

# Recent Advances in ML Models and Their Applications in Bioinformatics and Biomedical Engineering

Abedalrahman Busati\*

Information Technology Department, University of Fujairah, Fujairah, UAE

\*Corresponding Author

DOI: <https://doi.org/10.51244/IJRSI.2025.12030058>

Received: 21 March 2025; Accepted: 24 March 2025; Published: 11 April 2025

## ABSTRACT

The explosion of biological and biomedical data has opened up incredible possibilities for improving healthcare and understanding life itself. Machine learning (ML) has become a game-changer, helping us analyze complex datasets, predict diseases, and design personalized treatments. But it's not all smooth sailing, integrating ML into bioinformatics and biomedical engineering comes with its fair share of challenges. For instance, many advanced ML models are like "black boxes," making it hard to trust their decisions in critical areas like clinical diagnostics. Combining different types of biological data, such as genomics and proteomics, is another tough nut to crack. Add to that ethical concerns around data privacy and the sheer computational power needed to process massive datasets, and it's clear we have work to do. This review dives into these challenges, exploring how cutting-edge ML models like deep learning, reinforcement learning, and graph neural networks, are being used to decode genomes, automate medical imaging, speed up drug discovery, and even monitor health in real-time through wearable devices. It also proposed ways to make ML models more interpretable, integrate diverse biological data seamlessly, and ensure data privacy through federated learning. By tackling these challenges and fostering collaboration across disciplines, this work aims to make ML-driven healthcare solutions not only more effective but also fair and accessible to everyone. Together, we can unlock the full potential of ML to transform healthcare and improve lives worldwide.

**Keywords:** ML; Bioinformatics; Biomedical Engineering; DL; Personalized Medicine

## INTRODUCTION

The rapid growth of biological and biomedical data has created unprecedented opportunities for advancing healthcare and understanding complex biological systems. ML, a subset of artificial intelligence (AI), has become indispensable in analyzing these vast datasets, uncovering patterns, and making predictions that were previously unattainable. In bioinformatics, ML models are used to decode genomic sequences, predict protein structures, and identify biomarkers for diseases. In biomedical engineering, ML is revolutionizing medical imaging, wearable devices, and personalized medicine. This review explores recent developments in ML models and their transformative applications in these fields. Figure 1 shows the proposed research methodology.

ML has become a cornerstone of modern bioinformatics and biomedical engineering, with recent advancements in models like DL, reinforcement learning (RL), and graph neural networks (GNNs) driving innovation across these fields. DL, a subset of ML, has gained significant attention for its ability to process high-dimensional data, making it particularly effective in tasks such as medical imaging and genomics. For instance, convolutional neural networks (CNNs) are widely used to detect tumors and classify diseases in medical images (Esteva et al., 2017), while recurrent neural networks (RNNs) and their variants, such as long short-term memory (LSTM) networks, excel in analyzing genomic sequences and predicting gene expression (Alipanahi et al., 2015). More recently, transformers, originally developed for natural language processing (NLP), have been adapted for protein structure prediction, as demonstrated by groundbreaking tools like AlphaFold (Jumper et al., 2021). Beyond DL, reinforcement learning has shown promise in optimizing

treatment strategies and drug discovery, leveraging trial-and-error learning to adapt to dynamic environments such as personalized cancer therapy (Liu et al., 2022). Similarly, graph neural networks have emerged as powerful tools for analyzing biological networks, such as protein-protein interactions and gene regulatory networks, enabling applications like drug repurposing and disease gene identification (Jie et al., 2020).

