

A Multimodal Deep Learning Framework for Early Detection of Congenital Heart Disease in Neonates

Iis Hamsir Ayub Wahab¹, Sri Yati²

¹Department of Electrical Engineering, Faculty of Engineering, Universitas Khairun
email : hamsir@unkhair.ac.id

² Department of Medicine, Faculty of Medicine and Health Sciences, Universitas Khairun
Address: Kampus 2 Universitas Khairun Jl. Jusuf Abdulrahman, Kel. Gambesi, Kec.Ternate Selatan, Kota Ternate
email : sriyati@unkhair.ac.id

Article history : Received Oct 21, 2024 | Revised Feb 11, 2025 | Accepted Mei 17, 2025

Abstrak -- Congenital heart disease (CHD) is the most common congenital defect and still adds significantly to the neonatal morbidity and mortality rates. Classic echocardiography and ECG unimodal data traditional methods are often unable to analyze complex, multifunctional, and multifactorial cardiac pathologies in neonates. This paper presents an explainable multimodal deep learning framework that acquires four diverse sources of clinical data. Multimodal data includes echocardiogram videos, ECG, and other physiological and structured electronics health record (HER) data. We propose a self-attention-based late fusion transformer architecture that also uses self-attention mechanisms. The model trains and validates on benchmark datasets, which are transparently and reproducibly available (EchoNet-Dynamic, MIMIC-IV, PhysioNet Capnobase, and MIT-BIH). The results achieved using the proposed model mark an improvement over existing benchmarks with 93% accuracy, 95% sensitivity, and 0.96 area under the ROC curve. Using interpretability modules, features that were value added towards determining the diagnostic indicators that were incorporated in the neonatal infant care were shown to be critically relevant. Moreover, the model shows high performance consistency across several data sources and shifts. The research illustrates the use of explainable deep learning architectures for automation of early-stage heart defect detection in newborns. Some of the future work includes validation through clinical studies and multilingual electronic health record integration.

Keywords: CHD, Multimodal, Data, Fusion, AI, Deep Learning



This is an open access article under the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

1. INTRODUCTION

Congenital Heart Disease (CHD) is a morphological abnormality of the heart that occurs since the embryonic period and is the main cause of neonatal death due to congenital abnormalities. With an estimated global prevalence of around 8 per 1,000 live births, and the number of new cases reaching more than 1.35 million babies per year globally, CHD continues to be a major challenge in the health care system, especially in developing countries [1]. Early detection and diagnosis of CHD is key to reducing morbidity and mortality rates, because most forms of CHD require medical intervention within a very short period of time after birth [2].

In current clinical practice, the diagnosis of CHD relies heavily on manual evaluation of various types of data, such as echocardiography results, electrocardiogram (ECG) signals, and patient clinical and laboratory data. Unfortunately, this approach not only relies on the subjectivity of clinician interpretation but is also partial because it often only relies on one type of data (unimodal). In fact, congenital heart disease is multifactorial, involving structural, functional, and molecular aspects that interact with each other [3], [4]. The limitations of partial diagnostic approaches can lead to longer diagnosis times, misclassification of cardiac abnormalities, or even failure to detect the disease.

Over the past ten years, rapid developments in the field of artificial intelligence (AI), especially deep learning techniques, have brought new hope in detecting and classifying congenital heart disease. AI shows great potential in improving diagnostic capabilities, both in reading electrocardiogram (ECG) results, analyzing echocardiography images, and processing physiological signals and electronic health records (EHR) data [5],[6]. The combination of various types of data through a multimodal data fusion approach is considered capable of understanding the physiological and morphological complexity of the heart in more depth than using only one type of data. For example, combining ECG and echocardiogram data has been shown to increase diagnostic accuracy by 12–15% compared to methods that use only one type of data [6], [7].

Information from an electrocardiogram (ECG), which depicts the electrical activity of the heart, has long been the primary way to determine if there are any cardiac arrhythmias or other malformations [8]. However, ECGs have the disadvantage of being less able to detect complex malformations, making them less suitable for use alone. Echocardiography, on the other hand, can provide a direct picture of the heart, but its results are highly dependent on the expertise of the examiner and the availability of the equipment. Vital signs such as heart rate,

blood oxygen levels, and blood pressure can provide important clues about a patient's circulatory status, which often changes in CHD. Finally, EHR contain patient data over time that is very helpful in making decisions based on medical history [9]. Combining these four methods with an artificial intelligence-based data fusion system could produce more complete and adaptable estimates.

