International Journal of Engineering Research and Sustainable Technologies
Volume 4, No.2, June 2026, P 24 – 29
ISSN: 2584-1394 (Online version)

CARDIO TWIN-H: AN AI-INTEGRATED REAL TIME CARDIOVASCULAR DIGITAL TWIN SYSTEM FOR PROACTIVE RISK REDICTION AND CLINICAL DECISION SUPPORT

[1]Murshid R, [2]P.C.Akhila, [3]V.Rameshbabu
[1] UG Student , [2]Assistant Professor, [3]Professor,
[1,2,3]Department of DS, Dr.MGR Educational and Research Institute, India
*** Corresponding author email address**: akhila.cse@drmgrdu.ac.in

DOI: https://doi.org/10.63458/ijerst.v4i2.157

Cardiovascular diseases are still the topmost killers of people all around the world, claiming approximately 17.9 million lives each year. Traditional tools to assess the risk of developing cardiovascular diseases, including the Framingham Risk Score, do not incorporate interactivity, explainability, and real-time simulative capabilities that would enable doctors to provide proactive and personalized care. In this work, we present CardioTwin-H, an AI-Integrated Real-Time Cardiovascular Digital Twin System with Hardware Sensor Integration. CardioTwin-H is a combination of a MAX30102 pulse oximetry and heart rate sensor connected to a Raspberry Pi 4 microcomputer and a machine learning-based platform. Data are streamed to a FastAPI backend server, where an ensemble of Random Forest, Gradient Boosting, and Logistic Regression algorithms calculates the cardiovascular risk score in real time. Reports based on the SHapley Additive Explanations framework are used to interpret data and provide clinically meaningful explanations. Finally, a scenario simulation engine allows for projecting the influence of clinical intervention on the predicted health state of patients. Overall, the system can be deployed at hardware costs of less than six thousand Indian Rupees.

Keywords: artificial Digital Twin, Cardiovascular Risk Prediction, Machine Learning, SHAP Explainability, Raspberry Pi, MAX30102, Healthcare AI, Scenario Simulation, SDG 3.

1. Introduction

CVD causes an estimated 17.9 million deaths each year, as cited by the World Health Organization. In India, more than 28 percent of fatalities are because of CVDs, wherein CVDs predominantly affect people aged 25 to 64 years old [1]. Most CVD-related deaths can be avoided with timely diagnosis and continuous monitoring of the risk factors; however, primary healthcare facilities, especially in developing countries, do not have access to such technologies. Modern-day clinical risk assessment systems such as the Framingham Risk Score and QRISK3 do not change. They involve data input manually at each clinic visit, display one risk number without specifying the factors that make up the score, and lack the capability of allowing a physician to virtually evaluate the effects of treatment options before putting them into action. Wearable consumer devices have the ability to track a person's heartbeat and oxygen levels in real time but cannot be used to determine clinical risks.. However, the development of the digital twin, which is defined as "the virtual representation of the physical world that reflects its dynamics" and could be applied to simulate different possible situations, has provided healthcare with an innovative solution [3]. With the aid of inexpensive sensors integrated with machine learning algorithms, the application of the digital twin can help develop efficient systems for managing patients' cardiovascular risks in real time.

This paper proposes CardioTwin-H, an innovative technology that brings together five distinct functionalities in one low-cost medical device: physiological sensing in real-time using the MAX30102 sensor; cardiovascular risk prediction through artificial intelligence-enabled ensembles; SHapley Additive Explanation (SHAP)-based explainability; scenario-based simulations; and a doctor-oriented web dashboard. As far as the authors know, there is currently no product available on the market or academic prototype research in literature which incorporates all five features in one deployable device at similar hardware costs as CardioTwin-H.

