# A Comprehensive Review of Machine Learning Techniques with Mitigation Concept for Predictive Healthcare Analytics

*Ranjit Kumar*[1], *Ajay Kumar*[2]
[1] Maharaja Agrasen University, Baddi, Himachal Pradesh, India; [2] IILM University ,Greater Noida, India;
pdf.ranjitkumar@lincoln.edu.my, pdfsv.ajaykumar@lincoln.edu.my

---

**Abstract:** Artificial intelligence and machine learning are fundamentally reshaping the healthcare landscape by enabling predictive analytics, personalizing patient treatments, and improving clinical outcomes. Despite these theoretical advances, integrating predictive models into real-world clinical workflows remains hampered by challenges related to data privacy, model interpretability, and technical accessibility. This comprehensive review examines state-of-the-art machine learning techniques applied to predictive healthcare, exploring how recent innovations attempt to bridge the gap between algorithmic potential and bedside utility. Finally, the paper discusses critical ethical considerations, failure modes, and practical deployment challenges, providing a roadmap for future research in scalable and trustworthy clinical decision support systems
**Keywords**: Ovarian Cancer detection; Machine Learning; Healthcare; Prediction;

---

## Introduction

Artificial intelligence (AI) and machine learning (ML) have demonstrated tremendous success in modeling various complex healthcare prediction tasks, ranging from early disease diagnosis to dynamic patient treatment planning [1]. As medical data becomes increasingly digitized through electronic health records (EHRs) and continuous monitoring devices, there is a growing motivation to leverage predictive analytics to improve patient outcomes. However, the problem scope extends beyond merely maximizing predictive accuracy; it encompasses the critical need to design models that are clinically accessible, interpretable to medical professionals, and fundamentally secure. Many traditional ML workflows require highly specialized computational infrastructure and deep engineering expertise, which severely limits their accessibility for domain experts such as healthcare researchers and clinicians [2].
While current predictive architectures boast high accuracy on benchmark datasets, existing approaches are fundamentally insufficient for widespread clinical adoption for several prominent reasons. First, the opaque, "black-box" nature of advanced deep learning models restricts clinical trust, as practitioners cannot easily discern which specific features of the patient data drive the algorithmic decision-making process [3]. Second, existing methodologies often struggle with the stringent privacy and security requirements inherent to sensitive medical data, rendering centralized model training a massive legal and ethical liability [1]. Finally, conventional machine learning pipelines lack the adaptability to handle the inevitable phenomenon of concept drift, where shifting clinical practices or changing demographics degrade model reliability over time [4].

To address these systemic gaps in the literature, this the fundamental contributions of this review are to present a comprehensive taxonomy of current predictive healthcare analytics techniques, categorized by their approaches to interpretability, and clinical accessibility.

**Related work**

The major category of related work focuses on the democratization and accessibility of machine learning within healthcare analytics. Recent cloud-based solutions have sought to simplify the deployment pipeline by allowing healthcare researchers to build and evaluate predictive models using familiar querying languages like SQL [2]. The core idea of this approach is to eliminate the necessity for complex programming environments, thereby allowing clinicians to directly interact with patient data for localized tasks such as diabetes prediction [2]. While the primary strength of this paradigm is its profound reduction in the accuracy, a notable weakness is its limited flexibility for highly customized algorithmic architectures. Comparing review builds upon these accessible foundations but advocates for integrating more dynamic evaluation metrics to ensure long-term statistical rigor.The final review category encompasses techniques dedicated to model interpretability and generalized evaluation methodologies.

*Table 1. Recent research on ovarian cancer has focused extensively on leveraging ML classifications*

| Reference | Dataset | Sample Size | Classification | Accuracy | Remarks |
|---|---|---|---|---|---|
| [5] | Open Source | 214 | CNN | 92.06 % | Noisy |
| [6] | Generate | 235 | KNN | 67.16 | 10- fold cross validation |
| [7] | Open Source | 278 | DCNN | 75% | Low Accuracy |
| [8] | TCGA-OV | 282 | RED-CNN | 95% | Reduces CT radiation exposure But less accuracy |
| [9] | Hybrid | Customized | DNN | 98.96 | Overfitting |

**Key Contribution**

Researchers have developed physics-inspired approaches, mapping machine learning loss landscapes to identify conserved weights and relevant input features, which is highly beneficial for explaining decisions in sensitive areas like medicine [3]. A relative weakness is their mathematical complexity, which can be daunting for clinical end-users who require immediate diagnostic clarity. Our research bridges this gap by recommending automated mechanisms that abstract these complex interpretability metrics into actionable.