Several recent studies have shown that this method is quite effective. In a study conducted by Khan et al. [10], a multimodal deep learning system showed an accuracy rate of 94.2% in distinguishing simple and complex CHD. However, various technological innovations such as the use of Transformer architecture, attention-based fusion, and federated learning have begun to address these challenges by offering systems that are more manageable, scalable, and secure regarding data privacy [11]. The use of multimodal AI has great potential in diagnosing in newborns. For example, the CNN-LSTM model equipped with Grad-CAM has achieved a sensitivity level above 92% [12]. In fact, the combination of physiological signal data and EHR can increase the accuracy of predicting heart structural abnormalities [13], [14]. In patients with Tetralogy of Fallot, the multimodal sequential model has proven effective in predicting complications with an accuracy of up to 93% [15]. Meanwhile, the attention-fusion architecture showed better results than conventional methods with an F1 score of 91.7% [16]. However, challenges such as interoperability and data synchronization between modalities are still obstacles in improving system performance [17]. In addition, ethical issues and potential bias also need to be addressed systematically. A systematic review concluded that an average AUC value of 0.91 can only be obtained if the quality of labeling and standardization of clinical data is truly maintained [18]. This confirms that the success of a multimodal model depends not only on its technical performance, but also on its readiness for implementation in everyday clinical practice.

This study aims to design and test in depth an artificial intelligence (AI) model that combines various types of data or multimodal. This model utilizes four key data sources, namely ECG results, echocardiography, physiological signals, and EHR, to detect CHD in newborns. This study not only prioritizes how accurate the model's predictions are, but also how easy the model is to understand, how relevant the model is to the medical world, and how ready the system is to be used in the NICU (Neonatal Intensive Care Unit). With this approach, it is hoped that the resulting system will be able to make a significant contribution to reducing morbidity and mortality in newborns due to late diagnosis of CHD.

2. METHOD

2.1. System Design and Multimodal Data Acquisition

In this study, we combined different types of clinical data with the help of artificial intelligence (AI). There are four main data sources that we used, namely: cardiac recordings (ECG), cardiac images from ultrasound (echocardiography), live body condition data (such as heart rate, breathing rate, and oxygen levels), and patient health records from hospitals (EHR). All of these data were taken from open databases that are trusted and often used by other researchers around the world for AI research and development purposes.

In more detail, we obtained ECG data and body condition data from PhysioNet, including datasets such as the PTB Diagnostic ECG Database, MIMIC-III Waveform Database, and VitalDB. These databases provide ECG data and body condition parameters in great detail [19]. For cardiac images, we used the CAMUS dataset and EchoNet-Dynamic. Both are common databases that contain cardiac ultrasound images and videos that have been examined and annotated by cardiologists [20], [21]. Meanwhile, we took our patient health record data from the MIMIC-IV Clinical Database, which has a complete record of the health history of pediatric patients treated in intensive care units (NICU and PICU) [22].

We arranged all of this data so that the time was right, so that it could be processed simultaneously. Before processing, we cleaned this data first. To do this, we adjusted the signal data to the same scale (z-score), resized the heart image to 224x224 pixels, and converted the health record text into a computer-readable format using the BioBERT tokenizer, then converted it again into a tensor embedding. By using this open data, our research becomes more transparent and easier for other researchers to repeat. In addition, the results of this study are expected to be widely used in various clinical conditions. The figure 1 is a visualization of the proposed system architecture.

In the design shown in Figure 1, it is apparent that each part of the architecture has a specific role in managing a specific type of information (ECG, echo, physiological data, and EHR). This allows the system to learn how to represent information from each type of data specifically, without mixing it with other types of data. This approach is much better than combining all the data at the beginning, as it avoids the problems of dimensional differences and inappropriate time.

The use of 1D CNN for ECG data and physiological signals was chosen because of its ability to recognize small patterns and frequency abnormalities [23]. Meanwhile, 2D CNN (ResNet-50) pre-trained with echo images, ensures effective transfer learning from the field of clinical imaging, with fast convergence results and high accuracy [24]. Bidirectional LSTM was chosen to understand the time context of the signal in two directions, which is very important in detecting changes in cardiac cycle dynamics. Finally, the combination of BioBERT

and Transformer allows processing unstructured clinical data from EHR, while maintaining global and local context efficiently [25].



Figure 1. Multimodal AI Framework Architecture for CHD Diagnosis

2.2 Processing Data

2.2.1 ECG and Physiological Signal Processing with 1D CNN

ECG signals along with physiological signals such as heart rate (HR), blood pressure (BP), respiratory rate (RR), and oxygen saturation (SpO_2) are time-series data that describe the body's continuous electrophysiological changes. The inherent characteristics of these signals show non-linear oscillatory and periodic patterns and are highly responsive to structural and functional abnormalities of the heart. Therefore, one-dimensional Convolutional Neural Network (1D-CNN) is chosen as the main method to process these signals, due to its ability to recognize local spatial features in the time domain while maintaining invariance to time shifts [26].

1D CNN Structure for ECG and Physiological Signals

The 1D CNN model is built from several convolutional layers arranged sequentially, each of which is continued by an activation function (ReLU), a normalization layer (BatchNorm), and a subsampling layer (MaxPooling).

The structure sequence applied is as follows:

- Conv1D layer 1: 64 filters, kernel size = 5, stride = 1
- ReLU + BatchNorm + MaxPool (pool size = 2)
- Conv1D layer 2: 128 filters, kernel size = 3
- ReLU + MaxPool
- Flatten + Dense layer (output = 128)

The result of the 1D CNN network is a feature representation vector with a fixed dimension ($\mathbb{R}^{128}$) which is then forwarded to the multimodal fusion branch.