2. Related Work

The use of digital twins for the heart and artificial intelligence to predict risks has been investigated separately by many researchers. Kumar et al. [1] showed that it was possible to build a digital twin using smartwatch data that could detect abnormalities in real time. Although they have proved the feasibility of creating heart models using devices, they fail to incorporate risk predictions and simulations into their model. CardioTwin-XAI was proposed by Sharma et al. [2] to predict coronary artery disease risk through an innovative approach that combines IoT wearables, federated learning, and explainable artificial intelligence. This study highlighted the practical application of SHAP explainability in cardiac risk assessment. Nevertheless, due to its reliance on IoT wearable devices and federated learning models, the CardioTwin-XAI cannot be implemented in typical primary care centers in developing nations due to the lack of infrastructure.

The Digital Twin-Based Healthcare System was suggested by Patel et al. [3], wherein machine learning was used to predict diseases among remote patients based on their data. The concept created by these researchers proved the usefulness of integrating digital twins and ML-based risk prediction technologies. At the same time, this concept lacks a few features like its specialization in the cardiovascular field, an explainable nature, simulation opportunities, and the interface with physicians.

An examination of existing literature shows a clear discrepancy. There exists currently no existing system that incorporates all the features of real-time physiological sensors, cardiovascular risk prediction using AI specific to the domain, SHAP explainability, and scenario simulation within an affordable and doctor-friendly device. CardioTwin-H fills this gap completely

3. System Models

CardioTwin-H employs a three-tier client-server architecture with an additional hardware sensing layer. These four layers are: the Hardware Sensing Layer, the Presentation Layer, the Application Intelligence Layer, and the Data Persistence Layer. These layers interact with each other using REST API interface calls.

3.1. Hardware Sensing Layer

The Hardware Sensing Layer includes a MAX30102, which is an optical pulse oximeter and heart rate sensor, linked to a Raspberry Pi 4 Model B using I2C interface. This sensor works on the principle of Photoplethysmography (PPG) that involves shining the patient's finger tip with Red (660nm) and Infrared (880nm) LEDs and observing the changes in light absorption due to pulsations in the blood vessels.

Raw PPG data is constantly monitored from the MAX30102 FIFO buffer by the Raspberry Pi firmware at a rate of 100 Hz, filtered through a 4th-order Butterworth bandpass filter (0.5 – 4.0 Hz) to obtain cardiac waveforms, peak detection performed using scipy.signal.find_peaks method to calculate heart rate, ratio-of-ratios technique applied to determine SpO2, perfusion index analysis for assessing signal quality, and validated results communicated via HTTP POST at every second interval to the FastAPI server backend. Signal quality and live readings are displayed on a 0.96-inch SSD1306 OLED screen interfaced to the Raspberry Pi device.

3.2 Application Intelligence Layer

The Application Intelligence Layer refers to a FastAPI-based web service that hosts 13 REST APIs distributed across six functionalities: Authentication, Patient Management, Prediction, SHAP Explainability, Simulation Engine, and Report Generation. As such, the Prediction functionality loads pre-trained machine learning model artifacts on initialization and utilizes them on the combined vector of features based on live sensor measurements and clinical parameters input by the physician to compute cardiovascular risk scores. On the other hand, the SHAP explainability functionality relies on TreeExplainer in the case of ensembles and LinearExplainer in case of Logistic Regression to calculate the weighted ensemble SHAP scores for every prediction. Finally, the Simulation Engine receives updated clinical parameters input by the physician, merges them with the original feature vector, and computes predicted risk scores.*C. Presentation Layer*

Presentation layer is a React.js single page application consisting of the following six main pages: Login, Patient Directory, Patient Profile, Consultation view, Simulation panel, and Reports. Among these six main pages, the most important one is the consultation view, which shows live sensor readings in real-time through periodic polling, gives a clinical data entry form with sensor readings already populated, and provides risk scoring, risk categorization, and a SHAP horizontal bar graph showing features' importance.

3.3 Data Persistence Layer

Data Persistence Layer employs SQLite for prototyping purposes, with architectural flexibility to migrate to PostgreSQL in future production releases. There are seven tables used to maintain all system data: doctors, patients, clinical data, sensor readings, risk results, simulations, and reports.

4. Methodology

4.1 Dataset and Feature Set

The machine learning models are trained on a combined data set from the UCI Heart Disease Data Set (Cleveland subset, 303 samples) and the Framingham Heart Study Data Set (4,240 samples). After applying preprocessing techniques such as missing value imputation through median imputation method, standardization through StandardScaler method, encoding of categorical variables, and class balancing through SMOTE method, the final data set contains 17 attributes for training the models.

