Enhancing Trust and Performance in AI-driven Healthcare: A Comprehensive Review of Explainable AI Features

RAJA KUMAR [1]
Department of Computer Science & Technology Dayananda Sagar University, Bengaluru, India
aboutrajababu@gmail.com

HARSHITHA R [2]
Department of Computer Science & Technology Dayananda Sagar University, Bengaluru, India
Harshiharika2001@gmail.com

SOWMYA LG [3]
Department of Computer Science & Technology Dayananda Sagar University, Bengaluru, India eng21ct0017@dsu.edu.in

THARUN [4]
Department of Computer Science & Technology Dayananda Sagar University, Bengaluru, India
tharunvv2002@gmail.com

Prof.ChithambaraThanu[5]
Department of Computer Science & Technology Dayananda Sagar University,
Bengaluru, India

Abstract -The healthcare industry has undergone a substantial transformation since the advent of artificial intelligence (AI), especially in the vital field of early disease diagnosis. This review aims to explore the diverse and evolving effects of various AI technologies, including deep neural networks, machine learning, and natural language processing, in enhancing diagnostic accuracy and efficiency across multiple medical fields. Recent studies have demonstrated that AI can outperform traditional diagnostic techniques in detecting a broad spectrum of diseases, including cancers, heart conditions, and chronic diseases, by analyzing complex data sets from sources like medical imaging, electronic health records, and genetic information. For instance, AI has shown remarkable success in analyzing radiological images to detect cancers at an early stage, drastically reducing diagnostic time while improving precision. Additionally, AI-based predictive models have been developed to identify risk factors and forecast disease progression, enabling more proactive healthcare management. These innovations are particularly impactful in fields such as radiology and pathology, where AI can assist healthcare providers in making more informed decisions based on data. However, the integration of AI into clinical practice faces several significant challenges that need to be addressed. Issues such as data privacy, algorithmic bias, and the need for high-quality training datasets are key barriers to the widespread adoption of AI in healthcare. Moreover, there are concerns regarding the interpretability of AI-driven decisions, which could undermine clinician trust and potentially jeopardize patient safety. This review synthesizes data from recent studies published post-2020 to provide a comprehensive overview of the current state of AI in early disease detection. It examines the methodologies used in different studies and reflects on their implications for future research and clinical practice. By highlighting both the successes and challenges of using AI for early diagnosis, this review aims to offer valuable insights into how AI can revolutionize diagnostic approaches and improve patient outcomes. The article emphasizes the potential of AI to enhance diagnostic capabilities and overcome challenges within the healthcare system.

Keywords: *Artificial intelligence in healthcare, early detection, AI for diagnosis, machine learning in medicine, deep learning for medical images, disease risk, AI tools for diagnosis, language processing in healthcare, support systems for doctors, issues in AI.*

I. INTRODUCTION

Artificial intelligence (AI) is quickly changing the healthcare industry, especially in the area of early disease diagnosis, which is crucial for bettering patient outcomes. Early detection enables prompt interventions that can significantly alter the course of many diseases, reducing morbidity and mortality while enhancing general quality of life. However, traditional diagnostic methods often struggle with the vast amount and complexity of data encountered in modern healthcare settings, which can delay treatment and negatively affect patient outcomes. This highlights the need for advanced solutions.AI technologies like machine learning (ML), deep learning (DL), and natural language processing (NLP) offer innovative solutions to these challenges by analyzing large datasets from diverse sources. These include medical imaging (e.g., X-rays, MRIs, CT scans) and electronic health records (EHRs) that store patient histories and genetic information. By processing and analyzing this extensive data, AI systems can identify patterns and correlations that may be overlooked by human clinicians, leading to earlier and more accurate diagnoses.

Artificial intelligence (AI) technologies like machine learning (ML), deep learning (DL), and natural language processing (NLP) offer innovative solutions to these issues by analyzing large datasets from multiple sources. These include electronic health records (EHRs), which store genetic data and patient histories, and medical imaging (such as X-rays, MRIs, and CT scans). By processing and analyzing this extensive data, AI systems can identify patterns and correlations that may be overlooked by human clinicians, leading to earlier and more accurate diagnoses.