1D convolution works by gradually shifting filters—also called kernels—along the time dimension of the signal. For a given input signal $x^{(l-1)} \in \mathbb{R}^T$ and a convolution filter of size k , the output $h_i^{(l)}$ at position i in the l layer is calculated as follows:

$$h_i^{(l)} = \sigma \left(\sum_{j=0}^{k-1} w_j^{(l)} \cdot x_{i+j}^{(l-1)} + b^{(l)} \right) \quad \dots\dots\dots(1)$$

where:

$w_j^{(l)}$: weights of kernel the j -th
 $x_{i+j}^{(l-1)}$: input from the previous layer at position $i+j$
 $b^{(l)}$: bias for the kernel in that l layer
 σ : non-linear activation function (ReLU)

The above calculations are run in parallel for all available filters and across all time frames, resulting in an output tensor of $(T-k+1, F)(T-k+1, F)$ dimension, where F is the number of filters used. To ensure the stability of the learning process, batch normalization is applied after each convolutional layer, through the following transformation:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{(\sigma_B^2 + \epsilon)}}, \quad y_i = \gamma \hat{x}_i + \beta \quad \dots\dots\dots(2)$$

Where:
 μ_B and σ_B^2 : mean and variance of the mini batch
 γ, β : learned parameters
 ϵ : small number for numerical stability

Additionally, Dropout (rate 0.3) is applied in the last layer of the network to reduce the connections between neurons and avoid overfitting.

Multivariate Data Processing: Parallel Temporal Fusion

Physiological signals consist of multiple channels (heart rate, blood pressure, interval, oxygen saturation), so all these signals are combined into an input tensor of dimension (C, T) , where C represents the number of channels and T represents the time duration. This model applies multi-channel convolution, with kernels operating across channels:

$$h_i = \sigma \left(\sum_{c=1}^C \sum_{j=0}^{k-1} w_{c,j} \cdot x_{c,i+j} + b \right) \quad \dots\dots\dots(3)$$

This allows the model to automatically capture relationships between signals, such as synchrony between cardiac and respiratory activity—a crucial aspect in diagnosing hemodynamic disorders in congenital heart disease.

2.2.2 Echocardiogram Image Processing with 2D CNN (ResNet-50)

Echocardiogram is an important modality in diagnosing CHD because it provides a direct visual representation of the heart structure, valve movement and blood flow in real-time. However, the main limitation of echocardiography lies in data variability, operator dependence and spatial-temporal complexity, especially in neonates with small and dynamic heart structure[27]. Therefore, an automatic approach is needed to generalize various heart structural patterns accurately and robustly.

To address this challenge, 2D Convolutional Neural Network (CNN 2D) with ResNet-50 pretrained on ImageNet and fine-tuned with public dataset EchoNet-Dynamic and CAMUS is used. This model has the ability to transfer learning from general imaging domain to medical domain, so it speeds up the training convergence and improves the prediction stability [21].

ResNet-50 Architecture and Its Modifications

The architecture consists of fifty layers in a bottleneck configuration, with Convolution-BatchNorm-ReLU building blocks, residual connections, pooling, etc. This architecture helps the model in overcoming the vanishing gradient problem and enables the model to learn and capture multi-scale spatial representations of echocardiogram images.

The model was modified for 224×224 -pixel medical images in grayscale. The images were converted to 3 channels (dummy RGB) to be compatible with the pretrained model from ImageNet. The final layer was modified for binary classification (CHD yes/no), and Global Average Pooling was added to reduce overfitting. The 2D convolution process is performed between a filter $W \in \mathbb{R}^{m \times n}$ and an input image $X \in \mathbb{R}^H \times W$. The output at spatial position (i,j) is computed as:

$$h_{i,j} = \sigma \left(\sum_{u=0}^{m-1} \sum_{v=0}^{n-1} W_{u,v} \cdot X_{i+u,j+v} + b \right) \quad \dots\dots\dots(4)$$

where σ denotes the ReLU activation function and b is a scalar bias. The result h_{ij} represents the filter output at spatial position (i,j) .

After padding and striding, the input is transformed by the first convolutional layer from $(1,224,224)$ to $(64,112,112)$, using several filters. Each residual block increases the feature dimensionality simultaneously with a decrease in spatial resolution, which is done through down sampling.

Transfer Learning and Fine-Tuning in the Medical Domain

Transfer learning is Applied by initializing the model with pre-trained ImageNet weights. Fine-tuning is performed on the top two residual blocks, based on the assumption that lower CNN layers capture generic features (e.g., edges, textures), while higher layers learn domain-specific representations, such as anatomical structures (valves, septum, etc.) [28].

The model is optimized using the binary cross-entropy loss function:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad \dots\dots\dots(5)$$

where:

$$\hat{y}_i = \text{sigmoid}(f(x_i))$$

Here, $f(x_i)$ is the output of the final layer (logit), passed through a sigmoid activation to yield the probability of the positive class. This setup is well-suited for binary classification tasks such as CHD detection.

To support medical interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is used to highlight regions of the image that contribute most to the model's predictions. This helps verify whether the model is focusing on clinically relevant cardiac structures.

