Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An AI-Driven Multimodal Data Integration Framework for Holistic Patient Profiling in Healthcare Robotics

KOUSHIKI Koushiki

koushiki.verma47242@paruluniversity.ac.in

Parul University

Arun Udayasuriyan

Parul University

Manoj S Dikkatwar

Parul Institute of Pharmacy

Priya R. Swaminarayan

Professor & Dean Faculty of IT & CS

Research Article

Keywords: Artificial Intelligence (AI), Multimodal Data Integration, Electronic Health Records (EHR), Heart Disease Prediction, Healthcare Robotics, Machine Learning (ML), Holistic Patient Profiling

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11159021/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An AI-Driven Multimodal Data Integration Framework for Holistic Patient Profiling in Healthcare Robotics

Koushiki[1], Arun Udayasuriyan[2], Dr. Manoj S Dikkatwar[3], Dr. Priya R. Swaminarayan[4]

[1]Assistant Professor, Department of BCA, Parul Institute of Computer Application-BCA, Vadodara, Gujarat, koushiki.verma47242@paruluniversity.ac.in

[2]Assistant Professor, Department of MCA, Parul Institute of Engineering and Technology-MCA, Vadodara, Gujarat, arun.udayasuriyan29426@paruluniversity.ac.in

[3]Associate Professor & Head, Department of Pharmacy Practice, Parul Institute of Pharmacy, Vadodara, Gujarat, manoj.dikkatwar41518@paruluniversity.ac.in

[4]Professor & Dean Faculty of IT & CS, Vadodara, Gujarat, priya.swaminarayan@paruluniversity.ac.in

Abstract— With the fast development of healthcare data, there is a lot of opportunities are there for artificial intelligence (AI) to make improvement in disease diagnosis, patient monitoring, and clinical decision-making. While heterogeneous distribution of healthcare information makes comprehensive patient analysis challenging. This paper proposes a multimodal framework AIMDI-HR (AI-Driven Multimodal Data Integration for Holistic Patient Profiling in Healthcare Robotics), that integrates Electronic Health Records (EHR) and physiological vital-sign data which supports intelligent patient profiling and hearth disease prediction. This proposed framework concludes data pre-processing, feature normalization, multimodal data integration, prediction based on machine-learning, and feature importance analysis. In this paper three supervised machine learning models: Logistic Regression, Random Forest, and XGBoost, were estimated using a multimodal dataset which is built from healthcare datasets which are publicly available. After the experiment, the results signify that the Random Forest model got the highest performance with 90.16% accuracy, 89.29% F1-score, and 0.9448 ROC-AUC. Also feature importance analysis indicates contribution of both clinical and physiological features toward predication of disease. The proposed framework shows ability of multimodal AI techniques to increase predictive healthcare system and gives support for decisions for future healthcare robotic applications.

Index Terms— Artificial Intelligence (AI), Multimodal Data Integration, Electronic Health Records (EHR), Heart Disease Prediction, Healthcare Robotics, Machine Learning (ML), Holistic Patient Profiling.

I. INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative technology in modern healthcare by enabling intelligent dis-ease diagnosis, patient monitoring, clinical decision support, and personalized treatment planning [1]. The rapid acceptance of Electronic Health Records (EHRs), wearable sensors, Internet of Things (IoT) devices, medical imaging systems, and healthcare robotics has generated large volumes of heterogeneous healthcare data. Effectively integrating these diverse data sources can significantly improve disease prediction and support data-driven clinical decision-making [1],[8].

Traditional healthcare prediction systems generally rely on a single source of medical information, such as Electronic Health Records or physiological measurements [3],[6]. Although these approaches have demonstrated promising performance, they often fail to capture the complete clinical condition of a patient because valuable information from multiple healthcare modalities is ignored [6],[7]. Recent studies have shown that multimodal Artificial Intelligence, which combines heterogeneous healthcare data sources, provides more comprehensive patient profiling and improves predictive performance compared to unimodal approaches [2], [3].

In addition to multimodal learning, healthcare robotics has become an important application of

Artificial Intelligence for patient monitoring, rehabilitation, medication assistance, and intelligent clinical support. The integration of AI, IoT, robotics, and personalized healthcare technologies has opened new opportunities for developing patient-centered healthcare systems capable of continuous monitoring and real-time decision-making [4], [5].

Driven by these development, this paper proposes the **AI-Driven Multimodal Data Integration for Holistic Patient Profiling in Healthcare Robotics (AIMDI-HR)** frame-work. The proposed framework integrates Electronic Health Records (EHR) and physiological vital-sign data through data preprocessing, feature-level multimodal fusion, machine learning-based disease prediction, and feature importance analysis. Logistic Regression, Random Forest, and XGBoost are carried out and analyzed for heart disease prediction, while Random Forest feature importance analysis is done to check the most influential clinical and physiological features evolve to disease prediction and improve model interpretability.

Although the proposed AIMDI-HR framework is designed to support multiple healthcare data modalities, including medical images, ECG signals, EHR, and physiological vital signs, the current implementation utilizes publicly available EHR and vital-sign datasets because patient-aligned multimodal datasets containing all modalities are not readily available. However this research shows the feasibility of Multimodal healthcare data integration and it provides a base for future intelligent healthcare robotic systems.

The major evolution of this research are identified as follows:

- This research Proposed the AIMDI-HR framework for multimodal healthcare data integration and holistic patient profiling.
- Integrated Electronic Health Records and physiological vital-sign data using feature-level multimodal fusion.
- Implemented and compared Logistic Regression, Random Forest, and XGBoost models for heart disease prediction.
- Performed Random Forest feature importance analysis to get the most important clinical and physiological features contributing to heart disease prediction and improve model interpretability.

Demonstrated the applicability of multimodal Artificial Intelligence for intelligent healthcare decision support and future healthcare robotic systems.

II. Related Work

Artificial Intelligence (AI) has significantly transformed modern healthcare by allowing intelligent disease diagnosis, patient monitoring, and clinical decision support. Recent studies have demonstrated that multimodal AI frameworks, which combine heterogeneous healthcare data such as Electronic Health Records (EHR), medical images, physiological signals, and clinical notes, improve predictive performance and pro-vide comprehensive patient profiling compared with single-modality approaches [1], [3].

Multimodal data fusion has become a key research direction for developing intelligent healthcare systems. Several surveys have reviewed feature-level, decision-level, and hybrid fusion strategies for integrating heterogeneous medical data while highlighting challenges related to missing modalities, data heterogeneity, scalability, and clinical deployment [6],[8]. These studies highlight that effective multimodal integration can improve healthcare analytics and decision support significantly.

