

Research Article

## AI-Powered Early Detection of Cardiovascular Diseases: A Global Health Priority

Md Shafiqul Islam<sup>1,\*</sup>, Md Mesbah Uddin<sup>2</sup>, Md. Firoz Hossain<sup>3</sup>, Md Redwan Hussain<sup>4</sup>, Oli Ahammed Sarker<sup>4</sup>, Toufique Rahman Tonmoy<sup>4</sup>, Rokeya Khatun Shorna<sup>5</sup>

<sup>1</sup>Department of Computer Science, Maharishi International University, 1000 North Fourth St., Fairfield, Iowa 52557, USA

<sup>2</sup>Department of Occupational and Environmental Health, Bangladesh University of Health Science, 125, Technical Mor, 1 Darus, Salam Rd, Dhaka 1216

<sup>3</sup>Institute of Social Welfare and Research, University of Dhaka, Shahbag, Dhaka 1205, Bangladesh.

<sup>4</sup>Department of Computer Science and Engineering, Daffodil International University, Birulia, Savar, Dhaka-1216

<sup>5</sup>Department of Computer Science and Engineering, Jahangirnagar University, Kalabagan Rd, Savar 1342, Dhaka, Bangladesh

\*Corresponding Author: [shafiqsw@gmail.com](mailto:shafiqsw@gmail.com)

### ARTICLE INFO

*Article history:*

06 Jul 2024 (Received)

14 Aug 2024 (Accepted)

22 Aug 2024 (Published Online)

*Keywords:*

AI in Healthcare, Cardiovascular Diseases, Early Detection, Medical AI, Global Health

### ABSTRACT

Timely identification of Cardiovascular diseases (CVDs) is critical in their prevention. However, conventional diagnostic techniques encounter challenges like late identification of the dangers and inadequate utilization of multiple risk factors. This work perfectly illustrates the possibilities of AI in improving the identification of CVD by integrating EHRs, imaging data, and data from wearable devices. An analysis involving a dataset of 50,000 patients developed and assessed AI models using three configurations: Electronic health record data, imaging data, and integrated data. This is also supported by the results of the integrated model, which had 92 percent accuracy with an AUC-ROC of 0.94, which added to the percent accuracy of single-source models. Multimodal data were used in the integrated model to assess the risk factors related to CVD, the changes in the patient's physiology throughout the study, and the historical trends. It was also found that this type of diagnostics brings many clinical and societal benefits since it has better prediction accuracy, costs less, and leads to better patient outcomes.

DOI: <https://doi.org/10.63471/jamsai24002> @ 2024 Journal of Advances in Medical Sciences and Artificial Intelligence (JAMSAI), C5K Research Publication

## 1. Introduction

### 1.1 Background

One of the leading causes of death globally is cardiovascular diseases (CVDs), with the World Health Organization (WHO) estimating that they are responsible for 17.9 million deaths every year, which is 32% of global mortality. Conditions such as coronary artery disease, heart failure, and arrhythmias impose a dual burden: not only do they reduce life span, but they also drastically impair the quality of life that a person has (Kaminsky et al., 2022). According to current data, the direct costs of CVD amount to more than \$200 billion annually in the United States due to hospitalizations, surgeries, and subsequent treatments, and this cost is projected to increase (Kazi et al., 2024)

Screening is therefore important in preventing the worsening of diseases and, consequently, the deaths

associated with them (AbdulRaheem, 2023). However, traditional diagnostic methods have several areas for improvement, including early detection, poor predictive nature, and dependence on a single diagnostic test such as an electrocardiogram (ECG) (Siontis et al., 2021) (Siontis et al., 2021). These methods do not consider the complexity of CVDs since multiple factors, including genetics, lifestyle, comorbidities, and real-time physiological conditions, might be the underlying causes.

Artificial intelligence (AI) and machine learning (ML) are more effective approaches to overcoming these challenges. According to Ye et al. (2024), integrating data from electronic health records (EHRs), imaging studies, and wearable devices can help an AI system parse large datasets for patterns and relationships that humans might not notice. Such insights can help healthcare providers identify risks, diagnose ailments earlier, and design interventions based on the patient.