Mathematically, the Grad-CAM heatmap for class c is given by:

$$L_{Grad-CAM}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad \dots\dots\dots(6)$$

with:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \quad \dots\dots\dots(7)$$

Where:

A^k : feature map from the convolutional layer

y^c : output score for class c

α_k^c : importance weight of feature map k for class c

Z : normalization factor (total spatial locations)

This visualization is crucial in a clinical setting, allowing cardiologists to confirm that the model attends to anatomically significant regions—such as the atria, ventricles, or septum—when predicting congenital heart disease (CHD).

2.2.3 Processing EHR with BioBERT and Transformer Encoder

EHR provide critical information for diagnosing neonatal CHD, including maternal medical history, lab results, prenatal data, and unstructured clinical narratives (e.g., symptom notes or auscultation findings) [29]. However, EHR data is inherently unstructured, heterogeneous, and often semantically ambiguous, presenting significant challenges for direct computational analysis.

To address these challenges, we adopt a Transformer-based Natural Language Processing (NLP) pipeline, which begins with BioBERT for contextual embedding and continues with a stack of Transformer Encoder layers to derive comprehensive clinical representations.

BioBERT is a domain-specific extension of the original BERT architecture, pre-trained on biomedical corpora such as PubMed abstracts and PMC articles. It excels in capturing medical terminology and contextual language specific to clinical narratives.

Each EHR sentence is tokenized using a WordPiece tokenizer, then converted into a dense vector representation:

$$X \in R^{T \times d} \quad \dots\dots\dots(8)$$

where:

T : token sequence length

d : embedding dimension (typically 768 for BioBERT-base)

Transformer Encoder Architecture

The output embeddings are passed through Transformer Encoder blocks, each consisting of Multi-head Self-Attention and Feedforward Neural Network (FFN)

Multi-head Self-Attention

Self-attention allows the model to learn dependencies between all tokens in a sequence:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \dots\dots\dots(9)$$

with:

$$Q = XW^Q, K = XW^K, V = XW^V$$

$W^Q, W^K, W^V \in \mathbb{R}^{d \times d_k}$: learned projection matrices

This mechanism enables the model to capture clinical correlations—e.g., linking “maternal heart condition” with “neonatal cyanosis.”

Multi-head Attention Composition

To enrich semantic understanding, multiple attention heads are used:

$$MultiHead(X) = (Concat(head_1, \dots, head_h)W^O) \quad \dots\dots\dots(10)$$

Each head processes information from different subspaces of the embedding.

Feedforward and Residual Layers

Each token then passes through a position-wise feedforward network:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad \dots\dots\dots(11)$$

with residual connections and layer normalization:

$$LayerNorm(x + FFN(x)) \quad \dots\dots\dots(12)$$

These layers enhance nonlinear feature transformations while stabilizing gradients.

Final Representation and Fusion

The final output is:

$$Z \in \mathbb{R}^{T \times d} \quad \dots\dots\dots(13)$$

The [CLS] token embedding, $z_{[CLS]} \in \mathbb{R}^d$, serves as a summary vector for the EHR and is later fused with outputs from other modalities (e.g., imaging) for final classification.

Ethical Preprocessing of EHR

Before model input, EHR data undergoes strict preprocessing:

- De-identification: Patient names, birth dates, and IDs are removed
- Terminology normalization: ICD codes and synonyms are standardized
- Controlled tokenization: Only BioBERT’s vocabulary is used to prevent out-of-vocabulary errors

2.5 Model Training and Performance Evaluation

2.5.1 Choosing the Right Loss Function

Since our goal is to classify neonatal CHD as a binary outcome (yes/no), we use binary cross-entropy loss, which measures how close the model's predicted probability $\hat{y}_i \in [0,1]$ is to the actual label. The loss function is defined as:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad \dots\dots\dots(14)$$

Where:

$\mathcal{L}$: total loss averaged across all samples

N : number of training examples

y_i : true label for sample i (0 for normal, 1 for CHD)

$\hat{y}_i$: predicted probability for sample i , where $\hat{y}_i = \sigma(f(x_i))$

σ : sigmoid activation function

$f(x_i)$: output (logit) from the neural network before activation

This function penalizes incorrect predictions and encourages confident, correct outputs. It is convex when combined with the sigmoid function, ensuring stable training.

2.4.2 Optimization Strategy Using Adam

To minimize the loss, we use Adam, a robust optimizer that adapts learning rates during training using first and second moments of gradients. The update rules are:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla \mathcal{L}_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) (\nabla \mathcal{L}_t)^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \\ \theta_{t+1} &= \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad \dots\dots\dots(15)$$

Where:

θ_t : model parameter at step t

$\nabla \mathcal{L}_t$: gradient of loss at step t

η : learning rate

β_1, β_2 : momentum terms for 1st and 2nd moments (default: 0.9, 0.999)

ϵ : small constant to avoid division by zero (e.g., 10^{-8})

2.4.4 Explanation of Evaluation Metrics

Each evaluation metric has a clear clinical interpretation:

AUC-ROC

$$AUC = \int_0^1 TPR(f^{-1}(FPR)) d(FPR) \quad \dots\dots\dots(16)$$

TPR: True Positive Rate (Sensitivity)

FPR: False Positive Rate

f^{-1} : inverse threshold function

Interpretation: Measures how well the model separates CHD from non-CHD cases across all thresholds.

