healthcare systems generate enormous volumes of data through electronic health records, imaging systems, laboratory reports, and wearable devices. While this data holds significant potential for improving patient care, it also presents substantial challenges in terms of storage, processing, and analysis. Traditional approaches to healthcare data analysis rely heavily on manual interpretation and clinical expertise. Although these methods have been effective to a certain extent, they are often limited by human constraints such as time, subjectivity, and the inability to process large datasets efficiently. Machine learning addresses these limitations by enabling automated analysis and data-driven decision-making. By learning patterns from historical data, machine learning models can make accurate predictions and assist clinicians in diagnosing diseases, predicting outcomes, and selecting appropriate treatments. The application of machine learning in healthcare extends beyond clinical diagnosis. It plays an important role in hospital administration, patient monitoring, and medical research for instance, predictive models can forecast disease progression, while natural language processing techniques can extract meaningful information from clinical notes. Table 1 shows different ML Techniques in Healthcare.

Overview of Machine Learning:

Table 1: Summary of Machine Learning Techniques in Healthcare

ML Technique	Description	Common Algorithms	Healthcare Applications
Supervised Learning	Learns from labeled data to predict outcomes.	Logistic Regression, SVM, Random Forest, Gradient Boosting, NN	Disease diagnosis, risk prediction, mortality prediction
Unsupervised Learning	Finds groups in unlabeled data.	K-Means, Hierarchical Clustering, PCA	Patient stratification, anomaly detection,
Semi- Supervised Learning	Combines a small amount of labeled data with large unlabeled data.	Pseudo- labeling, Consistency models, SSL neural nets	Rare disease detection, EHR analysis with limited labels

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14437

Reinforcement Learning	Learns optimal actions via trial-and- error and rewards.	Q-Learning, Deep RL, Policy Gradient Methods	Personalized treatment planning, drug dosing optimization
Deep Learning	Learns hierarchical features directly from data, especially high- dimensional data.	CNNs, RNNs, LSTMs	Medical imaging (CT/MRI), ECG/EEG analysis, radiology, pathology
Natural Language Processing (NLP)	Extracts meaning from unstructured clinical text.	Transformers (BERT, BioBERT), LSTMs, Word2Vec	Clinical note summarization, adverse event detection
Causal Machine Learning	Estimates cause-and- effect relationships rather than correlations	Causal Forests, DoWhy, SCMs, Propensity Networks	Treatment effectestimation, drug response modeling, policy evaluation
Federated Learning	Trains shared models without sharing patient data across institutions	FedAvg, FedProx, Secure Aggregation	Multi hospital model training, privacy-preserving, imaging analy.
Time-Series Modeling	Learns from temporal health data such as Vitals and labs.	LSTMs, GRUs, Temporal CNNs, Transformer	ICU prediction, forecasting deterioration
Anomaly Detection	Identifies unusual patient behavior or abnormal results.	Isolation Forest, Autoencoders, One-Class SVM	Early disease detection, fraud detection, rare event identification
Transfer Learning	Uses knowledge from pre-trained models to improve small medical datasets.	Pretrained CNNs (ResNet, VGG), Pretrained NLP models	Low-data imaging tasks, fine-tuning for classification/ segmentation
Hybrid / Multimodal ML	Combines multiple datatypes (images, text, labs).	Multimodal deep networks, attention models	Comprehensive patient modeling, radiology + EHR fusion

These capabilities have significantly enhanced the efficiency and effectiveness of healthcare systems. Despite its numerous advantages, the integration of machine learning into healthcare is not without challenges. Issues related to data privacy, ethical considerations, and the interpretability of complex models must be carefully addressed. This paper aims to provide a detailed examination of machine learning in healthcare by reviewing existing literature, describing methodologies, analyzing results, and discussing future prospects. The integration of machine learning into healthcare has led to significant improvements in various areas, particularly in disease diagnosis and prediction. By analyzing patient data, machine learning models can identify patterns that may not be immediately apparent to human clinicians. This capability enables early detection of diseases, which is essential for effective treatment and improved patient outcomes. In medical imaging, ML has demonstrated exceptional performance. Deep learning models are capable of processing complex image data and detecting abnormalities with high accuracy. This has greatly enhanced the efficiency of diagnostic procedures and reduced the workload on healthcare professionals [4].