**Method, Experiments and Results**

The key design rationale behind this modular pipeline is to empower domain experts while strictly optimizing algorithmic integrity. By utilizing efficient large-scale approximation methods, accuracy can swiftly process hundreds of statistical moments [5]. Furthermore, to optimize the training phase, our approach integrates dynamic learning curve modeling to facilitate early stopping and optimal hyperparameter configuration [10]. This ensures that computational budgets are not wasted on algorithms that will fail to achieve the required clinical reference performance, thus streamlining the algorithm selection process [10]. Additionally, the study includes continuous statistical monitoring modules designed to detect changes in data sources and prevent concept drift over time [4].

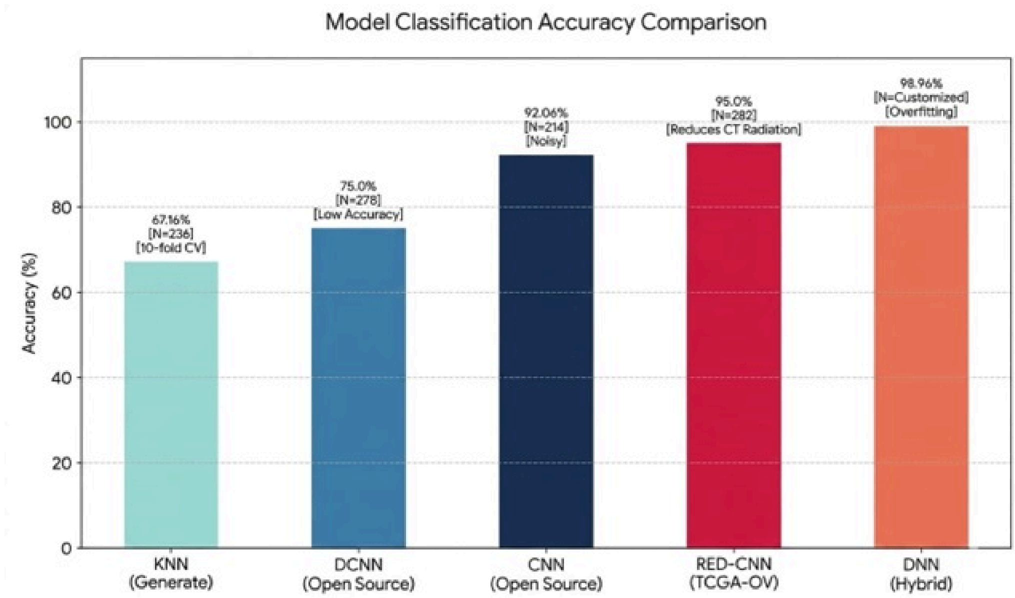


*Figure 1. Modal Classification Accuracy Comparison graph.*

**Discussions**

The practical implications of adopting advanced machine learning methodologies in predictive healthcare are inherently transformative for bedside clinical practice. By shifting toward SQL-based model generation and simplified cloud deployments, healthcare institutions can significantly accelerate the translation of raw clinical data into actionable bedside insights [2]. Furthermore, integrating continuous event-based sensor data with novel spatio-temporal ML pipelines could eventually enable high-fidelity, real-time patient monitoring in intensive care units [11]. This deployment paradigm allows non-technical clinical stakeholders to actively participate in the validation and refinement of predictive systems, thereby anchoring the technology in practical medical realities.

Despite the immense clinical potential, several critical limitations and failure modes must be explicitly acknowledged when deploying these systems.

First, predictive models remain highly vulnerable to changing data sources and concept drift, which can systematically invalidate previously learned clinical patterns and lead to erroneous patient predictions [4].

Second, the computational overhead associated with rigorous privacy-preserving techniques often introduces latency, which can be highly detrimental in acute care scenarios requiring instantaneous decision-making[1].

Third, an over-reliance on simplified, highly accessible ML tools might artificially constrain model expressivity, ultimately preventing the discovery of highly complex, non-linear biological relationships [2].

Furthermore, the integration of algorithmic decision-making into modern healthcare necessitates strict adherence to foundational ethical considerations.

Looking forward, several promising avenues for future scientific work have emerged from this comprehensive review.

Researchers should investigate the application of hybrid quantum machine learning architectures to tackle highly dimensional biological datasets, such as genomic sequencing, that currently overwhelm classical computing constraints [12].

Future studies must prioritize the standardization of physics-inspired interpretability methods, translating abstract loss landscape topographies into universally understood clinical risk scores [3].

