Explainable Large Language Models for Cardiovascular Disease Detection: Opportunities, Challenges, and Ethical Considerations

Manu Mathew Jiss

Department of Computer Science

University of the Pacific

Spring 2026

Abstract

Large Language Models (LLMs) are becoming an important part of modern healthcare, especially in supporting the early detection of cardiovascular diseases (CVDs). These models are capable of analyzing large volumes of complex medical data, such as electronic health records and clinical reports, and can assist clinicians in improving diagnostic accuracy and decision-making. However, despite their potential, LLMs often function as black-box systems, which raises serious concerns about transparency, trust, and ethical use in clinical environments. This study uses a mixed-method approach that includes a systematic review of 22 scholarly articles along with thematic analysis to explore how Explainable Artificial Intelligence (XAI) can help address these challenges. The results indicate that XAI techniques such as SHAP and LIME play a significant role in making LLM predictions more interpretable by highlighting the key factors influencing model outputs. This not only improves clinician understanding but also enhances confidence in AI-supported medical decisions. Overall, the findings suggest that combining LLMs with XAI is essential for developing transparent, trustworthy, and ethically responsible healthcare systems. Such integration is particularly valuable in cardiovascular disease detection, where explainability and reliability are critical for real-world clinical adoption.

Keywords: Large Language Models, Explainable AI, Cardiovascular Disease, Clinical Decision Support, Model Interpretability, Healthcare AI, Black-Box Models, AI Ethics

Author Note

This paper presents a literature-based review of Explainable Artificial Intelligence (XAI) and Large Language Models (LLMs) in cardiovascular disease detection. The study synthesizes current research to examine opportunities, challenges, and ethical considerations associated with the use of explainable AI in healthcare applications.

1 Introduction

Large Language Models (LLMs) are significantly advancing healthcare, particularly in detecting and managing cardiovascular diseases (CVDs). According to the British Heart Foundation, heart failure related hospital admissions in England increased by 33% from 65,025 in 2013/14 to 86,474 in 2018/19, highlighting the critical need for early detection strategies (Chen et al., 2024). Leveraging LLMs such as ChatGPT has demonstrated improvements in analyzing complex patient datasets like electrocardiograms (ECGs) and electronic health records (EHRs). For example, Chen et al. (2024) showed that an LLM-based pre training method enhanced heart failure risk prediction in patients with hypertension and myocardial infarction, achieving C-index scores of 0.6349 and 0.5805, respectively. Similarly, Guazzo et al. (2023) developed a natural language processing (NLP) model that identified CVD hospitalizations among diabetic patients with high precision and recall across multiple clinical scenarios.

Despite these promising applications, LLMs face significant limitations in healthcare due to their “black box” nature and a lack of transparency in decision-making processes that can undermine trust among clinicians. Opacity in AI models restricts users’ understanding of their inner workings and prevents critical evaluation of their predictions (Hassija et al., 2024). Explainable AI (XAI) has emerged to address these issues by providing tools to interpret and explain AI-generated results. XAI methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are designed to elucidate the factors influencing AI decisions, enabling clinicians to better understand predictions and build trust in these systems. However, integrating XAI into clinical workflows remains a challenge due to the computational resources required for these tools (Dwivedi et al., 2023).

In addition to technical challenges, ethical considerations, such as ensuring fairness and avoiding bias in AI systems, are crucial. For example, biases in training data can lead to unequal treatment outcomes, raising concerns about the equitable deployment of LLMs in healthcare. This paper explores the potential of LLMs in CVD detection while addressing the black-box effect and ethical implications of AI usage. The research questions include:

- How can LLMs enhance early detection of cardiovascular diseases?
- What challenges and ethical considerations arise from their black-box nature?
- How can XAI techniques improve trust and transparency in AI-driven healthcare systems?

1.1 Background

Large Language Models (LLMs) are advanced artificial intelligence systems designed to process and interpret extensive datasets, including complex medical records and test results. Their application in healthcare has shown potential to improve diagnostic precision and patient outcomes significantly. Yang et al. (2023) explored the application of LLMs such as ChatGPT-4 in analyzing electronic health records (EHRs). Their study demonstrated that LLMs could identify risk factors for cardiovascular diseases (CVDs) by synthesizing patient history, lab results, and demographic data, thereby streamlining clinical decision-making and reducing the cognitive load on healthcare professionals (Yang et al., 2023).

Similarly, Triantafyllidis et al. (2022) emphasized the potential of LLMs in mobile health (mHealth) applications, where they assist in chronic disease management such as diabetes and hypertension. Their systematic review found that mHealth solutions leveraging LLMs and deep learning achieved diagnostic accuracy exceeding 84% in most cases, especially in predicting cardiovascular events and managing diabetes through real-time monitoring and personalized feedback. These advancements, while not entirely transformative, illustrate the potential of LLMs to optimize resource use and improve healthcare delivery, particularly in low-resource settings (Triantafyllidis et al., 2022).

The integration of LLMs into healthcare, however, is hampered by their inherent “black-box” nature, which limits the interpretability of their decision-making processes. This lack of transparency can undermine trust among clinicians and patients. Dwivedi et al. (2023) highlighted the role of Explainable AI (XAI) in addressing this issue by making AI models more interpretable. XAI tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) offer insights into how specific variables, like patient history or test results, influence predictions. SHAP assigns a contribution score to each feature, while LIME generates interpretable local approximations of the model's predictions. These tools, while promising, require significant computational resources, making widespread clinical implementation challenging (Dwivedi et al., 2023).

Furthermore, Mavrepisa et al. (2024) described a conceptual framework for integrating XAI with LLMs in healthcare. Their study demonstrated how tailored LLMs could simplify complex XAI methods, enabling healthcare professionals to understand predictions more easily. By aligning XAI explanations with the clinical workflow, their approach enhanced usability and fostered trust among practitioners, illustrating a pathway toward broader adoption of these technologies (Mavrepisa et al., 2024).

These developments underscore the necessity of combining the predictive power of LLMs with the transparency provided by XAI to create ethical, effective, and user-friendly AI systems for healthcare applications.

1.2 Significance of Addressing the Black-Box Effect

The "black-box" effect in Large Language Models (LLMs) presents a critical challenge for their adoption in healthcare applications. This phenomenon refers to the opacity of these models, where their internal decision-making processes remain incomprehensible to end-users, particularly clinicians. Dwivedi et al. (2023) define the black-box nature as a significant barrier to trust, emphasizing that when clinicians cannot trace the reasoning behind AI-generated predictions, skepticism arises. This skepticism can have serious consequences, such as delayed or incorrect clinical decisions, especially in high-stakes scenarios like diagnosing cardiovascular diseases (Dwivedi et al., 2023). For instance, in a case where an AI model predicts an elevated risk of myocardial infarction without clear evidence from patient data, clinicians might disregard the prediction, potentially missing critical early interventions (Hassija et al., 2024).