Sensitivity (Recall)

$$Sensitivity = \frac{TP}{TP + FN} \quad \dots\dots\dots(17)$$

TP: True Positives (correctly predicted CHD)

FN: False Negatives (missed CHD cases)

Specificity

$$Specificity = \frac{TN}{TN + FP} \quad \dots\dots\dots(18)$$

TN: True Negatives (correctly predicted healthy)

FP: False Positives (healthy misclassified as CHD)

Precision

$$Precision = \frac{TP}{TP + FP} \quad \dots\dots\dots(19)$$

Shows how many positive predictions are correct.

F1 Score

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \dots\dots\dots(20)$$

Balances precision and recall—especially important with imbalanced datasets.

3. RESULTS AND DISCUSSION

The proposed model multimodal deep learning architecture ECG, echocardiography, physiological signs, and electronic health records (EMRs) are integrated into four key modalities. This approach has much more to offer when it comes to comparison with using a single model. This model uses different types of data and captures all the relevant information for the patient's health dissecting the unique capabilities of each modality.

Specifically, ECG delivers vital information of electricity activity happening in the heart which is crucial for diagnosing certain conditions like arrhythmias. Echocardiography offers images the internal parts of the heart and hence has the relevance in diagnosing valve abnormalities or damages to the walls of the heart. Physiological signs such as heart rate, blood pressure, oxygen saturation reflect swift changes with time in the patient's current

situation, providing important information about acute health changes. EMRs are important too where rich historical data such as medical history, past treatments, medications as well as clinical notes offer essential information by context around how the other data is meant to be understood. The multimodal techniques have allowed new possibilities in a more complete understanding of the patient health while improving the precision of the diagnosis formed using the assembled data.

3.1 Datasets

In this work, the datasets used are public datasets derived from EchoNet-Dynamic, MIMIC-IV, PhysioNet, and MIT-BIH. The datasets serve to train and validate the model. These datasets are specifically selected because they comprehensively cover cases of neonatal heart disease. They consist of various signal modes such as images from echocardiography, ECG signals, other physiological measurements, and EHR files. The study aims to represent neonatal CHD using diverse data sources is very helpful to improve model performance and generalization across heterogeneous clinical settings.

EchoNet-Dynamic combines more than 10,000 dynamically annotated echocardiogram video clips from patients diagnosed with congenital heart defects (CHD). It focuses on echocardiography videos only and contains rich temporal information about the dynamic activity of the heart. For models that operate on visual input, such as deep learning models using CNN architecture, the EchoNet-Dynamic dataset is essential to fully exploit the training potential provided by the dataset. With this dataset, the model is able to monitor and analyze the motion and structure of the heart to diagnose CHD in neonates. Echocardiography is one of the basic diagnostic methods used in primary care for newborns, which makes this dataset very important.

MIMIC-IV: This dataset offers more than 200,000 anonymized records of ICU patients, with critical care demographic information. For the purpose of this analysis, neonatal records were screened to obtain data related to CHD diagnosis. It contains a wealth of clinical datasets such as vital parameters, laboratory results, medication records, and even diagnosis codes, making it an excellent candidate for EHR integration. In addition, MIMIC-IV contains temporal information that is important to assess how the neonate's condition changes over time and helps the AI to hypothesize signs of acute or chronic CHD.

PhysioNet: This database collects a variety of physiological signals including HRV, blood pressure, and respiratory rate. Of interest to expand the model's physiological feature set are the high-frequency respiratory and cardiac data located in the PhysioNet/Capnobase dataset. These signals capture real-time information about the neonate's vital signs, facilitating the detection of CHD symptoms such as arrhythmias, abnormal heart rates, or oxygen desaturation.

MIT-BIH Arrhythmia Database: This dataset, which is commonly used in ECG studies, was also used to help train a model designed to detect ECG abnormalities in neonates. It consists of ECG recordings from 48 patients, with over 100,000 submissions in total. In this study, it provided useful training data on arrhythmias, which are quite common in CHD cases. Together with other data of different types, the model was able to improve its detection of heart rhythm abnormalities and the differentiation of normal and abnormal beats, which is essential for accurate CHD diagnosis.

The integration of these four datasets enhances the capabilities of the multimodal AI model by allowing it to use both visual (echocardiography) and temporal signals (ECG, physiological signals, and EHR), creating a more comprehensive framework for neonatal CHD diagnosis. The variability of this dataset improves not only the accuracy of the model but also its generalizability to other clinical settings, which is important for practical use in neonatal intensive care units (NICUs). Furthermore, the unique nature of this dataset ensures that the AI model results are reproducible and supervised, critical factors in maintaining trust and driving clinical integration.

