International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

A Systematic Review of Machine Learning Techniques for Disease Classification in Healthcare Systems

Priyanka Kushwah
Research Scholar, Department of Computer Science, Malwanchal University, Indore
Dr. Rajesh. D
Supervisor, Department of Computer Science, Malwanchal University, Indore

ABSTRACT

The increasing availability of healthcare data and advancements in computational technologies have accelerated the adoption of machine learning (ML) techniques for disease classification. This systematic review examines the application of various ML approaches in healthcare systems, focusing on their effectiveness in improving diagnostic accuracy and clinical decision-making. The study analyzes a wide range of algorithms, including supervised learning methods such as decision trees, support vector machines, and random forests, as well as deep learning models like artificial neural networks and convolutional neural networks. The review highlights how these techniques are applied across multiple disease domains, including cardiovascular diseases, cancer, diabetes, neurological disorders, and infectious diseases.

Furthermore, the paper critically evaluates the performance of these models based on key metrics such as accuracy, precision, recall, and computational efficiency. It also identifies major challenges, including data quality issues, class imbalance, lack of interpretability, and ethical concerns related to privacy and bias. The findings suggest that while machine learning significantly enhances disease classification capabilities, its integration into real-world healthcare systems requires improved transparency, standardized evaluation frameworks, and interdisciplinary collaboration. The study concludes by outlining future research directions, emphasizing the importance of explainable artificial intelligence and hybrid models to ensure reliable and scalable healthcare solutions.

Keywords:: Machine Learning, Disease Classification, Healthcare Systems, Deep Learning, Clinical Decision Support

1. INTRODUCTION

The integration of machine learning (ML) techniques into healthcare systems has transformed the landscape of disease classification, offering unprecedented opportunities for early diagnosis, personalized treatment, and improved patient outcomes. Disease classification, a critical component of clinical decision-making, involves categorizing medical conditions based on symptoms, diagnostic tests, and patient data. Traditionally, this process has relied heavily on physician expertise and rule-based systems, which, although effective, are often limited by subjectivity, variability, and scalability challenges. With the exponential growth of healthcare data generated from electronic health records (EHRs), medical imaging, wearable devices, and genomic sequencing, conventional approaches are increasingly inadequate to handle the complexity and volume of such data. Machine learning, a subset of artificial intelligence, addresses these limitations by enabling systems to learn patterns from large datasets and make data-driven predictions with high accuracy. Techniques such as supervised learning, unsupervised learning, and deep learning have been widely applied to classify diseases ranging from chronic conditions like diabetes and cardiovascular diseases to complex disorders such as cancer and neurological illnesses. These models not only enhance diagnostic precision but also facilitate early detection, thereby reducing mortality rates and healthcare costs.

Despite these advancements, the application of machine learning in disease classification presents several methodological and practical challenges that warrant systematic examination. Variability in data quality, class imbalance, lack of standardized datasets, and issues related to model interpretability often hinder the reliability and clinical adoption of ML-based systems. Furthermore, ethical concerns such as data privacy, bias in algorithmic decision-making, and transparency remain critical barriers in integrating these technologies into real-world healthcare settings. A systematic review of machine learning techniques for disease classification is therefore essential to synthesize existing research, evaluate the effectiveness of different algorithms, and identify gaps in current knowledge. Such a review enables a comparative analysis of models including decision trees, support vector machines, random forests, and neural networks in terms of accuracy, computational efficiency, and clinical applicability. Additionally, it provides insights into emerging trends such as hybrid models, explainable AI, and the use of big data analytics in healthcare. By critically analyzing the existing body of literature, this study aims to offer a comprehensive understanding of how machine learning techniques are reshaping disease classification and to propose future directions for research and practical implementation in healthcare systems.

2. BACKGROUND OF DISEASE CLASSIFICATION IN HEALTHCARE

Disease classification is a fundamental component of healthcare systems, forming the basis for accurate diagnosis, treatment planning, and epidemiological analysis. It involves the systematic categorization of diseases based on clinical symptoms, diagnostic findings, pathological characteristics, and underlying causes. Historically, disease classification has relied on standardized frameworks such as the International Classification of Diseases (ICD), which provides a globally accepted structure for recording and reporting diseases. These traditional classification systems have enabled healthcare professionals to maintain consistency in diagnosis and facilitate communication across medical institutions. However, conventional approaches are

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

largely dependent on manual interpretation and clinician expertise, which can introduce variability and potential diagnostic errors, particularly in complex or rare conditions. With the increasing burden of chronic diseases and the growing diversity of medical data, the limitations of traditional classification methods have become more evident, necessitating more efficient and scalable solutions.

