Sustainable Healthcare in Big data Analytics using Deep Learning: Reviews, Challenges, Predictions and Recommendations

Prajna Paramita Debata[1], Midhunchakkaravarthy[2], Brojo Kishore Mishra[3]
[1] Department of Engineering and Science, Lincoln University College
Selangor, Malaysia; [2] Faculty of AI Computing and Multimedia, Lincoln University College
Selangor, Malaysia; [3] Department of Computer Science and Engineering, NIST University
Berhampur, India

[1]pdf.prajna@lincoln.edu.my
[2]midhun@lincoln.edu.my
[3]brojokishoremishra@gmail.com

Abstract:
Sustainability in healthcare refers to the development and implementation of practices that reduce the environmental impact of healthcare systems while maintaining or improving the quality of care. A large amount of diverse medical data is now accessible from various healthcare institutions and devices. The handling and management of healthcare data have posed challenges due to issues like irregularity, high dimensionality, and sparsity.

An Electronic Health Record (EHR) or big data in healthcare analysis facilitates medical practice or supports healthcare functions. This review focuses three phases:

(i) Deep Learning applied in healthcare data which aims to find out the complexity of integrating various algorithms and systems for the health care data analysis.
(ii) Healthcare data used for validate the different suggested models.
(iii) Key findings of existing models.
(iv) Challenges of the existing models for healthcare analysis.

The outcome of this review is to analyse the challenges (Research Gap), opportunities, and future scope in sustainable healthcare analysis. The application of this analysis may include earlier identification of diseases, improved prognostic accuracy, accelerated clinical research progress, and enhanced patient management.

Keywords: Sustainability in healthcare; management of healthcare data; Deep Learning; Optimization algorithms; challenges

Introduction

Sustainability within the healthcare sector focuses on implementing eco-friendly practices while maintaining high-quality care delivery. As healthcare systems expand, it becomes vital to decrease their carbon emissions and waste production. One strategy involves utilizing data-driven approaches to improve resource efficiency and avoid superfluous treatments or procedures. The digital transformation facilitated by tools such as Electronic Health Records (EHRs) is essential in optimizing clinical workflows, minimizing paper use, and encouraging telemedicine to reduce patient travel and energy consumption.

The growth of digital healthcare has resulted in a surge of varied medical data obtained from hospitals, clinics, wearable technology, and remote monitoring systems. Managing this data is challenging due to its inconsistent structure, high dimensionality, and frequent sparsity. Irregularity arises from data being captured at varying intervals, while high dimensionality refers to the vast number of collected variables. Sparsity denotes the considerable absence or lack of recorded data within a dataset. Addressing these challenges requires robust data integration systems, advanced analytical techniques, and machine learning algorithms to extract meaningful insights, identify patterns, and support real-time clinical decision-making.

Electronic Health Records (EHRs) perform a major role in improving sustainability and patient outcomes in the healthcare field. EHRs enhance the comprehensive documentation of patients' medical histories, diagnostic results, and treatment plans, thereby promoting the early identification of diseases, improved prognostic assessments, and tailored treatment alternatives. Furthermore, they accelerate clinical research by providing extensive, anonymized datasets for in-depth analysis. EHRs also contribute to improved patient management by facilitating coordinated care, ensuring that healthcare professionals have access to complete and up-to-date information. This ensembled strategy not only encourages superior care but also aids in minimizing redundant procedures and maximizing the use of healthcare resources and main sustainability in healthcare.

By centralizing persistent data, EHRs advance facilitated care and move forward communication over diverse healthcare groups, eventually driving to more proficient and higher-quality care encounters for patients.

This review focuses three phases:

(v) Deep Learning applied in healthcare data which aims to find out the complexity of integrating various algorithms and systems for the health care data analysis.

(vi) Healthcare data used for validate the different suggested models.

(vii) Key findings of existing models.

(viii) Challenges of the existing models for healthcare analysis.

The whole structure of this review is elaborated in figure 1.