Table 1: Selected Feature Set for Cardiovascular Risk Prediction

No.	Feature	Source
1	Age (years)	Clinical input
2	Resting Blood Pressure (mmHg)	Clinical input
3	Serum Cholesterol (mg/dL)	Clinical input
4	Fasting Blood Sugar (mg/dL)	Clinical input
5	Maximum Heart Rate Achieved (BPM)	MAX30102 Sensor (live)
6	Blood Oxygen Saturation — SpO2 (%)	MAX30102 Sensor (live)
7	Body Mass Index (kg/m²)	Clinical input
8	Smoking Status (0=Never, 1=Former, 2=Current)	Clinical input
9	Physical Activity Level (0–3 scale)	Clinical input
10	Family History of CVD (binary)	Clinical input

Table 1 presents the ten primary features used in this study. The complete feature set of 17 features includes additional ECG-derived parameters from the UCI dataset. Features 5 and 6 are unique to CardioTwin-H — they are sourced directly from the MAX30102 hardware sensor in real time during each patient consultation, replacing manually entered values.

4.2 Machine Learning Models

A soft voting approach is utilized to merge the output of three machine learning models that are independently trained. A random forest classifier (n_estimators=200, max_depth=10, class_weight=balanced) is chosen because it is less affected by outliers and gives feature importance that is easy to interpret. An XGBoost gradient boosting classifier (n_estimators=300, learning_rate=0.05, max_depth=6) is chosen for its superior performance on tabular medical datasets. A logistic regression classifier (C=1.0, solver=lbfgs, class_weight=balanced) is chosen for providing calibrated probabilities of occurrence. The soft voting technique takes into consideration individual probabilities of each model using 0.35, 0.45, and 0.20 weights for Random Forest, XGBoost, and Logistic regression respectively.

Model validation is carried out using the stratified five-fold cross-validation technique on the training data set to get accurate predictions of its performance, after which a test data set of 20% is used for evaluating its performance. The decision threshold is tuned using the Youden Index method, with an emphasis on recall, since this is preferred in medical applications due to their sensitivity to false negatives.

4.3 SHAP Explainability

The SHAP Explainer Module makes use of TreeExplainer in the case of Random Forest and XGBoostalgorithms and LinearExplainer in the context of the Logistic Regression model, where LinearExplainer uses a background dataset containing 100 random samples from the training data. The calculation of weighted SHAP values of the ensemble algorithm is done as follows: SHAP_ensemble = (0.35 × SHAP_RF) + (0.45 × SHAP_XGB) + (0.20 × SHAP_LR). The top five features in terms of SHAP values will be provided alongside clinical explanations such as: 'High resting heart rate (98 BPM) increases the risk to the heart significantly.

4.4 Scenario Simulation Engine

The Simulation Engine accepts a modified parameter set from the physician — expressed as absolute values —

validates each parameter against physiologically plausible ranges, reconstructs the complete feature vector by merging modified values with unchanged original parameters, and reruns the full prediction pipeline. The engine returns the original risk score, the simulated risk score, the delta, and a risk category transition indicator — for example, 'Medium Risk → Low Risk' — enabling the physician to evaluate the projected clinical impact of a specific intervention before implementing it.

5. Hardware Design

5.1 Component Specifications

MAX30102 (Maxim Integrated) is the name of the sensor that offers an integrated solution for photoplethysmography featuring red and infrared LEDs, photodetector, and analog signal processing system. This device communicates over the I2C bus at address 0x57, works on a 3.3V supply, and can offer SpO2 readings with an accuracy of ±2% for 70-100% saturation level in a static state. The Raspberry Pi 4 Model B with BCM2711 processor featuring four cores of ARM Cortex-A72 processor running at 1.5 GHz and 4GB LPDDR4 RAM is used to run Rasbian Pi OS and native Python 3.11.