In recent years, the rapid digitization of healthcare through electronic health records, medical imaging, and wearable technologies has significantly expanded the volume and complexity of patient data. This evolution has paved the way for the adoption of advanced computational techniques, particularly machine learning, to enhance disease classification processes. Machine learning algorithms can analyze large and heterogeneous datasets to identify hidden patterns and correlations that may not be easily detectable by human clinicians. These techniques have demonstrated substantial potential in improving diagnostic accuracy, reducing time for disease detection, and supporting clinical decision-making. Furthermore, the integration of data-driven approaches with traditional medical knowledge has led to the development of intelligent healthcare systems capable of real-time disease monitoring and prediction. As healthcare continues to move toward precision medicine, disease classification is increasingly becoming a dynamic, data-centric process, highlighting the need for continuous research and innovation in this domain.

3. APPLICATIONS OF MACHINE LEARNING IN DISEASE CLASSIFICATION

1. Chronic Disease Classification

Machine learning techniques have been extensively applied in the classification of chronic diseases such as cardiovascular disorders, diabetes, and hypertension. These conditions typically involve multiple risk factors and long-term progression, making them suitable for predictive modeling. ML algorithms like decision trees, support vector machines (SVM), and random forests analyze structured clinical data—including blood pressure, cholesterol levels, glucose readings, and patient history—to classify disease presence and severity. For instance, in cardiovascular disease classification, ML models process electrocardiogram (ECG) signals and clinical parameters to detect abnormalities at early stages. Similarly, in diabetes, predictive models classify patients into risk categories, enabling preventive interventions. These applications improve early diagnosis, reduce complications, and support continuous patient monitoring, thereby enhancing overall healthcare efficiency.

2. Medical Imaging-Based Disease Classification

A significant advancement in disease classification has been achieved through the use of deep learning techniques in medical imaging. Convolutional neural networks (CNNs), a subset of deep learning, are particularly effective in analyzing complex visual data such as X-rays, magnetic resonance imaging (MRI), computed tomography (CT) scans, and histopathological images.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

These models automatically extract features and classify diseases such as cancer, tumors, lung infections, and retinal disorders with high accuracy. In oncology, ML-based systems assist in detecting and classifying tumor types, stages, and malignancy levels, which is crucial for treatment planning. Similarly, in neurology, ML models analyze brain imaging data to classify disorders like Alzheimer's disease and Parkinson's disease. The integration of ML in medical imaging reduces human error, speeds up diagnosis, and enables large-scale screening programs.

3. Infectious and Emerging Disease Classification

Machine learning has also proven highly effective in the classification of infectious and emerging diseases by utilizing both structured and unstructured healthcare data. During outbreaks, ML models can analyze clinical symptoms, laboratory results, and radiological data to rapidly classify diseases and predict their spread. For example, ML techniques were widely used in identifying and classifying COVID-19 cases using chest X-rays and CT scans, as well as symptom-based datasets. Additionally, natural language processing (NLP) methods are applied to clinical notes and health records to extract relevant information for disease classification. These systems support real-time surveillance, early detection, and efficient response to public health emergencies. As a result, machine learning plays a critical role in strengthening healthcare systems' ability to manage infectious diseases and respond to future pandemics.

4. LITERATURE REVIEW

The application of machine learning (ML) in healthcare has evolved significantly over the past decade, transforming traditional diagnostic paradigms into data-driven, predictive systems. Early foundational work by Rohit C. Deo (2015) established the conceptual basis for integrating ML into clinical practice, highlighting its ability to process high-dimensional medical data and uncover latent patterns that are not readily apparent through conventional statistical methods. This work emphasized that ML algorithms can support clinical decision-making by improving diagnostic accuracy and enabling personalized treatment strategies. Similarly, Andrew L. Beam and Isaac S. Kohane (2018) discussed the convergence of big data and ML in healthcare, noting that the rapid growth of electronic health records (EHRs) and biomedical datasets has created an ecosystem conducive to advanced analytical techniques. Their study underscored the importance of scalable ML models in handling heterogeneous healthcare data and improving disease classification outcomes.