Several researchers have researched AI-based integration of EHR (Electronic Health Records) with medical imaging and other treatment methods for personalized medicine. AI-driven fusion techniques have shown promising results in disease prediction, patient stratification, and clinical outcome analysis by combining complementary healthcare information from multiple sources [9],[12]. These approaches show the potential of multimodal learning for improving diagnostic accuracy.

Healthcare robotics has come out as an important application of Artificial Intelligence for patient monitoring, rehabilitation, medication assistance, and intelligent clinical support. Recent studies

have highlighted that integrating AI, IoT, wearable sensors, and robotics enables continuous patient monitoring and personalized healthcare delivery while introducing challenges related to interoperability, privacy, and ethical AI deployment [4]

Most healthcare studies are limited to types of care. They also talk about ideas and theories. They do not show how to actually use these ideas with healthcare data that everyone can access. The healthcare studies do not make use of publicly available datasets, which is a problem. Healthcare studies like these need to be more practical and use healthcare data that's available, to the public. There is not a lot of research that brings together types of healthcare data and uses machine learning to make predictions. This type of research also needs to look at which features most important and use healthcare robots. The AIMDI-HR framework is an idea that combines Electronic Health Records and physiological vital-sign data. It uses machine learning. Looks, at which features are most important to figure out what makes a patient's profile. The AIMDI-HR framework uses a way of combining data called feature-level multi-modal fusion and a method called Random Forest to find the most important clinical and physiological features. This helps with patient profiling. The AIMDI-HR framework is a way to do things that uses Machine Learning (ML) and healthcare robots to make patient care better.

A. Framework Overview

The rapid advancement of Artificial Intelligence (AI), machine learning, and healthcare robotics has significantly transformed modern healthcare by enabling intelligent disease diagnosis, patient monitoring, clinical decision support, and personalized healthcare services [1],[4]. The increasing availability of Electronic Health Records (EHRs), physiological vital-sign data, wearable sensors, and medical imaging technologies has generated large volumes of heterogeneous healthcare information. However, conventional healthcare systems often analyze these data modalities independently, resulting in fragmented patient information and limiting the ability to obtain a com-prehensive understanding of an individual's health condition. Recent advances in multimodal Artificial Intelligence (AI) have identified that integrating heterogeneous healthcare data can improve predictive performance and support more in-formed clinical decision-making. Motivated by these developments, this paper proposes the **AI-Driven Multimodal Data Integration for Holistic Patient Profiling in Healthcare Robotics (AIMDI-HR)** framework. The proposed framework is designed to integrate multiple healthcare data modalities into a unified patient representation that supports intelligent disease prediction, explainable decision-making, and future healthcare robotic applications. The overall framework design is inspired by recent multimodal healthcare architectures and explainable AI systems proposed in the literature [1], [13].

Although the AIMDI-HR framework is conceptually de-signed to support four major healthcare modalities, namely Electronic Health Records (EHR), medical images, physio-logical signals (ECG), and physiological vital signs, the cur-rent implementation utilizes only publicly available Electronic Health Record and vital-sign datasets because patient-aligned multimodal datasets containing all modalities are not readily available. The inclusion of medical images and ECG within the framework provides a scalable architecture that can be extended in future research without modifying the overall system design.

The proposed AIMDI-HR framework consists of eight interconnected modules: Patient Data Acquisition, Data Pre-processing, Feature Extraction, Feature-Level Multimodal Fusion, Explainable Artificial Intelligence (XAI), Holistic Patient Profiling, Clinical Decision Support, and Healthcare Robot Interaction. Each of the modules has its own role, but all of them work towards the development of a smart and scalable healthcare ecosystem.