Explainable AI (XAI) offers a solution by providing interpretability tools that help demystify the decision-making processes of LLMs. Tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have been instrumental in elucidating the importance of input variables on predictive outcomes (Dwivedi et al., 2023). For example, SHAP assigns contribution scores to factors such as blood pressure or cholesterol levels, highlighting their roles in determining a diagnosis. Similarly, LIME generates visual explanations, enabling clinicians to validate and understand predictions in real time. Chen et al. (2024) demonstrated that integrating SHAP with LLMs in cardiovascular risk prediction enhanced clinicians' confidence by aligning AI-driven insights with established medical knowledge.

However, the integration of XAI with LLMs is not without challenges. Mavrepisa et al. (2024) noted that while XAI improves interpretability, it often requires advanced computational resources, which can limit its implementation in resource-constrained settings. Furthermore, Sebin et al. (n.d.-c) highlighted the need for audience-specific explanations, ensuring that outputs are comprehensible to diverse users, including non-specialist clinicians. This alignment not only fosters trust but also ensures that predictions are actionable across various healthcare environments (Sebin et al., n.d.-c).

Addressing the black-box effect is pivotal for the ethical and effective deployment of AI in healthcare. Transparent AI systems enable clinicians to make informed decisions, reduce diagnostic errors, and promote equitable patient care. By combining the predictive power of LLMs with the interpretability of XAI, healthcare can achieve a balance between technological advancement and clinical reliability.

2 Methodology

The research methodology for this paper followed a systematic approach to explore the role of Large Language Models (LLMs) in cardiovascular disease (CVD) detection and the application of Explainable AI (XAI) to address the black-box problem. Academic databases, including PubMed, IEEE Xplore, and Google Scholar, were utilized to identify relevant literature. In PubMed, journals like Journal of Biomedical Informatics and Artificial Intelligence in Medicine provided insights into the integration of AI in clinical workflows, while IEEE Xplore contributed papers from IEEE Transactions on Biomedical Engineering and IEEE Access that examined the technical advancements of LLMs and XAI methods in healthcare.

To ensure comprehensive coverage, the search included keywords such as “Explainable AI in Healthcare,” “Black Box Effect in AI,” “LLMs in Cardiovascular Disease Detection,” and “Interpretable AI for CVD Risk Prediction.” These keywords were applied to titles, abstracts, and, when necessary, full-text content. Filters were used to prioritize peer-reviewed articles published between 2022 and 2024, focusing on those with high citation metrics and relevance to the research themes. After the initial search, 31 articles were identified.

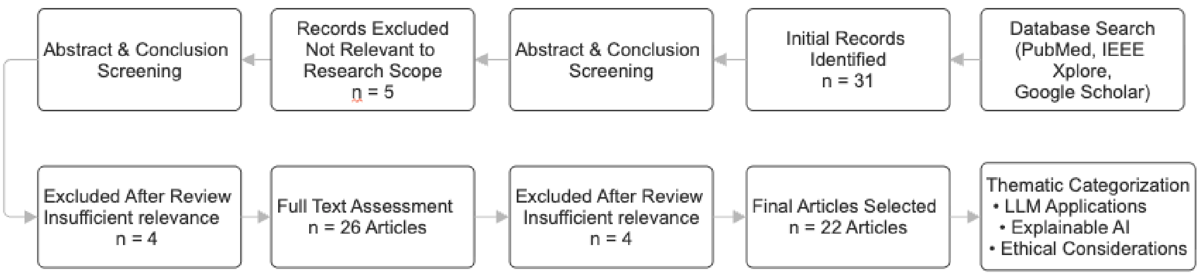

Figure 1: Article Selection and Screening Process Used for the Literature Review

Figure 1 illustrates the systematic literature search, screening, exclusion, and thematic categorization process used to identify the final set of articles included in this study.

The selection process involved a two-stage screening. First, abstracts and conclusions were reviewed to evaluate relevance to the study’s scope. Second, full texts were screened for alignment with three primary themes:

- The potential of LLMs in healthcare applications.
- The technical and ethical challenges posed by the black-box nature of LLMs.
- The role of XAI in making AI-driven healthcare systems more interpretable and trustworthy.

Through this process, 22 high-quality articles were selected for in-depth analysis. Articles were selected based on relevance to the research questions, publication quality, citation impact, and direct discussion of LLMs, explainable AI, or cardiovascular healthcare applications. Full-text reviews were conducted for these articles to extract detailed information, ensuring a robust understanding of their methodologies, findings, and implications.

This structured methodology enabled the identification of knowledge gaps and informed the development of practical recommendations for integrating LLMs and XAI in healthcare workflows. The articles were categorized into thematic clusters based on their focus, including the use of LLMs for cardiovascular disease detection, report generation, and ethical considerations in AI adoption. This approach ensured a thorough exploration of the research objectives while maintaining clarity and relevance.

3 Literature Review

Large Language Models (LLMs) have gained prominence in healthcare for their ability to enhance diagnostic accuracy, improve patient-provider communication, and streamline workflows. However, their integration presents significant technical, ethical, and operational challenges. This section reviews advancements, integration with diverse data types, challenges, ethical considerations, and gaps identified in existing research.

3.1 Advancements in LLMs for Healthcare

LLMs have shown transformative potential in early disease detection and clinical support. Guo et al. (2024) introduced the HEART model, which leverages retrieval-augmented techniques to analyze structured and unstructured medical data for precise diagnostics. Cambria et al. (2024) highlighted the potential of domain specific LLMs, like ClinicalBERT, in improving decision-making by integrating pre-trained models with clinical workflows.

Sarangi et al. (2024) demonstrated how LLMs, when applied to cardiovascular imaging, provided actionable insights, improving diagnostic efficiency. These advancements emphasize the significant role of LLMs in reducing clinician workload and enhancing patient outcomes.

3.2 Integration of LLMs with Diverse Data Types

One of the strengths of LLMs is their ability to process diverse datasets, such as textual records, imaging data, and ECG readings. Wu et al. (2024) discussed strategies for utilizing XAI with multimodal data to enhance interpretability in healthcare applications. Similarly, Sarangi et al. (2024) highlighted the integration of textual and imaging data for radiological diagnoses, demonstrating improved accuracy and patient outcomes.

Guo et al. (2024) explored how retrieval-augmented LLMs synthesized multimodal datasets, enabling personalized care. Cambria et al. (2024) further noted the use of ECG and imaging data in cardiovascular diagnostics, showcasing LLMs' versatility in uncovering hidden patterns in clinical data.