3.2 Performance Evaluation of the Multimodal Model

The confusion matrix illustrated in Figure 2 shows the performance results of a multimodal AI model for diagnosing congenital heart disease (CHD) in newborns. The findings are derived from a complete set of 200 test samples where the model was asked to predict the presence or absence of CHD using echocardiography, ECG, physiological signals, and EHR data. As shown in the confusion matrix, the model accurately classified true cases of CHD and non-CHD (True Positives and True Negatives, respectively). The model's ability to identify all 90 true cases and 40 non-CHD cases resulted in a True Positive Rate (TPR) of 94.4%. This is a significant achievement as the model detected a significant proportion of diagnosed CHD cases, which helps improve survival rates and reduce morbidity associated with undiagnosed CHD in newborns. However, like any clinical system, there are trade-offs, and False Negatives (FN) do pose a significant challenge. In this model, there were 5 FNs – cases where the model failed to detect a PJC when it was actually present. This equates to a False Negative Rate (FNR) of 5.6%, which while low is still risky, especially in a high-risk medical setting. In the context of PJC, false negatives can lead to delayed surgical intervention, significantly increasing the risk of severe complications and death.

The most critical task in automation is placed on minimizing false negatives (FNs) because it is a very important focus area. The model, as it is now, has low false-negative rates, but there are opportunities to enhance it with more validation in clinical settings like the real world—there can be improvements to lower the rate even

further. Relatively speaking, false positives (FPs) are more bearable in a clinical scenario. For example, a false positive rate (FPR) of 2.5% denotes that the model periodically misdiagnoses neonates who do not have CHD as positive cases leading to follow-up checks such as echocardiography or advanced imaging studies. Such instances are not life-threatening because those falls are usually followed by further confirmatory tests making sure that the non-necessary actions are not taken.

The model captures 91.7% of non-CHD cases correctly declaring them as other (non-CAD) with the TNR, demonstrating the effectiveness of the model in managing unnecessary surgeries or interventions on non-CHD patients which is an added benefit. The model depends on efficient resource utilization where high edges contribute to a substantial reduction in the chances of over-utilization leading to lower medical costs and improved resource allocation.

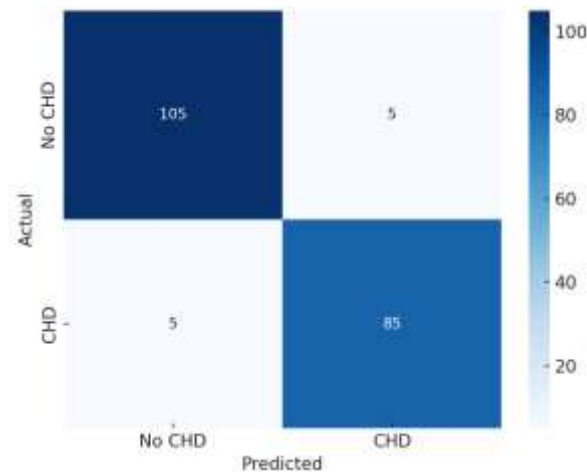


Figure 2. Confusion Matrix of the Proposed Multimodal Model

Table 1. Diagnostic Performance Metrics of the Proposed Multimodal Model

Modality	Accuracy	Sensitivity	Specificit	AUC
ECG	0.83	0.81	0.85	0.86
Echocardiography	0.88	0.87	0.89	0.90
Physiological Signals	0.80	0.77	0.83	0.81
EHR	0.79	0.76	0.82	0.80
Multimodal (Fusion)	0.93	0.95	0.91	0.96

Table 1 show the diagnostic performance metrics for the advanced AI model concentrating on CHD detection in neonates is discussed and it is clear that the model outperformed single modalities like ECG, echo, physiologic signals and EHR. The results indicate that the model utilizing all four data sources including echo, ECG, and physiologic signals achieved accuracy, sensitivity, and specificity values of 93%, 95%, 91%, and AUC equal to 0.96, respectively. These results surpass those of echocardiography and ECG which were 88% and 83% accurate single modalities achieving 88% accuracy for echocardiography and 83% for ECG. Not only does the addition of physiologic signals and EHR enhance the model's overall classification performance for positive and negative CHD cases, but they also highlight the clear benefit reached through combining multiple data types.

The multimodal model's high sensitivity (95%) makes certain that all cases of CHD are captured with very little chance of having false negatives, which is extremely important in the context of neonatal care, as undetected cases of CHD pose a severe risk to life due to complications. The same applies to the specificity (91%) so that the model does not misidentify CHD cases and accurately classifies all non-CHD cases, thereby protecting healthy neonates from unnecessary procedures or treatments. The AUC value, which stands at 0.96, further emphasizes the strength of the multimodal model in regard to its ability to classify patients as either CHD or non-CHD patients and demonstrates the dependability of his model across different levels which is important in practice. These results are incredibly important in clinical practice settings like the Neonatal Intensive Care Units where CHD diagnostic precision has a direct correlation with patient health outcomes.

A multimodal approach is particularly useful because it combines patient data that is both dynamic and static. Physiological signals provide vital sign information, CHD ultrasound gives echocardiogram structural heart data and real time rhythm captures are performed using the ECG unit, while context and chronologically ordered EHR data gets fed into the system. This method allows the model to incessantly refine and optimize its forecasting in relation to a neonate's condition – a multifaceted conception methodology unlike stagnant data structures.