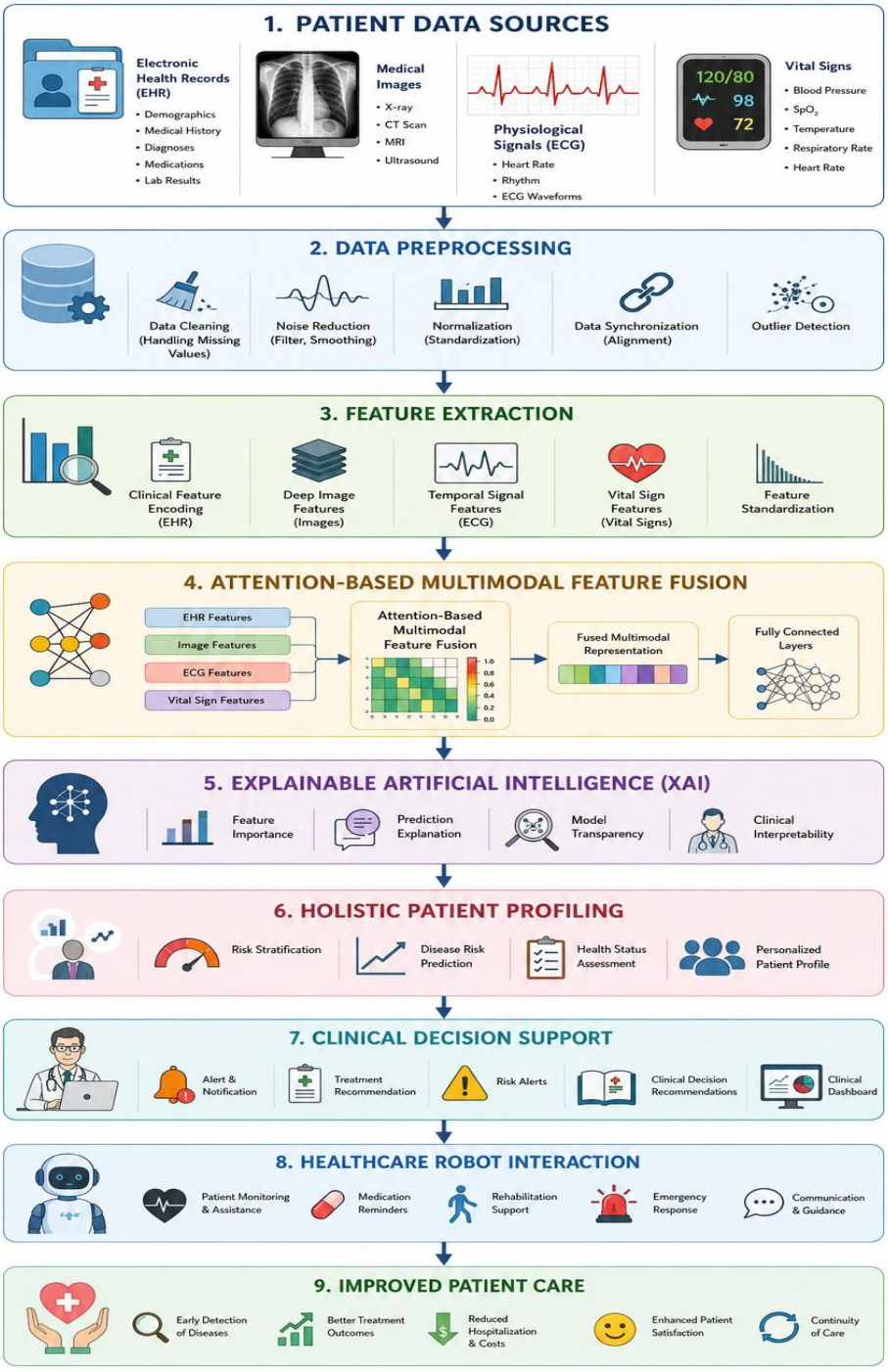


Fig. 1. Overall structure of the proposed AIMDI-HR framework.

The first aim of the suggested framework is to develop a patient profile using health information from different sources in a complementary way. This patient profile will then be used in machine learning-based disease prediction and explainable decision support for further healthcare robotics application.

Figure 1 indicates the structure of the proposed framework of AIMDI-HR, where the workflow is shown in a sequential way from the data acquisition stage up to healthcare robotic decision support stage.

B. Patient Data Acquisition

Patient data acquisition is the first stage of the proposed AIMDI-HR framework, where healthcare information from multiple modalities is collected to facilitate comprehensive patient profiling. The framework is conceptually designed to support four major categories of healthcare data, namely Electronic Health Records (EHR), medical images, physio-logical signals (ECG), and physiological vital signs. These heterogeneous data sources provide complementary clinical information that enables a more complete understanding of a patient's health condition and supports intelligent healthcare decision-making [1], [3].

Electronic Health Records contain structured clinical information, including demographic characteristics, medical his-tory, laboratory test results, clinical observations, and disease-related attributes. In the present study, the Cleveland Heart Disease dataset is employed as the primary EHR data source for heart disease prediction [9]. This dataset contains 303 patient records with 13 clinical attributes and one target variable indicating that whether there is heart disease or not.

Physiological vital-sign data provide real-time details regarding a patient's physical condition through measurements such as heart rate, respiratory rate, body temperature, oxygen saturation, systolic blood pressure, diastolic blood pressure, body mass index (BMI), and other derived physiological indicators. In this work, the Human Vital Signs Dataset 2024 is utilized to complement the clinical information available in the EHR dataset. Since the vital-sign dataset contains substantially more records than the EHR dataset, a representative subset of 303 samples is randomly selected to maintain a balanced multimodal dataset for experimental evaluation.

Although the AIMDI-HR framework is designed to accommodate medical images and ECG signals, these modalities are not included in the current implementation because publicly available patient-aligned multimodal datasets containing all four modalities are not readily available. Their inclusion in the conceptual architecture provides a scalable framework that can be extended in future research without modifying the overall system design.

By integrating structured Electronic Health Records with physiological vital-sign information, the proposed framework demonstrates the feasibility of multimodal healthcare data integration while establishing a foundation for future incorporation of medical imaging, ECG signals, wearable sensors, and additional healthcare modalities [10].

C. Data Preprocessing

Data collected from different healthcare sources often varies in format, scale, and representation, making preprocessing an essential step before multimodal data integration. In the proposed AIMDI-HR framework, preprocessing is performed to improve data quality by reducing inconsistencies, standardizing feature values, and preparing heterogeneous healthcare datasets for subsequent machine learning analysis and multimodal fusion [3],[6]. In this study, two publicly available healthcare datasets were used, namely the Cleveland Heart Disease dataset representing Electronic Health Records (EHR) and the Human Vital Signs Dataset 2024 representing physiological measurements. Initially, both sets of data were checked for missing values, duplicates, data type consistency, and data integrity. Following this, the analysis experiment revealed that the chosen sets of data had neither missing nor duplicated values; hence, no further cleansing was required in the next stage.

Label encoding was used to represent all categorical variables from the physiological data set, such as Gender and Risk Category, in numbers. As algorithms in machine learning work on numbers, label encoding helped convert the categorical variables in the required format.

In order to get rid of unnecessary data, Patient ID and Time Stamp were stripped off the physiological data set because these two features are not useful for predicting any diseases. This step minimizes unnecessary computational work.

Normalization of the feature was then carried out using Min-Max Normalization method to convert all the numerical values to a value between 0 and 1. Normalization makes sure that features with wide numerical range do not dominate the learning process. The EHR dataset and physiological

vital-sign dataset were normalized independently to preserve the characteristics of each modality before multimodal integration [10].

Once the process of normalization was done, representative physiological data was randomly selected such that there was equal representation as in the case of EHR data sets; this led to the creation of two datasets of 303 patients each.

The pre-processing stage yields standardized, normalized, and machine learning ready datasets that will form the basis of multimodal feature fusion and disease prediction in the AIMDI-HR framework.

D. Feature Extraction

Feature extraction is performed to obtain meaningful health-care attributes that accurately represent a patient's clinical condition and can be effectively utilized by machine learning algorithms. Since healthcare data originate from different modalities, extracting relevant features is essential for im-proving predictive performance and enabling multimodal data integration [3], [10].

In the current implementation of the AIMDI-HR framework, feature extraction is performed on two healthcare modalities: Electronic Health Records (EHR) and physiological vital-sign data. The EHR dataset contains structured clinical attributes including age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels, thalassemia status, and the target variable indicating the presence or absence of heart disease. The physiological vital-sign dataset contributes complementary patient information, including heart rate, respiratory rate, body temperature, oxygen saturation, systolic blood pressure, diastolic blood pressure, age, gender, body weight, height, heart rate variability (HRV), pulse pressure, body mass index (BMI), mean arterial pressure (MAP), and risk category. These physiological measurements provide continuous indicators of a patient's physical situation and enhance the overall representation of patient health.

Prior to multimodal integration, all extracted features undergo categorical encoding, normalization, and feature selection to make sure numerical consistency and compatibility with machine learning algorithms. Irrelevant attributes such as Patient IDand Timestamp are excluded because they do not contribute to disease prediction. The resulting standardized feature vectors are subsequently used for multimodal feature-level fusion.