3.3 Challenges in Implementing Explainable AI

The black-box nature of LLMs presents a significant barrier to their adoption in healthcare. Cambria et al. (2024) discussed how tools like SHAP and LIME can partially mitigate this issue by offering interpretable outputs, though they require computational expertise. Wu et al. (2024) proposed strategies to bridge the gap between usability and transparency in XAI systems, emphasizing the importance of user-friendly solutions.

Guo et al. (2024) highlighted the computational demands of XAI frameworks, particularly in resource-limited clinical settings. Addressing these challenges is crucial for aligning LLM outputs with real-world healthcare workflows.

3.4 Ethical Considerations in AI-Driven Healthcare

The deployment of LLMs raises ethical concerns, including algorithmic bias and data privacy. Guo et al. (2024) noted that biased datasets could lead to inequitable outcomes, particularly for underserved populations. Cambria et al. (2024) advocated for fairness and equity in AI systems, emphasizing the need for transparent governance practices.

Sarangi et al. (2024) emphasized the importance of stringent data protection measures to safeguard patient confidentiality while enabling effective AI-driven diagnostics. Ethical adherence is essential to foster trust among clinicians and patients.

3.5 Gaps Identified in Existing Literature

Despite their potential, LLMs face barriers to seamless integration in clinical environments. Wu et al. (2024) identified the lack of scalable XAI frameworks as a critical challenge. Guo et al. (2024) emphasized the need for adaptable LLMs that balance complexity with practicality in resource-constrained settings.

Cambria et al. (2024) suggested future research focus on designing intuitive XAI tools to bridge technical gaps and improve adoption rates. Addressing these gaps is pivotal to maximizing the benefits of AI technologies in healthcare.

4 Potential Applications of LLMs in Healthcare

Large Language Models (LLMs) are advancing the landscape of healthcare with innovative applications in diagnostics, risk stratification, patient engagement, and personalized treatment. Their ability to process vast datasets and provide actionable insights has proven invaluable for healthcare professionals. This section explores these applications, supported by recent studies.

4.1 Advancing Diagnostics with LLMs

LLMs have demonstrated their ability to enhance diagnostic precision. Han et al. (2023) evaluated LLMs for cardiovascular risk prediction, demonstrating effective pattern recognition in structured and unstructured data, albeit with some calibration challenges. Wu et al. (2024) emphasized the importance of usability in explainable AI (XAI) frameworks integrated with LLMs, enabling clinicians to understand the reasoning behind diagnostic outputs.

4.2 Risk Stratification and Predictive Analytics

LLMs have been pivotal in predictive analytics, particularly in stratifying patient risk. Mishra et al. (2024) highlighted how LLM-generated explanations simplified cardiovascular disease information, enabling healthcare providers to identify high-risk individuals effectively. The integration of electronic health records and real-time data analysis enhances risk predictions and supports preventive care strategies.

4.3 Enhancing Patient Engagement

LLMs like GPT-4 have transformed patient-doctor communication by simplifying complex medical terms and generating clear, actionable reports. Sarangi et al. (2024) demonstrated the effectiveness of LLMs in providing accurate yet patient-friendly descriptions of medical conditions, improving treatment adherence and patient satisfaction.

4.4 Personalized Treatment Plans

The use of LLMs in personalized medicine has been transformative. By analyzing genetic, lifestyle, and historical medical data, Han et al. (2023) showed that these models provide tailored treatment recommendations, reducing reliance on trial-and-error methods. Wu et al. (2024) also discussed how XAI methodologies ensure that treatment suggestions remain interpretable and actionable for clinicians.

Table 1: Evidence-Based Applications of Large Language Models in Healthcare

Application Area	**Use Case**	**Outcomes**	**Challenges**	**Models**	**Source**
Diagnostics	Predicting 10-year CVD risk from structured data	Comparable to Framingham Risk Score	Reliability and hallucination risks	GPT-3.5, GPT-4	Han et al., 2023
Risk Stratification	Classifying patients into low, medium, high risk	Identifies high-risk patients for early prevention	Limited generalizability across populations	GPT-3.5, GPT-4	Han et al., 2023

Application Area	Use Case	Outcomes	Challenges	Models	Source
Personalized Treatment	Generating simplified prevention guidance	Improved readability and completeness of responses	Output varies with prompt and data quality	GPT-4	Mishra et al., 2024
Clinical Decision-Making	Supporting differential diagnosis in imaging	Moderate agreement with expert radiologists	Lack of transparency in reasoning	GPT-3.5, GPT-4	Sarangi et al., 2024
Explainability (XAI)	Applying SHAP and LIME for interpretation	Improved transparency of predictions	High computational cost	SHAP, LIME	Wu et al., 2024

This table summarizes key applications of LLMs in healthcare, focusing on their use cases, outcomes, challenges, and the models involved. Data from multiple sources were synthesized to provide an overview of their potential and limitations.

5 Case Studies: Successful Applications of LLMs in CVD Detection

The integration of Large Language Models (LLMs) into cardiovascular disease (CVD) detection has demonstrated transformative potential. By leveraging advancements in AI, researchers have developed innovative tools for early diagnosis, risk assessment, and clinical support. This section presents key case studies highlighting the application of LLMs in CVD detection, diagnosis, and patient management.

5.1 Integration with Clinical Notes and EHRs

LLMs have been effectively used to process unstructured data, such as clinical notes and electronic health records (EHRs), to improve diagnostic outcomes. Fernández-Ruiz (2019) utilized a deep neural network model trained on ECG data to classify arrhythmias with an average area under the receiver operating characteristic curve (AUC) of 0.97, matching the performance of cardiologists. This model also demonstrated a higher sensitivity in detecting rhythm abnormalities, showcasing the utility of integrating LLMs with clinical datasets (Fernández-Ruiz, 2019).

5.2 Predictive Analysis of Cardiovascular Risk Factors

LLMs like ChatGPT-3.5 have shown mixed results in quantitative cardiovascular risk assessment. Marafino et al. (2023) evaluated ChatGPT-derived Pooled Cohort Equation (PCE) scores for atherosclerotic cardiovascular disease (ASCVD) risk estimation. The study found a moderate correlation (Pearson coefficient 0.46) between ChatGPT-derived and true PCE scores, but highlighted poor calibration and reproducibility, underscoring the need for model refinement (Marafino et al., 2023).

5.3 Integration with Wearable Technology Data

Wearable technologies combined with LLMs have advanced real-time health monitoring. Qiu et al. (2023) introduced a model leveraging ECG embeddings and language-based features for zero-shot cardiovascular disease detection and automated ECG report generation, achieving competitive accuracy compared to supervised baselines. This innovation underscores the feasibility of transferring knowledge from LLMs to clinical cardiology (Qiu et al., 2023).

5.4 Explainable AI in Cardiovascular Diagnostics

Ensuring transparency in LLM-based systems is crucial for clinical adoption. Linardatos et al. (2021) provided a comprehensive review of explainable AI methods, emphasizing their relevance in enhancing clinician trust and understanding of model outputs. These methods are pivotal in critical healthcare applications where interpretability is a prerequisite for implementation (Linardatos et al., 2021).

