SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I3P111

Volume 12 Issue 3, 116-124, March 2025
© 2025 Seventh Sense Research Group®

Original Article

A Comparative Analysis of Deep Learning Models in Diverse AI Healthcare Applications

S. Hemalatha[1], Kiran Mayee Adavala[2], S.N. Chandra Shekhar[3], Pullela SVVSR Kumar[4], A.R. Venkataramanan[5], D. Naga Malleswari[6]

[1]*Department of Computer Science and Business Systems, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu, India.*
[2]*(AI&ML), Kakatiya Institute of Technology and Science, Warangal, Telangana, India.*
[3]*Department of ECE, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India.*
[4]*Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India.*
[5]*Department of Mechanical Engineering, Sona College of Technology, Salem, Tamil Nadu, India.*
[6]*Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fileds, Vaddeswaram, Andhra Pradesh, India.*

[1]*Corresponding Author : pithemalatha@gmail.com*

Received: 07 January 2025 Revised: 08 February 2025 Accepted: 10 March 2025 Published: 29 March 2025

Abstract - *This review article made the comparative analysis quantitatively evaluates four applications of deep learning in healthcare: tuberculosis detection in chest radiography, skin cancer classification, Electronic Health Record (EHR) analysis, and drug discovery. The studies exhibit distinct model architectures and metrics, with tuberculosis and skin cancer classification achieving AUC values of 0.99 and above 0.91, respectively, using CNNs optimized with transfer learning and data augmentation. EHR analysis, utilizing CNN and RNN hybrids, reports AUCs between 0.70 and 0.85 for tasks like disease progression and patient readmission, demonstrating the variability introduced by heterogeneous, sequential data. Drug discovery models employ RNNs for molecular sequence prediction, highlighting a conceptual framework rather than specific performance metrics. The findings indicate that image-based models excel in quantitative performance and scalability, while models for EHR and molecular data face challenges in standardization and interpretability. This analysis underscores the need for data harmonization and explainability to enhance the clinical readiness of deep learning across diverse healthcare domains.*

Keywords - *Deep Learning, Healthcare, Tuberculosis detection, Skin cancer classification, Electronic Health Records (EHR), CNNs, RNNs, Artificial Intelligence (AI).*

1. Introduction

AI has transformed the healthcare sector by improving disease analysis, treatment scheduling, and patient care management. AI-powered solutions, especially deep learning models, have performed exceptionally well in medical imaging, predictive analytics, and precision medicine [1-3]. CNNs and RNNs have been highly effective in automating tasks like disease classification, drug discovery, and EHR analysis, minimizing human intervention while enhancing accuracy [4-6].

AI applications, including virtual health assistants, telemedicine, and robotic-assisted surgeries, continue to advance patient care and accessibility [7]. Despite these innovations, challenges remain in achieving scalable, interpretable, and clinically deployable AI models in healthcare settings [8, 9]. While numerous studies have explored AI's role in healthcare, existing research primarily focuses on isolated applications of deep learning without an in-depth comparative evaluation of different models across diverse clinical use cases [10, 11].

Most studies assess the performance of AI models within specific domains, such as tuberculosis detection, skin cancer classification, or predictive analytics, but few provide a holistic analysis of how these models compare in terms of accuracy, generalizability, interpretability, and real-world feasibility [12].

Furthermore, deep learning models often face domain-specific limitations that affect their effectiveness and clinical adoption. One major challenge is the variability in medical data, as the performance of AI models heavily depends on the quality and quantity of training data. While image-based

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

AI models benefit from large, annotated datasets, text-based models used for Electronic Health Record (EHR) analysis often struggle with unstructured and heterogeneous data formats, making standardization difficult [13, 14]. Another critical issue is the lack of interpretability, as deep learning predictions often function as "black boxes," providing limited explanations for their decisions.

This lack of transparency makes it challenging for clinicians to trust and validate AI-driven recommendations in medical practice [15]. Additionally, scalability and deployment remain significant hurdles. Although deep learning models perform well in controlled settings, their real-world implementation is constrained by regulatory approval processes, data privacy concerns, and high computational requirements. These factors collectively hinder the seamless integration of AI into clinical workflows and patient care [16].