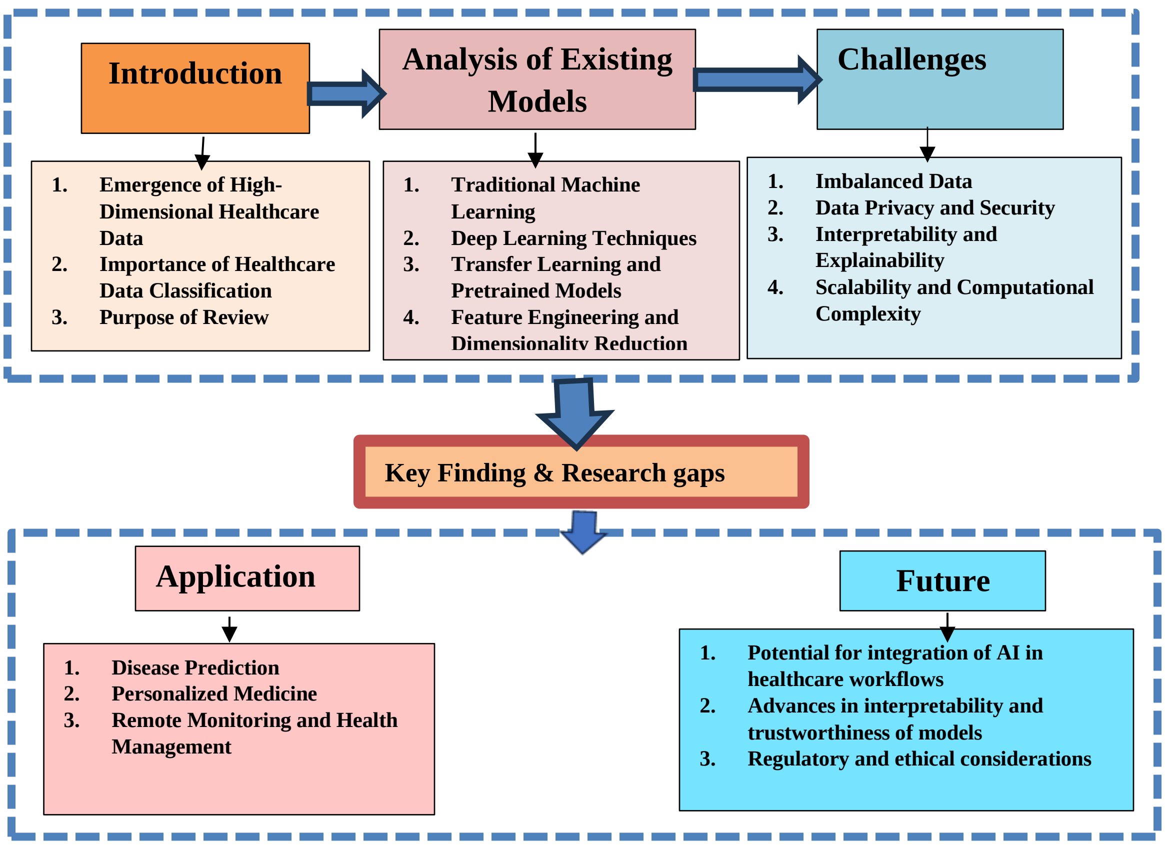

Figure 1. Structure of the review

Related work

In recent years, many dreadful diseases are threatening human beings due to rapidly defiled environment. Therefore, a robust classification model is required to diagnose these diseases with high accuracy and less computational complexity. Generally, biopsy helps to diagnose cancer like deadly diseases. But this is a very painful and time-consuming process. In recent days, with the advent of technology, a robust machine learning based classification model will reduce the human errors happened by inexperience or fatigue and assist to consider a decision before the biopsy in different deadly diseases. However, in case of high dimensional healthcare data, a metaheuristic algorithm is needed to select the significant genes and simultaneously analyses them by an efficient machine learning algorithm [4-6].

Here, the survey on healthcare data classification is extended in four parts, i.e., clinical datasets, preprocessing techniques, Machine Learning/Deep Learning techniques, optimization algorithms. Moreover, the key findings and the research gap of each model has been discussed. Table 1 gives the detailed description of the survey.

