From Data to Decisions: Machine Learning in Medical Informatics

Furkan ŞAKİROĞLU[1*] **Review Article**

[1] İbni Sina Family Health Center, Erzurum, Türkiye

Received: December 27, 2024 **Accepted:** December 30, 2024 **Published:** December 31, 2024. **DOI: 10.5281/zenodo.14582709**
***Corresponding author:** Furkan Şakiroğlu, İbni Sina Family Health Center, Erzurum, Türkiye. Email: fsakiroglu@gmail.com
Cite This Article as: Sakiroglu F. From Data to Decisions: Machine Learning in Medical Informatics. Essentials Frontline Med J. 2024:1(1); 35-38.
Externally Peer-Reviewed.

ABSTRACT

Objective: Medical informatics is an interdisciplinary domain that integrates information technology to enhance the quality and efficiency of healthcare services. Recent advancements in artificial intelligence (AI) and its subset, machine learning (ML), have significantly influenced this field. This article aims to examine the role of machine learning in medical informatics, focusing on its applications, benefits, limitations, and future potential, particularly in improving healthcare delivery and decision-making processes.

Methods: Machine learning, a branch of AI, enables computational systems to predict outcomes or make decisions based on data without explicit programming. Its implementation in medicine spans various domains, including disease diagnosis, personalized treatment planning, clinical decision support systems, epidemiological research, and clinical trials.

Results: Machine learning contributes to faster and more accurate diagnoses, facilitates early disease detection, alleviates healthcare professionals' workload, and optimizes resource utilization. Despite these benefits, its application faces challenges such as data privacy concerns, ethical dilemmas, and technical constraints.

Conclusion: The adoption of machine learning in medical informatics has the potential to revolutionize healthcare systems. While challenges remain, advancements in this technology are expected to lead to more widespread and impactful applications in the future.

Keywords: Medical informatics, machine learning, artificial intelligence

INTRODUCTION

Medical informatics is an interdisciplinary field that uses information technology to improve the quality and efficiency of healthcare services. In recent years, artificial intelligence and its subfield, machine learning, have started to play a significant role in this area (1). Machine learning is a subfield of artificial intelligence that enables computers or software to make predictions or decisions based on new, unknown data without direct programming. It allows systems to learn from their experiences through statistical models and algorithms. Machine learning has become an effective tool in overcoming many challenges in the medical field due to its ability to extract information from large datasets, recognize patterns, and make predictions (2).

MACHINE LEARNING MODELS

In supervised learning, the model is trained using data with known correct answers, while in unsupervised learning, the model is trained with unlabeled data to discover the structure and patterns within the data. In reinforcement learning, the model learns by interacting with its environment and receiving feedback from its outcomes (3).

1. Supervised Learning Models

Logistic Regression: This is a linear model used for binary classification. It is suitable for predicting the probabilities between two classes (e.g., yes/no, success/failure). The model computes the probability of an event happening using input features, typically providing an output value between 0 and 1 (probability). It can be used for disease prediction.

Decision Trees: A decision tree classifies or regresses by branching data according to its features. Each branch represents a decision based on a feature. This model allows the data to be divided meaningfully by identifying significant split points and is easily interpretable and visually representable (4).

Random Forest: Random Forest is an ensemble method made by combining multiple decision trees. Each tree is trained on a random subset of the dataset, which increases the model's generalization ability. This method is resistant to overfitting and often provides high accuracy (5).

Support Vector Machine (SVM): SVM creates an optimal hyperplane that separates data into different classes. It tries to find the widest margin between data points. While it may not be linear, it can handle non-linear boundaries using kernel functions, making it effective with complex datasets, especially in high-dimensional, non-linear data. The kernel trick allows the creation of non-linear boundaries (6).

K-Nearest Neighbors (KNN): To classify a data point, KNN looks at its K closest neighbors. The class of the new data point is assigned based on the majority class of its neighbors. The main advantage of KNN is that it is non-parametric, meaning the model does not make assumptions about the data. However, it may perform poorly on high-dimensional datasets (7).