5.2 Circuit Design

Both the MAX30102 sensor breakout board and the SSD1306 OLED display use the I2C communication protocol through the GPIO Pin 3 (SDA) and GPIO Pin 5 (SCL) pins of the Raspberry Pi, each at their assigned addresses of 0x57 and 0x3C, respectively. The 3.3V voltage rail is utilized to power both devices using GPIO Pin 1. Both components have been designed by manufacturers with all necessary capacitors and pull-up resistors already mounted onto the integrated circuits. The entire circuit will be constructed using a solderless breadboard prototype and a 3D-printed box casing for manufacturing purposes.

5.3 Signal Processing Pipeline

The firmware-based signal processing flow converts raw data from 18 bits ADC readings in MAX30102 FIFO into clinically useful physiological parameters in 10 steps: FIFO read at 100 Hz, ADC decoding, DC offset subtraction by 200-points moving average, fourth-order Butterworth filter of 0.5 to 4.0 Hz range using scipy.signal.sosfilt, detection of R-peak using minimum inter-peak distance of 30 points, heart rate calculation from mean of inter-peak time difference of last 10 peaks, calculation of SpO2 using ratio-of-ratios method (SpO2=110 – 25R), PI calculation using formula (PI=AC_IR/DC_IR*100%), signal quality gating using threshold of PI<0.05, and REST API data transfer at intervals of 1 second for accepted signals.

6. Result and Discussions

6.1 Machine Learning Performance

Ensemble is created, trained, and validated on the combined UCI-Framingham dataset. The Table 2 below shows cross-validated results of each model, as well as the performance of the final ensemble model on the unseen data set. Soft Voting Ensemble model scores the maximum accuracy at 89.2% and ROC-AUC of 0.94 compared to the rest of the individual models. This high recall rate of 91.4% of the positive samples indicates the use of Youden Index for threshold selection to minimize risk misclassification of cases.

Table 2: Machine Learning Model Performance on Holdout Test Set

Model	Accuracy (%)	Recall (%)	ROC-AUC
Random Forest	86.4	87.2	0.91
XGBoost (Gradient Boosting)	88.1	89.6	0.93
Logistic Regression	81.7	83.1	0.88
Soft Voting Ensemble (Proposed)	89.2	91.4	0.94

Table 2 demonstrates that the soft voting ensemble consistently outperforms individual models across all evaluation metrics. The ensemble's higher recall compared to individual models is particularly significant in the clinical context — it ensures that a greater proportion of truly high-risk patients are correctly identified, supporting the primary clinical objective of early, proactive cardiovascular risk detection.

6.2 Hardware Validation

Hardware verification is performed through comparison of CardioTwin-H firmware's heart rate and SpO2 results against a verified reference pulse oximeter at 15 test points (3 patients × 5 tests each). The heart rate values are within ±10 BPM from the reference at 13 out of 15 test points (86.7%) – exceeding the set goal of 80% accuracy. SpO2 results are within ±3% from the reference at 14 out of 15 test points (93.3%). All readings rejected due to Perfusion Index below 0.05 are gated correctly, and firmware runs stably for four consecutive hours.

6.3 System Performance

System performance tests conducted from end to end prove that the average time taken from the POST /api/predict request to the display of risk scores on the dashboard is 1.84 seconds (sample size of 10), which satisfies the non-functional requirements set at three seconds. The latency between the Raspberry Pi device and the backend server using local Wi-Fi is an average of 320 milliseconds, which satisfies the five-hundred-millisecond requirement. SHAP value calculations take an average of 441 milliseconds per prediction.

6.4 SHAP Explainability Analysis

Analysis using SHAP for the entire test dataset identifies the five most important risk factors for cardiovascular disease, which are, in descending order of mean absolute SHAP value, maximum heart rate attained (thalach), cholesterol levels, resting blood pressure, age, and the number of major arteries colored by fluoroscopy (ca). The real-time heart rate measurement from the live sensor always comes out as the single most important predictor, thus proving the importance of including the MAX30102 hardware sensor in the prediction algorithm, instead of manually inputting the heart rate values.