Table 1. Analysis of existing models in the healthcare field

Literature	Healthcare Data	Machine Learning/Deep Learning techniques	Key findings	Research Gap
Bahrami et al., 2025 [7]	Cardiac surgery ICU patient data	Linear Discriminant Analysis, CatBoost, Artificial Neural Networks, XGBoost (Ensemble Model)	Hyperparameter tuning enhanced ensemble model sensitivity to 85.84% in ventilator need prediction	Application to other ICU settings and exploration of additional metaheuristic algorithms
Singh et al., 2024 [8]	Multi-disease patient datasets	Decision Tree, Extra Tree Classifier, Extreme Gradient Boosting, LightGBM, Random Forest, Artificial Neural Network	Random Search optimizer achieved 100% accuracy in disease prediction	Validation on real-world clinical data and assessment of model generalizability
Saputra et al., 2023 [9]	Cardiovascular disease prognostic data	KNN, SVM, Random Forest, ANN, NB, LR, Stochastic Gradient Descent, AdaBoost	Stochastic Gradient Descent and Artificial Neural Network models achieved highest performance in CVD prognosis	Exploration of specific optimization methods and validation on larger, diverse populations
Obayya et al., 2023[10]	Histopathological breast cancer images	Convolutional Neural Network (CNN)	Optimized CNN model improved accuracy in breast cancer diagnosis	Application of specific optimization algorithms and validation on larger image datasets

Awotunde et al., 2022 [11]	Leukemia patient data	Convolutional Neural Network (CNN)	Hyperparameter-optimized CNN achieved 99.9% accuracy in leukemia diagnosis	Exploration of other optimization techniques and validation on diverse datasets
Alhhazmi et al., 2024[12]	COVID-19 case data	Deep Learning models	Combined deep learning and Bayesian optimization effectively forecasted COVID-19 confirmed cases	Application to other infectious diseases and assessment of model adaptability
Humayun et al., 2023 [13]	Mammographic images for breast cancer risk assessment	Deep Learning models with CNN architecture	Framework achieved high accuracy in assessing breast cancer risk	Validation on larger, diverse populations and exploration of transfer learning techniques
Wu et al., 2023 [14]	Patient data for differential diagnosis of secondary hypertension	Deep Learning models including CNNs and RNNs	Deep Learning models improved accuracy in differential diagnosis of secondary hypertension	Need for integration with electronic health records for comprehensive analysis
Ali Ahmed et al., 2022 [15]	CT images of COVID-19 patients	Comparative analysis of 2D and 3D Deep Learning models	3D models provided better accuracy in detecting COVID-19 from CT images	Exploration of ensemble methods combining 2D and 3D approaches
Dairi et al., 2021 [16]	COVID-19 transmission data across various regions	Comparative study of Machine Learning models including SVM, Random Forest, and Neural Networks	Identified Neural Networks as superior in forecasting COVID-19 transmission rates	Need for integration of real-time data and continuous model updating

Song et al., 2020 [17]	Histopathological images of gastric cancer	Deep Learning Model with CNN architecture	Model demonstrated high accuracy in detecting gastric cancer from histopathological images	Requirement for large-scale validation and assessment of model generalizability
Das et al., 2024 [18]	Various medical datasets across multiple diseases	Convolutional Neural Networks (CNNs), Logistic Regression, Support Vector Machines (SVMs), Random Forest, Gradient Boosting, AdaBoost	CNNs excel in image-based diagnostics; ensemble methods enhance predictive accuracy in diverse medical conditions	Need for hybrid models combining deep learning and traditional methods for complex healthcare analytics
Shankar et al., 2020 [19]	Fundus images for diabetic retinopathy detection	Synergic Deep Learning Model combining CNNs and Recurrent Neural Networks (RNNs)	Model achieved high accuracy in early detection of diabetic retinopathy	Need for real-time implementation and testing in diverse clinical settings
Khandakar et al., 2021 [20]	Thermogram images of diabetic patients' feet	Convolutional Neural Network (CNN)	Developed a machine learning model achieving high accuracy in early detection of diabetic foot complications	Need for larger-scale studies and real-time implementation in clinical settings
Bacchi et al., 2020 [21]	Ischaemic stroke patient data including imaging and clinical outcomes	CNN+ANN	Demonstrated potential of deep learning in predicting functional outcomes post-thrombolysis in ischaemic stroke patients	Requirement for larger cohort studies and integration with diverse clinical data
Vincent Paul et al., 2022 [22]	Heart disease patient datasets	BP-NN with mRmR	Developed an intelligent framework utilizing deep learning for accurate heart disease prediction	Need for real-time application and validation across diverse populations
Dourado et al., 2019 [23]	Skull computed tomography images for stroke detection	CNN	Proposed an online stroke detection system using deep learning, facilitating timely diagnosis	Need for real-time deployment and validation in emergency clinical environments
Liu et al., 2019	Cerebral stroke	Deep Neural	Developed a hybrid	Need for validation on