Naive Bayes: Naive Bayes assumes that each feature is independent of the class label. By using Bayes' theorem, it computes conditional probabilities for each class and selects the class with the highest probability. It is often very fast and particularly effective on large datasets. It is simple, efficient, and works well with limited data (8).

Artificial Neural Networks: Artificial neural networks consist of multiple layers and attempt to mimic the functioning of the human brain. Connections between layers are computed using weights, with each connection representing a neuron. Neural networks have the capacity to learn complex relationships but require significant computational power (9).

Multilayer Perceptron: A multilayer perceptron is a neural network with multiple hidden layers. Input data is processed with weights between the layers, and classification occurs at the output layer. It is typically used for classification and regression problems (10).

2. Unsupervised Learning Models

Clustering Analysis: Clustering divides data into groups with similar characteristics from an unlabeled dataset. The most popular clustering algorithm is K-means. Clustering algorithms aim to discover the natural structure of the data and work with unlabeled data for exploration (11).

Reinforcement Learning: In reinforcement learning, an agent (model) learns by interacting with its environment. The agent learns the best actions based on rewards or penalties received from the environment. This type of learning is used in decision-making processes and allows for dynamic, complex decisions based on interactions with the environment (12).

Principal Component Analysis (PCA): PCA is used to reduce the dimensions (variables) of a dataset by creating new components that capture the most variance in the data. The data is then re-expressed around these components, reducing computational power and storage costs (13).

APPLICATIONS OF MACHINE LEARNING IN MEDICINE

1. Disease Diagnosis and Personalized Treatment

Machine learning is used to diagnose diseases. By using data from imaging techniques, machine learning algorithms assist doctors in identifying pathological conditions in medical images through image processing. For instance, deep learning algorithms are used in the detection and classification of masses in medical images such as X-rays, magnetic resonance imaging or computed tomography scans (14). Personalized medicine aims to create treatment plans tailored to individuals by considering genetic, environmental, and lifestyle factors. Machine learning helps suggest patient-specific treatment strategies based on information derived from patient data. Machine learning techniques used in genomic data analysis can assist in predicting a patient's response to specific treatments (15).

2. Clinical Decision Support Systems

Machine learning plays a fundamental role in the development of clinical decision support systems. These systems analyze patient data and provide recommendations for diagnosis and treatment. For example, systems that identify potential risks and recommend early interventions using data from electronic health records can improve the quality of healthcare services (16).

3. Epidemiology

In epidemiological research, machine learning enables the extraction of patterns and inferences about disease spread from large datasets. For example, during the COVID-19 pandemic, machine learning models were used to predict transmission rates and guide effective control measures (17).

4. Clinical Research

In clinical research, machine learning accelerates drug development processes and reduces costs. It helps make critical decisions, such as identifying patient groups for clinical trials and determining drug targets. Machine learning is also used to create new drugs and improve the efficacy of existing ones by analyzing large datasets to identify potential drug components and expedite clinical trials (18).

BENEFITS OF MACHINE LEARNING

Machine learning offers numerous benefits in the medical field, enhancing healthcare quality. It provides faster and more accurate diagnoses compared to traditional methods, enabling early detection of diseases. Furthermore, it reduces the workload of healthcare professionals, improving resource efficiency, as machines support routine tasks, allowing medical staff to focus on more critical work. Machine learning converts complex data into meaningful insights, facilitating more effective decision-making. It is also a powerful tool for predicting disease progression and identifying potential risks in advance. Finally, it assists in creating personalized treatment plans tailored to individual patient needs.