Subsequent research has focused on the application of specific ML techniques for disease classification across various medical domains. Konstantina Kourou et al. (2015) provided one of the earliest comprehensive analyses of ML applications in cancer prognosis and prediction, demonstrating how algorithms such as support vector machines (SVM), artificial neural networks (ANN), and decision trees can classify cancer types with high accuracy. This was further expanded by Geert Litjens et al. (2017), who conducted a seminal survey on deep

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

learning in medical image analysis. Their study highlighted the effectiveness of convolutional neural networks (CNNs) in extracting hierarchical features from imaging data, significantly improving the classification of diseases such as tumors and lesions. In a landmark contribution, Andre Esteva et al. (2017) demonstrated that deep neural networks could achieve dermatologist-level accuracy in skin cancer classification, marking a critical milestone in the clinical applicability of AI.

The integration of deep learning into healthcare systems has further been explored by Riccardo Miotto et al. (2018), who reviewed the opportunities and challenges associated with deep learning in healthcare. Their findings indicated that deep learning models are particularly effective in processing unstructured data such as clinical notes and medical images, although issues related to interpretability and data quality persist. Complementing this, Benjamin Shickel et al. (2018) examined the application of deep learning techniques to EHR data, introducing the concept of "Deep EHR" systems capable of capturing temporal patterns in patient records for disease classification. These studies collectively highlight the shift from traditional ML approaches to more sophisticated deep learning architectures capable of handling complex healthcare datasets.

Research has also emphasized the practical implications and clinical integration of ML-based disease classification systems. Jenna A. M. Sidey-Gibbons and Christopher J. Sidey-Gibbons (2019) provided a practical introduction to ML in medicine, focusing on its real-world applications and limitations. Their work highlighted the need for robust validation frameworks and clinician involvement to ensure the reliability of ML models. Similarly, Eric J. Topol (2019) introduced the concept of high-performance medicine, emphasizing the synergy between human expertise and artificial intelligence. Topol argued that ML should augment, rather than replace, clinical decision-making, thereby enhancing the quality of care.

In the context of disease-specific applications, Muntasir I. Al-Janabi et al. (2018) reviewed ML techniques for heart disease prediction, demonstrating the effectiveness of classification algorithms in identifying cardiovascular risk factors. Their findings indicated that ensemble methods such as random forests often outperform single classifiers in terms of accuracy and robustness. This aligns with broader research trends emphasizing the importance of hybrid and ensemble models in improving disease classification performance. Additionally, Ajay Rajkomar et al. (2019) highlighted the transformative potential of ML in clinical medicine, particularly in automating diagnostic processes and enabling precision medicine.

Recent studies have shifted toward comprehensive and systematic analyses of ML applications in healthcare. Rohit R. Chandan et al. (2023) conducted an extensive review of ML techniques for disease diagnosis and prognosis, identifying key trends such as the increasing use of deep learning and the integration of multimodal data sources. Similarly, Katarzyna Kolasa and

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Grzegorz Kozinski (2023) examined existing systematic reviews in the field, providing a meta-level understanding of research trends and methodological approaches. Their work highlighted the growing emphasis on evidence synthesis and the need for standardized evaluation metrics in ML-based healthcare studies.

Furthermore, Rashedul Islam et al. (2023) focused on chronic disease prediction, demonstrating the effectiveness of ML algorithms in managing long-term health conditions. Their study emphasized the importance of feature selection, data preprocessing, and model optimization in achieving high classification accuracy. Complementing this, Qiang An et al. (2023) provided a comprehensive review of ML applications in healthcare, highlighting emerging trends such as explainable AI (XAI), federated learning, and real-time disease monitoring systems.

The literature indicates that machine learning has significantly enhanced disease classification capabilities across various healthcare domains. From traditional algorithms to advanced deep learning models, ML techniques have demonstrated their ability to improve diagnostic accuracy, reduce healthcare costs, and support clinical decision-making. However, challenges such as data heterogeneity, model interpretability, and ethical considerations remain critical areas for future research. The reviewed studies collectively underscore the need for interdisciplinary collaboration, robust validation frameworks, and the integration of domain knowledge to ensure the successful implementation of ML-based disease classification systems in real-world healthcare settings.

5. RESEARCH PROBLEM

The rapid advancement of machine learning (ML) techniques has created significant opportunities for improving disease classification in healthcare systems; however, several critical challenges limit their effective implementation and clinical adoption. One of the primary research problems lies in the heterogeneity and quality of healthcare data, which often includes missing values, noise, and class imbalance. These issues can negatively impact the performance, reliability, and generalizability of ML models across diverse patient populations. Additionally, while complex models such as deep learning architectures demonstrate high accuracy, they often lack interpretability, making it difficult for healthcare professionals to trust and validate their predictions in clinical settings.