[24]	patient data	Network (DNN)	machine learning model for cerebral stroke prediction, achieving improved accuracy on imbalanced medical datasets	larger, diverse populations and exploration of real-time applicability
Scrutinio et al., 2020 [25]	Rehabilitation data of severe stroke patients	Machine Learning models for mortality prediction	Machine learning models effectively predicted mortality after rehabilitation in severe stroke patients	Requirement for integration with clinical workflows and assessment of model interpretability
Zoabi et al., 2021 [26]	Symptomatic data of individuals tested for COVID-19	Machine Learning-based prediction models	Developed a machine learning model predicting COVID-19 diagnosis based on reported symptoms, aiding in early detection	Need for continuous model updates with emerging data and validation across different populations

The citations of the above reviewed papers are mentioned through a pie chat in the figure 2.

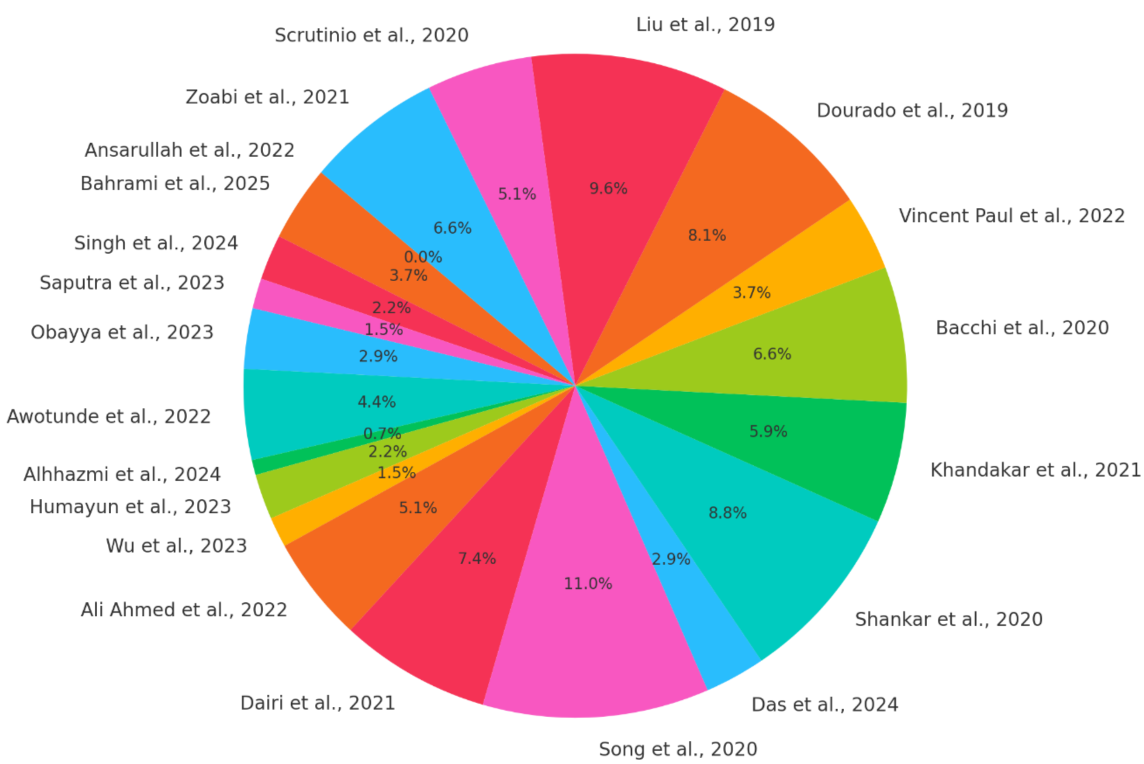

Figure 2. Citation distribution of the selected papers from 2019 to 2024

Analysis of Key findings and Research Gaps

This review extends the literature survey in four parts such as, healthcare datasets, existing Machine Learning/Deep Learning techniques, key findings, and the research gap of each model. After reviewing more than twenty papers within 2019 to 2024, we have listed out the key finding below:

1. Different intelligent framework utilizing deep learning techniques are applied for healthcare analysis.
2. Various ensembled algorithms are used to design a robust model in healthcare.
3. Optimized Machine Learning/Deep Learning models are applied to achieve highest accuracy in early disease detection.
4. Various models are assessed by both high dimensional genomic data and image data.
5. Different preprocessing techniques are used to solve the below issues in healthcare data:
    i. Curse of high dimensionality,
    ii. imbalanced data,
    iii. Redundant gene problem,
    iv. Extraction of biological information
    v. Bias from various factors
    vi. Challenges in cross-platform comparisons
6. Various evaluation metrices like Accuracy, Precision, Recall, F1 Score, ROC-AUC, and Calibration and Interpretability Metrics are used to validate the existing model.

The research gaps of the existing models are also discussed below:

1. Investigating Sophisticated Optimization Methods for Bigger Datasets.
2. To evaluate model generalizability, further optimization techniques must be researched and their effectiveness verified on sizable, varied datasets—especially in actual clinical situations.