To bridge this research gap, this study presents a comparative analysis of four key deep learning applications in healthcare. The first application is tuberculosis detection in chest radiography. Convolutional Neural Network (CNN) architectures such as AlexNet and GoogLeNet are utilized to automate X-ray screening, improving diagnostic accuracy and accessibility in resource-limited settings [13]. The second application focuses on skin cancer classification, where deep learning models have demonstrated dermatologist-level accuracy in identifying malignant skin lesions, aiding in early detection and reducing the need for invasive diagnostic procedures [1].

The third analysis area is EHR analysis, which employs hybrid deep learning models, combining CNNs and RNNs, to enhance patient risk prediction and clinical decision-making by extracting valuable insights from complex and unstructured medical data [12].

Finally, the study examines drug discovery, where deep learning accelerates molecular sequence analysis and predictive modeling, identifying novel therapeutic compounds and optimizing the drug development process [17]. By systematically evaluating these applications, this study aims to highlight the strengths, limitations, and clinical viability of deep learning models across diverse healthcare domains.

This study systematically evaluates these applications based on key performance metrics, including model architecture, dataset characteristics, interpretability, validation techniques, and scalability. The analysis aims to provide actionable insights into the strengths and limitations of different deep learning models, addressing crucial factors that influence their clinical readiness and real-world deployment.

This article is in the field of AI-driven healthcare. First, it offers a comprehensive comparative analysis, distinguishing itself from existing studies focusing on individual AI applications. By evaluating multiple deep learning models side by side, this study provides a broader perspective on their effectiveness across different healthcare domains. Second, it includes a critical evaluation of interpretability and scalability, examining how AI models perform on both structured and unstructured medical data.

This assessment directly addresses the black-box problem, a significant challenge in deep learning that affects model transparency and clinical trust. Lastly, the study provides guidance for clinical implementation, emphasizing the need for data harmonization, model explainability, and regulatory compliance to facilitate AI adoption in real-world healthcare settings. This research contributes valuable insights for improving AI-driven medical decision-making and enhancing patient outcomes by addressing these crucial aspects.

The rest of this article is organized as follows: Section II offers a comprehensive review of related research on AI-driven healthcare applications, summarizing key advancements, methodologies, and existing challenges. Section III presents a detailed comparative analysis of deep learning models, emphasizing their strengths, limitations, and clinical feasibility across various healthcare domains. Finally, Section IV concludes the study by summarizing key findings and discussing future research directions to improve AI's scalability, interpretability, and real-world adoption in healthcare.

2. Related Works

AI and DL have significantly transformed healthcare by enhancing diagnostics, disease prediction, patient management, and drug discovery. Various studies have explored the integration of machine learning, deep learning, and hybrid AI techniques across different healthcare domains. However, a thorough comparison of their effectiveness across multiple applications is still lacking, particularly in terms of generalizability, scalability, and clinical feasibility. This section reviews existing research on AI-driven healthcare applications, highlighting key advancements, challenges, and research gaps.

2.1. AI in Medical Imaging and Disease Detection

Medical imaging has been one of the most successful applications of AI, with deep learning models achieving near-human performance in detecting diseases from radiographic images. Esteva et al. [1] developed a CNN-based model for skin cancer classification, demonstrating dermatologist-level accuracy in distinguishing between malignant and benign lesions. Similarly, Lakhani and Sundaram [13] applied deep learning for Tuberculosis (TB) detection in chest radiography, using pretrained CNNs (AlexNet and GoogLeNet) to achieve an AUC of 0.99, indicating high sensitivity and specificity.

Notwithstanding these improvements, challenges remain in ensuring the generalizability and interpretability of these models across varied datasets and imaging conditions. Mazurowski et al. [14] emphasized the need for transfer learning and data augmentation to enhance model robustness in medical imaging. Jin et al. [15] explored explainable AI techniques to improve clinician trust in AI-driven diagnostic tools.

2.2. AI in EHR and Predictive Analytics

AI has also been leveraged in Electronic Health Records (EHRs) to enhance patient prediction, progression modeling, and decision support. Shickel et al. [12] provided a comprehensive review of DL techniques for EHR analysis, emphasizing the role of CNNs and RNNs in extracting valuable insights from both structured and unstructured clinical data. Rajkomar et al. [2] developed an AI-driven EHR analysis system that is achieved in predicting patient readmission risks. However, data standardization, privacy, and bias mitigation remain significant challenges. Desai et al. (2022) [18] highlighted the difficulty of integrating AI models with existing hospital workflows due to inconsistent EHR formats and missing data issues. The AI-driven EHR analytics is highly on data, necessitating advanced data pre-processing and feature engineering techniques.