Another major concern is the absence of standardized frameworks for evaluating and comparing ML models in disease classification tasks. Variations in datasets, evaluation metrics, and experimental designs make it challenging to identify the most effective techniques. Furthermore, ethical and legal issues, including data privacy, security, and algorithmic bias, pose significant barriers to real-world deployment. Despite numerous studies exploring ML applications in healthcare, there remains a gap in systematically analyzing these challenges and identifying scalable, transparent, and clinically applicable solutions. Therefore, this research seeks to address

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

these limitations by critically reviewing existing ML techniques and highlighting areas for improvement in disease classification systems.

6. DISCUSSION

The reviewed literature demonstrates that machine learning (ML) has emerged as a transformative tool in disease classification, significantly enhancing diagnostic accuracy, efficiency, and scalability in healthcare systems. Across various studies, supervised learning algorithms such as support vector machines, decision trees, and random forests have shown strong performance in handling structured clinical datasets, particularly for chronic disease prediction. At the same time, deep learning techniques—especially convolutional neural networks—have proven highly effective in analyzing unstructured data such as medical images and electronic health records. This dual capability highlights the versatility of ML approaches in addressing diverse healthcare challenges. Furthermore, the integration of ML into clinical decision support systems has enabled early disease detection and personalized treatment planning, which are critical for improving patient outcomes and reducing healthcare costs.

However, despite these advancements, several challenges persist that limit the widespread adoption of ML-based disease classification systems. A major concern is the quality and availability of healthcare data, as incomplete, noisy, or imbalanced datasets can significantly affect model performance. Additionally, the lack of interpretability in complex models, particularly deep learning systems, poses a barrier to clinical trust and acceptance. Healthcare professionals often require transparent and explainable models to validate diagnostic decisions. Ethical and legal concerns, including data privacy, algorithmic bias, and regulatory compliance, further complicate the implementation of ML in real-world settings. Moreover, many studies lack standardized evaluation metrics, making it difficult to compare model performance across different applications. Therefore, while ML offers substantial potential, its successful integration into healthcare systems requires addressing these technical, ethical, and operational challenges through interdisciplinary collaboration and robust validation frameworks.

7. CONCLUSION

Machine learning techniques have significantly advanced the field of disease classification in healthcare systems by enabling accurate, efficient, and data-driven diagnostic processes. The literature reveals that both traditional machine learning algorithms and advanced deep learning models play a crucial role in analyzing complex medical data and improving disease detection across various domains, including cardiovascular diseases, cancer, and neurological disorders. These technologies contribute to early diagnosis, personalized treatment, and enhanced clinical decision-making, ultimately leading to better patient outcomes and optimized healthcare delivery.

Despite these promising developments, challenges such as data quality issues, lack of model interpretability, and ethical concerns must be carefully addressed to ensure the reliability and acceptance of ML-based systems in clinical practice. Future research should focus on developing explainable and transparent models, improving data standardization, and integrating domain knowledge with advanced computational techniques. Additionally, the adoption of emerging approaches such as explainable artificial intelligence and hybrid modeling can further strengthen disease classification systems. Overall, machine learning holds immense potential to revolutionize healthcare, but its success depends on responsible implementation, continuous evaluation, and collaboration between healthcare professionals and technology experts.

REFERENCES

1. Al-Janabi, M. I., Qutqut, M. H., & Hijjawi, M. (2018). Machine learning classification techniques for heart disease prediction: A review. International Journal of Engineering & Technology, 7(4), 5373–5379.
2. Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930.
3. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.
4. 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.
5. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.
6. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246.
7. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.
8. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.
9. Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: A practical introduction. BMC Medical Research Methodology, 19(1), 64.
10. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

11. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.
12. Chandan, R. R., Singh, J., Ravi, V., Shivahare, B. D., Alahmadi, T. J., Singh, P., & Diwakar, M. (2023). Reviewing the impact of machine learning on disease diagnosis and prognosis: A comprehensive analysis. Current Medical Imaging.
13. Kolasa, K., & Kozinski, G. (2023). Systematic reviews of machine learning in healthcare: A review of review studies. Expert Review of Pharmacoeconomics & Outcomes Research.
14. Islam, R., et al. (2023). A comprehensive review for chronic disease prediction using machine learning algorithms. BMC Medical Informatics and Decision Making.
15. An, Q., et al. (2023). A comprehensive review on machine learning in healthcare applications. Healthcare Analytics Journal.