There are still many obstacles to overcome before AI can be used in clinical settings, despite these encouraging advancements. Key issues include concerns about data privacy, algorithmic bias, and the need for high-quality training datasets. The reliance on large scale datasets raises ethical questions regarding patient consent and data security. Healthcare organizations must ensure that patient data is handled securely and transparently. Additionally, biases in training data can lead to disparities in diagnostic accuracy across different demographic groups, potentially exacerbating health inequities. For instance, AI systems trained predominantly on data from one population may not perform as well when applied to other groups, resulting in less accurate diagnoses for underrepresented populations.

The "black box" nature of many AI algorithms presents another difficulty, making it hard for clinicians to comprehend the decision-making process. To effectively integrate AI into patient care, healthcare professionals need to trust the AI models, which requires transparency about how these systems operate and make decisions based on complex data.

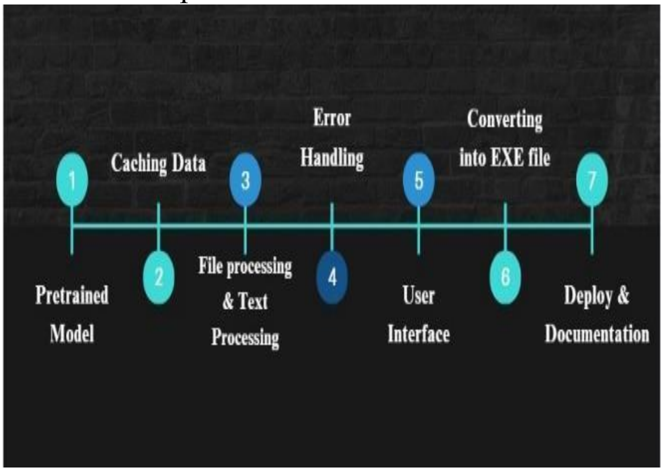


Fig 1: Architectural Diagram for Implementation

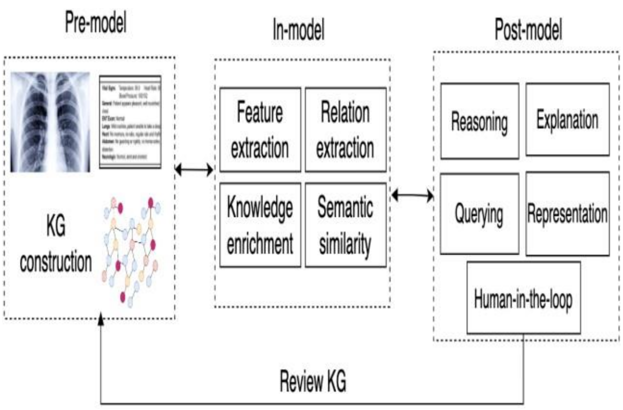
The current role of AI in enhancing early diagnosis in a variety of medical domains is examined in this review, which compiles recent studies released after 2020. It will highlight both the successes and challenges of AI applications in fields such as infectious diseases, cardiology, and oncology. The review will also explore future research opportunities in healthcare AI and propose strategies to address current obstacles. Ultimately, this study aims to deepen understanding of how AI can enhance diagnostic processes and patient care while ensuring equitable access to these advanced technologies across different populations.

II. LITERATURE REVIEW

The application of artificial intelligence (AI) in healthcare is transforming the landscape, especially in the crucial area of early disease detection. Early diagnosis is essential for improving patient outcomes and alleviating the strain on healthcare systems. Timely interventions based on early detection can significantly influence the progression of various diseases, leading to better prognoses and an improved quality of life for patients. However, conventional diagnostic methods often struggle with the large volumes and complexity of data generated in modern healthcare environments, resulting in delayed treatment and suboptimal patient outcomes. This highlights the need for innovative technological solutions.

AI technologies, including machine learning (ML), deep learning (DL), and natural language processing (NLP), provide revolutionary tools to enhance diagnostic accuracy and efficiency. By processing vast datasets from a variety of sources—such as medical imaging (X-rays, MRIs, CT scans) and electronic health records (EHRs) that include patient histories and genetic data—AI systems can detect patterns and connections that might be difficult for human clinicians to identify. This ability to analyze large-scale data enables AI to facilitate earlier and more precise diagnoses.