2.3. AI in Drug Discovery and Personalized Medicine

DL has accelerated drug discovery and personalized medicine by enabling faster identification of molecular targets and predicting drug interactions. Mak and Pichika [5] demonstrated how AI models reduce drug development time by 50%, enhancing molecular docking and compound screening efficiency. Soni et al. [17] applied RNN-based architectures to genomic data analysis, improving precision in personalized cancer therapy.

Despite these advancements, interpretability and data quality issues hinder AI adoption in clinical pharmacology. Ghosh et al. [21] emphasized that AI-based drug discovery models require large, high-quality training datasets, often unavailable due to privacy regulations. Furthermore, Yan et al. [23] pointed out that continuous monitoring and validation are necessary to ensure the reliability of AI-driven predictions in drug response modeling.

2.4. Challenges and Gaps in AI-Driven Healthcare

While AI has demonstrated significant potential, several challenges limit its clinical deployment. Data remains a major concern, as handling patient data raises ethical issues and requires robust encryption methods and federated learning approaches to ensure confidentiality and compliance [11]. Another critical challenge is model interpretability, as most DL models function, making it difficult for clinicians to validate AI-driven diagnoses and trust automated decision-making [15]. Additionally, generalizability poses a significant issue, as AI models often struggle to maintain performance across datasets from different populations and geographic regions.

This necessitates improved transfer learning techniques to enhance adaptability and reliability [14]. Furthermore, regulatory and ethical concerns present barriers to AI adoption in healthcare, as compliance with strict regulations such as HIPAA and GDPR is mandatory, often slowing down the integration of AI-driven solutions into clinical practice [16].

2.5. Summary of Literature Review and Need for this Study

The literature highlights significant advancements in AI-driven medical imaging, EHR analytics, and drug discovery. However, most studies focus on individual applications, lacking a comparative evaluation of different deep learning approaches. Existing research does not provide a clear assessment of model performance across diverse healthcare domains, making it difficult to identify the most effective AI methodologies for clinical use. This study addresses this gap by systematically comparing four major deep-learning applications in healthcare tuberculosis detection, skin cancer classification, EHR analysis, and drug discovery. The comparative approach will offer valuable insights into model strengths, weaknesses, scalability, and clinical feasibility, helping researchers and practitioners make informed decisions on AI adoption in healthcare.

3. Comparative Analysis

Deep learning has transformed healthcare by enhancing accuracy and efficiency in diagnostics, personalized medicine, patient outcome prediction, and drug discovery. This study conducts a comparative analysis of four key applications: tuberculosis detection in radiography [13], skin cancer classification [1], Electronic Health Record (EHR) analysis [12], and drug discovery [17]. Each employs specialized deep learning architectures, including CNNs for image-based tasks, RNNs for sequential data, and hybrid models for multimodal data. The analysis evaluates their performance, architectural design, dataset characteristics, clinical relevance, interpretability, and scalability. By identifying strengths and limitations, this study offers the role of DL in healthcare and its potential for real-world clinical integration.

3.1. Four Methods Comparative Analysis

The diverse applications of deep learning in healthcare, from Table 1 (Appendix), diagnostic radiology and drug discovery to EHR analysis and dermatology. Each study highlights deep learning and treatment personalization but also notes the current limitations in data availability across diverse patient populations and clinical environments.

3.2. Dataset Characteristics and Data Pre-processing

The Tuberculosis Detection study [13] utilizes Montgomery County and Shenzhen datasets, ensuring model generalizability across diverse populations. Pre-processing includes image normalization and data augmentation to enhance training data quality. In Drug Discovery [17], the study integrates omics data from clinical trials and preclinical research, which is essential for biomarker identification but demands extensive pre-processing due to dataset heterogeneity. EHR Analysis [12] faces challenges due to the unstructured, multimodal nature of patient records, requiring data normalization, handling of missing values, and embeddings for categorical features. Meanwhile, the Skin Cancer Classification study [1] employs a large dataset of over 129,000 dermatologist-labeled images, including dermoscopic and biopsy-confirmed cases, ensuring high reliability but still facing real-world variability in image quality. Comparative Insight: The effectiveness of deep learning in healthcare strongly depends on data quality and diversity. While EHR and molecular datasets require extensive pre-processing to manage unstructured and heterogeneous data, medical imaging datasets benefit from augmentation and normalization to address variations in imaging conditions and device settings.