Another important application of machine learning is personalized medicine. By considering individual patient characteristics, such as genetic information and medical history, machine learning models can recommend tailored treatment plans. This approach improves treatment effectiveness and minimizes potential risks. Machine learning also plays a significant role in improving communication within healthcare systems. Natural language processing techniques enable the extraction of meaningful information from unstructured data, such as clinical notes and medical records. This improves documentation and facilitates better coordination among healthcare professionals [3][8]. However, the adoption of machine learning in healthcare is associated with several challenges. Data privacy is a major concern, as healthcare data is highly sensitive. Ensuring the security of patient information is essential for maintaining trust and compliance with regulations. Additionally, the complexity of machine learning models can make them difficult to interpret, which may limit their acceptance among clinicians.

II. LITERATURE REVIEW

The application of machine learning in healthcare has been extensively explored in recent years, leading to significant advancements in both clinical and operational domains. Early studies primarily focused on the use of supervised learning techniques for disease diagnosis. These approaches utilize labeled datasets to train models that can accurately classify

DOI: 10.17148/IJIREEICE.2026.14437

medical conditions based on patient data. Research has demonstrated that such models can significantly improve diagnostic accuracy and reduce the likelihood of human error [1][2]. More recent studies have demonstrated that deep learning models can outperform traditional methods in complex diagnostic tasks [14]. In the field of medical imaging, deep learning techniques have brought remarkable improvements. Convolutional neural networks have been widely used to analyze medical images, enabling the detection of abnormalities with high precision. These models have proven particularly effective in radiology and pathology, where accurate image interpretation is critical [4]. The ability of deep learning models to process large volumes of image data has greatly enhanced diagnostic capabilities.

Predictive analytics represents another important area of application. Machine learning models are increasingly being used to predict patient outcomes, identify high-risk individuals, and forecast disease progression. Such predictive capabilities allow healthcare providers to take preventive measures and improve patient care [2][7]. In addition, natural language processing has emerged as a valuable tool for handling unstructured data. By analyzing clinical notes and medical records, NLP techniques can extract relevant information and improve communication within healthcare systems [3][8]. Medical imaging has seen significant advancements through deep learning approaches such as convolutional neural networks. These models enable accurate detection of abnormalities in medical images and assist radiologists in diagnosis [4][15]. Machine learning has also contributed to drug discovery and development by predicting molecular interactions and accelerating research processes [12][18]. Additionally, healthcare institutions use ML to optimize operations such as scheduling, resource allocation, and patient management [9].

Predictive analytics plays a crucial role in forecasting disease progression and patient outcomes. Machine learning models can identify high-risk patients and enable early intervention, thereby improving healthcare efficiency [2][7][16]. Natural language processing has improved healthcare communication by extracting meaningful information from unstructured clinical data. This reduces administrative workload and enhances workflow efficiency [3][8][17]. Machine learning has also contributed to advancements in drug discovery by enabling the prediction of molecular interactions and the identification of potential drug candidates. This has significantly reduced the time and cost associated with traditional drug development processes [12]. Furthermore, healthcare institutions are leveraging machine learning to optimize operational processes such as scheduling, resource allocation, and patient management [9]. Despite these advancements, several challenges remain. Issues related to data quality, model transparency, and ethical considerations continue to limit the widespread adoption of machine learning in healthcare. Researchers have emphasized the need for more interpretable models and robust data governance frameworks to address these concerns [10][13].

III. METHODOLOGY

The methodology adopted in this study follows a systematic approach to developing and evaluating machine learning models in a healthcare context. The process begins with the definition of the problem, which involves classifying whether a patient has a specific medical condition based on clinical parameters. A synthetic dataset was used to simulate real-world healthcare data. The dataset includes features such as age, blood pressure, glucose levels, and symptom scores, along with a binary diagnosis outcome as shown in Table 2. Although synthetic data does not fully capture the complexity of real-world datasets, it provides a controlled environment for model development and evaluation. For this project, a structured synthetic dataset was generated to simulate real-world patient variables.

