Pertanika Proceedings 1 (5): 131 - 138 (2025)

PERTANIKA
JOURNALS

PERTANIKA PROCEEDINGS

Journal homepage: http://www.pertanika.upm.edu.my/

The Role of Big Data and Deep Learning in Proactive Healthcare: Insights, Challenges, and Future Directions

Karimulla Sha Shaik[1]*, K. Hari Krishna[2] and K. K. Baseer[3]

[1]*Department of Computer Science and Engineering, Mohan Babu University, Tirupathi, Andhra Pradesh 517102, India*
[2]*Department of Artificial Intelligence and Machine Learning, Mohan Babu University, Tirupathi Andhra Pradesh 517102, India*
[3]*GITAM School of Technology, GITAM (Deemed to be University), Bengaluru, Karnataka 561203, India*

ABSTRACT

The combination of deep learning technology and big data analytics is becoming increasingly important in proactive healthcare, which emphasizes the early identification, prediction, and prevention of diseases. Innovative approaches to illness management have been made possible by rapidly expanding genomic and other healthcare data, wearable data, electronic health records (EHRs), and medical images. Processing these massive datasets to extract valuable insights that enable prompt interventions and individualized therapies depends heavily on deep learning, specialized models such as Deep Belief Networks (DBNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs), which are essential to the analysis of these massive datasets to derive insightful information that enables prompt interventions and individualized care. This review explores the significant role of big data and deep learning in proactive healthcare, with a focus on chronic diseases such as chronic kidney disease (CKD). These technologies enable early diagnosis through medical imaging and predictive models based on clinical data, thus enhancing the ability to manage disease progression. Notwithstanding the many benefits, there are still issues with data privacy, data quality and standards, AI model interpretability, and integrating these models into current healthcare processes. The effective implementation of AI-driven solutions in clinical settings depends on resolving these problems. In the future, developments in real-time predictive analytics, federated learning, and explainable AI should improve preventative healthcare procedures even more. To shed light

ARTICLE INFO

Article history:
Received: 29 August 2025
Published: 31 October 2025

DOI: https://doi.org/10.47836/pp.1.5.024

E-mail addresses:
karimullashah@gmail.com (Karimulla Sha Shaik)
khk396gmail.com (K. Hari Krishna)
drkkbaseer@gmail.com (K. K. Baseer)
* Corresponding author

on how big data and deep learning are transforming healthcare into a more proactive, effective, and patient-centered system, this paper offers a thorough assessment of existing uses, problems, and future perspectives.

Keywords: AI Integration, Chronic Kidney Disease (CKD), convolutional neural networks (CNN), deep belief networks, deep learning, healthcare workflows, medical imaging, predictive modelling, proactive healthcare

INTRODUCTION

Big data and deep learning have transformed many industries, and one of the most promising fields for these technologies is healthcare. The capacity to analyze vast and intricate datasets has made proactive healthcare which prioritizes the early identification, prediction, and prevention of diseases more possible. The delivery of healthcare could be revolutionized by these datasets, which include EHRs, medical imaging, genomic data, and real-time sensor data from wearables (Zhang et al., 2021). Better patient outcomes and lower healthcare costs can result from healthcare providers using various data sources to better understand a patient's health, anticipate the beginning of diseases, and optimize treatment regimens.

A subfield of machine learning called deep learning has shown itself to be an effective tool in the medical field, particularly in processing large volumes of unstructured data like patient records and medical imaging. In applications like disease identification, progression prediction, and customized treatment regimens, models such as CNNs, RNNs, and DBNs have demonstrated great potential (Debal & Sitote, 2022). For example, CNNs are very good at analyzing medical images; they can identify anomalies like tumors, lesions, or early indicators of long-term diseases like chronic kidney disease (CKD) in retinal scans, MRIs, and X-rays (Sabanayagam et al., 2020). RNNs are excellent at evaluating time-series data from EHRs, which helps predict future health outcomes and the progression of diseases (Bai et al., 2022).