3.3. Deep Learning Architectures and Model Design

The Tuberculosis Detection study [13] utilizes AlexNet and GoogLeNet, both optimized for image classification. These networks, pretrained on ImageNet, benefit from transfer learning, enhancing their performance on medical imaging tasks. An ensemble model combining both networks and a radiologist-augmented approach further improves accuracy in uncertain cases. In Drug Discovery [17], Recurrent Neural Networks (RNNs) and transfer learning are crucial in handling sequential omics data, allowing models to recognize patterns in molecular interactions, pharmacokinetics, and toxicology profiles that require time-dependent analysis. EHR Analysis [12] explores a mix of architectures, including CNNs for feature extraction, RNNs for time-series analysis, and autoencoders for dimensionality reduction. These approaches help address EHR-specific challenges, such as temporal dependencies and feature sparsity. In Skin Cancer Classification [1], the study employs GoogleNet Inception v3, leveraging transfer learning to handle the complexity of high-resolution dermatological images. The model is trained end-to-end, capturing subtle variations in lesion appearances directly from raw pixel data.

3.4. Model Validation and Evaluation Techniques

In Tuberculosis Detection [13], validation relied on AUC as the primary metric, with the model achieving an AUC of 0.99. To improve reliability, radiologists reviewed ambiguous cases, increasing specificity to 100% in those instances. Cross-validation on multiple datasets further demonstrated the model's generalizability across different populations. In Drug Discovery [17], evaluation focused on qualitative assessments of molecular interaction predictions and disease pathways. While standard validation metrics like AUC for classification or RMSE for regression could be applied, their use depends on the specific prediction tasks being benchmarked.

For EHR Analysis [12], validation emphasized prediction accuracy and interpretability, but the lack of universal benchmarks makes standardization difficult. However, using large EHR datasets, such as MIMIC-III, has enabled consistent cross-study comparisons.

In Skin Cancer Classification [1], the study used sensitivity and specificity as key performance metrics. The CNN model achieved dermatologist-level accuracy on biopsy-confirmed images, with an AUC exceeding 91%. Cross-validation with biopsy-proven test images further supported the model's reliability and clinical relevance.

3.5. Performance Metrics and Accuracy

The Tuberculosis Detection study [13] achieved an AUC of 0.99, with sensitivity and specificity improving to 97.3% and 100%, respectively, in a radiologist-augmented setup. This underscores the high predictive capability of CNNs in radiological applications, making them well-suited for TB screening in clinical settings. The near-perfect AUC indicates strong reliability in binary classification tasks. In Drug Discovery [18], the study does not rely on a direct metric, as its focus is on the framework for applying deep learning to molecular discovery rather than classification-based evaluation. However, if precision-recall or Root Mean Square Error (RMSE) were applied to molecular property prediction, it could provide more quantifiable comparisons.

The study highlights drug efficacy, requiring the interpretation of complex molecular relationships. For EHR Analysis [12], the study employs prediction accuracy and AUC, but the values vary significantly based on the specific prediction task (e.g., mortality prediction, readmission prediction). AUC scores in EHR applications typically range from 0.70 to 0.85, with some models achieving 0.9 for specific outcomes. This variability reflects the challenges of heterogeneous data structures and feature engineering requirements in health informatics.

In Skin Cancer Classification [1], the model achieved dermatologist-level performance, with an AUC exceeding 91% for keratinocyte carcinoma and melanoma detection. The high AUC, combined with strong sensitivity and specificity, demonstrates its real-world applicability, particularly when tested on biopsy-confirmed samples.

3.6. Data Size and Quality

The Tuberculosis Detection study [13] utilized 1,007 radiographs from different regions, making it one of the smaller datasets among the studies. Although the dataset included images from four distinct sources, enhancing generalizability, its limited sample size remains a constraint. In contrast, Drug Discovery [17] leverages massive, multi-source molecular datasets, often comprising millions of molecular entries from various drug databases.