5.5 Advanced Multimodal Analysis

Zhao et al. (2024) proposed ECG-Chat, a multimodal LLM designed for ECG data analysis and medical report generation. This model integrates text and signal modalities through contrastive learning, achieving superior performance in classification and retrieval tasks. ECG-Chat highlights the potential of multimodal LLMs to address complex diagnostic challenges in cardiology (Zhao et al., 2024).

Table 2: Evidence-Based Applications of Large Language Models in Healthcare

Data Source	**Model**	**Use Case**	**Outcome**	**Challenges**	**Source**
ECG data	Deep neural network	Arrhythmia detection	AUC ≈ 0.97, cardiologist-level performance	Scalability, dataset diversity	Fernández-Ruiz, 2019
Synthetic patient data	ChatGPT-3.5	ASCVD risk estimation	Moderate correlation with true risk scores (Pearson $\approx$ 0.46)	Poor calibration, reproducibility	Marafino et al., 2023
ECG signals	ECG embeddings + LLM	Zero-shot CVD detection	Competitive classification accuracy	Limited dataset coverage	Qiu et al., 2023
Clinical datasets	Explainable AI models	Model interpretability	Improved clinician trust	Implement complexity	Linardatos et al., 2021
ECG + medical records)	ECG-Chat	Multimodal ECG analysis and reporting	Superior performance in classification and retrieval	Stability in long-text generation	Zhao et al., 2024

This table summarizes key applications of LLMs in cardiovascular healthcare, highlighting their use cases, outcomes, challenges, and the models employed. Insights were drawn from multiple studies to illustrate the transformative potential and limitations of these AI-driven systems in clinical settings.

The graph below highlights the trends in diagnostic performance of Large Language Models (LLMs) in cardiovascular disease (CVD) detection, focusing on key metrics such as diagnostic accuracy (AUC), classification accuracy, and correlation with clinical risk scores, showcasing their advancements from 2021 to 2024.

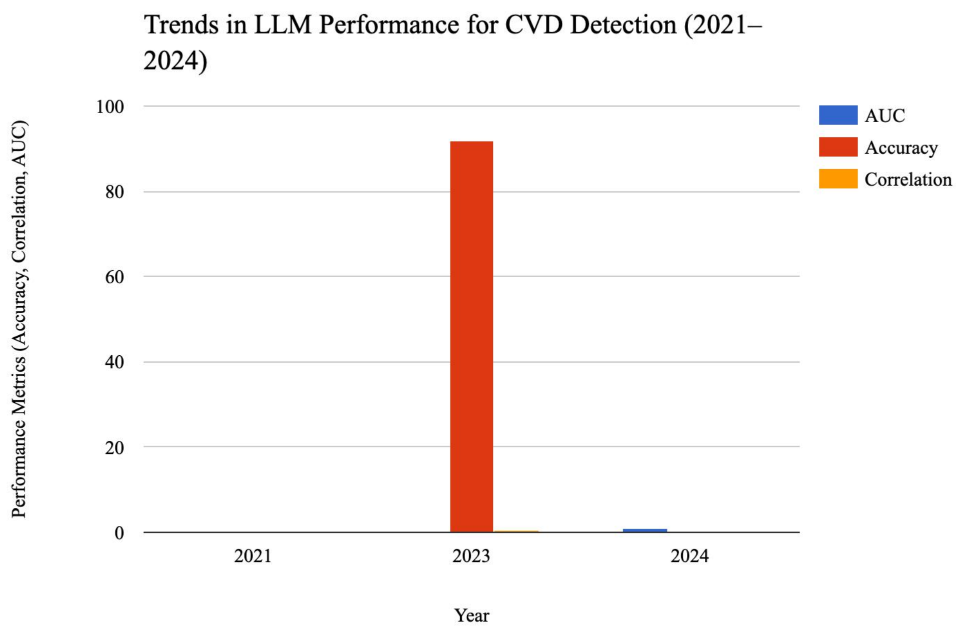

Figure 2: Trends in LLM Performance for CVD Detection (2021–2024)

Keywords (Definitions)

- **AUC (Area Under the Curve):** A performance metric for classification models, representing the ability to distinguish between classes. A higher AUC indicates better performance.
- **Accuracy:** A measure of correct classifications made by a model out of the total predictions, demonstrating its reliability in identifying CVDs like atrial fibrillation and myocardial infarction.
- **Pearson Correlation:** A statistical measure indicating the strength and direction of the relationship between predicted and true risk scores, used in assessing model calibration for cardiovascular risk

factors.

- **ECG-Chat:** A multimodal LLM specifically designed for ECG data analysis and report generation, showcasing the integration of text and signal modalities for advanced diagnostics.
- **Explainable AI (XAI):** AI methodologies that provide transparent and interpretable outputs, fostering trust and usability among clinicians in high-stakes applications like CVD detection.
- **Zero-Shot Classification:** A capability of certain LLMs to classify data accurately without prior exposure to specific training examples, as demonstrated in ECG-related tasks.

Insights were drawn from multiple studies to illustrate the transformative potential and limitations of these AI-driven systems in clinical settings.

6 Discussion

6.1 Synthesis of Findings

The integration of Large Language Models (LLMs) in healthcare has shown promising advancements in predictive capabilities and clinical decision support. Studies highlight the potential of LLMs in analyzing large datasets, enabling accurate and efficient risk predictions for cardiovascular diseases (Yang et al., 2023). Han et al. (2023) demonstrated how GPT-4, when applied to UK Biobank data, performed comparably to traditional models like the Framingham Risk Score (FRS), with improved flexibility in handling incomplete datasets. These findings align with the broader consensus that LLMs can complement traditional risk assessment tools. While these models offer groundbreaking potential, further optimization is essential for their adoption across diverse clinical settings.

A key insight from these studies is that LLMs are not merely improving predictive accuracy, but are reshaping how clinical reasoning itself may be supported by AI systems. Rather than replacing traditional models, LLMs appear best suited as complementary tools that enhance decision-making through contextual understanding and flexible data integration. This suggests a shift from static risk models toward more adaptive, AI-assisted clinical workflows.

6.2 Addressing the Black-Box Problem

The inherent opacity of LLMs remains a significant barrier to their acceptance in healthcare. Explainable AI (XAI) frameworks, such as SHAP and LIME, have been proposed to enhance the interpretability of these models (Linardatos et al., 2021). Han et al. (2023) emphasized that providing transparent outputs is crucial, especially for high-stakes applications like cardiovascular risk prediction. Similarly, Bearman and Ajjawi (2023) underscore the importance of educating clinicians on navigating AI outputs. Enhancing explainability without compromising performance will be pivotal in addressing the trust deficit associated with AI in medicine.