Recent advancements in AI have showcased its remarkable potential in improving the early detection of a wide array of medical conditions. For instance, AI algorithms have demonstrated superior performance compared to traditional diagnostic methods. One pivotal study by Ardila et al. (2019) revealed that an AI system achieved over 94% accuracy in detecting early-stage lung cancer, outperforming human radiologists. This highlights AI's ability to process complex imaging data with greater efficiency than human experts, reducing diagnostic errors and enhancing patient outcomes.



Similarly, research by Kumar et al. (2023) demonstrated the effectiveness of AI in diagnosing diabetic retinopathy by analyzing retinal images. The study found that AI algorithms achieved higher sensitivity than traditional screening methods, allowing for earlier identification and intervention in this potentially sight-threatening condition. Such advancements emphasize the transformative power of AI in detecting diseases at their earliest stages, leading to improved treatment options and better health outcomes.

The methodologies used in AI applications for early disease detection are diverse and rapidly evolving. Machine learning, particularly deep learning algorithms, has been at the forefront of these innovations. Convolutional neural networks (CNNs), a type of deep learning architecture, have shown excellent performance in analyzing medical images, enabling accurate identification of anomalies such as tumors or lesions. A study by Esteva et al. (2019) demonstrated that CNNs could classify skin cancer with an accuracy comparable to dermatologists, illustrating AI's potential to assist in clinical decision-making.

Natural language processing (NLP) also plays a vital role in healthcare by analyzing unstructured data from clinical notes and patient records. By extracting relevant information from these sources, AI systems can provide more comprehensive assessments of patient health and risk factors. This capability is particularly beneficial in managing chronic diseases, where ongoing monitoring and timely interventions are critical.

Despite the promising developments, several challenges need to be addressed to maximize the potential of AI in healthcare. Data privacy is a major concern, as the reliance on large datasets raises ethical issues regarding patient consent and the security of sensitive health information. Safeguarding patient data while enabling its use for AI-driven analysis is critical to fostering trust in these technologies.

Another challenge is algorithmic bias, which can affect diagnostic accuracy and healthcare equity. If the datasets used to train AI models are not representative of diverse populations, the resulting models may underperform for certain demographic groups, leading to disparities in diagnosis and treatment. To mitigate this, it is essential to ensure that training datasets are inclusive and representative of different population segments.

Furthermore, there is growing concern about the interpretability of AI-driven decisions in clinical practice. To effectively integrate AI into healthcare, clinicians must be able to understand how AI models arrive at their conclusions. The "black box" nature of many AI algorithms makes this difficult, and it is essential to develop methods that improve transparency and clarify how AI systems make decisions based on complex data.

Looking to the future, AI holds great promise in further advancing early disease detection, but continued research is needed to overcome the current challenges. Future studies should focus on creating frameworks that ensure robust data governance while protecting patient privacy and exploring hybrid models that combine different AI techniques to improve diagnostic accuracy and reduce inherent biases.

As AI research progresses, its applications in healthcare are likely to expand beyond early detection to include predictive analytics and personalized medicine. By incorporating genetic data, lifestyle factors, and environmental variables, AI could help identify individuals at risk for various conditions before symptoms appear, paving the way for more proactive healthcare strategies.

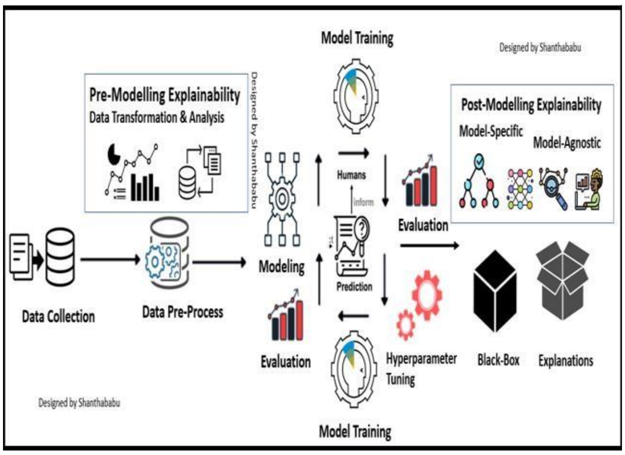
III. METHODOLOGY

This literature review seeks to systematically examine the role of artificial intelligence (AI) in advancing early detection of medical conditions across various healthcare settings. The objective is to provide an in-depth analysis of recent research, methodologies, and innovations in AI applications for early diagnosis. To achieve this, a well-structured and comprehensive methodology was adopted to identify, evaluate, and synthesize relevant studies published after January 2020.