The applications of these ML models span a wide range of domains within bioinformatics and biomedical engineering. In genomics and proteomics, ML has revolutionized tasks like genome sequencing, variant calling, and functional annotation, with deep learning models predicting the impact of genetic mutations on protein function (Gulshan et al., 2016) and identifying post-translational modifications. In medical imaging, ML, particularly, DL has transformed diagnostics by automating tasks such as tumor segmentation, disease classification, and anomaly detection. For example, CNNs have been used to detect diabetic retinopathy from retinal images (Menze et al., 2015) and classify brain tumors from MRI scans (Zhavoronkov et al., 2019), (Qawqzeh, 2019). In drug discovery, ML accelerates the identification of drug-target interactions, optimizes drug candidates, and predicts potential side effects, with generative adversarial networks (GANs) enabling the design of novel molecules (Attia et al., 2019). Beyond these applications, ML is also reshaping personalized medicine through wearable devices, where algorithms analyze real-time data to monitor vital signs, detect anomalies, and provide tailored health recommendations, such as predicting cardiovascular events using smartwatch data (Scott and Su-In, 2017). Together, these advancements highlight the transformative potential of ML in advancing healthcare and biological research.

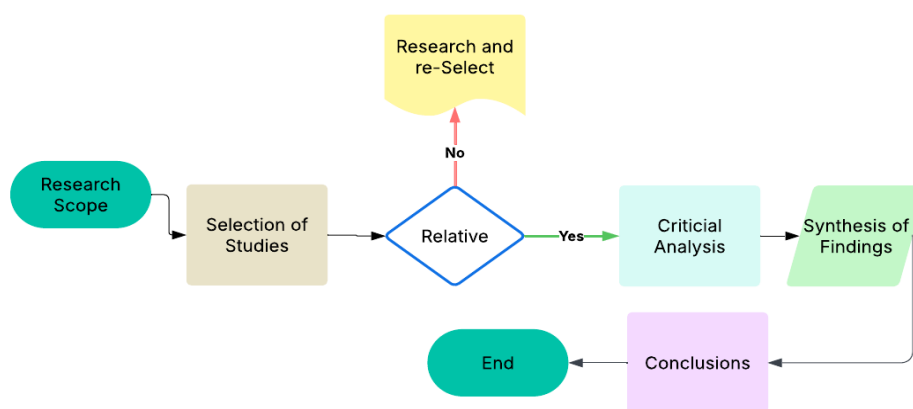


Figure 1. Research Methodology

## Related Works

The integration of ML into bioinformatics and biomedical engineering has been extensively studied in recent years, with numerous researchers contributing to the development and application of ML models in these fields. This section highlights key studies and advancements that have shaped the current landscape of ML in bioinformatics and biomedical engineering.

### ML in Bioinformatics

**Genomics and Sequence Analysis:** Early work by Alipanahi et al., 2015, demonstrated the potential of DL for predicting DNA- and RNA-binding protein specificities. More recently, models like (Gottesman et al., 2019), (Otoom et al., 2019), Qawqzeh (2019) have revolutionized protein structure prediction, achieving unprecedented accuracy in the Critical Assessment of Structure Prediction (CASP) competitions. **Proteomics and Protein Interaction Prediction:** Researchers have employed GNNs to model protein-protein interaction networks, enabling the identification of novel drug targets and disease biomarkers (Rappoport and Shamir (2019). For example, Zitnik et al., (2018) developed a GNN-based framework for drug repurposing, which has been applied to various diseases, including COVID-19. **Multi-Omics Data Integration:** Rappoport and Shamir (2018) reviewed multi-omics clustering algorithms, emphasizing the importance of integrating heterogeneous

datasets to uncover complex biological relationships. Their work has inspired the development of multi-task learning frameworks for omics data analysis.

## ML in Biomedical Engineering

**Medical Imaging:** The application of CNNs in medical imaging has been widely studied. Esteva et al., (2017) demonstrated the use of CNNs for skin cancer classification, achieving performance comparable to dermatologists. Similarly, Gulshan et al., (2017) developed a DL algorithm for diabetic retinopathy detection, showcasing the potential of ML in automating diagnostic tasks. Moreover, drug discovery gained lot of attention. For instance, Zhavoronkov et al., (2019), pioneered the use of GANs for drug discovery, designing novel molecules with desired properties. Their work has inspired further research into generative models for drug design and optimization. Wearable devices and personalized medicine also represent a good application area in which Attia et al., (2019) developed an ML model for predicting atrial fibrillation using data from wearable devices, highlighting the potential of ML in personalized healthcare. Their work has paved the way for the integration of ML into continuous monitoring systems.