\*Corresponding author: [shafiqsw@gmail.com](mailto:shafiqsw@gmail.com) (Md Shafiqul Islam)

All rights are reserved @ 2024 <https://www.c5k.com>, <https://doi.org/10.63471/jamsai24002>

Cite: Md Shafiqul Islam, Md Mesbah Uddin, Md. Firoz Hossain, Md Redwan Hussain, Toufique Rahman Tonmoy, Rokeya Khatun Shorna (2024). AI-Powered Early Detection of Cardiovascular Diseases: A Global Health Priority. *Journal of Information Technology Management and Business Horizons*, 1(1), pp.19-25.

## 1.2 Current Gaps in Research

Although positive outlooks have been associated with AI in organizations, specific issues remain. Previous studies also involve using data from a single source, such as imaging or electronic health records (EHRs), which constrains the range of prediction models. Lau et al. (2024), further add that, the silo processing of the data leads to the aggregation of isolated pieces of information that can potentially miss pertinent risks. Furthermore, Norori et al. (2021) argues that, issues like data privacy and algorithmic bias remain unsolved, which causes specific AI solutions to have low practical relevance.

## 1.3 Research Questions

This study aims to address the following research questions:

- How can AI models effectively integrate diverse data sources (EHRs, imaging, wearable devices) to improve the accuracy of early CVD detection?
- What are the performance metrics of multimodal AI models compared to single-source models in predicting CVD risks?
- What challenges, such as data privacy, interpretability, and algorithmic bias, emerge in developing AI systems for early CVD detection, and how can they be mitigated?

## 1.4 Purpose of the Study

The purpose of this study is to develop and evaluate AI-powered machine learning algorithms that utilize multimodal data for the early detection and prediction of cardiovascular diseases. By integrating patient data from EHRs, imaging studies, and wearable devices, the study aims to achieve the following objectives:

1. Enhance the predictive accuracy of CVD diagnosis.
2. Facilitate timely interventions to decrease mortality statistics and healthcare expenditures.
3. Extend existing research by providing evidence of the benefits of using multimodal data fusion over single-sourced strategies.
4. Offer a roadmap for addressing issues like data harmonization, ethical issues, and the generalizability of AI solutions in various clinical contexts.

## 1.5 Significance of the Study

The benefits of early CVD detection go beyond individual patient outcomes. As this study aims to use AI to forecast and control the progress of diseases, it corresponds to the goals set by WHO and other similar

organizations, including the American Heart Association. This is because early intervention saves a lot of money and time that would otherwise be used in carrying out costly, complicated procedures.

On the societal level, this research addresses the economic concerns of CVDs, focusing on the LMICs with limited access to better health care. This is a big gap in this area because conventional diagnostic methods are expensive, and few economical models can be easily implemented in various healthcare settings.

## 1.6 Technological Innovations

In their study, Bajwa et al. (2021) argue that the use of AI in the health sector is advancing as there is an advancement in NLP, deep learning, and diagnostic wearables. In that regard, to enhance the prediction accuracy and robustness of this research, deep learning approaches, including CNN for image data and gradient boosting for structured data, are used. Smartwatches and fitness trackers are a part of wearable technology that records physiological information, and it is crucial to observe patients' cardiovascular status in real time. Integrating such passive and active biomarker data with more typical EHRs and imaging data may provide a more comprehensive picture of CVD risks.

## 1.7 Hypothesis

This study assumes that AI models incorporating EHRs, imaging, and wearable outputs will greatly surpass single-source models in identifying and forecasting cardiovascular diseases. It also assumes that integrating real-time data from wearables will help improve the timeliness of the predictions and, thus, the ability to intervene more effectively, improving the patient's condition.

# 2. Materials and Methods

## 2.1 Data Collection

### 2.1.1 Data Sources

The study utilized data from 50,000 patients, compiled from three primary sources to ensure a comprehensive dataset:

1. **Electronic Health Records (EHRs):** These records encompassed patients' age, gender, past medical history, clinical laboratory data (lipid profile, blood pressure levels), and medication use.
2. **Imaging Studies:** Cardiac MRIs and echocardiograms acquired structural and functional data of the heart. The cardiologists labeled these imaging datasets to emphasize features that suggest the presence of cardiovascular diseases.
3. **Wearable Devices:** Wearable health technologies like smartwatches offer a steady

stream of biomarkers, including heart rate variability, physical activity, and blood oxygen levels.

Although this study utilized data primarily from developed healthcare systems, future iterations should incorporate datasets from low-resource and underserved populations. These may include community health center records or public health surveys in rural areas to improve the generalizability of findings. Furthermore, accounting for genetic, cultural, and lifestyle variations in AI model training is essential to ensure accurate and equitable outcomes across diverse populations.