Although the AIMDI-HR framework is conceptually de-signed to support additional modalities such as medical images and electrocardiogram (ECG) signals, these modalities are not included in the current implementation because publicly available patient-aligned multimodal datasets containing all healthcare modalities are not readily available. Their inclusion within the framework provides a scalable architecture for future extensions involving deep learning-based feature extraction and multimodal healthcare analytics [2], [12].

E. Feature-Level Multimodal Data Fusion

Multimodal data fusion is the core component of the pro-posed AIMDI-HR framework because it combines complementary information from multiple healthcare data sources into a unified patient representation. Integrating heterogeneous healthcare modalities enables machine learning models to capture richer clinical relationships than can be achieved using a single data source alone, thereby improving predictive performance and supporting comprehensive patient profiling [1], [6], [13].

In the present study, feature-level multimodal fusion is employed to integrate Electronic Health Records (EHR) with physiological vital-sign data. After completing data preprocessing, categorical encoding, normalization, and feature se-lection, both datasets are transformed into standardized numerical feature vectors. In order to maintain dimensional consistency, a random sample of 303 physiological vital sign records is obtained to equalize the size of patient records in the EHR dataset.

The multimodal dataset is created by concatenating horizontally the normalized EHR feature

matrix and the normalized physiological vital sign feature matrix. The feature-level fusion technique ensures that the full set of clinical and physiological features from both modalities are retained while simultaneously ensuring a single representation for each patient. The multimodal dataset is comprised of 303 patients' records and 29 features, which includes the target variable for predicting heart disease. This fused dataset becomes the main input for the following machine learning models.

Some of the benefits of using feature-level fusion for structured healthcare datasets are: Finally, the unified feature representation simplifies model development while maintaining scalability for incorporating additional healthcare modalities in future implementations [8], [10].

Although more sophisticated multimodal fusion techniques such as attention-based learning, transformer architectures, and deep neural network fusion have recently been proposed, feature-level fusion was selected in this work because it is computationally efficient, interpretable, and well suited for structured tabular healthcare datasets. Furthermore, this approach provides a reliable baseline for evaluating the effective-ness of multimodal healthcare data integration while allowing future extensions to incorporate advanced deep learning-based fusion strategies [2], [14].

F. Explainable Artificial Intelligence

Artificial Intelligence models have created extraordinary predictive performance in healthcare; however, many machine learning algorithms which operates as black-box systems, which make it is tough to understand the logic behind their predictions. it difficult for clinicians to understand the reasoning behind their predictions. In healthcare applications, transparency and interpretability are essential because clinical decisions directly influence patient diagnosis, treatment planning, and healthcare management. Consequently, Explainable Artificial Intelligence (XAI) has become an important component of modern intelligent healthcare systems by improving trust, accountability, and model interpretability [13], [15].

In the proposed AIMDI-HR framework, feature importance analysis is performed using the best-performing Random Forest classifier to improve model interpretability. The Random Forest algorithm computes feature importance scores based on the work of each feature toward reducing impurity across the ensemble of decision trees. These importance scores provide a quantitative measure of the influence of individual clinical and physiological attributes on the prediction of heart disease.

Following model training, feature importance scores are took out from the trained Random Forest model and ranked in descending order. A plot for the feature importance is created to show the relative importance of the healthcare attributes which make up the highest importance. This way, the clinical and physiological variables that are the most influential for disease prediction are found, allowing more insight into the way the model works.

The feature importance analysis makes the AIMDI-HR framework more transparent by showing the healthcare attributes which play the most important role in predicting outcomes. This kind of explanation is useful in the clinical settings, where healthcare experts need proof of what influences the predictions made by the model. Furthermore, identifying the most important features supports future improvements in multi-modal healthcare systems and facilitates the integration of Artificial Intelligence into intelligent healthcare robotic applications for patient monitoring and clinical decision support [12], [14].

The proposed framework of AIMDI-HR can attain both accurate prediction performance and explainability through the combination of Explainable Artificial Intelligence (XAI) in the process of multimodal healthcare data analysis.

G. Holistic Patient Profiling

Holistic patient profiling is one of the most important purposes of the suggested AIMDI-HR model. Contrary to traditional health-care organizations that make use of a singular source of

patient data, the suggested model utilizes supplementary data from the clinical and physiological domain to build up a holistic representation of the patient's health status. This multimodal patient profile enables machine learning models to capture complex relationships among healthcare attributes and supports more accurate disease prediction and intelligent clinical decision-making [1], [3].

The present approach incorporates a comprehensive patient profile using a combined multimodal dataset of the EHRs and physiological vital sign characteristics. The multimodal dataset includes demographic data, clinical data, lab data, cardiovascular data, body composition data, and physiological data. In comparison to either mode alone, the use of the multimodal data set gives a wider insight into patient health.

The multimodal patient profile forms the key input data set for the Machine Learning (ML) algorithms used in this research work, such as Logistic Regression, Random Forest and XGBoost. The use of data from several healthcare modalities allows the learning algorithms to establish relationships among clinical and physiological attributes, which enhances the accuracy and robustness of heart disease prediction. The resulting multimodal patient profile also facilitates random forest feature selection, which helps healthcare experts to understand the key clinical and physiological attributes involved in predicting heart diseases.

Despite the fact that the current implementation relies on public datasets for EHR and physiological vital signs, it is designed in a way to include other health care modalities such as medical imaging, electrocardiograms (ECGs), wearables data, genomics, and clinical notes. The incorporation of these modalities in future work will enable more comprehensive patient profiling and further improve personalized healthcare services [11], [12].

The derived holistic patient profile represents an interpretable and scalable representation of patient health, enabling intelligent disease prediction, individualized treatment decisions, and future use in healthcare robots. Thus, the suggested framework serves as a foundation for next-generation AI-based healthcare systems.

H. Clinical Decision Support

Clinical Decision Support (CDS) system is a crucial element of the proposed AIMDI-HR architecture, in which Artificial Intelligence aids medical practitioners with intelligent predictions based on the combination of patient data. By using multimodal medical data, the framework helps in evidence-based decision making to achieve early disease diagnosis and treatment. AI-assisted clinical decision support has demonstrated considerable potential for improving diagnostic accuracy while reducing clinical workload and supporting patient-centered healthcare delivery [5], [16].

As per the proposed model, the patient's profile created using data fusion at feature level becomes the input for the Supervised Machine Learning (ML) models. The three common classification methods which have been used to predict the heart disease are logistic regression, random forest, and Extreme Gradient boosting (XGBoost). The combination of clinical and physiological features helps the model in recognizing the inter-relationship between the various healthcare features.