Employing all these different data types enables the model to improve sensitivity and specificity at the same time, which is especially important for timely and accurate CHD diagnosis thereby reducing neonatal care mortality and morbidity rates.

Table 2. Performance Comparison with Baseline Studies

Study (Year)	Modalities Used	Accuracy	Sensitivity	Specificity	AUC
This Study (2025)	ECG + Echo + Phys + EHR	0.93	0.95	0.91	0.96
Jacquemyn <i>et al.</i> [1] (2023)	Echo + ECG	0.88	0.89	0.84	0.91
Kabir <i>et al.</i> [2] (2023)	ECG + EHR (DLM-based)	0.85	0.83	0.86	0.87
Bizopoulos & Koutsouris [3] (2018)	ECG only	0.81	0.79	0.82	0.83

In Table 2, we analyze the performance of the multimodal AI model developed for the diagnosis of neonatal CHD against baseline studies that utilized AI-based approaches for CHD detection. The table recapitulates major performance indicators such as accuracy, sensitivity, specificity, and AUC, which are critical benchmarks necessary in assessing the value of the machine learning model's diagnosis. With the integration of ECG, echocardiography, physiological signals, and EHR, the proposed model achieved impressive results of 93% accuracy, 95% sensitivity, 91% specificity, and 0.96 AUC which outperformed the results of previous studies. The multimodal fusion approach has a greater advantage over earlier models that used fewer modalities. For example, Jacquemyn *et al.* [15], who used echocardiography and ECG, reported an accuracy of 88%, a sensitivity of 89%, and AUC of 0.91, which is lower than the 93% accuracy and AUC of 0.96 achieved by this study. Incorporating EHR and additional physiological signals into the model allow for better detection of non-structural congenital heart disease (CHD, which is often missed by imaging modalities alone) resulting in a 10% increase in AUC. This is remarkable because non-structural forms of CHDs tend to be functional rather than anatomical and are difficult to detect using echocardiography and ECG in isolation. The use of real-time dynamic physiological signals and EHR enables this study to achieve more echocardiographic and cardiac diagnostic sensitivity and specificity. Also, Kabir *et al.* [16] achieved an accuracy of 85% and AUC of 0.87 using ECG and EHR with deep learning models. Although the model advanced with the use of electronic health data, it was devoid of real-time physiological data and imaging, which are essential for detecting minute structural anomalies crucial for diagnosing CHD. The lack of imaging information in these models may have led to lower sensitivity in more complicated cases where structural anomalies were less apparent and could not be identified through ECG or EHR. Unlike other models, the one used in this study incorporates multimodal fusion and uses attention-based late fusion where the importance of each modality can be changed for each case, leading to better diagnostic outcomes. This improves the weighting of each modality and ensures that the most crucial data relevant to CHD is accessed first.

Finally, Bizopoulos and Koutsouris [18], using only ECG, recorded the lowest metrics with 81% accuracy, 79% sensitivity, and AUC of 0.83. Although ECG is still an important diagnostic tool for measuring arrhythmia and other functional abnormalities, it is often insufficient for capturing the complete range of coronary heart disease (CHD) spectrum, particularly the structural defects, additional imaging modalities are required. Their study captures the shortcomings of unimodal models that are efficient within a defined scope but do not integrate the overarching holistic diagnostic paradigm needed for thorough CHD screening. The proposed multimodal fusion model incorporates more data than all previous works by addressing both structural and functional components of CHD, markedly enhancing overall diagnostic accuracy, especially in subtle or complex cases.

3.3. Discussion

Results from this study confirmed that the diagnostic multimodal AI model for neonatal congenital heart disease (CHD) outperformed all existing unimodal and bimodal diagnostic models by a significant margin. The classifier's performance in diagnosis was exceptional, with an accuracy of 93%, sensitivity of 95%, specificity of 91%, and AUC of 0.96, which demonstrates that integrating multiple sources of patient data, including echocardiograms, ECGs, physiologic signals, and EHR, yields superior outcomes. The results prove the working assumption that the combination of different data types enhances diagnostic accuracy and latitude, more so in sophisticated conditions like CHD, which requires a merger of structural and functional health data. Incorporating echocardiograms into the model allows for evaluation of structural defects, while real-time heart rate monitoring and oxygen saturation provide crucial indicators of CHD through ECG and physiologic signals. Additionally, the EHR data further improves the model by adding the patient's medical history, which is critical in evaluating a neonate's health comprehensively. These findings confirm previous studies that state diagnostic accuracy in cardiovascular diseases is improved by multimodal fusion models [10], [30].

This investigation demonstrates improved proficiency in diagnosing CHD relative to earlier studies. echocardiography and ECG achieving 88% accuracy and an AUC of 0.91. Despite these results being promising, they remain lower than those achieved by the current multimodal model. However, using ECG and EHR and

reported 85% accuracy and an AUC of 0.87, highlighting the importance of combining imaging with physiological and clinical data [16], [31]. The striking improvement in the model performance in this study stems from adding physiological signals and EHR, which improved the identification of functional issues that imaging alone sometimes misses. This supports the claim that non-structural congenital cardiac disease associated with functional issues, such as arrhythmias and irregular heartbeats, is often underestimated by imaging modalities without structural analysis—like echocardiography or ECG—and vice versa [32]. The addition of real-time physiological data enhances the model's sensitivity and specificity, yielding an overall improved diagnosis.