Finally, exploring methods to seamlessly integrate asynchronous, event-based sensor data into standard clinical diagnostic models could redefine real-time patient monitoring paradigms [11].

findings

**Conclusions**

This comprehensive review highlights the rapid evolution and lingering challenges of machine learning techniques in predictive healthcare analytics. While predictive modeling has historically struggled with a strict dichotomy between high statistical accuracy and clinical opacity, robust evaluation metrics, and accessible deployment environments offer a highly viable path forward. By moving away from isolated, black-box algorithms and embracing transparent, democratized tools, the healthcare industry can safely harness the profound power of artificial intelligence to improve longitudinal patient outcomes.

This review demonstrates that scalable healthcare analytics must intrinsically balance technical innovation with ethical and legal responsibility. Mitigating concept drift, ensuring patient data privacy, and fostering clinician trust are not merely supplementary features, but fundamental prerequisites for the real-world deployment of clinical decision support systems. Ultimately, sustained interdisciplinary collaboration between data scientists, medical professionals, and health policy-makers will be essential to fully realize the promise of predictive healthcare analytics in the modern clinical era.

**References**


1. A. Guerra-Manzanares, J. Lechuga, Michail Maniatakos, and F. E. Shamout, “Privacy-Preserving Machine Learning for Healthcare: Open Challenges and Future Perspectives,” *Lecture Notes in Computer Science*, pp. 25–40, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-39539-0_3.
2. M. A. Salari and R. B, “Machine Learning for Everyone: Simplifying Healthcare Analytics with BigQuery ML,” *Journal of Artificial Intelligence, Machine Learning and Data Science*, vol. 3, no. 1, pp. 2407–2416, Mar. 2025, doi: https://doi.org/10.51219/jaimld/salari-ma/520.
3. M. P. Niroomand, C. T. Cafolla, J. W. R. Morgan, and D. J. Wales, “Characterising the area under the curve loss function landscape,” *Machine Learning: Science and Technology*, vol. 3, no. 1, p. 015019, Jan. 2022, doi: https://doi.org/10.1088/2632-2153/ac49a9.
4. Redefining Data Products in the Age of Artificial Intelligence and Deep Learning,” *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, vol. 3, 2022, doi: https://doi.org/10.63282/3050-9262.ijaidsml-v3i4p105.
5. L. Zhang, J. Huang, and L. Liu, “Improved Deep Learning Network Based in combination with Cost-sensitive Learning for Early Detection of Ovarian Cancer in Color Ultrasound Detecting

System," *Journal of Medical Systems*, vol. 43, no. 8, Jun. 2019, doi: https://doi.org/10.1007/s10916-019-1356-8.

6. U. R. Acharya *et al.*, "Use of Nonlinear Features for Automated Characterization of Suspicious Ovarian Tumors Using Ultrasound Images in Fuzzy Forest Framework," *International Journal of Fuzzy Systems*, vol. 20, no. 4, pp. 1385–1402, Feb. 2018, doi: https://doi.org/10.1007/s40815-018-0456-9.
7. M. Wu, C. Yan, H. Liu, and Q. Liu, "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks," *Bioscience Reports*, vol. 38, no. 3, p. BSR20180289, Mar. 2018, doi: https://doi.org/10.1042/bsr20180289.
8. Yasuyo Urase *et al.*, "Simulation Study of Low-Dose Sparse-Sampling CT with Deep Learning-Based Reconstruction: Usefulness for Evaluation of Ovarian Cancer Metastasis," *Applied Sciences*, vol. 10, no. 13, pp. 4446–4446, Jun. 2020, doi: https://doi.org/10.3390/app10134446.
9. Ashwini Kodipalli, S. L. Fernandes, and Santosh Dasar, "An Empirical Evaluation of a Novel Ensemble Deep Neural Network Model and Explainable AI for Accurate Segmentation and Classification of Ovarian Tumors Using CT Images," *Diagnostics*, vol. 14, no. 5, pp. 543–543, Mar. 2024, doi: https://doi.org/10.3390/diagnostics14050543.
10. F. Mohr and Jan, "Learning curves for decision making in supervised machine learning: a survey," *Machine Learning*, vol. 113, no. 11–12, pp. 8371–8425, Dec. 2024, doi: https://doi.org/10.1007/s10994-024-06619-7.
11. S.-I. Hou and J. Park, "Bridging the Digital Divide: Factors Affecting Asynchronous and Synchronous Telehealth Use," *Innovation in Aging*, vol. 9, no. Supplement_2, Dec. 2025, doi: https://doi.org/10.1093/geroni/igaf122.1143.
12. Quantum Machine Learning Algorithms for Complex Optimization Problems,'" *International Journal of Intelligent Systems and Applications in Engineering*, 2024, doi: https://doi.org/10.17762/ijisae.v12i4.6663.