However, privacy concerns prevent the disclosure of exact sample sizes, and data heterogeneity necessitates extensive pre-processing. For EHR Analysis [12], datasets like MIMIC-III contain hundreds of thousands of patient records, making it one of the largest data sources in this comparison. The high volume of data allows for comprehensive studies, but standardization challenges and data-cleaning complexities remain significant hurdles. Meanwhile, the Skin Cancer Classification study [1] used a large dataset of 129,450 images, including dermoscopic and biopsy-confirmed cases. This dataset, one of the largest dermatology image collections, improves model robustness and compensates for real-world variability in lighting, image quality, and device differences.

3.7. Model Complexity and Architecture Selection

The Tuberculosis Detection study [13] employs AlexNet and GoogLeNet, which, while relatively simpler CNN architectures by modern standards, are optimized with ensemble methods to improve accuracy. These models are well-suited for structured, single-modality image data but may struggle with multimodal data integration. In Drug Discovery [17], the study incorporates RNNs and transfer learning, both essential for processing molecular data sequences. Given the need to analyze drug interactions and pharmacodynamics, these models are architecturally and computationally complex compared to traditional CNNs.

For EHR Analysis [12], the study utilizes RNNs for temporal data processing and CNNs for feature extraction, effectively addressing the heterogeneous nature of EHR data, including numeric, categorical, and unstructured text formats. The combination of RNNs and CNNs reflects the need for advanced architectures to capture both sequential dependencies and structured data variability. Meanwhile, the Skin Cancer Classification study [1] leverages the GoogleNet Inception v3 architecture with transfer learning, allowing it to handle high-resolution images efficiently. While its single CNN setup is ideal for large, labeled datasets, it may lack flexibility in dealing with multimodal or longitudinal data.

3.8. Interpretability and Clinical Integration

In Tuberculosis Detection [13], the radiologist-augmented model enhances interpretability by incorporating human oversight, allowing it to achieve high specificity and sensitivity. This setup makes it well-suited for clinical deployment, particularly in resource-limited settings where radiologist availability is scarce. In Drug Discovery [17], interpretability is a significant experiment omics data. The study emphasizes the need for explainable AI techniques, such as attention mechanisms, to improve clinician trust and regulatory compliance by revealing drug mechanisms at a molecular level. For EHR Analysis [12], interpretability remains a major limitation, as DL models. While techniques like attention layers can help, the heterogeneous nature of EHR data and the variety of predictive tasks make it difficult to establish standardized interpretability metrics. The clinical adoption of these models depends on their ability to provide transparent and justifiable predictions. Meanwhile, in Skin Cancer Classification [1], interpretability is achieved primarily through accuracy and specificity in biopsy-proven images, making the model reliable for dermatologist-level classification. However, additional interpretability techniques could improve end-user trust and decision-making confidence for clinical deployment.

3.9. Generalizability and Scalability

In Tuberculosis Detection [13], the model demonstrates good generalizability across multiple regions, but its small dataset size limits scalability. Training on larger, more diverse datasets could improve its applicability in broader clinical settings. In Drug Discovery [17], scalability is a strength due to the high volume of available molecular data. However, data privacy and interpretability issues present major hurdles. While deep learning models in drug discovery can generalize across multiple disease areas, they require disease-specific validation to ensure reliability.

For EHR Analysis [12], scalability is challenged by variations in EHR formats and standards across institutions. While models may scale effectively within standardized healthcare systems, they often struggle with cross-institutional generalizability due to the lack of data harmonization. In Skin Cancer Classification [1], the model benefits from a large and diverse dataset, supporting generalizability. Additionally, its potential integration into mobile health applications enhances scalability in under-resourced areas. However, deployment across different devices would require further optimization to account for variations in camera quality and image conditions.

4. Conclusion

From this comparative analysis of DL in healthcare analysis, this article reveals both the usefulness and the challenges of applying DL in healthcare. Image-based tasks, such as tuberculosis and skin cancer detection, demonstrate high quantitative performance with near-perfect AUC values, making them readily deployable in clinical settings, especially in resource-limited areas. These models benefit from CNN architectures and are relatively straightforward to interpret, enhancing clinical applicability. On the other hand, drug discovery and EHR analysis tackle more complex, heterogeneous data, where RNNs and hybrid architectures are essential but introduce challenges in model interpretability and data standardization. While these models offer transformative potential in personalized medicine and clinical prediction, they require significant advancements in explainability and data harmonization to achieve widespread clinical acceptance. Future research must prioritize scalability, interpretability, and privacy to harness the DL in healthcare, ensuring that AI-driven tools can be trusted and utilized effectively across diverse healthcare settings.