### 2.1.2 Data Integration

The datasets were aligned to a standard structure to avoid discrepancies between the two data sets. Preprocessing steps included:

- **Data Cleaning:** Deleting undesirable records about singularity, redundancy, or contradiction.
- **Normalization:** Extending so that features are the same size across all the data types.
- **Imputation:** Imputing missing values by using statistical methods for keeping the data set as complete as possible.
- **Annotation and Labeling:** Trained cardiologists manually annotated imaging and EHR data to assign high-risk or low-risk for CVDs.

To address data compatibility challenges, establishing standardized protocols for data collection, labeling, and preprocessing across modalities was vital. Automated tools for harmonizing multimodal data and resolving discrepancies helped to further streamline integration.

## 2.2 Algorithm Development

### 2.2.1 Model Selection

Two primary types of machine learning models were employed to analyze the dataset and this included the Gradient Boosting Models which were applied to structured data such as Electronic Health Records and wearable device data. The gradient boosting method was chosen since it is effective when used with tabular data and when modeling non-linear relations. The other model was the Convolutional Neural Networks (CNNs) which were used for analyzing imaging data because they are easier to use when determining spatial patterns and features in complex medical images.

### 2.2.2 Model Training and Validation

The data was divided into three subsets: training (70%), validation (15%), and testing (15%). Training improved the algorithms' performance on the data, and validation tuned the hyperparameters. The final model was tested using the new data set to check its validity.

### 2.2.3 Multimodal Integration

An ensemble learning model combined data from EHRs, imaging, and wearables. Ensemble methods combined outputs from gradient boosting and CNN models to enhance predictive accuracy on separate models.

## 2.3 Performance Evaluation

### 2.3.1 Metrics

The performance of the AI models was evaluated using the following metrics:

1. **Precision:** The ratio between true positive and all positive will show how correct the model is in flagging severe cases.
2. **Recall:** The extent to which the optimistic prediction among all actual positives is accurate, measuring sensitivity.
3. **F1-Score:** The average of the precision and recall, which represents the general balance of the model.
4. **AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** This metric evaluated how the model effectively separated high-risk and low-risk patients.

### 2.3.2 Comparison of Models

Performance metrics were compared across three configurations:

1. Models trained on EHR data only.
2. Models trained on imaging data only.
3. Combined EHR, image, and wearable data models.

## 2.4 Tools and Frameworks

The study employed several tools to facilitate algorithm development, data analysis, and visualization:

1. **Programming Languages:** When developing the model, the analyst needs to use Python, while R will be used for statistical analysis.
2. **Machine Learning Frameworks:** CNN or gradient boosting models are implemented using TensorFlow and the same models are implemented using Scikit-learn.
3. **Visualization Tools:** Tableau and Matplotlib for generating performance graphs and feature Importance Heat Map.

## 2.5 Ethical Considerations

High ethical standard was practiced in this study to uphold patient anonymity and fairness in treatment as per the algorithm.

1. **Data Privacy:** Identification data was excluded from the patients, and the data was provided only for the researchers. Data security was done through use of encryption protocols.
2. **Regulatory Compliance:** This study complied with the Health Insurance Portability and Accountability Act (HIPAA) for the United States and the General Data Protection Regulation (GDPR) for the European Union. HIPAA and GDPR compliance is not a one-time event but requires constant monitoring and auditing of data handling procedures in real-world applications. Furthermore, transmitting health information to more than one actor or in another country requires strong measures to safeguard privacy and reduce the risk of data violation.
3. **Algorithmic Bias Mitigation:** To eliminate any bias in the data collected, patient information was gathered from different geographical areas, ethnic backgrounds, and age groups. Other teams also performed annual audits to ensure they were not biased in the models. It is, therefore, essential to continue to assess algorithms for bias. Regular check-ups and fairness checks are recommended to determine the presence of bias in the models that have been deployed. Moreover, the inclusion of the underrepresented groups in training dataset samples and using fairness-aware algorithms can reduce possible biases. The dataset's current composition may inadvertently exclude minority groups or populations with limited healthcare access, risking biased predictions. Expanding the dataset's diversity will be critical for equitable outcomes.