LIMITATIONS OF MACHINE LEARNING

Despite the significant potential of machine learning technologies in medical informatics, several limitations and challenges remain. First, the confidentiality and security of medical data are critical, as this information is highly sensitive, and data breaches could lead to ethical and legal issues. Additionally, the success of machine learning models depends on the quality and accuracy of the data used. Incorrect, incomplete, or biased data can negatively impact the performance of the models, and systematic errors in medical data collection exacerbate this issue. Moreover, machine learning models are often described as “black boxes,” making it difficult for doctors and patients to understand and question the decisions made by these algorithms. This raises concerns regarding trust and limits the clinical application of these algorithms. Furthermore, the accountability for decisions made by machine learning systems is an important ethical and legal issue. In cases of malpractice, it is unclear whether responsibility lies with the manufacturer, the healthcare institution using the algorithm, or the individual users. Finally, the implementation and development of these technologies require significant technical expertise, strong infrastructure, and considerable resources, which may limit their widespread use. Overcoming these challenges will enable more extensive and effective application of machine learning technologies in the medical field.

FUTURE EXPECTATIONS

The impact of machine learning in medical informatics will continue to increase rapidly, leading to greater advancements. AI-assisted technologies, such as autonomous diagnostic systems, could make accurate diagnoses without doctor intervention. Patient data can be analyzed in real time, providing immediate contributions to treatment processes. These innovations could be facilitated by wearable technologies and the Internet of Things (IoT). Moreover, explainable AI models will foster greater trust in these systems. Integrating genetic, proteomic, and metabolomic data with machine learning will provide more comprehensive treatment solutions, leading to personalized and efficient healthcare (15).

CONCLUSION

Machine learning in medical informatics holds great potential for enhancing the quality of healthcare services, reducing costs, and improving individual patient care. However, to use this technology effectively, several issues, such as data privacy, ethical responsibilities, and technical challenges, need to be addressed. In the future, the use of machine learning in medicine is expected to become more widespread, leading to revolutionary changes in healthcare systems.

DECLARATIONS

Financial Support: No specific funding was received for this study.

Conflicts of Interest: The author declares no conflicts of interest in relation to this manuscript.

Ethical Approval Statement: This review study did not involve human or animal subjects, and thus, ethical approval was not applicable.

REFERENCES

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine. 2019;25(1):44-56.
2. Shinde PP, Shah S, editors. A review of machine learning and deep learning applications. 2018 Fourth international conference on computing communication control and automation (ICCUBEA); 2018: IEEE.
3. Song C, Ristenpart T, Shmatikov V, editors. Machine learning models that remember too much. Proceedings of the 2017 ACM SIGSAC Conference on computer and communications security; 2017.
4. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. Behavior therapy. 2020;51(5):675-87.
5. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research. 2021;70:245-317.
6. Chandra MA, Bedi S. Survey on SVM and their application in image classification. International Journal of Information Technology. 2021;13(5):1-11.
7. Kramer O, Kramer O. K-nearest neighbors. Dimensionality reduction with unsupervised nearest neighbors. 2013:13-23.
8. Vembandasamy K, Sasipriya R, Deepa E. Heart diseases detection using Naive Bayes algorithm. International Journal of Innovative Science, Engineering & Technology. 2015;2(9):441-4.

9. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. Elsevier; 2013. p. 47-58.

10. Taud H, Mas J-F. Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios. 2018:451-5.

11. Ahuja R, Chug A, Gupta S, Ahuja P, Kohli S. Classification and clustering algorithms of machine learning with their applications. Nature-inspired computation in data mining and machine learning. 2020:225-48.

12. Hahne F, Huber W, Gentleman R, Falcon S, Gentleman R, Carey V. Unsupervised machine learning. Bioconductor case studies. 2008:137-57.

13. Kurita T. Principal component analysis (PCA). Computer vision: a reference guide: Springer; 2021. p. 1013-6.

14. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94.

15. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nature medicine. 2019;25(1):24-9.

16. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics. 2017;22(5):1589-604.

17. Hamilton AJ, Strauss AT, Martinez DA, Hinson JS, Levin S, Lin G, et al. Machine learning and artificial intelligence: applications in healthcare epidemiology. Antimicrobial Stewardship & Healthcare Epidemiology. 2021;1(1):e28.

18. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends in pharmacological sciences. 2019;40(8):577-91.