Acknowledgments

Each author has made significant contributions to the research, analysis, and writing of this article. Their collective expertise has strengthened the study's depth and impact. The specific contributions are as follows:

S. Hemalatha: Conceptualization of the research, methodology design, manuscript drafting, and overall study supervision.

Kiran Mayee Adavala: Data collection, model implementation, and experimental validation of deep learning approaches.

S.N. Chandra Shekhar: Literature review, statistical analysis, and comparative study of performance metrics across different applications.

Pullela SVVSR Kumar: Preprocessing techniques, optimization of deep learning models, and visualization of results (AUC and ROC curves).

A.R. Venkataramanan: Interpreting findings, critical revisions, and ensuring alignment with clinical applicability.

D. Naga Malleswari: Manuscript editing, formatting, and final proofreading to maintain clarity and coherence.

References

[1] Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature*, vol. 542, pp. 115-118, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[2] Alvin Rajkomar, Jeffrey Dean, and Isaac Kohane, "Machine Learning in Medicine," *The New England Journal of Medicine*, vol. 380, no. 14, pp. 1347-1358, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Eric J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," *Nature Medicine*, vol. 25, pp. 44-56, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Salman Razzaki et al., "A Comparative Study of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis," *arXiv Preprint*, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kit-Kay Mak, and Mallikarjuna Rao Pichika, "Artificial Intelligence in Drug Development: Present Status and Future Prospects," *Drug Discovery Today*, vol. 24, no. 3, pp. 773-780, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Daniel A. Hashimoto et al., "Artificial Intelligence in Surgery: Promises and Perils," *Annals of Surgery*, vol. 268, no. 1, pp. 70-76, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[7] Fei Jiang et al., "Artificial Intelligence in Healthcare: Past, Present and Future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230-243, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Deekshaa Khanna, "Use of Artificial Intelligence in Healthcare and Medicine," *International Journal of Innovations in Engineering Research and Technology*, vol. 5, no. 12, pp. 1-4, 2018. [Google Scholar] [Publisher Link]

[9] Sirina Keesara, Andrea Jonas, and Kevin Schulman, "Covid-19 and Health Care's Digital Revolution," *The New England Journal of Medicine*, vol. 382, no. 23, pp. 1-3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Reed T. Sutton et al., "An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success," *NPJ Digital Medicine*, vol. 3, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ziad Obermeyer, and Ezekiel J. Emanuel, "Predicting the Future-Big data, Machine Learning, and Clinical Medicine," *The New England Journal of Medicine*, vol. 375, no. 13, pp. 1216-1219, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[12] Benjamin Shickel et al., "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589-1604, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] Paras Lakhani, and Baskaran Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," *Radiology*, vol. 284, no. 2, pp. 574-582, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] Guang-Zhong Yang et al., "Medical Robotics-Regulatory, Ethical, and Legal Considerations for Increasing Levels of Autonomy," *Science Robotics*, vol. 2, no. 4, pp. 1-2, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Alexander Selvikvåg Lundervold, and Arvid Lundervold, "An Overview of Deep Learning in Medical Imaging Focusing on MRI," *Journal of Medical Physics*, vol. 29, no. 2, pp. 102-127, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yogesh K. Dwivedi et al., "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management*, vol. 57, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yasunari Matsuzaka, and Ryu Yashiro, "Applications of Deep Learning for Drug Discovery Systems with BigData," *BioMedInformatics*, vol. 2, no. 4, pp. 603-624, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Elias Hossain et al., “Natural Language Processing in Electronic Health Records in Relation to Healthcare Decision-Making: A Systematic Review,” *Computers in Biology and Medicine*, vol. 155, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Badawy, Mohammed, Nagy Ramadan, and Hesham Ahmed Hefny et al., “Healthcare Predictive Analytics Using Machine Learning and Deep Learning Techniques: A Survey,” *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, pp. 1-45, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Mohamed Khalifa, and Mona Albadawy, “AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency,” *Computer Methods and Programs in Biomedicine Update*, vol. 5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Nicholas J. Schork, *Artificial Intelligence and Personalized Medicine*, Precision Medicine in Cancer Therapy, vol. 178, pp. 265-283, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Ahmed M. Alhuwaydi, “Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions-A Narrative Review for a Comprehensive Insight,” Risk Management and Healthcare Policy, vol. 17, pp. 1339-1348, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Muhammad Rafiq et al., “Predictive Analytics Support for Complex Chronic Medical Conditions: An Experience-Based Co-Design Study of Physician Managers’ Needs and Preferences,” *International Journal of Medical Informatics*, vol. 187, 2024. [CrossRef] [Google Scholar] [Publisher Link]