Explainability should not be treated as a technical add-on, but as a core design principle for healthcare AI systems. Without meaningful interpretability, even highly accurate models risk being rejected by clinicians. This highlights the need for future LLMs to embed explainability directly into their architectures rather than relying solely on post-hoc explanation tools.

6.3 Implications for Clinical Practice

LLMs like GPT-4 have demonstrated their potential to streamline clinical workflows by providing rapid and reliable risk assessments. Han et al. (2023) observed that LLMs, when integrated with patient datasets, could accurately stratify patients into low, medium, and high-risk categories for cardiovascular events. Additionally, Yang et al. (2023) reported how LLMs improve operational efficiency in resource-intensive healthcare settings. However, Marafino and Liu (2023) caution that computational demands and integration barriers remain challenges, especially in under-resourced healthcare systems. Addressing these issues through resource-efficient algorithms and infrastructure investment will broaden the impact of AI in clinical practice.

The true value of LLMs in clinical practice lies not only in automation, but in their potential to reduce cognitive burden on clinicians. By summarizing complex patient histories and highlighting risk patterns, LLMs can act as intelligent clinical assistants rather than mere prediction engines.

6.4 Ethical and Societal Considerations

The ethical implications of LLM deployment in healthcare are multifaceted. Data privacy, bias in training datasets, and equitable access remain significant concerns (Bearman & Ajjawi, 2023). Han et al. (2023) noted the importance of maintaining data security, particularly when working with sensitive patient information. Marafino and Liu (2023) further highlighted the risk of exacerbating healthcare inequalities due to algorithmic biases. Fostering interdisciplinary collaboration between ethicists, AI researchers, and healthcare professionals is crucial for creating equitable and ethical AI solutions.

Ethical challenges represent one of the greatest barriers to real-world adoption of LLMs. Technical improvements alone will not be sufficient unless accompanied by strong governance frameworks that ensure fairness, accountability, and transparency. This reinforces the idea that AI in healthcare is as much a social problem as it is a computational one.

6.5 Limitations of Current Approaches

While Large Language Models (LLMs) such as ChatGPT have shown significant promise in healthcare, several limitations hinder their widespread adoption. One key issue is the dependency on high-quality datasets and extensive computational resources. Han et al. (2023) emphasized that even advanced models like GPT-4 struggle in real-world scenarios, particularly in resource-limited settings where infrastructure is insufficient to support their computational demands. This limitation underscores the need for lightweight models that can operate efficiently with constrained resources (Han et al., 2023).

Another critical challenge is the variability in performance due to demographic disparities in training datasets. Marafino and Liu (2023) observed that ChatGPT's predictions for cardiovascular risk exhibited inconsistencies, especially when applied to diverse populations. For example, ChatGPT's sensitivity to sociodemographic indicators led to inappropriate risk adjustments, which could undermine its reliability in clinical decision-making (Marafino & Liu, 2023). These findings highlight the importance of developing inclusive datasets that represent varied demographics, ensuring that AI models produce equitable outcomes across diverse patient groups.

Furthermore, reproducibility and calibration remain problematic. Studies by Marafino and Liu (2023) found that repeated prompts with identical data produced inconsistent results, raising questions about the stability of LLMs in critical applications. While these issues do not entirely negate the utility of LLMs, they indicate a need for improved training techniques, such as fine-tuning and robust evaluation frameworks, to enhance model reliability and consistency.

Addressing these limitations requires a multi-faceted approach. Beyond inclusive datasets and efficient model designs, integrating mechanisms to ensure calibration and reproducibility is essential. Additionally, fostering collaboration between AI developers and healthcare professionals can lead to tailored solutions that balance innovation with practical constraints. These measures will be pivotal in unlocking the full potential of LLMs while ensuring their responsible and effective deployment in healthcare.

Many of the limitations observed in current LLMs stem from a mismatch between research environments and real clinical settings. Models are often evaluated under ideal conditions, whereas hospitals operate under constraints such as time pressure, incomplete data, and diverse patient populations. Bridging this gap should be a priority for future research.

6.6 Opportunities for Future Research

6.7 Research Limitations

This study is limited by its reliance on published literature between 2022 and 2024, which may not capture the most recent developments in large language models and explainable AI. Additionally, the findings are based on secondary sources rather than experimental validation, which may limit the generalizability of some conclusions.

Future research must focus on advancing interpretability techniques and exploring innovative applications of LLMs in healthcare. Han et al. (2023) suggested integrating multimodal datasets to enhance prediction accuracy across diverse patient populations. Bearman and Ajjawi (2023) emphasized the need for interdisciplinary efforts to develop AI tools that are user-friendly and adaptable. Prioritizing these areas will ensure that LLMs achieve their full potential in revolutionizing healthcare.

Future research should prioritize building smaller, domain-specific LLMs trained on high-quality clinical datasets, rather than relying solely on general-purpose models. Additionally, combining LLMs with symbolic reasoning or rule-based systems may offer a promising path toward more reliable and interpretable healthcare AI.

7 Research Contributions

This paper makes three primary contributions:

- It synthesizes current literature on the use of Large Language Models in cardiovascular disease detection.
- It examines how Explainable Artificial Intelligence techniques address transparency challenges associated with black-box AI systems.
- It identifies ethical, technical, and practical challenges that must be addressed before widespread clinical adoption.

8 Conclusion

The study of Large Language Models (LLMs) and Explainable AI (XAI) in healthcare, particularly in the context of cardiovascular disease (CVD) detection, highlights a rapidly growing and influential area of research. This paper examined how LLMs can support clinicians by improving diagnostic accuracy, streamlining healthcare workflows, and addressing challenges related to their black-box nature. While these models demonstrate strong potential, their real-world adoption depends on effectively managing ethical concerns, system scalability, and computational demands. Overall, the findings emphasize that technological progress must be balanced with responsible, patient-centered implementation for AI to be truly beneficial in healthcare.

One of the most important gaps discussed in this paper is the lack of focus on underserved populations and healthcare systems with limited resources. Although many LLM-based systems perform well in advanced and well-equipped clinical environments, their practical value becomes uncertain in regions where there is limited access to data, infrastructure, and technical support. This situation suggests that current research may unintentionally widen existing healthcare inequalities instead of helping to close them. Future research should place greater emphasis on developing scalable, low-cost, and context-aware models that can work effectively across different global settings, so that fair and inclusive AI-driven healthcare becomes not just a technical aim, but also an ethical responsibility.

The ethical concerns surrounding LLMs represent another major area that requires deeper attention. Although recent progress in data security and bias reduction is encouraging, these solutions are still at an early stage. This highlights the need for stronger ethical frameworks that focus on protecting patient privacy, promoting fairness, and ensuring inclusivity in AI systems. For instance, closer collaboration between AI researchers, clinicians, and policymakers could help develop more realistic and practical ethical guidelines. Addressing these ethical challenges is essential for building long-term trust in AI tools and encouraging their responsible use in real healthcare environments.