## Challenges and Ethical Considerations

Interpretability and explainability is considered a challenge in this field. Lundberg and Lee (2017) introduced SHAP (SHapley Additive exPlanations), a framework for interpreting ML model predictions. Their work has been instrumental in advancing explainable AI (XAI) in healthcare. Moreover, data privacy and federated learning witnessed some efforts from several researchers, for example, Rieke et al., (2020), explored the use of federated learning (FL) for training ML models on decentralized datasets, addressing privacy concerns in biomedical research. Their work has inspired the development of FL frameworks for multi-center clinical trials. Ethical AI in healthcare showed a good progress in recent research output. Char et al., (2018) discussed the ethical challenges of implementing ML in healthcare, emphasizing the need for transparency, fairness, and accountability. Their work has informed the development of ethical guidelines for AI in medicine.

## Challenges and Limitations

Despite the remarkable potential of ML in bioinformatics and biomedical engineering, its integration into these fields is not without significant hurdles. One of the most pressing challenges lies in the quality and availability of data. ML models thrive on large, high-quality datasets, but biomedical data is often fragmented, noisy, or incomplete. For instance, medical imaging datasets may lack diversity in patient demographics, leading to models that perform well for some groups but poorly for others. Additionally, the sensitive nature of this data raises privacy concerns, making it difficult to share across institutions due to strict regulations like GDPR and HIPAA. Even when data is available, the process of annotating and labeling it, essential for supervised learning, is time-consuming and requires specialized expertise, creating bottlenecks in model development. These data-related challenges are compounded by the "black-box" nature of many advanced ML models, particularly DL. Clinicians and researchers need interpretable models to understand how predictions are made, but achieving this transparency remains a significant obstacle. Without it, trust in ML-driven decisions is limited, hindering adoption in critical areas like clinical diagnostics and treatment planning (Qawqzeh et al., 2023).

Beyond data and interpretability, there are broader ethical, technical, and practical challenges that must be addressed. Algorithmic bias is a persistent issue, as models trained on biased datasets can perpetuate or even amplify existing inequalities. For example, a model trained primarily on data from one ethnic group may fail to generalize to others, leading to unfair or inaccurate outcomes. Ethical concerns also extend to issues of informed consent and transparency, as patients may not fully understand how their data is used or the implications of ML-driven decisions. On the technical side, the computational complexity of training and deploying ML models can be prohibitive, especially in resource-constrained settings like low-income countries or rural healthcare facilities. Even when models are successfully developed, integrating them into real-world clinical practice is fraught with challenges, from resistance to change among healthcare providers to the need for extensive validation and regulatory approval. These barriers highlight the need for interdisciplinary

collaboration (Qawqzeh et al., 2010), innovative solutions, and a commitment to ethical AI development to ensure that ML-driven advancements are both effective and equitable.

## Future Directions

The rapid evolution of ML in bioinformatics and biomedical engineering has opened up exciting new possibilities for advancing healthcare and biological research. While significant progress has been made, several challenges remain, and the field continues to evolve with emerging technologies and methodologies. This section outlines key future directions that hold the potential to drive innovation, address current limitations, and ensure that ML-driven solutions are both effective and equitable. From improving model interpretability to integrating multi-omics data and addressing ethical concerns, these advancements will shape the next generation of ML applications in healthcare and life sciences.

One of the most pressing challenges in ML is the "black-box" nature of many advanced models, particularly deep learning. While these models achieve high accuracy, their lack of interpretability limits their adoption in clinical settings where understanding the decision-making process is critical. Future research should focus on developing explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), to provide insights into how models arrive at their predictions (Alejandro et al., 2020). Additionally, the integration of multi-omics data, spanning genomics, transcriptomics, proteomics, and metabolomics, holds immense potential for understanding complex biological systems. However, combining these heterogeneous datasets remains a challenge due to differences in scale, dimensionality, and noise levels (Adadi and Berrada, 2020). Techniques such as multi-task learning, transfer learning, and graph-based approaches, including GNNs, can help address these challenges and enable a more holistic understanding of disease mechanisms (Zhang et al., 2024). Furthermore, federated learning (FL) offers a promising solution to data privacy concerns by allowing models to be trained across multiple datasets without transferring the data itself (Basmah et al., 2024). This approach is particularly valuable for multi-center clinical trials and global health initiatives, where data sharing is logistically or ethically challenging (Khan et al., 2020). Furthermore, the summarized future directions of the study can be visualized in figure 2.