3. It is necessary to conduct more research on metaheuristic algorithms and apply them to different intensive care unit settings, with an emphasis on confirming these techniques using clinical data.
4. Optimization Techniques Adapted for Varying Populations: To make sure the techniques are widely applicable and efficient across all demographics research is required to examine certain optimization algorithms and validate them on a variety of patient groups.
5. Model Adaptability for Infectious Diseases: An Evaluation Current models on different infectious diseases have the potential to be used and validated, with an emphasis on evaluating
6. Transfer Learning Applied to Diverse Populations: It may be possible to enhance model performance in a variety of contexts by incorporating transfer learning strategies into healthcare models for bigger, diverse populations.

Application and Future direction

The major application in the field of healthcare data analysis is listed below:

1. Disease Prediction:
   Early disease prediction in different deadly diseases like diabetes, cardiovascular disease, cancer etc.
2. Personalized Medicine:
   Treatment Response Prediction is also possible.
3. Remote Monitoring and Health Management:
   Wearable devices and IoT-based predictions can be possible to reduce the human error.

In the future direction, this review can be extended in the below direction:

i. Potential for integration of AI in healthcare workflows
ii. Advances in interpretability and trustworthiness of models
iii. Regulatory and ethical considerations.

Conclusion

To summarize the finding of this review, we address the research gaps in healthcare optimization and AI applications which holds significant ability to enhance healthcare decision-making, improve patient care, and model accuracy.

Our future research work will focus on exploring advanced optimization strategies, validating models across diverse populations, and adapting methods for real-world healthcare settings, mainly in critical fields such as intensive care of a patient and infectious disease management.

To develop dynamic, healthcare systems will require seamless integration of real-time data and EHRs. Moreover, various ensemble techniques and integrating deep learning with nature inspired optimization algorithms can lead to more robust, reliable, and scalable solutions.

To address these issues, future healthcare models can become more adaptable, generalizable, and capable of improving clinical outcomes in the field of healthcare analytics.