From the experimental results analysis, it is clear that all three machine learning models are effective in predicting heart disease classification. From the three, the Random Forest Classifier model gave the best results with an accuracy of 90.16%, precision of 89.29%, recall of 89.29%, f1-score of 89.29% and ROC-AUC of 0.9448. It gives the best results because of the use of ensemble learning which is capable of dealing with non-linearity of the multimodal healthcare data without overfitting.

In order to increase transparency and allow for interpretation within the clinical setting, a Random Forest Feature Importance Analysis is carried out after the creation of the model. The importance scores that emerge from this process help clinicians determine the features that have the most impact on the prediction of heart disease.

Although the current implementation is evaluated using publicly available datasets, the proposed clinical decision support framework provides a scalable foundation for future deployment in intelligent healthcare environments. The frame-work can be extended to support real-time hospital

information systems, wearable monitoring devices, telemedicine platforms, and healthcare robotic systems for continuous patient monitoring and personalized clinical assistance [4], [17].

I. Healthcare Robot Interaction

Healthcare robotics has emerged as an important application of Artificial Intelligence by assisting healthcare professionals in patient monitoring, rehabilitation, medication management, elderly care, and intelligent clinical support. The integration of multimodal Artificial Intelligence with robotic healthcare systems enables robots to perceive patient conditions more accurately, provide personalized assistance, and support timely clinical interventions. Recent advances have demonstrated that AI-enabled healthcare robots can improve healthcare efficiency, reduce clinician workload, and enhance the quality of patient-centered care [4], [18].

In the proposed AIMDI-HR framework, healthcare robot interaction represents the final stage of intelligent healthcare decision support. The multimodal patient profile generated from Electronic Health Records and physiological vital-sign data, together with machine learning predictions and Random Forest feature importance analysis, provides valuable information that can be utilized by future healthcare robotic systems. Instead of allowing robots to make autonomous medical decisions, the system will be able to utilize this information to help doctors through the continuous assessment of patients' condition, recognition of possible medical issues, and timely alerting of the professionals for further analysis.

The ability to provide explainable predictions of the proposed model is especially important for healthcare robotics because such predictions allow for understanding the reasons of patients' possible illnesses, both for doctors and robotic systems, which allows for safe cooperation between them.

While the present solution does not involve any physical robots, the proposed AIMDI-HR architecture is intended to be scalable enough to facilitate the future incorporation of health care robots, sensors, Internet of Things (IoT), and smart hospital settings. Future implementations may enable robotic systems to utilize real-time multimodal patient information for continuous health monitoring, personalized patient assistance, rehabilitation support, medication reminders, and intelligent clinical decision support [17], [18].

As such, the suggested theoretical framework provides a viable basis for leveraging multimodal Artificial Intelligence, Explainable AI, and healthcare robotics technologies in developing intelligent healthcare systems of the future that can deliver dependable and patient-centric clinical services.

III. Methodology

The suggested AIMDI-HR framework has been implemented via Python to create a multimodal healthcare prediction system using Electronic Health Records (EHR) and vital signs. The methodology utilized in this research involves patient data acquisition, preprocessing, multimodal data integration, machine learning model creation, feature importance analysis, and performance evaluation.

At first, two public healthcare datasets have been chosen for experimentation purposes. The Cleveland Heart Disease database is used as an Electronic Health Record (EHR), and the Human Vital Signs Dataset 2024 is considered as physiological patient data. Since both datasets originate from independent public sources, a representative subset of physiological records was randomly selected to match the number of EHR records, thereby enabling multimodal feature integration [9], [10].

The dataset was preprocessed after collecting them to make sure that they would be of good quality and compatible with the machine learning models. The data preprocessing steps included missing values check, duplicate checking, encoding categorical values using Label Encoding, eliminating attributes like patient ID and timestamp that are irrelevant, and normalization using Min-Max normalization. These pre-processing operations transformed heterogeneous healthcare in-formation into standardized numerical feature representations suitable for multimodal analysis

[6], [7].

Following preprocessing, feature-level multimodal fusion was performed by concatenating the normalized Electronic Health Record features with the normalized physiological vital-sign features. The resulting multimodal dataset consisted of 303 patient samples containing both clinical and physiological information. This integrated dataset served as the input for the predictive machine learning models [1].

Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) algorithms were build using supervised learning method for the prediction of heart disease. The dataset used for the experiment was divided into training and testing subsets in an 80:20 train test ratio while ensuring class balance using stratified sampling method. Each algorithm was trained on the training dataset and then tested on the testing subset.

Model performance was measured in terms of Accuracy, Precision, Recall, F1-score, Confusion Matrix, Classification Report, and ROC-AUC curve. The use of these performance measures is useful for comprehensively assessing the performance of classification models and comparing among various machine learning algorithms.

For enhancing the interpretability of the model, feature importance analysis was done on the best Random Forest model. The feature importance values were calculated for the trained classifier to find out the significance of different features in predicting heart disease. The resulting feature importance plot identifies the most influential healthcare attributes, enabling clinicians to better understand the factors that contribute to model predictions and supporting more informed clinical decision-making [13], [14].

The proposed methodology combines multimodal health-care data integration, supervised machine learning, and explainable Artificial Intelligence to establish an interpretable and scalable framework for intelligent healthcare decision support and future healthcare robotic applications.

IV. Experimental Design

The experimental evaluation of the proposed AIMDI-HR framework is being done to determine the efficacy of multimodal data integration in predicting heart diseases. The experimental evaluation is done through Python in which Electronic Health Records (EHR) along with physiological vital sign data are integrated to form a multimodal dataset. The experimental process involves the following processes: dataset collection, preprocessing of data, multimodal data integration, creation of machine learning model, performance evaluation, and Explainable Artificial Intelligence (XAI).

A. Experimental Environment

The proposed AIMDI-HR framework was implemented using Python 3.13 within the Jupyter Notebook environment. Data preprocessing, machine learning model development, and experimental evaluation were performed using several open-source Python libraries, including Pandas, NumPy, Scikit-learn, XGBoost, and Matplotlib. The experiments were con-ducted on a Windows 11 operating system. These software tools provided an efficient platform for multimodal health-care data preprocessing, predictive model development, performance evaluation, visualization, and Explainable Artificial Intelligence analysis.

B. Datasets

The proposed AIMDI-HR framework was evaluated using two publicly available healthcare datasets representing different patient data modalities. The first dataset is the Cleveland Heart Disease Dataset, which serves as the Electronic Health Record (EHR) source. It contains 303 patient records with 13 clinical attributes, including age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels, thalassemia, and the target variable indicating the presence or absence of heart

disease [9], [10].