In conclusion, research on LLMs and XAI for CVD detection remains at an early but highly promising stage. Although these technologies show strong potential to improve healthcare outcomes, their real-world deployment still faces several technical, ethical, and organizational challenges. Addressing issues such as model interpretability, fairness, data limitations, and system reliability will be essential before widespread adoption can occur. Future studies should therefore focus on developing AI systems that are not only accurate, but also transparent, resource-efficient, and inclusive. This area of research holds significant long-term value, and Continued advancements in explainable and ethical AI are expected to contribute significantly to the development of trustworthy and impactful healthcare solutions.

References

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology, 54*(5), 1160–1173. https://doi.org/10.1111/bjet.13337

Cambria, E., Malandri, L., Mercorio, F., Nobani, N., & Seveso, A. (2024, July 21). XAI meets LLMs: A Survey of the Relation between Explainable AI and Large Language Models. *arXiv.org*. https://arxiv.org/abs/2407.15248

Chen, C., Li, L., Beetz, M., Banerjee, A., Gupta, R., & Grau, V. (2024, March 15). Large language model-informed ECG Dual Attention Network for heart failure risk prediction. *arXiv.org*. https://arxiv.org/abs/2403.10581

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2022). Explainable AI (XAI): core ideas, techniques, and solutions. *ACM Computing Surveys, 55*(9), 1–33. https://doi.org/10.1145/3561048

Fernández-Ruiz, I. (2019). Artificial intelligence to improve the diagnosis of cardiovascular diseases. *Nature Reviews Cardiology, 16*(3), 133. https://doi.org/10.1038/s41569-019-0158-5

Guazzo, A., Longato, E., Fadini, G. P., Morieri, M. L., Sparacino, G., & Di Camillo, B. (2023). Deep-learning-based natural-language-processing models to identify cardiovascular disease hospitalisations of patients with diabetes from routine visits' text. *Scientific Reports, 13*(1). https://doi.org/10.1038/s41598-023-45115-1

Guo, J., Shan, X., Wang, G., Chen, D., Lu, R., & Tang, S. (n.d.). HEART: Heart Expert Assistant with ReTrieval-augmented. *OpenReview*. https://openreview.net/forum?id=fGVQgxvrzI

Han, C., Kim, D. W., Kim, S., You, S. C., Bae, S., & Yoon, D. (2023, January 1). Large-language-model-based 10-year risk prediction of cardiovascular disease: Insight from the UK Biobank Data. *medRxiv*. https://www.medrxiv.org/content/10.1101/2023.05.22.23289842v1

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting Black-Box Models: A review on Explainable Artificial intelligence. *Cognitive Computation, 16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of Machine Learning Interpretability Methods. *Entropy, 23*(1), 18. https://doi.org/10.3390/e23010018

Marafino, B. J., & Liu, V. X. (2023, January 1). Performance of a large language model (chatgpt-3.5) for pooled cohort equation estimation of atherosclerotic cardiovascular disease risk. *medRxiv*. https://www.medrxiv.org/content/10.1101/2023.08.11.23293957v1

Mavrepis, P., Makridis, G., Fatouros, G., Koukos, V., Separdani, M. M., & Kyriazis, D. (2024, January 23). XAI for All: Can Large Language Models Simplify Explainable AI? *arXiv.org*. https://arxiv.org/abs/2401.13110

Mishra, V., Sarraju, A., Kalwani, N. M., & Dexter, J. P. (2024). Evaluation of prompts to simplify cardiovascular disease information using a large language model: Cross-Sectional Study (Preprint). *Journal of Medical Internet Research, 26*, e55388. https://doi.org/10.2196/55388

Qiu, J., Han, W., Zhu, J., Xu, M., Rosenberg, M., Liu, E., Weber, D., & Zhao, D. (n.d.). Transfer knowledge from natural language to electrocardiography: Can we detect cardiovascular disease through language models?. *ACL Anthology*. https://aclanthology.org/2023.findings-eacl.33/

Sarangi, P. K., Irodi, A., Panda, S., Nayak, D. S. K., & Mondal, H. (2023). Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: Perspectives of four large language models. *Indian Journal of Radiology and Imaging*. https://doi.org/10.1055/s-0043-1777289

Sebin, B., Taskin, N., & Mehdiyev, N. (n.d.). Exploring the intersection of Large Language models (LLMs) and explainable AI (XAI): A systematic literature Review (Research-in-Progress). *AIS Electronic Library (AISeL)*. https://aisel.aisnet.org/confirm2024/25/

Triantafyllidis, A., Kondylakis, H., Katehakis, D., Kouroubali, A., Koumakis, L., Marias, K., Alexiadis, A., Votis, K., & Tzovaras, D. (2022). Deep Learning in MHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review. *JMIR Mhealth and Uhealth, 10*(4), e32344. https://doi.org/10.2196/32344

Wu, X., Zhao, H., Zhu, Y., Shi, Y., Yang, F., Liu, T., Zhai, X., Yao, W., Li, J., Du, M., & Liu, N. (2024, March 13). Usable XAI: 10 Strategies Towards Exploiting Explainability in the LLM Era. *arXiv.org*. https://arxiv.org/abs/2403.08946

Yang, R., Tan, T. F., Lu, W., Thirunavukarasu, A. J., Ting, D. S. W., & Liu, N. (2023). Large language models in health care: Development, applications, and challenges. *Health Care Science, 2*(4), 255–263. https://doi.org/10.1002/hcs2.61

Zhao, Y., Zhang, T., Wang, X., Han, P., Chen, T., Huang, L., Jin, Y., & Kang, J. (2024, August 16). ECG-Chat: a large ECG-Language model for cardiac disease diagnosis. *arXiv.org*. https://arxiv.org/abs/2408.08849

Appendix A: Summary of Reviewed Articles

The following table summarizes 20+ peer-reviewed articles reviewed for this paper. The table provides details on the year of publication, authors, objectives, methodologies, results, conclusions, and thematic clusters.