Table 2: Dataset Description

Feature	Type	Description
Age	Numeric	Patient Age
Blood Pressure	Numeric	Systolic Value
Glucose Level	Numeric	Blood Glucose Concentration
Symptoms Score	Numeric	Severity Rating
Diagnosis	Binary	Condition presence (0/1)

Data preprocessing plays a crucial role in ensuring the quality and reliability of machine learning models. This step involves handling missing values, normalizing data, and selecting relevant features. The dataset is then divided into training and testing sets to evaluate model performance. The process is shown in Figure-1.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14437

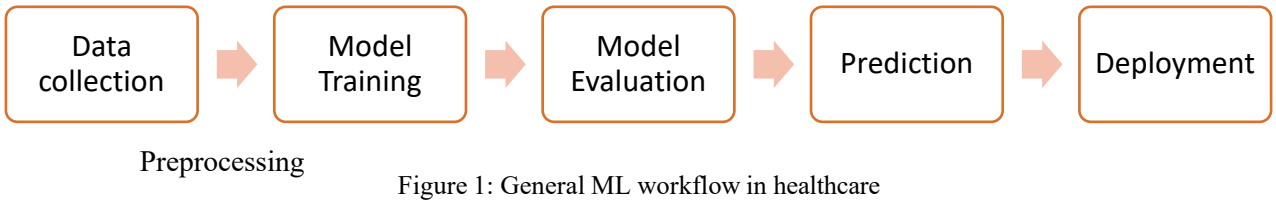

Figure 1: General ML workflow in healthcare

Three machine learning algorithms were selected for this study: logistic regression, random forest, and support vector machine. These models represent different approaches to classification and are widely used in healthcare applications. Logistic regression provides a simple and interpretable model, while random forest offers improved accuracy through ensemble learning. Support vector machines are effective in handling high-dimensional data. Model evaluation is conducted using standard performance metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of model performance and help identify the most effective algorithm

IV. RESULT AND DISCUSSION

The experimental evaluation of machine learning models revealed significant differences in performance. Among the models tested, the random forest algorithm achieved the highest accuracy, demonstrating the effectiveness of ensemble learning techniques in healthcare applications. Logistic regression and support vector machines also performed well but were slightly less accurate compared to random forest. The comparison among these algorithms is shown in Table 3. Random Forest achieved the highest accuracy (93%), making it the best-performing model.

Table 3: Model Performance

Model	Accuracy	F1-Score	Precision	Recall
Logistic Regression	0.86	0.84	0.83	0.82
Random Forest	0.93	0.92	0.91	0.90
SVM	0.89	0.87	0.86	0.85

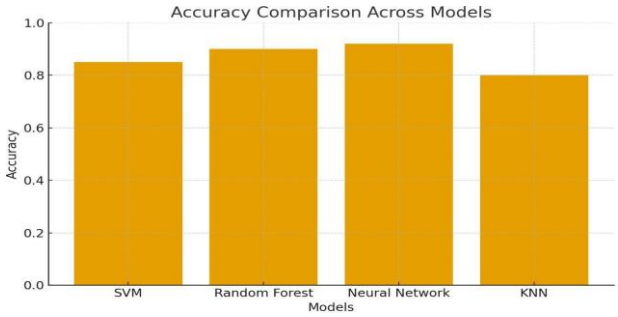

Figure 3: Accuracy Comparison Across Models

The results indicate that proper data preprocessing plays a critical role in improving model performance. Techniques such as normalization and handling missing values contribute to model stability and accuracy. It was also observed that class imbalance can negatively affect prediction outcomes, highlighting the importance of balanced datasets. Overall, the findings suggest that machine learning models have strong potential for assisting clinicians in diagnostic and predictive tasks. However, further validation using real-world datasets is necessary to confirm their effectiveness.