The second dataset is the Human Vital Signs Dataset 2024, which contains physiological measurements such as heart rate, respiratory rate, body temperature, oxygen saturation, systolic blood pressure, diastolic blood pressure, body mass index (BMI), heart rate variability (HRV), pulse pressure, mean arterial pressure (MAP), and patient risk category. These physiological measurements provide continuous patient monitoring information that complements the structured clinical information available in the EHR dataset [1].

As both the data sets are derived from separate public domains and consist of varying number of patient records, therefore for the purpose of ensuring an equal number of patient records in both data sets, 303 physiological records were randomly selected which is similar to the number of patient records present in the EHR data set.

TABLE I

SUMMARY OF EXPERIMENTAL DATASETS

Dataset	Type	Samples	Features
Cleveland HD	EHR	303	14
Human Vital Signs	Vital Signs	303	16

A. Data Preprocessing

The collected datasets were preprocessed to improve data quality and ensure compatibility for multimodal integration.

Initially, both datasets were checked for missing values, duplicate records, and data consistency. The experimental analysis confirmed that is there any missing value or duplicate value were present in either dataset.

Categorical attributes within the physiological vital-sign dataset, including Gender and Risk Category, were trans-formed into numerical representations using Label Encoding. Non-informative attributes such as Patient ID and Timestamp were removed because they do not contribute to disease prediction and may introduce unnecessary computational complexity [6].

Both the EHR dataset and the physiological vital-sign dataset were normalized independently using the Min-Max normalization technique, which scales all numerical attributes to the range of 0 to 1. Feature normalization prevents variables with larger numerical values from dominating the learning process and improves the convergence and stability of machine learning algorithms [7].

After normalization, 303 physiological samples were selected randomly to be equal in number to the EHR samples. The above-mentioned data preprocessing approach yielded two balanced datasets which were then used for multimodal feature fusion.

C. Multimodal Data Integration

In this research, the proposed AIMDI-HR model has incorporated the use of EHR and physiological vital signs using feature-level multimodal fusion. Both datasets have been aligned after pre-processing and normalization by choosing an equal number of patient samples. The normalized EHR dataset with 303 patient samples was combined with the normalized physiological vital sign dataset with 303 patient samples.

The obtained multi-modal dataset contains 303 instances and 29 features, which include clinical factors, physiological parameters, and the target variable for predicting heart disease. This integrated feature representation enables the machine learning models to simultaneously utilize complementary information obtained from both healthcare modalities, thereby providing a more comprehensive understanding of patient health than single-modality approaches [8], [13].

D. Machine Learning Models

Three supervised learning techniques have been used to assess the performance of the proposed AIMDI-HR framework in predicting heart diseases. The logistic regression model was chosen as a classification algorithm due to its simplicity and extensive use in classifying medical conditions into two classes. The random forest algorithm was used as an ensemble technique that is able to model nonlinear relationships with reduced overfitting due to many decision trees. Extreme Gradient Boosting (XGBoost) was utilized because of its high predictive accuracy, computational efficiency, and robustness in healthcare prediction tasks [5].

The multimodal data was split into training and testing subsets using an 80/20 split. Stratified sampling was used to ensure that the proportions of heart disease classes are maintained in the split.

The training subset was used for the creation of the prediction models, whereas the test subset was used to evaluate the general capability of the models on unseen(new) data.

Apart from building the prediction models, Random Forest feature importance analysis was done to make the model more transparent in terms of contribution of individual features towards heart disease prediction. Feature importance scores were extracted from the best-performing Random Forest classifier to determine the relative influence of each healthcare attribute on the prediction outcome [13].

E. Performance Evaluation Metrics

The predictive performance of the created machine learning models was tested through different classification measures to make a complete assessment of the efficiency of the models. The accuracy metric was used to evaluate the percentage of accurately classified patient records. The Precision metric was used to find out the percentage of accurately predicted positive records among the total number of predicted positive records. The recall (sensitivity) measure showed how well the models were able to detect the patient having heart disease.

Moreover, Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) was used for testing the discriminative ability of each classifier using different classification criteria. Furthermore, confusion matrix and classification report of each machine learning algorithm were used for analyzing true positives, true negatives, false positives, and false negatives.

For enhancing the interpretability of the model, Random Forest feature importance was analyzed using the most effective classifier. The scores for feature importance were obtained using the Random Forest model to determine which attributes of the patients play an important role in heart disease prediction. The resulting feature importance plot provides a clear visualization of the relative importance of healthcare variables, thereby improving model transparency and enhancing clinician confidence in AI-assisted healthcare decision-making [14], [15].

V. Results and Discussion

Experimental evaluation of the AIMDI-HR framework involved the use of three supervised machine learning models: Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). The models were trained using the multimodal dataset that was created by fusing EHRs with physiological vital sign information using the multimodal feature level fusion method. Model performance was measured using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrix, classification report, and Random Forest feature importance measure. Table II presents an overview of the performance of the machine learning models employed.

TABLE II
Performance Comparison of Machine Learning Models

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Regression	83.61	78.12	89.29	83.33	0.9340

Random Forest	**90.16**	**89.29**	**89.29**	**89.29**	**0.9448**
XGBoost	88.52	86.21	89.29	87.72	0.9361

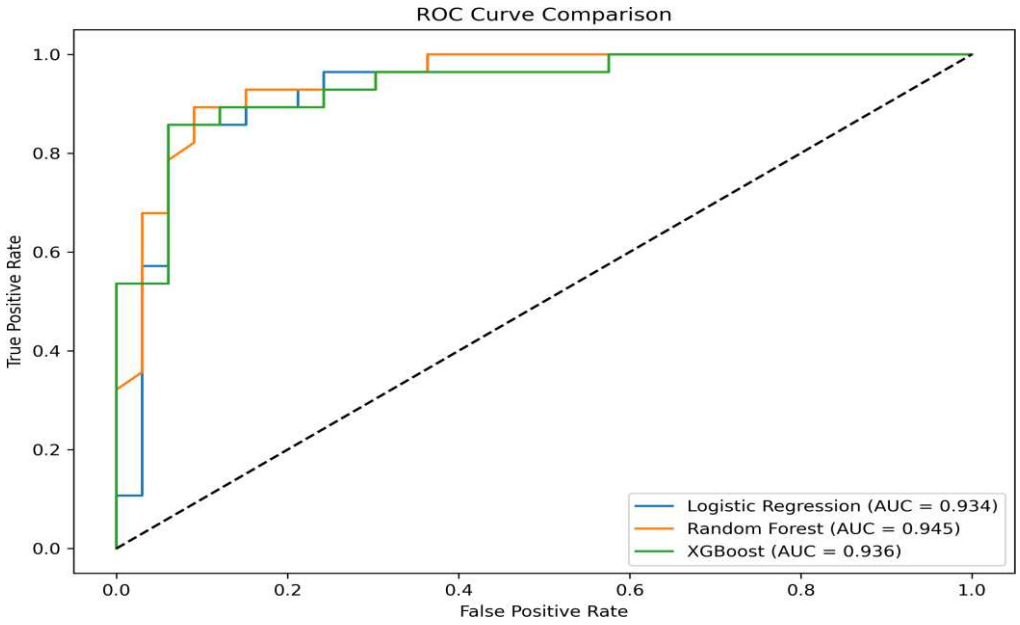


Fig. 2. ROC curve comparison of Logistic Regression, Random Forest, and XGBoost models for heart disease prediction using the proposed AIMDI-HR framework.