The incorporation of deep learning and big data into healthcare systems is fraught with difficulties, despite its enormous potential. Strong regulatory frameworks and encryption techniques are necessary to resolve issues with data security and privacy since health data is sensitive (Debal & Sitote, 2022). Furthermore, predictive models' precision and dependability can be impacted by the consistency and quality of healthcare data, which frequently originates from multiple sources. Furthermore, deep learning models are frequently viewed as "black boxes," and healthcare practitioners who require clear decision-making tools may find it difficult to interpret them (Elkholy et al., 2021).

The Table 1 presents a comparative analysis of various deep learning and machine learning approaches used for CKD prediction, highlighting the datasets, merits, performance metrics, and limitations of each model.

Table 1
DLL models analysis

Author & Reference	Approach	Dataset used	Merits	Performance metrics	Limitation
Elkholy et al. (2021)	DNN	Publicly available datasets	Increased performance with classifiers	Accuracy, precision, F-measure, recall	Small sets of data were used
Ma et al. (2020)	ANN	UCI CKD	Improved accuracy	Sensitivity, Specificity, AUC	Complexity in the segmentation process
Nallarasan et al. (2024)	Deep K-Net	TCIS	High accuracy	Accuracy, precision, F-measure, recall	High computational complexity
Lakshmi et al. (2024)	CNN	Renal PathNet	Accuracy = 92 %	Accuracy, precision, F-measure, recall	Time consumption
Yun et al. (2024)	DL	UCI	Accuracy = 83 % AUC value = 0.83	Accuracy, precision, F-measure, false positive rate, sensitivity	Less robustness due to traditional algorithms
Rao et al. (2023)	GNN and tabular data model	India UCI ML Repository dataset	95.089% accuracy.	Accuracy, F-measure, recall	Degraded illness prediction outcomes.
Kriplani et al. (2019)	DNN	UCI ML repository dataset.	97.7% of accuracy, 0.955 kappa value, 0.121 RMSE.	Accuracy, Kappa, RMSE	Lesser accuracy.

Elkholy et al. (2021) used a Deep Neural Network (DNN) approach with publicly available datasets. The model showed increased performance with classifiers, and its key metrics included accuracy, precision, F-measure, and recall. However, the model's limitation was that it used a small set of data. Ma et al. (2020) employed an Artificial Neural Network (ANN) on the UCI CKD dataset, achieving improved accuracy. The model's performance was evaluated using sensitivity, specificity, and AUC. Nallarasan et al. (2024) utilized a Deep K-Net approach on the TCIS dataset, which resulted in high accuracy. The performance metrics included accuracy, precision, F-measure, and recall, but the model was limited by its high computational complexity. Lakshmi et al. (2024) used a CNN with the Renal PathNet dataset, achieving an accuracy of 92%. The performance metrics included accuracy, precision, F-measure, and recall, but time consumption emerged as a limitation. Yun et al. (2024) applied Deep Learning (DL) on the UCI dataset, achieving an accuracy of 83% and an AUC value of 0.83. The model performed well on accuracy, precision, F-measure, false positive rate, and sensitivity, but faced challenges with less robustness compared to traditional algorithms. Rao et al. (2023) utilized a GNN and tabular data model with the India UCI ML Repository dataset, achieving 95.089% accuracy. The model's performance metrics included accuracy, F-measure, and recall. However, it was limited by degraded illness prediction outcomes. Kriplani et al. (2019) used a DNN with the UCI ML repository dataset, achieving 97.7% accuracy, a 0.955 Kappa value, and a 0.121 RMSE. While it performed well on metrics like accuracy and Kappa, it had lower accuracy compared to other models.

INSIGHTS INTO BIG DATA AND DEEP LEARNING IN PROACTIVE HEALTHCARE

Big Data and Deep Learning in Healthcare

Predictive models that can evaluate illness risk and assist healthcare practitioners in making well-informed decisions have been developed as a result of the integration and analysis of these data kinds. Machine learning's deep learning subfield has proven useful in automating the interpretation of complex data, particularly in fields like time series and image data. RNNs, CNNs, and DBNs are now often utilized in the healthcare industry for tasks like picture identification, illness diagnosis prediction, and treatment planning.