6.5Comparative Analysis

Table 3 presents a comparative analysis of CardioTwin-H against existing cardiovascular risk tools and the three reviewed IEEE papers along five key capability dimensions.

Table 3: Comparative Analysis of CardioTwin-H Against Existing Systems

System	Real-Time Sensing	AI Risk Prediction	SHAP Explainability	Scenario Simulation	Cost
Framingham Risk Score	No	No	No	No	Free
Consumer Wearables	Yes	No	No	No	High
Kumar et al. [1] — 2024	Yes	No	No	No	N/A
Sharma et al. [2] — 2025	Yes	Yes	Yes	No	N/A
Patel et al. [3] — 2023	Partial	No	No	No	N/A
CardioTwin-H (Proposed)	Yes	Yes	Yes	Yes	<₹6K

Table 3 confirms that CardioTwin-H is the only system in the comparison that delivers all five capabilities simultaneously within an affordable hardware cost. This comprehensive capability profile, combined with a hardware cost accessible to primary care settings in India, distinguishes CardioTwin-H as a novel contribution to the field of affordable healthcare AI.

7. Conclusion

In this paper, we introduce CardioTwin-H, an AI-Integrated Real-Time Cardiovascular Digital Twin System with hardware-sensor integration intended to be used in primary healthcare settings. This system features five abilities all built into a single low-cost device: real-time physiological monitoring using the MAX30102 sensor connected to Raspberry Pi 4, ensemble machine learning risk prediction of cardiovascular disease with an accuracy of 89.2%

and ROC-AUC of 0.94, SHAP-based model interpretability with layman clinical interpretations, scenario simulation with which intervention strategies can be formulated prior to clinical action, and a clinical interface for doctors via React.js web application.

Validation of hardware demonstrates that heart rate is accurate to within ±10 BPM in 86.7% of cases, and SpO2 is accurate to within ±3% in 93.3% of cases compared with the reference system. System testing validates the overall prediction latency time at 1.84 seconds, comfortably meeting the 3-second criterion for clinical utility. The SHAP study validates the importance of live sensor heart rate as the most important cardiovascular risk predictor.

The cardio twin-H software solution contributes directly to SDG number three – Good health and well-being – by facilitating proactive and personalized risk assessment for the prevention of cardiovascular disease in environments where the use of any commercial cardiovascular monitoring equipment is unaffordable. Further development of this technology will incorporate features such as multidisease risk prediction involving metabolic and renal systems, integration of wearable devices for continuous non-clinical assessment, and conducting clinical trials through the cooperation of medical professionals. Provisional patenting will be considered after successful prototype validation.

References

1. Kumar, A., et al., "Enhancing Cardiac Health: Digital Twin Technology for Real-Time Monitoring and Abnormality Detection of Human Heart," IEEE Conference Publication, 2024. DOI: 10.1109/10674146.2024
2. Sharma, R., et al., "CardioTwin-XAI: A Consumer-Centric Digital Twin Framework for Predictive Risk Stratification and Personalized Management of Coronary Artery Disease in Healthcare 5.0," IEEE Journals and Magazine, 2025. DOI: 10.1109/11370951. 2025
3. Patel, S., et al., "Digital Twin-Based Healthcare System (DTHS) for Earlier Disease Identification and Diagnosis Using Machine Learning," IEEE Journals and Magazine, 2023. DOI: 10.1109/10239395. 2023
4. World Health Organization, "Cardiovascular Diseases (CVDs) — Key Facts," WHO Global Health Observatory, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). 2023
5. Ministry of Health and Family Welfare, Government of India, "India: Health of the Nation's States," Public Health Foundation of India, New Delhi, 2017.
6. Lundberg, S. M., and Lee, S.-I., "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
7. D'Agostino, R. B., et al., "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study," Circulation, vol. 117, no. 6, pp. 743–753, 2008.
8. Dua, D., and Graff, C., "UCI Machine Learning Repository — Heart Disease Dataset," University of California, Irvine, School of Information and Computer Sciences, 2019. [Online]. Available: https://archive.ics.uci.edu/dataset/45/heart+disease. 2019.