Figure 2 represents the ROC curve of the machine learning models assessed in this paper. The Random Forest classifier obtained the highest value of ROC-AUC of 0.9448, whereas the value for the XGBoost classifier was 0.9361, and the value for logistic regression was 0.9340. The high values of ROC-AUC reflect the good discriminating ability of the AIMDI-HR framework among the patients with and without heart diseases at varying thresholds of classification.

These experimental results show that the proposed AIMDI-HR approach is quite efficient in integrating multiple healthcare modalities for intelligent heart disease prediction. The integration of structured data from Electronic Health Records along with vital sign attributes allows machine learning algorithms to identify more clinical patterns than when using a single healthcare modality. These findings are consistent with recent studies reporting that multimodal healthcare data integration improves predictive performance and enhances intelligent clinical decision support systems [1], [8].

In terms of performance among all the classifiers, the Random Forest model performed best in terms of accuracy, which was 90.16%, precision of 89.29%, recall of 89.29%, F1-score of 89.29%, and ROC-AUC score of 0.9448. The excellent performance of the Random Forest algorithm is because of the ensemble learning approach, which models complex nonlinear relationships and avoids overfitting by combining several decision trees. These results indicate the effectiveness of the ensemble learning algorithms in multimodal healthcare prediction applications where different types of patient data are involved. Logistic Regression achieved 83.61% accuracy and served as the baseline classifier. Although Logistic Regression had a good recall of 89.29%, poor precision in terms of predicting the true positives led to a decrease in its performance. XG-Boost also showed competitive performance, which is characterized by 88.52% accuracy, 86.21% precision, 89.29% recall, 87.72% F1-score, and 0.9361 ROC-AUC score. Even though XGBoost had an excellent predictive performance, it is still inferior to the Random Forest classifier in most evaluation

metrics.

The ROC-AUC analysis is yet another means to assess the discriminating power of the developed framework. In the case of ROC-AUC, the Random Forest method produced the best result of 0.9448, followed by the XGBoost method with 0.9361 and the Logistic Regression method with 0.9340. High ROC-AUC results of all the three methods confirm the presence of very informative heart disease features in the multimodal dataset.

To be able to determine the significance of individual health care features, the analysis of feature importance was conducted using the best performing Random Forest classifier. Figure 3 shows the top ten features that were determined by the trained model.

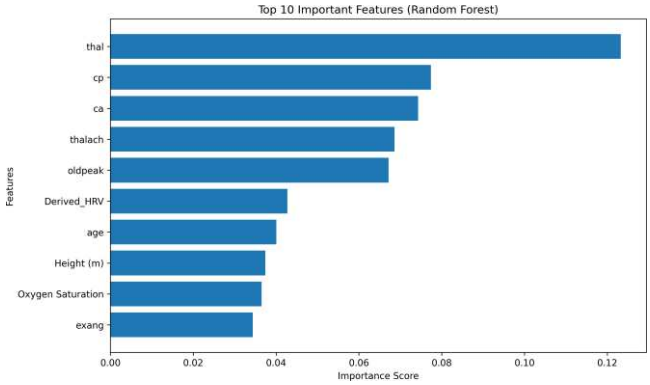


Fig. 3. Top 10 Important Features Identified by the Random Forest Classifier

From the feature importance analysis, it is clear that various clinical features like thal, type of chest pain (cp), major vessel (ca), maximum heart rate (thalach), old peak, and exercise-induced angina (exang) have played an important role in predicting heart diseases. In addition, physiological variables obtained from the vital-sign dataset, such as heart rate variability (Derived_HRV), oxygen saturation, pulse pressure, and mean arterial pressure, also exhibited notable im-portance. These observations demonstrate that combining heterogeneous clinical and physiological information enables the proposed AIMDI-HR framework to capture complementary patient characteristics, thereby improving predictive performance and supporting intelligent healthcare decision-making. Similar findings have been reported in recent studies on multimodal healthcare data fusion and predictive analytics [7], [13].

Overall, the experimental evaluation demonstrates that multimodal healthcare data integration combined with ensemble machine learning provides an effective approach for intelligent heart disease prediction. The proposed AIMDI-HR framework achieved high predictive performance while identifying the most influential healthcare features responsible for disease prediction. These results indicate that the framework has strong potential for future intelligent clinical decision support systems and healthcare robotic applications.

VI. Conclusion

This paper presented the AIMDI-HR (AI-Driven Multi-modal Data Integration for Holistic Patient Profiling in Health-care Robotics) framework, which integrates Electronic Health Records (EHR) and physiological vital-sign data to support intelligent healthcare decision-making through multimodal Artificial Intelligence. The proposed framework combines structured clinical information with physiological measurements using feature-level multimodal fusion, enabling the development of a comprehensive patient representation for heart disease prediction. In addition to predictive modeling, the framework incorporates Random Forest feature importance analysis to improve model transparency by identifying the clinical and physiological features that contribute most significantly to heart disease prediction and facilitate clinical interpretation.

The experimental evaluation was conducted using two publicly available healthcare datasets comprising Electronic Health Records and physiological vital-sign data. Three supervised machine learning algorithms, namely Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost), were implemented and compared. In terms of performance, among the aforementioned models, the Random Forest classifier produced the most optimal results with the following metrics: 90.16\% of accuracy, 89.29\% of precision, 89.29\% of recall, 89.29\% of F1-score, and 0.9448 of ROC-AUC score. In addition, the Random Forest feature importance analysis showed that both clinical and physiological features played significant roles in heart disease prediction, showing the effectiveness of using multiple types of healthcare data for providing an intelligent clinical decision support. While at the moment the implementation uses independent publicly available datasets that have been aggregated via a representative sampling approach, the suggested AIMDI-HR framework shows the potential of aggregating heterogeneous healthcare data in order to provide a comprehensive patient description. Moreover, due to its modular architecture, the suggested AIMDI-HR framework is easy to extend by adding such types of healthcare data as medical imaging, ECG, wearable sensor data, genomics, and clinical notes.