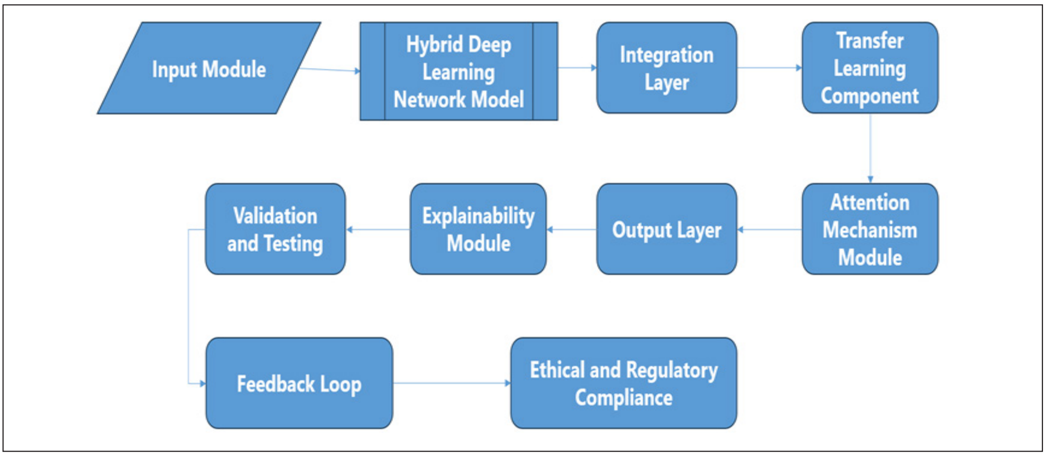

Figure 1. Deep learning network model and healthcare workflow integration (Courtesy: Google)

A system called the Deep Learning Network Model and Healthcare Workflow Integration uses deep learning technology to enhance every step of the healthcare process, from gathering data to treating patients, as per Figure 1.

Input Module

Collects data from various sources like environmental factors, wearables, clinical records, and medical imaging. **Goal:** Provide relevant data for analysis, enabling the model's forecasting and decision-making.

Hybrid Deep Learning Network Model

Analyzes input data using deep learning techniques (RNNs, CNNs) for pattern recognition, disease classification, and predictions. **Goal:** Processes large datasets to extract valuable insights, such as risk factors and suggested treatments.

Validation and Testing

Validates and tests the model on unseen data to ensure accuracy and reliability. **Goal:** Prevents overfitting and ensures robustness for real-world healthcare scenarios.

Explainability Module

Makes the deep learning model's predictions transparent and interpretable. **Goal:** Enhance trust by helping medical professionals understand the reasoning behind decisions.

Output Layer

Converts deep learning findings into actionable information, like treatment options and risk evaluations. **Goal:** Assists healthcare providers in making informed decisions based on model predictions.

Integration Layer

Integrates AI-driven insights with existing healthcare systems (CDSS, EHR). **Goal:** Ensures seamless incorporation of predictions and suggestions into clinical workflows.

Transfer Learning Component

Adapt models trained on one dataset to work on similar datasets with fewer data. **Goal:** Reduce the amount of labeled data needed and enable use across diverse healthcare scenarios.

Attention Mechanism Module

Prioritizes important information in sequence-based tasks (e.g., disease progression forecasting). **Goal:** Enhance prediction accuracy by focusing on relevant data and managing complex inputs.

Feedback Loop.

Continuously improves the model based on patient outcomes and healthcare decisions. **Goal:** Allows the model to adapt over time, improving its performance with new data.

Ethical and Regulatory Compliance

Ensures the AI system meets ethical standards and regulations like HIPAA and GDPR. **Goal:** Maintains patient privacy, security, and confidence, while ensuring fairness and transparency in decisions.

HEALTHCARE WORKFLOW INTEGRATION

Deep learning models enhance healthcare workflows across five stages:

- Data Collection: Information entered by patients and medical professionals.
- AI Data Processing: Deep learning models analyze data to generate insights.
- Clinical Decision Support: AI outputs assist in guiding treatment choices.
- Treatment: AI insights are used to create personalized patient treatments.
- Ethical Compliance: Ensures safe and ethical use of patient data.