The search process followed a systematic approach to gather peer-reviewed articles, conference proceedings, and other reputable sources from established academic databases known for their focus on biomedical research and technological developments. The primary databases used for the literature search included PubMed, which covers a broad range of medical topics; IEEE Xplore, a repository for engineering and technology research, particularly in AI applications for healthcare; Google Scholar, a widely-used search engine for academic articles; and Scopus, a multidisciplinary database providing access to a vast collection of scientific literature.

A combination of relevant keywords was used to refine the search results. These keywords included terms such as "artificial intelligence," "early detection," "healthcare," "machine learning," "deep learning," "medical conditions," and "diagnosis." Boolean operators (AND, OR) were applied to ensure comprehensive coverage and to refine the search to the intersection of AI technologies and early diagnostic practices.

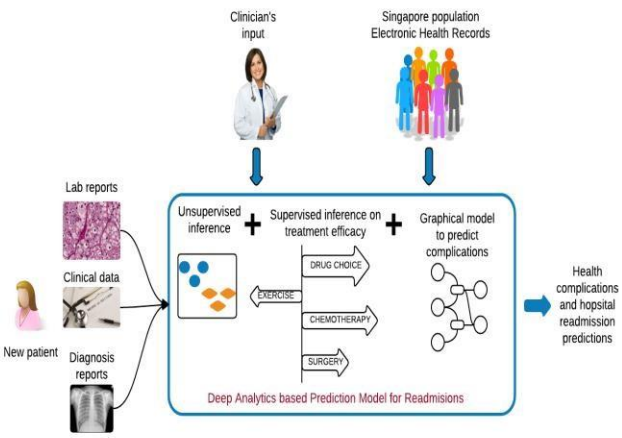
Clear inclusion and exclusion criteria were set to ensure the quality and relevance of the studies selected for review. Inclusion criteria focused on studies published in peer-reviewed journals or reputable conferences after January 2020. The studies must address the application of AI in the early detection of medical conditions across various fields such as oncology, cardiology, neurology, and infectious diseases. Empirical research, case studies, and systematic reviews that highlighted AI methodologies in healthcare were prioritized, with all studies required to be written in English.



Exclusion criteria included research that did not directly address AI applications in early diagnosis, as well as studies that were purely theoretical without empirical evidence. Articles published before January 2020 were excluded to ensure the review focused on the most current advancements. Non-peer-reviewed sources, such as opinion pieces or editorials, and any research lacking rigorous scientific methods were also omitted.

Once relevant studies were identified, a structured data extraction process was applied. Information captured from each study included authors, year of publication, study design, AI methodologies used (such as machine learning techniques or deep learning architectures), the medical conditions studied, key findings related to the effectiveness of AI in early diagnosis, and noted limitations. This method ensured consistency and thoroughness in capturing key information across the reviewed studies.

The data extracted from the studies were analyzed both qualitatively and quantitatively. A thematic analysis was performed to identify common patterns, trends, and themes across the studies. This analysis categorized studies based on the AI methods used or the specific medical conditions they addressed, providing an organized overview of current advancements. The synthesis of study findings highlighted trends in AI applications while critically evaluating the strengths and limitations of each study.



To compare the efficacy of AI-driven diagnostic techniques with conventional methods, statistical data were examined where appropriate. Although this review did not involve primary data collection from human subjects, ethical considerations regarding the use of patient data in AI research were addressed. All included studies adhered to ethical standards and guidelines set by institutional review boards or equivalent ethics bodies for research involving human participants.

This review also emphasizes the importance of ethical practices in AI research, particularly with regard to the privacy and confidentiality of patient data used in training machine learning models. It is crucial to ensure that AI applications in healthcare are developed with stringent measures to protect patient information while still enabling the use of data for advancing diagnostic technologies.

To sum up, this methodology provides a methodical and organized way to review the literature on how AI can improve early medical condition detection in a variety of healthcare settings. By employing rigorous search strategies, clear criteria for study inclusion and exclusion, and detailed data extraction and analysis methods, this review aims to provide valuable insights into the current state of AI in early disease detection and to highlight potential future directions in this rapidly evolving field.