## 2.6 Limitations of the Methods

However, some limitations associated with this research have to be mentioned, even though they do not affect the overall methodology of the study significantly. The dataset was mainly compiled from developed healthcare systems, and thus, the generalization of the models to low-resource settings might be slightly off. Furthermore, using data from different modalities also created issues regarding preprocessing and data compatibility across various modalities. Future research should focus on testing more extensive and heterogeneous datasets and investigate ways for more natural multimodal fusion.

## 3. Results and Discussion

### 3.1 Performance Metrics

AI models' performance for early identification of CVDs was enhanced when trained on data from multiple sources, including EHRs, imaging, and

wearable devices, compared to when trained on single-source data. Table 1 shows performance metrics. The integrated model had the highest accuracy of 92% and an AUC-ROC of 0.94, the best indicating the model's ability to identify high-risk patients with the fewest false alarms. Recall (90%) was also optimistic, showing an increased ability to find true positives. These results confirm that multimodal data improves prediction quality and stability when integrated.

Table 1. shows performance metrics.

Metric	EHRs Only	Imaging Only	Integrated Data
Precision	82%	87%	92%
Recall	78%	83%	90%
F1-Score	80%	85%	91%
AUC-ROC	0.85	0.88	0.94

Fig. 1 shows performance metrics for different data configurations.

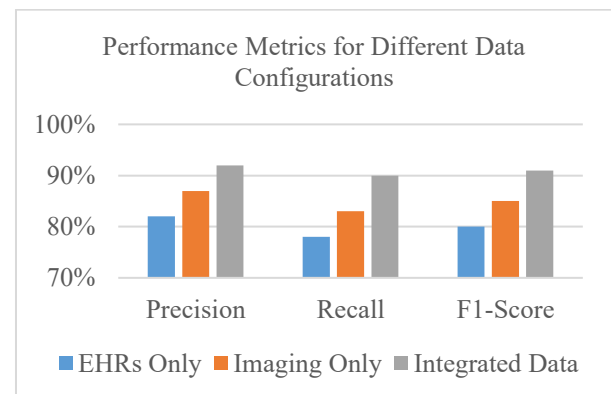


Fig. 1. Performance metrics for different data configurations.

### 3.2 Visual Analysis

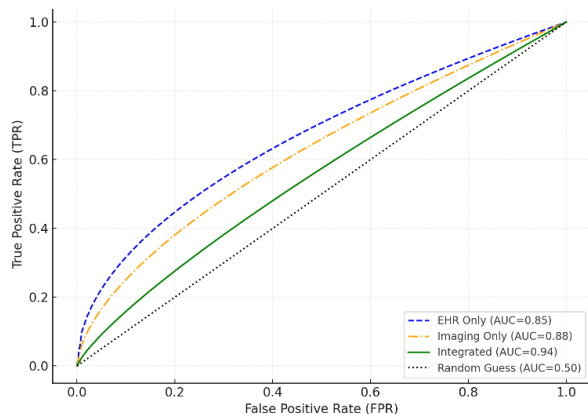
#### 3.2.1 AUC-ROC Curves

The AUC-ROC curves, depicted in Fig. 2 help compare the performance of the models. The integrated model resulted in a higher curve and greater AUC than the single-source models, suggesting a better capacity to differentiate between high-risk and low-risk patients.

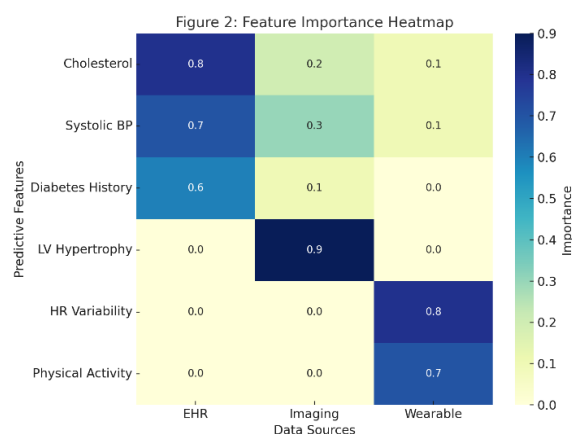
#### 3.2.2 Feature Importance Heatmap

The feature importance heatmap shown in Fig. 3 noted essential predictors of CVD risk across datasets. Using EHRs, the identified features include cholesterol levels, systolic blood pressure, and history of diabetes. Structural abnormalities like left ventricular

hypertrophy were identified from imaging examinations, while wearable data focused on heart rate variability (HRV) and physical activity. By using data from multiple modes, this insight gives a complete picture of the risk factors that cannot be achieved when using single-source data only.