The proposed framework aligns with recent advances in multimodal Artificial Intelligence, explainable healthcare analytics, and intelligent clinical decision support systems, demonstrating a practical implementation of multimodal healthcare prediction using Electronic Health Records and physiological vital-sign data [1], [8]. Future work will focus on utilizing patient-aligned multimodal datasets, investigating advanced multimodal fusion techniques based on deep learning and transformer architectures, integrating privacy-preserving federated learning, and extending the framework to real-time healthcare robotic systems capable of continuous patient monitoring, personalized clinical assistance, and intelligent decision support [2], [13], [14].

DECLARATIONS

Data Availability

The datasets used in this study are publicly available. The Cleveland Heart Disease dataset and the Human Vital Signs dataset were obtained from publicly accessible sources and used for research purposes in accordance with their respective terms of use. The data were used for preprocessing, multimodal data integration, and machine learning analysis in this study. The processed data and code generated during the current study are available from the corresponding author on reasonable request.

Dataset links

Cleveland Heart Disease dataset: UCI Machine Learning Repository – Heart Disease dataset: https://archive.ics.uci.edu/dataset/45/heart+disease

Human Vital Signs Dataset 2024: Kaggle – Human Vital Sign Dataset: https://www.kaggle.com/datasets/nasirayub2/human-vital-sign-dataset/data

Ethical Approval and accordance

Not applicable. This study used secondary data from publicly available datasets and did not involve the recruitment or direct participation of human subjects.

Consent to participate

Not applicable. This study used publicly available secondary datasets and did not involve direct participation of human subjects.

Consent to publish

Not applicable. This study used publicly available secondary datasets and did not involve direct participation of human subjects.

Funding

No funding was received for this research.

Clinical Trial Number

Not applicable.

REFERENCES

[1] L. Soenksen, Y. Ma, C. Zeng, L. Boussioux, K. V. Carballo, L. Na, H. Wiberg, M. Li, I. Fuentes, and D. Bertsimas, "Integrated multimodal artificial intelligence framework for healthcare applications," *NPJ Dig-ital Medicine*, vol. 5, 2022.

[2] D. Schouten, G. Nicoletti, B. Dille, C. Chia, P. Vendittelli, M. Schu-urmans, G. Litjens, and N. Khalili, "Navigating the landscape of multimodal ai in medicine: A scoping review on technical challenges and clinical applications," *Medical Image Analysis*, vol. 105, p. 103621, 2025.

[3] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, "Multimodal biomedical ai," *Nature Medicine*, vol. 28, pp. 1773–1784, 2022.

[4] N. Bout, G. Moukhliss, H. Belhadaoui, N. Afifi, and M. Abik, "Inte-grating emotional ai, iot, and robotics for patient-centered healthcare: Challenges and future directions," *Discover Internet of Things*, vol. 5, no. 1, p. 75, 2025.

[5] R. Lin, "Ai-driven personalized healthcare: Leveraging multimodal data for precision medicine," in *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2024, pp. 693–697.

[6] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Vela´squez, "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom," *Information Fusion*, vol. 102, p. 102040, 2023.

[7] X. Chen, H. Xie, X. Tao, F. Wang, M. Leng, and B. Lei, "Artificial intelligence and multimodal data fusion for smart healthcare: Topic modeling and bibliometrics," *Artificial Intelligence Review*, vol. 57, 2024.

[8] J. R. Teoh, J. Dong, X. Zuo, K. W. Lai, K. Hasikin, and X. Wu, "Advanc-ing healthcare through multimodal data fusion: A comprehensive review of techniques and applications," *PeerJ Computer Science*, vol. 10, 2024.

[9] F. Mohsen, H. Ali, N. Hajj, and Z. Shah, "Artificial intelligence-based methods for fusion of electronic health records and imaging data," *Scientific Reports*, vol. 12, 2022.

[10] L. Tong, W. Shi, M. Isgut, Y. Zhong, P. Lais, L. Gloster, J. Sun, A. Swain, F. O. Giuste, and M. D. Wang, "Integrating multi-omics data with ehr for precision medicine using advanced artificial intelligence," *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 80–97, 2023.

[11] S. N. Khan, Danishuddin, M. W. A. Khan, L. Guarnera, and S. F. Akhtar, "Multi-modal ai in precision medicine: Integrating genomics, imaging, and ehr data for clinical insights," *Frontiers in Artificial Intelligence*, vol. 8, 2026.

[12] N. Parvin, S. Joo, J. H. Jung, and T. K. Mandal, "Multimodal ai in biomedicine: Pioneering the future of biomaterials, diagnostics, and personalized healthcare," *Nanomaterials*, vol. 15, 2025.

[13] M. A. Archana, A. Jarina Raihan, B. Jothilakshmi, C. Parthasarathy, T. Kumar, and T. Lakshmibai, "Multimodal fusion for ai-driven health-care: Integrating genomic, ehr, and wearable sensor data in a federated 6g framework with explainable ai," in *Proceedings of the 4th Interna-tional Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2025, pp. 1894–1900.

[14] J. Liu, X. Cen, C. Yi, F. Wang, J. Ding, J. Cheng, Q. Wu, B. Gai, Y. Zhou, R. He, F. Gao, and Y. Li, "Challenges in ai-driven biomedical multimodal data fusion and analysis," *Genomics, Proteomics & Bioin-formatics*, vol. 23, 2025.

[15] R. Zhang, Y. Chen, W. Yue, Y. Zhang, X. Li, S. Feng, F. Yuan, and M. Luo, "Multimodal artificial intelligence in medicine: A task-oriented framework for clinical translation," *Frontiers in Medicine*, vol. 12, 2026.

[16] F. Yeasmin, S. Arefeen, R. Zakaria, and A. Hasan, "An integrative artificial intelligence framework for the diagnosis of multiple diseases in clinical settings," *Journal of Computer Science and Technology Studies*, 2025.

[17] D. Patel, "Ai-driven automation of patient data integration in complex healthcare systems," *European Journal of Biology and Medical Science Research*, 2025.

[18] M. A. Rahman, S. U. Jamil, M. S. Hossain, M. A. Khan, T. Zia, M. Paracha, M. Alrashoud, M. Chen, and S. A. F. Al-Hazzaa, "An llm-enabled multimodal agentic ai framework for the medical internet of things (miot)," *ACM Transactions on Multimedia Computing, Commu-nications and Applications*, 2026.