CONCLUSIONS AND FUTURE WORK

By facilitating proactive approaches for disease detection, prediction, and customized therapy, big data and deep learning technologies have the potential to drastically change the healthcare industry. A new era of proactive healthcare is being welcomed by the combination of big data and deep learning, which has enormous promise for improved patient outcomes, individualized treatment, and early illness diagnosis. Even though there are still issues with data quality, privacy, and model interpretability, these problems should be resolved as these technologies continue to progress. To prevent and cure chronic illnesses like CKD, these technologies will become increasingly important as they develop, which will eventually result in more effective and efficient healthcare systems. Focuses on deep learning models tailored to individual genetic and environmental factors for precise disease prediction and treatment in the future.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their respective institutions for providing the resources and encouragement to carry out this research. We also extend our appreciation to our colleagues and peers for their constructive feedback and insightful discussions that helped shape this study.

We further acknowledge the contributions of researchers and practitioners in the domains of Big Data and Deep Learning, whose work has been a guiding reference for this paper. Finally, we are deeply thankful to our families for their constant motivation, patience, and moral support throughout the course of this work.

REFERENCES

Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports*, *12*(1), Article 8377. https://doi.org/10.1038/s41598-022-11715-9

Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, *9*(1), 109. https://doi.org/10.1186/s40537-022-00588-5

Elkholy, S. M. M., Rezk, A., & Saleh, A. A. E. F. (2021). Early prediction of chronic kidney disease using deep belief network. *IEEE Access*, *9*, 135542-135549. https://doi.org/10.1109/ACCESS.2021.3112627

Kriplani, H., Patel, B., & Roy, S. (2019). Prediction of chronic kidney diseases using deep artificial neural network technique. In J. D. Peter, S. L. Fernandes, C. E. Thomaz & S. Viriri (Eds.) *Computer aided intervention and diagnostics in clinical and medical images* (pp. 179-187). Springer International Publishing. https://doi.org/10.1007/978-3-030-19618-5_19

Lakshmi, B. S., Omotunde, H., Birundadevi, M., & Brahmam, M. G. (2024). Deep Learning for Automated Classification of Kidney Lesions in Chronic Kidney Disease Patients. In *International Conference on Integrated Circuits and Communication Systems (ICICACS)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICICACS52316.2024.00123

Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, *111*, 17-26. https://doi.org/10.1016/j.future.2020.04.019

Nallarasan, V., Ponnusamy, V., Lakshminarayanan, R., Vigneshwari, S., & Vinoth, R. (2024). Prediction of kidney disease utilizing a hybrid deep learning methodology. In *2nd International Conference on Computer, Communication and Control (IC4)* (pp. 1-8). IEEE. https://doi.org/10.1109/IC4.2024.001

Rao, P. K., Chatterjee, S., Nagaraju, K., Khan, S. B., Almusharraf, A., & Alharbi, A. I. (2023). Fusion of graph and tabular deep learning models for predicting chronic kidney disease. *Diagnostics*, *13*(12), Article 1981. https://doi.org/10.3390/diagnostics13121877

Sabanayagam, C., Xu, D., Ting, D. S., Nusinovici, S., Banu, R., Hamzah, H., Lim, C., Tham, Y. C., Cheung, C. Y., Tai, E. S., Wang, Y. X., Jonas, J. B., Cheng, C. Y., Lee, M. L., Hsu, W., & Wong, T. Y. (2020). A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *The Lancet Digital Health*, *2*(6), e295-e302. https://doi.org/10.1016/S2589-7500(20)30129-0

Yun, C., Tang, F., Gao, Z., Wang, W., Bai, F., Miller, J. D., Liu, M., Lee, Y., & Lou, Q. (2024). Construction of risk prediction model of type 2 diabetic kidney disease based on deep learning. *Diabetes & Metabolism Journal*, *48*(4), 771-779. https://doi.org/10.4093/dmj.2023.0033

Zhang, K., Liu, X., Xu, J., Yuan, J., Cai, W., Chen, T., Wang, K., Gao, Y., Nie, S., Xu, X., Qin, X., Su, Y., Xu, W., Olvera, A., Xue, K., Li, Z., Zhang, M., Zeng, X., Zhang, C. L., Li, O., … & Wang, G. (2021). Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nature biomedical engineering*, *5*(6), 533-545. https://doi.org/10.1038/s41551-021-00655-6