Appendix

Table 1. Comparative analysis of deep learning case studies

Methods	Key Points	Methodology	Results	Strengths	Limitations
Deep Learning for Tuberculosis Detection on Chest Radiography [13]	The study evaluates deep DCNNs for detecting TB in chest X-rays. Automated TB detection is especially beneficial for regions with limited access to radiologists.	The study utilizes two DCNN architectures-AlexNet and GoogLeNet-with both pretrained and untrained models on ImageNet. The researchers trained the networks on deidentified datasets and used data augmentation techniques. An ensemble approach of both networks was applied, and in cases of classifier disagreement, radiologists reviewed the images to improve accuracy.	The best-performing model achieved 0.99, significantly improving accuracy when using pretrained models and data augmentation. The radiologist-augmented workflow achieved a sensitivity of 97.3% and a specificity of 100%.	High accuracy and sensitivity suggest this model suits TB screening in low-resource settings. The radiologist-augmented approach leverages human expertise for ambiguous cases for the model.	This method is dataset-dependent, and the efficacy might vary with different populations or image qualities. Moreover, generalizability could be constrained by the variations in radiograph standards across regions.
DLDR Medicine [17]	This review discusses the role of DL in precision medicine and drug discovery, emphasizing its potential to individualize treatment strategies based on patient-specific molecular data.	The paper reviews various deep learning frameworks, including RNNs and transfer learning, applied to analyze omics data and predict drug efficacy. It highlights how deep learning can automate data mining from vast molecular databases and assist in de novo drug design.	The authors outline how deep learning enhances drug discovery by identifying biomarkers and therapeutic targets more efficiently than traditional methods. They cite examples like patient-specific epigenetic signatures for personalized cancer therapies.	The review underscores the versatility of deep learning in integrating diverse datasets for more comprehensive drug discovery. Its application in precision medicine holds promise for transforming treatment plans across various medical fields, including oncology and nephrology.	The paper highlights limitations in model interpretability and the need for vast, high-quality training data, which can be a barrier to drug discovery due to privacy concerns and dataset heterogeneity. Additionally, deep learning models for drug discovery require substantial computational resources.
Deep EHR for EHR Analysis [12]	This article surveys advancements in deep learning for EHR analysis, covering applications in information extraction, outcome prediction, and patient phenotyping.	The paper reviews DL techniques, such as CNNs, RNNs, and autoencoders, applied to structured and unstructured EHR data for clinical tasks. It identifies model architectures tailored to EHR complexities, including health records' heterogeneous and sequential nature.	The paper finds that deep learning models outperform traditional methods in predictive accuracy and efficiency across a range of EHR-, such as disease prediction and clinical decision support.	This survey provides a comprehensive overview of DL applications in EHR, addressing how these methods enhance clinical workflow and decision-making. The models' ability to automate feature extraction directly from raw data is advantageous for clinical informatics.	Challenges include limited model interpretability and the lack of standardized benchmarks across institutions, which impede model generalizability. Privacy and data security are also critical concerns, as EHR data is sensitive and prone to biases from missing or incomplete records.

CNN for Skin Cancer Classification [1]	This study develops a CNN to classify skin cancer using a large dataset of skin lesion images. It focuses on automating the differentiation between benign and malignant skin lesions, aiming for early cancer detection.	A CNN model based on the GoogleNet Inception v3 architecture was trained on over 129,000 images and validated against the performance of dermatologists on biopsy-proven images. The model used end-to-end training without hand-crafted feature extraction, benefiting from transfer learning.	CNN performed on par with 21 board-certified dermatologists in classifying keratinocyte carcinomas and melanomas. Sensitivity and specificity metrics were high, demonstrating the model's potential for clinical application.	The study leverages a massive and diverse dataset, enhancing the model's robustness against image variations. The model’s capability to perform dermatologist-level classifications suggests a significant impact on accessible cancer screening.	The CNN's performance may be limited in real-world clinical settings where image quality and patient diversity vary. Additionally, the reliance on labeled biopsy data could limit the model's scalability in areas with limited biopsy-proven images.