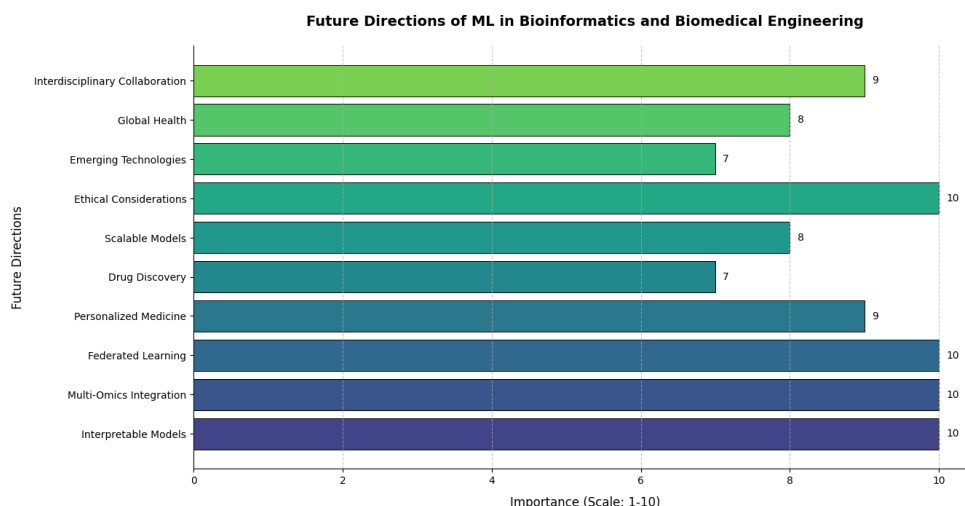


Figure 2. Summarized future directions of the study

Looking ahead, ML has the potential to revolutionize personalized medicine by tailoring treatments to individual patients based on their genetic, molecular, and clinical profiles. Dynamic ML models that adapt to real-time data from wearable devices or continuous monitoring systems, combined with RL for optimizing treatment strategies, can significantly improve patient outcomes (Al-Hamadani et al., 2024). In drug discovery, ML can accelerate the identification of drug-target interactions and optimize drug candidates using generative models like variational autoencoders (VAEs) and GANs (Wu et al., 2023). However, as the volume of biomedical data grows, scalable and computationally efficient ML models will be essential. Techniques such as model compression, distributed computing, and quantization can help reduce the computational burden,

while lightweight models for edge devices (e.g., smartphones, wearables) will enable real-time analysis in resource-constrained settings (Fei et al., 2021). Ethical and regulatory considerations must also remain at the forefront, with frameworks needed to address issues such as data privacy, algorithmic bias, and informed consent (Abbas et al., 2024), (Qawqzeh et al., 2012). Finally, the integration of ML with emerging technologies like quantum computing, blockchain, and the Internet of Medical Things (IoMT) will further enhance its capabilities, enabling solutions for global health challenges and fostering interdisciplinary collaboration to ensure equitable access to ML-driven healthcare advancements (Jafari and Adibnia, 2025).