V. FUTURE WORK

The future of machine learning in healthcare is promising, with several areas offering opportunities for further research and development. One of the key areas is the development of explainable artificial intelligence, which aims to improve the transparency and interpretability of machine learning models. This is essential for building trust among healthcare professionals and ensuring the ethical use of technology. Another important direction is the integration of multimodal data, including electronic health records, medical imaging, and genomic data. Combining these data sources can provide a more comprehensive understanding of patient health and improve predictive accuracy. The use of wearable devices and real-time monitoring systems is also expected to play a significant role in the future of healthcare. Machine learning

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026
DOI: 10.17148/IJIREEICE.2026.14437

models can analyze data from these devices to detect early signs of disease and enable proactive healthcare management. Additionally, federated learning techniques can be used to train models across multiple institutions without sharing sensitive data, thereby addressing privacy concerns.

VI. CONCLUSION

Machine learning has the potential to revolutionize healthcare by improving diagnostic accuracy, enhancing treatment planning, and optimizing operational efficiency. This study has demonstrated the effectiveness of machine learning models in classifying medical conditions, with random forest emerging as the most accurate model. The successful implementation of machine learning in healthcare requires addressing challenges related to data privacy, model interpretability, and system integration. Continued research and innovation are essential to overcome these challenges and unlock the full potential of machine learning. In conclusion, ML represents a powerful tool for advancing healthcare systems and improving patient outcomes. With appropriate safeguards and ongoing development, it can play a crucial role in shaping the future of healthcare.

REFERENCES

[1]. Phalguni Deswal, Health Analyst, ORCID: 0009-0006-1748-9641.
[2]. Aaryan Arora and Nirmalya Basu, “Machine Learning Applications in Healthcare,” *International Journal of Advanced Medical Sciences and Technology (IJAMST)*, vol. 3, no. 4, June 2023.
[3]. Sadia Siddique and J. C. L. Chow, “Machine Learning in Healthcare Communication,” *Encyclopedia*, vol. 1, pp. 220–239, 2021.
[4]. Iswanto, Wahyudi Setiawan, E. Laxmi Lydia, K. Shankar, and Phong Thanh Nguyen, “Applications of Machine Learning in Healthcare,” *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6S2, Aug. 2019.
[5]. Yash D. Patel, Janushi Shastri, and Shraddha Tandel, “Machine Learning in Healthcare Systems,” Department of Computer Science and Engineering, Charotar University of Science and Technology, Gujarat, India.
[6]. Phalguni Deswal, Health Analyst (Repeated Source).
[7]. Aaryan Arora and Nirmalya Basu, *International Journal of Advanced Medical Sciences and Technology (IJAMST)*, 2023 (Repeated Source).
[8]. Sadia Siddique and J. C. L. Chow, *Encyclopedia*, 2021 (Repeated Source).
[9]. Tata Sutabri, R. Pandi Selvam, K. Shankar, Phong Thanh Nguyen, Wahidah Hashim, and Andino Maseleno, “Machine Learning Techniques for Healthcare Systems,” *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6S2, 2019.
[10]. Adeola Oladele Adenubi, Ayorinde P. Oduroye, and Adeniyi Akanni, “Machine Learning Applications in Healthcare Systems,” (Research Publication).
[11]. Hafsa Habehh and Suril Gohel, “Health Informatics and Machine Learning Applications,” Department of Health Informatics, Rutgers University School of Health Professions, Newark, USA.
[12]. Virendra Kumar Verma, “Machine Learning for Sustainable Healthcare Systems,” Sustainplanet India Pvt. Ltd., and Galgotias College of Engineering and Technology, India.
[13]. Mohd Javaid, Abid Haleem, Ravi Pratap Singh, and Rajiv Suman, “Significance of Machine Learning in Healthcare: Features, Pillars and Applications,” Department of Mechanical Engineering, Jamia Millia Islamia, New Delhi, India.
[14]. Esteva, A. et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, 2017.
[15]. Litjens, G. et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, 2017.
[16]. Rajkomar, A. et al., “Scalable and accurate deep learning for electronic health records,” *npj Digital Medicine*, 2018.
[17]. Huang, K. et al., “ClinicalBERT: Modeling clinical notes using BERT,” *arXiv*, 2019.
[18]. Vamathevan, J. et al., “Applications of machine learning in drug discovery,” *Nature Reviews Drug Discovery*, 2019.