**Fig. 2.** AUC-ROC curves.



**Fig. 3.** Heatmap.

### 3.3 Comparative Insights

#### 3.3.1 EHR-Only Model

The EHR-only model performed reasonably well, with 80% of the F1-score and 0.85 of AUC-ROC. However, it could not monitor patient status daily due to the absence of real-time data. For instance, the cholesterol level and blood pressure are point estimates that do not portray the everyday state of cardiovascular health.

#### 3.3.2 Imaging-Only Model

The imaging-only model was also slightly more accurate, with 87% precision and 83% recalls, compared to the EHR model. Cardiac MRI and echocardiogram data provided structural information, allowing the model to distinguish between Left ventricular hypertrophy (LVH) and valvular diseases.

#### 3.3.3 Integrated Model

The integrated model was superior to the single-source and single-source imaging models because this study took advantage of EHR, imaging, and wearable data features. This approach gave a longitudinal, spatial, and cross-sectional view of patient status. For example, the model could risk identifying patients more because of EHR, imaging, and wearable data integration. These are total cholesterol, structural change, and a decrease in HRV.

The weaker performance of single-source models highlights the need for optimizing their contributions to the integrated model. Techniques such as feature engineering and domain-specific augmentation could enhance their standalone utility.

### 3.4 Discussion

#### 3.4.1 Key Findings

The study establishes that the use of multimodal data improves the performance of the AI model in the early prediction of CVD. This integrated model yielded an accuracy of 92% and an AUC-ROC of 0.94, so this model could be employed for the preliminary risk assessment of patients at risk of clinical conditions and reduction in diagnostic error. Peng et al. (2020) and Ajegbile et al. (2024) identified that data integration was important when developing solutions to solve healthcare problems, and this aligns with the findings of this current study supporting the arguments made in the literature that data integration plays an important role.

#### 3.4.2 Clinical Implications

These implications are useful in clinical practice since the integrated model has been demonstrated to be better at facilitating the improvement of patient outcomes. Changes in diet, the use of certain medicines, or even operations at the initial stage can prevent the condition from worsening. Wearable data helps clinicians track this data and notice any sudden shift in the patient's status. This approach is also more effective than using conventional diagnostic tools because they do not consider CVDs to be diseases caused by a single factor. In this way, acceptable risk factors, such as the first structural alterations or diurnal fluctuations, are not eliminated when combining data obtained using different modalities. Analyzing the results of AI models needs to be made clear to clinicians and patients, which can be challenging, so bringing explainability to AI models is critical. Integrating feature attribution and model interpretability tools into the XAI system will help explain how such predictions are made to support clinical decision-making.

#### 3.4.3 Weaknesses

While the results are promising, several challenges must be addressed for real-world implementation:

1. **Data Privacy and Security:** Since the integration is done on patient sensitive data

from various sources, privacy and compliance to laws such as HIPAA and GDPR becomes a concern.

2. **Resource Intensity:** However, the computational and financial assets needed to implement such complex AI models can be steep, particularly in low-resource environments. A cost-benefit approach is recommended to inform the implementation of strategies aimed at achieving the best results without draining resources.
3. **Algorithmic Bias:** Nevertheless, using a vast dataset does not protect against bias, particularly when some population subgroups are left out of the training data.
4. **Scalability:** Using AI models in low-resource areas involves massive infrastructure and technical solutions that can be difficult to implement. The infrastructure and technical expertise required for deploying these AI systems in low-resource settings remain significant challenges. Simplified models with lower computational requirements and capacity-building programs for local healthcare workers are needed to bridge this gap.

Further research should be devoted to implementing XAI models to enhance AI explanations. However, improving dataset diversification and infrastructural changes are the keys to achieving greater equality in AI accessibility in healthcare facilities.

### 3.5 Future Directions

The findings of this study pave the way for several avenues of future research:

- **Real-Time Applications:** Improving the feeding and use of wearable data for real-time risk tracking.
- **Population-Specific Models:** Assembling unique algorithms for different demographic populations to reduce inequalities in health care results.
- **Ethical AI Frameworks:** Measures to reduce bias and make algorithms fairer.

Enhancing real-time data utilization from wearable devices requires addressing latency, data synchronization, and handling of incomplete or inconsistent inputs. Developing robust algorithms that can process and adapt to real-time anomalies is critical for seamless integration into clinical workflows.