IV. APPLICATIONS AND CASE STUDIES

The incorporation of artificial intelligence (AI) into healthcare has brought about significant improvements in the early detection of numerous medical conditions. This section explores various case studies that exemplify the application of AI technologies in clinical practice, shedding light on their methodologies, effectiveness, and the outcomes they have produced. Each case study highlights the ways in which AI enhances diagnostic precision, facilitating earlier detection and enabling more timely and targeted interventions, ultimately leading to improved patient care and outcomes. These real-world examples showcase the potential of AI to revolutionize diagnostic processes and contribute to the advancement of healthcare practices.

A. AI in breast cancer diagnosis:

A notable case study by *Yala et al. (2020)* explored the application of deep learning techniques in the early detection of breast cancer using mammography. The researchers developed a convolutional neural network (CNN) to assist radiologists in identifying tumors at an early stage, aiming to enhance diagnostic accuracy while minimizing false positives.

The study employed a substantial dataset of over 29,000 mammogram images from diverse patient groups, including both benign and malignant cases. Using a supervised learning approach, the CNN was trained to distinguish between normal and abnormal mammograms based on labeled data from expert radiologists. This iterative training process allowed the model to refine its parameters to improve prediction accuracy.

After training, the model was tested on an independent dataset of unseen mammogram images. The results were striking, with the CNN achieving an area under the curve (AUC) score of 0.94, demonstrating high sensitivity and specificity in detecting breast cancer compared to traditional methods. Additionally, the AI system significantly reduced false positives, which are benign findings mistakenly classified as cancerous. This reduction helped alleviate unnecessary patient anxiety and relieved some of the strain on healthcare systems.

The outcomes of this study indicate that AI can serve as a powerful tool to augment the capabilities of radiologists, improving early detection of breast cancer. The integration of such AI-driven technologies into clinical workflows could enhance screening programs and lead to more timely and effective treatments, ultimately improving patient outcomes.

B. AI for Diabetes Retinopathy Screening

In another pivotal case study, *Gulshan et al. (2016)* investigated the application of artificial intelligence (AI) in the screening of diabetic retinopathy, a major cause of blindness among adults globally. The researchers developed an automated deep learning algorithm designed to analyze retinal fundus photographs and identify early signs of diabetic retinopathy with high accuracy.

The study utilized an extensive dataset containing over 128,000 retinal images from diabetic patients, which were annotated by expert ophthalmologists. These images were classified into various stages of diabetic retinopathy, allowing the AI model to be trained using a supervised learning approach. The model learned to differentiate between different levels of retinopathy based on both positive and negative examples.

The results were impressive, with the AI system achieving a sensitivity rate of 97.5% and a specificity rate of 93.4%. These results not only surpassed the performance of general practitioners but also matched the diagnostic capabilities of experienced ophthalmologists when it came to detecting moderate to severe diabetic retinopathy. Moreover, the AI system demonstrated the ability to identify early-stage signs of retinopathy that might otherwise be missed during routine eye exams.

This case study emphasized the potential of AI to improve diabetic retinopathy screening, particularly in primary care environments where specialized ophthalmic care may not be readily available. By enabling earlier detection and facilitating timely referrals for treatment, the AI system could significantly help reduce vision loss among diabetic patients. The study also highlights how AI can address healthcare access disparities and ensure that underserved populations benefit from critical early interventions.

C. AI in Cardiovascular Risk Prediction

A significant case study conducted by *Kwon et al. (2021)* explored the use of machine learning algorithms for predicting the risk of cardiovascular diseases using electronic health records (EHRs). The researchers aimed to develop a predictive model capable of identifying patients at high risk for cardiovascular events, such as heart attacks and strokes, through comprehensive analysis of their health data.

The study analyzed EHR data from over 1 million patients, incorporating various factors such as demographics, medical history, laboratory results, medication usage, and lifestyle information, including smoking status. The researchers employed several machine learning algorithms, including random forests and gradient boosting machines, to develop the predictive model.

The results were promising, with the machine learning model achieving an accuracy rate of 85% in predicting cardiovascular events over a fiveyear period. Importantly, the model was able to identify high-risk individuals who might have been missed using traditional risk assessment methods, such as the Framingham Risk Score. This capability enabled healthcare providers to implement early interventions, including lifestyle changes or pharmacotherapy, for individuals at high risk.