Year Published	First Author	Title	Objective	Methodology	Results	Conclusions	Links	Cluster
2024	Guo	HEART: Heart Expert Assistant with Retrieval-augmented Generation	Develop a model, HEART, to assist in early cardiovascular disease detection using Large Language Models (LLMs) and echocardiographic data.	HEART uses a dual-component structure with a Diagnostic Module (trained on cardiac ultrasound) and a Case Retrieval Module, which retrieves similar cases to aid diagnosis.	The model demonstrated strong potential in predicting congenital heart disease with promising results on a specific dataset.	HEART could become foundational for predicting cardiovascular diseases, especially useful in regions with limited medical resources	openreview.net/forum?id=fGVQgxvrzI	Using LLM and other AI models for CVD detection
2024	Chen	Large Language Model-informed ECG Dual Attention Network for Heart Failure Risk Prediction	Develop a novel ECG network to predict heart failure (HF) risk by integrating LLM-informed pre-training for enhanced feature learning.	The dual-attention network, trained with UK Biobank data, incorporates cross-lead and lead-specific temporal attention to capture ECG features.	The network showed superior predictive performance, with significant improvements in interpretability and accuracy over traditional methods.	The approach advances HF risk prediction through enhanced feature learning and interpretability, demonstrating the benefits of LLM pre-training in healthcare applications	arxiv.org/abs/2403.10581	Using LLM and other AI models for CVD detection

2021	An	High-Risk Prediction of Cardiovascular Diseases via Attention-Based Deep Neural Networks	Develop a deep learning model (DeepRisk) to predict high-risk cardiovascular diseases using attention-based neural networks.	DeepRisk utilizes EHR data and attention mechanisms to capture temporal patterns and heterogeneous data from various medical sources.	The model outperformed existing methods, significantly improving prediction accuracy for cardiovascular disease risk.	Attention-based models like DeepRisk can aid in identifying at-risk patients, enhancing preventive care in cardiovascular disease management	ieeexplore.ieee.org/abstract/document/8798764	Using LLM and other AI models for CVD detection
2023	Qiu	Transfer Knowledge from Natural Language to Electrocardiography	Explore the potential of transferring LLM knowledge to electrocardiography (ECG) for cardiovascular disease diagnosis and report generation.	The approach includes aligning ECG and language embeddings with Optimal Transport to facilitate zero-shot cardiovascular disease detection and ECG diagnosis report generation.	The method generated high-quality diagnostic reports and showed competitive zero-shot classification performance against supervised baselines.	Transferring LLM knowledge to ECG is feasible and beneficial, supporting advancements in cardiac diagnostics	aclanthology.org/2023.findings-eacl.33/	Using report generation and ECG data

2024	Zhao	ECG-Chat: A Large ECG-Language Model for Cardiac Disease Diagnosis	Create ECG-Chat, an LLM tailored to generate ECG diagnostic reports and facilitate patient-doctor dialogues.	ECG-Chat employs contrastive learning to align ECG and text data, building a robust dialogue and report generation system trained on 19,000 ECG diagnosis samples and 25,000 dialogues.	The model outperformed baselines in ECG classification, retrieval, and report generation, offering comprehensive diagnostics and interaction capabilities.	ECG-Chat effectively bridges ECG and textual modalities, making it a valuable tool for patient engagement and diagnostic support in cardiology	arxiv.org/abs/2408.08849	Using report generation and ECG data
2019	Hannun	Artificial Intelligence to Improve the Diagnosis of Cardiovascular Diseases	Develop AI models to enhance ECG interpretation for detecting cardiac arrhythmias and early-stage heart disease.	A deep neural network was trained on over 91,000 ECG records to classify arrhythmias, with additional models for detecting left ventricular dysfunction.	The AI models matched or exceeded cardiologist performance, with high AUC values (up to 0.97), showing strong diagnostic capability.	AI-driven ECG interpretation could streamline arrhythmia detection, improving diagnostic accuracy and efficiency in cardiology	nature.com/articles/s41569-019-0158-5	Using report generation and ECG data

2024	Lu	Enhancing Clinical Relevance of Pretrained Language Models Through Integration of External Knowledge: Case Study on Cardiovascular Diagnosis From Electronic Health Records	Improve the adaptability and clinical usefulness of pretrained language models (PLMs) by integrating domain-specific knowledge for cardiovascular diagnosis.	The study utilized knowledge adapters to inject diverse medical knowledge into PLMs, enabling more accurate cardiovascular diagnoses from clinical narratives.	PLMs with integrated knowledge adapters outperformed Clinical-BERT and similar models, providing better interpretability and performance in diagnostic tasks.	This approach highlights the potential of PLMs with integrated medical knowledge to enhance diagnostic accuracy in healthcare	ai.jmir.org/2024/1/e56932	Pre training with LLMs
2024	Cheema	AI and Heart Failure: Present State and Future with Multimodal Large Language Models	Examine the role of AI and multimodal large language models (LLMs) in managing and diagnosing heart failure.	The study discussed current AI applications, including LLMs, that process multimodal data (e.g., ECG, echocardiography) to support heart failure management.	LLMs show promise in accurately identifying heart failure symptoms and predicting patient outcomes, although implementation challenges remain.	While AI offers potential in heart failure management, further integration into clinical workflows is necessary to realize its benefits fully	sciencedirect.com/science/article/pii/S2772963X24002199	Uses different data types like images for diagnosis

2024	Panagoulias	Evaluating LLM-Generated Multimodal Diagnosis from Medical Images and Symptom Analysis	Assess the accuracy and correctness of LLMs in multimodal medical diagnosis using images and symptom data.	The study discussed current AI applications, including LLMs, that process multimodal data (e.g., ECG, echocardiography) to support heart failure management.	The model achieved approximately 84% accuracy in diagnoses, identifying specific strengths and weaknesses.	The proposed evaluation paradigm reveals the potential and limitations of LLMs in multimodal medical diagnostics, providing a framework applicable to other LLMs	arxiv.org/abs/2402.01730	Uses different data types like images for diagnosis
2024	Boonstra	Artificial Intelligence: Revolutionizing Cardiology with Large Language Models	Review the impact of large language models (LLMs) on cardiology, particularly their applications in clinical tasks and patient interaction.	This state-of-the-art review examines LLM applications, including patient-physician communication, clinical data processing, and task automation.	LLMs improve patient outcomes by facilitating automated workflows, enhancing data extraction, and supporting clinical decision-making.	LLMs hold transformative potential in cardiology, but ethical and practical challenges, like data privacy and algorithm bias, must be addressed	academic.oup.com/eurheartj/article/45/5/332/7505599	patient interaction and communication.