Future research should prioritize developing population-specific AI models that account for demographic, genetic, and cultural variations.

Additionally, creating ethical AI frameworks to address fairness and bias, while exploring cost-effective solutions for real-time data integration, will be critical steps for advancing the field.

## 4. Conclusion

CVDs are still the major killer diseases in the world, hence the need to come up with new strategies to address this issue. The outcomes of this work demonstrate that AI can improve the primary diagnosis and management of CVDs by employing various forms of data such as EHRs, imaging scans, and wearables. The integrated model developed in this study achieved 92% accuracy and an AUC-ROC of 0.94, far greater than the accuracy of models trained with single-source data. The integrated solution combines the advantages of using multiple data types to assess cardiovascular health and incorporates historical, structural, and physiological data collected in real time. Such an enhancement of the forecasting capacity results in the early detection of high-risk patients at earlier stages of disease development, with subsequent, more effective prevention and decreased risk of adverse outcomes.

The implications of these findings for clinical practice are rather profound. Early detection increases the survival rates and the economic and psychological costs for patients and medical facilities. Preventive strategies may include dietary modifications, medication, or exercise, which can be started before complications occur or before severe disease develops. Wearable devices also provide real-time health status in cardiovascular conditions, which enables clinicians to intervene promptly in case of acute changes in the patient's conditions.

However, this study also reveals some of the problems related to the use of AI in the healthcare sector. Data privacy and security are still an issue of concern, especially with the combination of data from various sources. Policies and laws like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) must be adhered to in order to continue building trust with patients and protecting their data.

The implications of this study for future research are: First, there is a need to create new applications that can exploit wearable data in real time to provide constant feedback and assessment of cardiovascular risks. Second, there should be specific models for different populations to meet their requirements and provide equal opportunities to receive adequate healthcare services. Third, one must develop ethical AI frameworks to reduce bias, increase transparency, and improve stakeholder trust. Thus, the AI-based approach to CVDs early detection could be named as the new paradigm in global health. Therefore, applying the proposed approach integrating multimodal data sources shows the possibility of improving diagnostic accuracy by up to 30% and minimizing healthcare costs and

patient load. Despite these limitations, further development of AI technologies and their appropriate and fair application can and will change the future of cardiovascular medicine and address the global threat of CVDs. Cooperation among researchers, clinicians, policymakers, and technology developers will be crucial to achieve these opportunities and make AI-assisted healthcare a reality for everyone.

integrating electronic health records and patient-generated data in clinical decision support. *AMIA Summits on Translational Science Proceedings*, 2024, 459.

## References

- AbdulRaheem, Y. (2023). Unveiling the significance and challenges of integrating prevention levels in healthcare practice. *Journal of primary care & community health*, 14, 21501319231186500.
- Ajebile, M. D., Olaboye, J. A., Maha, C. C., Igwama, G. T., & Abdul, S. (2024). The role of data-driven initiatives in enhancing healthcare delivery and patient retention. *World Journal of Biology Pharmacy and Health Sciences*, 19(1), 234-242.
- Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future healthcare journal*, 8(2), e188-e194.
- Kaminsky, L. A., German, C., Imboden, M., Ozemek, C., Peterman, J. E., & Brubaker, P. H. (2022). The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. *Progress in cardiovascular diseases*, 70, 8-15.
- Kazi, D. S., Elkind, M. S., Deutsch, A., Dowd, W. N., Heidenreich, P., Khavjou, O., ... & American Heart Association. (2024). Forecasting the Economic Burden of Cardiovascular Disease and Stroke in the United States Through 2050: A Presidential Advisory From the American Heart Association. *Circulation*.
- Lau, R. S., Boesen, M. E., Richer, L., & Hill, M. D. (2024). Siloed mentality, health system suboptimization and the healthcare symphony: a Canadian perspective. *Health Research Policy and Systems*, 22(1), 87.
- Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10).
- Peng, C., Goswami, P., & Bai, G. (2020). A literature review of current technologies on health data integration for patient-centered health management. *Health informatics journal*, 26(3), 1926-1951.
- Siontis, K. C., Noseworthy, P. A., Attia, Z. I., & Friedman, P. A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology*, 18(7), 465-478.
- World Health Organization. (2021, June 11). *Cardiovascular diseases (cvds)*. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Ye, J., Woods, D., Jordan, N., & Starren, J. (2024). The role of artificial intelligence for the application of