## CONCLUSIONS

The integration of ML into bioinformatics and biomedical engineering has ushered in a new era of innovation, transforming how we analyze biological data, diagnose diseases, and develop treatments. From decoding the complexities of the human genome to enabling real-time health monitoring through wearable devices, ML has demonstrated its potential to address some of the most pressing challenges in healthcare and biological research. However, as we continue to push the boundaries of what is possible with ML, it is crucial to acknowledge and address the challenges that accompany these advancements. One of the most significant hurdles is the need for high-quality, diverse datasets to train robust and generalizable models. The sensitive nature of biomedical data further complicates this issue, as privacy concerns and regulatory restrictions limit data sharing and collaboration. Additionally, the "black-box" nature of many ML models, particularly DL, poses a barrier to their adoption in clinical settings, where interpretability and transparency are paramount. Overcoming these challenges will require innovative solutions, such as federated learning for secure data sharing and explainable AI techniques to build trust in ML-driven decisions.

Ethical considerations must also remain at the forefront of ML development. Algorithmic bias, informed consent, and equitable access to ML-driven healthcare solutions are critical issues that demand attention. By developing frameworks for ethical AI and fostering interdisciplinary collaboration, we can ensure that ML technologies are deployed responsibly and equitably. Furthermore, the integration of ML with emerging technologies, such as quantum computing and blockchain, holds immense potential for solving complex problems in drug discovery, genomics, and personalized medicine. Looking ahead, the future of ML in bioinformatics and biomedical engineering is bright, with numerous opportunities for innovation and impact. The development of interpretable models, the integration of multi-omics data, and the adoption of federated learning are just a few of the directions that promise to drive the field forward. As we continue to explore these avenues, it is essential to prioritize collaboration across disciplines and sectors, ensuring that ML-driven advancements are accessible and beneficial to all. In conclusion, ML has the potential to revolutionize healthcare and biological research, but realizing this potential will require addressing current challenges and embracing a commitment to ethical, equitable, and interdisciplinary innovation. By doing so, we can unlock the full potential of ML to improve human health and advance our understanding of complex biological systems.

Table 1. ML Models and Their Applications

| Domain          | ML Model                  | Application  | Advantages                                   | Limitations  |
|-----------------|---------------------------|--|--|--|
| Genomics        | DL (LSTM)                 | Gene expression prediction, variant calling              | High accuracy, handles sequential data       | Requires large datasets, computationally expensive |
| Proteomics      | Graph Neural Networks     | Protein-protein interaction prediction, drug repurposing | Captures complex relationships, scalable     | Interpretability challenges, data sparsity         |
| Medical Imaging | Convolutional Neural Nets | Tumor detection, diabetic retinopathy classification     | High precision, automated feature extraction | Requires annotated datasets, black-box nature      |



|                         |                        |  |   |   |
|-------------------------|------------------------|--|---|---|
| <b>Drug Discovery</b>   | Reinforcement Learning | Drug candidate optimization, personalized therapy  | Adapts to dynamic environments, optimizes outcomes          | High computational cost, limited interpretability       |
| <b>Wearable Devices</b> | Ensemble Learning      | Cardiovascular event prediction, anomaly detection | Combines multiple models for robustness, real-time analysis | Data privacy concerns, requires continuous data streams |

## REFERENCES

1. 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. <https://doi.org/10.1038/nature21056>
2. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), 831–838. <https://doi.org/10.1038/nbt.3300>
3. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
4. Liu, M., Shen, X., & Pan, W. (2022). Deep reinforcement learning for personalized treatment recommendation. *Statistics in Medicine*, 41(20), 4034–4056. <https://doi.org/10.1002/sim.9491>
5. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
6. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
7. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024. <https://doi.org/10.1109/TMI.2014.2377694>
8. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., & Aladinskaya, A. V. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
9. Qawqzeh, Y. K. (2019). Neural network-based diabetic type II high-risk prediction using photoplethysmogram waveform analysis. *International Journal of Advanced Computer Science and Applications*, 10(12), 1–7.
10. Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., ... & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *The Lancet*, 394(10201), 861–867. [https://doi.org/10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0)
11. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, 4768–4777.
12. Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Research*, 46(20), 10546–10562. <https://doi.org/10.1093/nar/gky889>
13. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119. <https://doi.org/10.1038/s41746-020-00323-1>
14. Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine*, 25(1), 16–18. <https://doi.org/10.1038/s41591-018-0310-5>
15. Otoom, M. M., Jemmali, M., Qawqzeh, Y., Sa, K. N., & Al Fay, F. (2019). Comparative analysis of different machine learning models for estimating the population growth rate in data-limited areas. *International Journal of Computer Science and Network Security*, 19(12), 96–102.

16. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13), i457–i466. <https://doi.org/10.1093/bioinformatics/bty294>
17. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
18. Qawqzeh, Y. K., Alourani, A., & Ghwanmeh, S. (2023). An improved breast cancer classification method using an enhanced AdaBoost classifier. *International Journal of Advanced Computer Science and Applications*, 14(1), 1–10.
19. Qawqzeh, Y. K., Mohd, A. M. A., Reaz, M., & Maskon, O. (2010). Photoplethysmography analysis of artery properties in patients presenting with established erectile dysfunction. *Proceedings of the International Conference on Computer Science and Network Technology*, 165–168.
20. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
21. Adadi, A., & Berrada, M. (2020). Explainable AI for healthcare: From black box to interpretable models. In V. Bhateja, S. Satapathy, & H. Satori (Eds.), *Embedded Systems and Artificial Intelligence* (pp. 123–135). Springer. [https://doi.org/10.1007/978-981-15-0947-6\\_31](https://doi.org/10.1007/978-981-15-0947-6_31)
22. Zhang, F., Kreuter, D., Chen, Y., et al. (2024). Recent methodological advances in federated learning for healthcare. *Patterns*, 5(6), 101006. <https://doi.org/10.1016/j.patter.2024.101006>
23. Khan, F. A., Rahman, A., Alharbi, M., & Qawqzeh, Y. K. (2020). Awareness and willingness to use PHR: A roadmap towards cloud-dew architecture-based PHR framework. *Multimedia Tools and Applications*, 79, 8399–8413. <https://doi.org/10.1007/s11042-018-6692-z>
24. Alotaibi, B. K., Khan, F. A., Qawqzeh, Y., Jeon, G., & Camacho, D. (2024). Performance and communication cost of deep neural networks in federated learning environments: An empirical study. *International Journal of Interactive Multimedia and Artificial Intelligence*, 12(1), 1–10. <https://doi.org/10.9781/ijimai.2024.12.001>
25. Al-Hamadani, M. N. A., Fadhel, M. A., Alzubaidi, L., & Balazs, H. (2024). Reinforcement learning algorithms and applications in healthcare and robotics: A comprehensive and systematic review. *Sensors*, 24(8), 2461. <https://doi.org/10.3390/s24082461>
26. Wu, X., Liu, C., Wang, L., & Bilal, M. (2023). Internet of things-enabled real-time health monitoring system using deep learning. *Neural Computing and Applications*, 35(20), 14565–14576. <https://doi.org/10.1007/s00521-021-06440-6>
27. Fei, C., Liu, R., Li, Z., Wang, T., & Baig, F. N. (2021). Machine and deep learning algorithms for wearable health monitoring. In A. K. Manocha, S. Jain, M. Singh, & S. Paul (Eds.), *Computational Intelligence in Healthcare* (pp. 123–135). Springer. [https://doi.org/10.1007/978-3-030-68723-6\\_6](https://doi.org/10.1007/978-3-030-68723-6_6)
28. Abbas, A., Alroobaea, R., Krichen, M., et al. (2024). Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things. *Personal and Ubiquitous Computing*, 28, 59–72. <https://doi.org/10.1007/s00779-021-01583-8>
29. Qawqzeh, Y., Reaz, M. B. I., Ali, M. A. M., Kok, B. G., Zulkifli, S. Z., & Noraidatulakma, A. (2012). Assessment of atherosclerosis in erectile dysfunction subjects using second derivative of photoplethysmogram. *Scientific Research and Essays*, 7(25), 2230–2236.
30. Jafari, M., & Adibnia, F. (2025). Securing IoMT healthcare systems with federated learning and BigchainDB. *Future Generation Computer Systems*, 165, 107609. <https://doi.org/10.1016/j.future.2024.107609>