2024	Ogunpola	Machine Learning-Based Predictive Models for Detection of Cardio-vascular Diseases	Evaluate different machine learning models to enhance the detection of cardiovascular diseases, with a focus on myocardial infarction.	The study used models like XG-Boost, CNN, and Logistic Regression, addressing challenges of imbalanced datasets for improved accuracy.	An optimized XG-Boost model achieved high performance with 98.50% accuracy, 99.14% precision, and 98.71% F1 score.	Properly tuned machine learning models can significantly improve cardiovascular disease detection accuracy	mdpi.com/2075-4418/14/2/144	General Reviews and Surveys on LLMs in Healthcare
2023	Gala	The Utility of Language Models in Cardiology: A Narrative Review of the Benefits and Concerns of ChatGPT-4	Explore ChatGPT-4's applications in cardiology, focusing on its benefits and limitations for clinical use.	The narrative review assesses ChatGPT-4's capabilities in diagnosis support, patient education, and administrative automation.	ChatGPT-4 is beneficial for patient engagement and streamlining clinical tasks but faces issues related to accuracy, reliability, and data privacy.	While ChatGPT-4 can enhance healthcare delivery, ethical oversight is crucial to avoid misinformation and maintain patient trust	mdpi.com/1660-4601/20/15/6438	General Reviews and Surveys on LLMs in Healthcare

2022	Triantafyllidis	Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review	Review deep learning applications in mobile health (mHealth) for chronic disease management, focusing on cardiovascular disease, diabetes, and cancer.	The systematic review analyzes studies using mHealth data with deep learning models for diagnosing and monitoring chronic conditions.	Deep learning models show promising results in accuracy and management, especially in cardiovascular disease diagnosis, using mHealth data.	mHealth-based deep learning can revolutionize chronic disease management, although real-life clinical validation is needed	mhealth.jmir.org/2022/4/e32344	General Reviews and Surveys on LLMs in Healthcare
2024	The Lancet Digital Health	Promises and Challenges of Digital Tools in Cardiovascular Care	Highlight the impact of digital tools and AI in improving cardiovascular care, with a focus on accessibility and equitable treatment.	This editorial discusses innovations like AI-based imaging and large language models, along with challenges in regulation and data privacy.	Digital tools offer transformative potential in cardiovascular care but are hindered by data biases, accessibility gaps, and regulatory limitations.	Bridging the gap in digital health benefits for underserved populations and refining regulatory frameworks are critical to equitable cardiovascular care	pubmed.ncbi.nlm.nih.gov/39214758/	General Reviews and Surveys on LLMs in Healthcare

2019	Su	Extraction of Risk Factors for Cardiovascular Diseases from Chinese Electronic Medical Records	Develop an information extraction system to identify cardiovascular disease (CVD) risk factors from Chinese electronic medical records.	The system uses a pipeline with Named Entity Recognition (NER) and textual classification, employing BLSTM for risk factor recognition and CNN/SVM for time and assertion classification.	The system achieved high performance with F1 scores of 0.9609 for risk factor recognition and 0.9812 for time classification, demonstrating efficiency in extracting CVD risk factors.	This extraction system provides a valuable tool for early CVD risk identification, which could positively impact prevention efforts	sciencedirect.com/science/article/abs/pii/S0169260718311489	General Reviews and Surveys on LLMs in Healthcare
2023	Sarangi	Radiological Differential Diagnoses Based on Cardiovascular and Thoracic Imaging Patterns: Perspectives of Four Large Language Models	Assess the ability of four large language models (ChatGPT, Google Bard, Microsoft Bing, Perplexity) in providing differential diagnoses for cardiovascular and thoracic imaging patterns.	The study compared model-generated differential diagnoses with expert radiologists' consensus for 15 imaging patterns.	Perplexity had the highest concordance (66.67%) and acceptance (90.67%) rates, while Bard scored the lowest in both metrics.	LLMs offer promise in aiding radiologists but show variability in accuracy, emphasizing the need for careful model selection in clinical contexts	thieme-connect.de/products/ejournals/abstract/10.1055/s-0043-1777289	General Reviews and Surveys on LLMs in Healthcare

2023	Chen	Artificial Intelligence for Risk Assessment on Primary Prevention of Coronary Artery Disease	Review the use of AI in enhancing primary prevention strategies for coronary artery disease (CAD) through improved risk assessment.	The review examines AI's role in processing multimodal data, including biomarkers and genetic information, to predict CAD risk more precisely.	AI models have shown potential to outperform traditional risk assessment tools, but gaps remain in clinical validation and actionable implementation.	AI-enhanced risk assessment could transform CAD prevention, yet requires rigorous validation to integrate effectively into clinical settings	link.springer.com/article/10.1007/s12170-023-00731-4	General Reviews and Surveys on LLMs in Healthcare
2024	Houssein	Adapting Transformer-Based Language Models for Heart Disease Detection and Risk Factors Extraction	Adapt transformer models for detecting heart disease and extracting related risk factors from electronic health records.	Five transformer models (BERT, RoBERTa, BioClinicalBERT, XLNet, BioBERT) were fine-tuned using the i2b2 dataset for heart disease risk factor identification.	RoBERTa achieved the highest performance with an F1 score of 94.27%, while the ensemble model showed superior overall accuracy.	Transformer-based models effectively automate risk factor extraction, demonstrating high potential for clinical NLP applications	link.springer.com/article/10.1186/s40537-024-00903-y	General Reviews and Surveys on LLMs in Healthcare

2023	Guazzo	Deep-learning-based Natural Language Processing Models to Identify Cardiovascular Disease Hospitalizations of Patients with Diabetes from Routine Visits' Text	Develop deep-learning NLP models to identify cardiovascular disease (CVD) hospitalizations from unstructured clinical text in diabetic patients' electronic health records (EHRs).	The models were evaluated across four time windows, varying from an unlimited window to a six-month limit before each clinical visit.	The models performed best with an unlimited or 24-month time window, achieving high precision and recall, but performance declined with shorter time windows.	NLP models can effectively identify CVD history in clinical text, enhancing data usability in EHRs for retrospective analysis and clinical decision-making	nature.com/articles/s41598-023-45115-1	General Reviews and Surveys on LLMs in Healthcare
2024	Quer	The Potential for Large Language Models to Transform Cardiovascular Medicine	Examine the transformative potential of large language models (LLMs) in cardiovascular medicine, especially for early diagnosis and patient care.	This paper reviews applications of LLMs, including patient communication, EHR summarization, and risk prediction through multimodal AI.	LLMs show significant promise for automating tasks and supporting diagnostic accuracy but face challenges like data bias and patient privacy concerns.	LLMs could revolutionize cardiovascular medicine by enhancing patient care and efficiency, though regulatory and ethical issues need addressing for safe clinical integration	thelancet.com/journals/landig/article/PIIS2589-7500(24)00151-1	General Reviews and Surveys on LLMs in Healthcare

2023	Saridena	A Supervised Deep Learning Model for the Detection of Cardiovascular Disease	Develop a supervised deep learning model to detect cardiovascular disease (CVD) using patient data.	The model was trained on a dataset of 70,000 records using various layers, activation functions, and optimizers to achieve high accuracy in predicting CVD.	The model achieved up to 73% accuracy, demonstrating its potential for reliable CVD detection.	AI-based models like this could transform CVD diagnosis, enabling healthcare professionals to identify at-risk patients effectively	www.jsr.org/hs/index.php/path/article/view/5178	General Reviews and Surveys on LLMs in Healthcare