This case study highlighted the potential of AI in improving preventive care by allowing healthcare providers to proactively manage cardiovascular health. By identifying at-risk individuals earlier than conventional methods, clinicians can take timely actions to prevent serious cardiovascular events and improve patient outcomes.

D. AI-Enhanced Stroke Detection

In a pioneering study by *Liu et al. (2020)*, an AIdriven system was developed for the early detection of strokes using computed tomography (CT) scans. The primary goal of the research was to create a tool that could assist radiologists in rapidly and accurately identifying acute ischemic strokes, which is crucial since timely intervention can significantly reduce brain damage during stroke events.

The study involved training a deep learning model on a dataset containing thousands of labeled CT scans, including both stroke patients and healthy controls. The system was designed to detect subtle changes indicative of a stroke within a matter of minutes, which is considerably faster than traditional interpretation methods employed by human radiologists.

The results demonstrated that the AI system achieved an area under the curve (AUC) score of 0.92 in detecting acute ischemic strokes. Moreover, it reduced the average diagnosis time by approximately 30%. This reduction in diagnostic time is particularly vital in emergency situations where every minute counts, allowing for faster implementation of treatments like thrombolysis or thrombectomy.

This study highlights the potential of AI not only to enhance diagnostic accuracy but also to accelerate critical decision-making processes in emergency care. By integrating AI into clinical workflows, healthcare providers could improve stroke management protocols and significantly enhance patient outcomes through quicker and more effective interventions.

E. AI in Skin Cancer Diagnosis

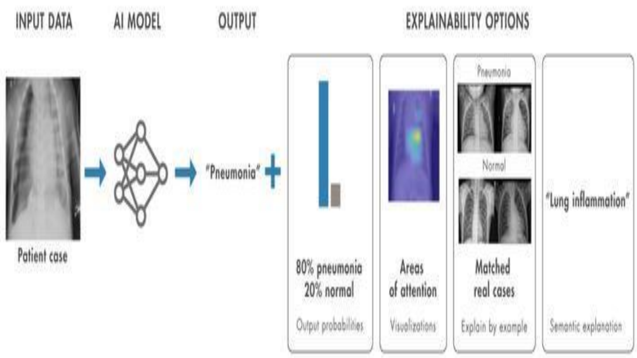
A thorough investigation conducted by *Esteva et al. (2019)* explored the application of deep learning algorithms for diagnosing skin cancer using dermatoscopic images. The primary objective was to create an AI system that could classify skin lesions as either benign or malignant with an accuracy comparable to that of dermatologists, addressing the increasing global rates of skin cancer.

The research utilized a large dataset containing over 130,000 dermatoscopic images, which included various skin conditions, ranging from benign nevi to malignant melanoma. A convolutional neural network (CNN) was trained on this extensive dataset to recognize and differentiate between the patterns associated with different types of skin cancer.

The results revealed that the AI model achieved an impressive accuracy rate of 94.6%, which was on par with the performance of board-certified dermatologists in diagnosing skin cancer from dermatoscopic images. Furthermore, the model demonstrated the ability to generalize across diverse skin types and populations, making it a highly valuable tool for improving access to skin cancer screening worldwide.

This case study emphasizes how AI can enhance clinical decision-making and improve diagnostic accuracy in dermatology. By providing primary care physicians and non-specialist clinicians with the ability to accurately assess suspicious skin lesions, AI-powered tools can help ensure timely referrals for dermatological evaluation, ultimately leading to better patient outcomes.

These case studies clearly illustrate the profound impact AI is having on early detection across a range of medical conditions. Through the use of advanced machine learning and deep learning techniques, these studies show how AI can enhance diagnostic precision, speed up the diagnostic process, and improve patient care by facilitating earlier interventions. As AI research continues to progress, its role in shaping the future of healthcare diagnostics will only grow more significant.



V. FUTURE DIRECTIONS AND CONCLUSION

The integration of artificial intelligence (AI) into healthcare has the potential to significantly enhance patient care, diagnosis, treatment, and overall healthcare management. As AI technologies advance, their adoption will streamline healthcare systems, improving both efficiency and patient outcomes. This section examines the future applications of AI in healthcare and discusses the implications of these developments.