References

1. Karatas, Mumtaz, et al. "Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives." Expert Systems with Applications 200 (2022): 116912.
2. Uddin, Shahadat, et al. "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction." Scientific Reports 12.1 (2022): 6256.
3. Batko, Kornelia, and Andrzej Ślęzak. "The use of Big Data Analytics in healthcare." Journal of big Data 9.1 (2022): 3.
4. Adnan, Mohammed, et al. "Federated learning and differential privacy for medical image analysis." Scientific reports 12.1 (2022): 1953.
5. Omidi, Fatemeh, et al. "Influenza vaccination and major cardiovascular risk: a systematic review and meta-analysis of clinical trials studies." Scientific Reports 13.1 (2023): 20235.
6. Nochaiwong, Surapon, et al. "Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: a systematic review and meta-analysis." Scientific reports 11.1 (2021): 10173.
7. Bahrami, Ali, et al. "Enhancing machine learning performance in cardiac surgery ICU: Hyperparameter optimization with metaheuristic algorithm." PloS one 20.2 (2025): e0311250.
8. Singh, Jagandeep, Jasminder Kaur Sandhu, and Yogesh Kumar. "Metaheuristic-based hyperparameter optimization for multi-disease detection and diagnosis in machine learning." Service Oriented Computing and Applications (2024): 1-20.
9. Saputra, Jayson, et al. "Hyperparameter optimization for cardiovascular disease data-driven prognostic system." Visual Computing for Industry, Biomedicine, and Art 6.1 (2023): 16.
10. Obayya, Marwa, et al. "Hyperparameter optimizer with deep learning-based decision-support systems for histopathological breast cancer diagnosis." Cancers 15.3 (2023): 885.
11. Awotunde, Joseph Bamidele, et al. "An enhanced hyper-parameter optimization of a convolutional neural network model for leukemia cancer diagnosis in a smart healthcare system." Sensors 22.24 (2022): 9689.
12. Alhhazmi, Areej, et al. "Artificial intelligence in healthcare: combining deep learning and Bayesian optimization to forecast COVID-19 confirmed cases." Frontiers in Artificial Intelligence 6 (2024): 1327355.
13. Humayun, Mamoona, et al. "Framework for detecting breast cancer risk presence using deep learning." Electronics 12.2 (2023): 403.
14. Wu, Lin, et al. "Differential diagnosis of secondary hypertension based on deep learning." Artificial Intelligence in Medicine 141 (2023): 102554.
15. Ahmed, Sara Atito Ali, et al. "Comparison and ensemble of 2D and 3D approaches for COVID-19 detection in CT images." Neurocomputing 488 (2022): 457-469.
16. Dairi, Abdelkader, et al. "Comparative study of machine learning methods for COVID-19 transmission forecasting." Journal of biomedical informatics 118 (2021): 103791.
17. Song, Zhigang, et al. "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning." Nature communications 11.1 (2020): 4294.

18. Das, Surajit, et al. "Machine Learning in Healthcare Analytics: A State-of-the-Art Review." Archives of Computational Methods in Engineering (2024): 1-40.
19. Shankar, K., et al. "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model." Pattern Recognition Letters 133 (2020): 210-216.
20. Khandakar, Amith, et al. "A machine learning model for early detection of diabetic foot using thermogram images." Computers in biology and medicine 137 (2021): 104838.
21. Bacchi, Stephen, et al. "Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study." Academic radiology 27.2 (2020): e19-e23.
22. Vincent Paul, Sofia Mary, et al. "Intelligent framework for prediction of heart disease using deep learning." Arabian Journal for Science and Engineering 47.2 (2022): 2159-2169.
23. Dourado Jr, Carlos MJM, et al. "Deep learning IoT system for online stroke detection in skull computed tomography images." Computer Networks 152 (2019): 25-39.]
24. Liu, Tianyu, Wenhui Fan, and Cheng Wu. "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset." Artificial intelligence in medicine 101 (2019): 101723.
25. Scrutinio, Domenico, et al. "Machine learning to predict mortality after rehabilitation among patients with severe stroke." Scientific reports 10.1 (2020): 20127.
26. Zoabi, Yazeed, Shira Deri-Rozov, and Noam Shomron. "Machine learning-based prediction of COVID-19 diagnosis based on symptoms." npj digital medicine 4.1 (2021): 1-5.