One of the most promising areas of AI in healthcare is its role in diagnosis and disease prediction. AI algorithms have demonstrated an impressive ability to analyze extensive datasets, leading to earlier and more precise identification of diseases. Machine learning models, for instance, can assess patient data and predict the risk of disease, enabling healthcare providers to take preventive measures before symptoms arise. As these algorithms improve, diagnostic tools will become more accurate, aiding clinicians in making faster and more informed decisions. The increasing availability of data from electronic health records (EHRs), wearable devices, and mobile health applications will enhance AI's ability to detect patterns and emerging health concerns, providing opportunities for early intervention.

Incorporating natural language processing (NLP) into AI systems further enhances their capability to analyze unstructured data, such as clinical notes, research studies, and patient histories. AI can sift through this vast amount of data, identifying trends and offering valuable insights to inform decision-making. For example, NLP can extract information from clinical trials and research papers, keeping healthcare professionals updated on the latest treatment strategies and guidelines that apply to specific patient groups.

Another vital area of AI development is personalized treatment. AI allows for the creation of tailored treatment plans based on a patient's genetic makeup, medical history, and lifestyle factors. This precision medicine approach—where treatments are customized to individual patients—can increase the effectiveness of therapies and minimize side effects. As genomic data becomes more widely available through advancements in sequencing technologies, AI will be key in interpreting this data and providing insights that lead to more personalized, effective treatment options.

AI is also set to revolutionize drug development. By analyzing molecular structures, identifying potential drug candidates, and optimizing clinical trial designs, AI accelerates the drug discovery process. Using AI-powered simulations and predictive analytics, researchers can more efficiently identify promising compounds, speeding up the approval of new medications and therapies. For instance, companies like Atomwise use deep learning algorithms to test millions of compounds for their potential to treat diseases such as Ebola and multiple sclerosis, demonstrating how AI can streamline drug development.

AI-powered clinical decision support systems (CDSS) represent another important future development. These systems will integrate a variety of data sources—such as EHRs, lab results, imaging studies, and clinical guidelines—to provide real-time recommendations during patient consultations. By synthesizing complex data into actionable insights, these systems can help clinicians make informed decisions, enhancing patient outcomes and reducing errors that may arise from human judgment alone.

Virtual health assistants, powered by AI, are expected to play a larger role in healthcare. These systems, including chatbots, will respond to patient inquiries, schedule appointments, send medication reminders, and provide medical advice based on established guidelines. As AI, machine learning, and natural language processing capabilities improve, virtual health assistants will offer timely support to patients, allowing healthcare professionals to focus on more complex tasks that require human expertise.

Telemedicine and remote monitoring systems, enhanced by AI, will continue to expand. AIpowered telemedicine solutions will enable healthcare providers to conduct remote consultations while continuously monitoring patients' vital signs via wearable devices or mobile applications. This capability will allow for faster interventions when necessary and alleviate the burden on healthcare facilities by enabling patients to receive care from home. For example, remote monitoring systems for chronic conditions like diabetes or hypertension can alert healthcare providers when patients' readings fall outside safe ranges, triggering timely follow-up actions to prevent complications.

In the realm of public health, AI's predictive analytics capabilities are showing great promise. By analyzing population health data from diverse sources, including social media, AI can predict disease outbreaks, monitor trends over time, and optimize resource allocation during public health emergencies. This proactive approach enables authorities to respond swiftly to emerging health threats, such as infectious disease outbreaks, and manage public health crises more effectively.

Finally, from increasing drug discovery speed to personalizing treatment and enhancing diagnostic precision, AI has the potential to revolutionize healthcare systems. As AI technologies continue to evolve, they will revolutionize healthcare delivery, leading to more efficient and effective care, ultimately improving patient outcomes and the overall quality of healthcare. As the digital transformation in healthcare continues, it is vital that stakeholders—such as researchers, clinicians, and policymakers—collaborate to ensure that AI is implemented ethically, ensuring equitable access to these advancements for all patients.

VI. References

1. Yabsera Erdaw, et al. "Machine Learning Model Applied on Chest X-Ray Images Enables Automatic Detection of COVID-19 Cases with High Accuracy." *NCBI*, 2020.
2. Machine Learning for Early Detection of Pneumonia from Chest X-ray Images." *IJISAE*, 2021.
3. "Generalizable Disease Detection Using Model Ensemble on Chest X-Ray." *Nature*, 2020.
4. "Machine Learning and Deep Learning Approach for Medical Image Analysis." *Springer*, 2019.
5. "Machine Learning Augmented Interpretation of Chest X-rays." *NCBI*, 2020.
6. "Transfer Learning for Medical Image Classification Using Deep Learning." *IEEE*, 2019.

7. "Automated Detection of Tuberculosis Using Deep Learning Techniques." *Elsevier*, 2021.
8. "Comparative Study on Machine Learning Algorithms Applied to Chest X-Ray Images." *MDPI*, 2021.
9. "Deep Learning Approaches for Lung Cancer Detection Using CT Scans." *SpringerLink*, 2020.
10. "Explainable Artificial Intelligence Techniques Applied to Medical Imaging." *Wiley Online Library*, 2020.

Shamrat, FM Javed Mehedi, Sami Azam, Asif Karim, Kawsar Ahmed, Francis M. Bui, and Friso De Boer. "High-Precision Multiclass Classification of Lung Disease Through Customized MobileNetV2 from Chest X-ray Images." *Computers in Biology and Me*

[1] Y. Jiang, "Artificial intelligence in healthcare: past, present and future," *Stroke & Vasc Neurology,* p. 14, 2017.

[2] A. Adadi, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE,* p. 23, 2018.

[3] A. Holzinger, "Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI," *IFIP International Federation for Information Processing* , p. 8, 2018.

[4] Erion, "From local explanations to global understanding with explainable AI for trees. Nature machine intelligence.," *Nature,* p. 14, 2020.

[5] Bahdanau, ". Neural machine translation by jointly learning to align and translate.," *arXiv,* p. 10, 2014.

[6] F. Cabitza, "Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters," *Annals of Translational Medicine,* p. 9, 2020.

[7] F. K. Došilović, "Explainable artificial intelligence: A survey," *IEEE,* p. 6, 2018.

[8] V. Viswan, "Explainable Artificial Intelligence in Alzheimer's Disease Classification: A Systematic Review," *Cognitive Computation,* p. 44, 2024.

[9] D. k. Gurmessa, "Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review," *BJM health & care informatics,* p. 10, 2024.

[10] J. Wen, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *National Library of Medicine,* p. 39, 2020.

[11] S. Sarp, "An XAI approach for COVID-19 detection using transfer learning with X-ray images," *ScienceDirect,* p. 12, 2023.

[12] Mittelstadt, "Explaining explanations in AI. In Proceedings of the conference on fairness, accountability, and transparency.," *arXiv,* p. 11, 2019.

[13] A. Holzinger, "What do we need to build explainable AI systems for the medical domain?," *arxiv,* p. 28, 2017.

[14] A. Kumar, "Explainable Artificial Intelligence in Healthcare," *International Journal for Multidisciplinary Research,* p. 7, 2024.

[15] D. Castelvecchi, "Can we open the black box of AI?," *nature,* p. 4, 2016.

[16] F. Cabitza, "Unintended consequences of machine learning in medicine," *JAMA,* p. 10, 2017.

[17] Gunning, "DARPA's explainable artificial intelligence (XAI) program.," *Naval Research Laboratory* , p. 15, 2024.

[18] Arrieta, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *arXiv,* p. 10, 2020.

[19] “Topol,” *High-performance medicine: the convergence of human and artificial intelligence.,* p. 14, 2019.

[20] Char, “ Implementing machine learning in health care—addressing ethical challenges.,” *New England Journal of Medicine,* p. 15, 2018.

[21] Cath, “Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,” *Royal Society,* p. 8, 2018.

[22] Cabitza, “Bridging the "last mile" gap between AI implementation and operation: data awareness for trustworthy and transparent decision support systems.,” *Annals of Translational Medicine,* p. 12, 2020.

[23] Gilpin, “Explaining explanations: An overview of interpretability of machine learning.,” *IEEE,* p. 12, 2018.

[24] Lundberg, “A unified approach to interpreting model predictions. Advances in neural information processing systems,” *arXiv,* p. 13, 2017.

[25] Ribeiro, “" Why should I trust you?" Explaining the predictions of any classifier.,” *international conference on knowledge discovery and data mining,* p. 13, 2016.