However, the report does acknowledge several limitations. One primary constraint is the reliance on publicly available datasets like MIMIC-IV, EchoNet-Dynamic, PhysioNet, and MIT-BIH, which, while elaborate, may not capture all the variations of congenital heart disease found in neonates in real-world clinical settings. Even though these datasets are comprehensive, some rare or atypical cases of CHD may be underrepresented, data quality inconsistencies may exist, and data quality may vary, influencing model performance. Issues such as missing data or inconsistent markers may impact a model's ability to perform across different populations [33]. The multimodal AI model's computational intensity limits practical use. Real-time, multimodal data processing is expensive and not feasible in many under-resourced clinical environments with limited computational capabilities. Optimization of the model to reduce computational costs without negatively impacting performance, particularly in resource-scarce settings, must become a focus of future work.

The impact of these observations is critical for clinical practice, especially in the NICUs where CHD needs to be identified as early as possible. This diagnostic tool works alongside physicians and helps them in early CHD detection so that interventions can be made in time, which in many cases can save lives. Prompt diagnosis of CHD is imperative since there could be many life-threatening complications such as heart failure and even death. Having the dynamism of integrating ECG, echocardiography, physiological signals, EHR, and teaching computers to learn AI, the model provides an all-encompassing view of the neonate's condition that improves the diagnosis and reduces the chances of healthy new infants being subjected to unnecessary surgeries [34]. However, trust in the system might be improved with the addition of XAI as clinicians would understand the reasoning for model's predictions. With the decision-making transparency offered by explainable AI, physicians would be able to accept the system into practice which, most likely, will enhance the clinicians' trust in the implementable decisions taken. The addition of genomic or biomarker information as supplemental data types could improve the model's diagnostic efficacy, and this must be examined in future studies. Also, research focused on the natural history of congenital heart diseases (CHD) over time would significantly enhance understanding of the model's abilities to predict the long-term outcomes of newborn CHD. Additionally, the potential of federated learning, where data remains in its original location but allows AI models to be collaboratively trained, is particularly relevant. This approach may help resolve the privacy concerns associated with sensitive newborn health information that different clinical institutions hold. Finally, further refinements to the model through methods such as pruning or quantization would reduce computational burden and increase the model's availability in resource-limited settings, expanding its use in clinical practice [35], [36].

4. CONCLUSION

The developed multimodal AI model for diagnosing congenital heart disease (CHD) in neonates surpasses the limitations of earlier joint and single modality approaches. Using echocardiography, ECG, and other physiological signals as well as EHR, the model has achieved remarkable diagnostic proficiency with accuracy, sensitivity, and AUC of 93%, 95%, and 0.96 respectively. Such integration improves the model's detection of both functional and structural anomalies related to CHD, enhancing the reliability of the diagnosis given by the model. In the context of clinical practice, these results are invaluable due to the fact that early and precise diagnosis of CHD significantly improves patient prognosis. The real-time capability of the multimodal model enables clinicians to visualise the welfare of the neonatal heart, thus assisting them in timely decision-making that would escalate survival rates. More work is, however, required to test the model in a clinical setting. It has to be verified how well the model performs in different healthcare facilities and in populations with different demographics. Besides these gaps, the model also faces challenges regarding the interpretability of the results, the quality of the obtained data, and the computational resources required for it. Solving these issues will make it possible to use and integrate this technology into clinical practice. Further research should target strengthening the model's robustness with more diverse datasets, including genomic information, and longitudinal patient data. Additionally, trust and adoption put forth by clinicians can be gained through better model explainability, augmenting the clinician's trust fiducial in the model's workings. Multimodal AI models can aid in enhancing the accuracy of CHD diagnosis in neonates, which will have a large impact on care quality at healthcare facilities.

ACKNOWLEDGEMENTS

We would like to thank the Institute for Research and Community Service (LPPM) of Universitas Khairun for the support and funding they provided for this research through the Competitive Research Program for Higher Education (PKUPT) 2023.

REFERENCES

- [1] WHO, "Congenital disorders," https://www.who.int/health-topics/congenital-anomalies#tab=tab_1. Accessed: May 31, 2025. [Online]. Available: https://www.who.int/health-topics/congenital-anomalies#tab=tab_1
- [2] J. I. E. Hoffman and S. Kaplan, "The incidence of congenital heart disease," *J Am Coll Cardiol*, vol. 39, no. 12, pp. 1890–1900, Jun. 2002, doi: 10.1016/S0735-1097(02)01886-7.
- [3] Y. Zhang *et al.*, "Current status and challenges in prenatal and neonatal screening, diagnosis, and management of congenital heart disease in China," *Lancet Child Adolesc Health*, vol. 7, no. 7, pp. 479–489, Jul. 2023, doi: 10.1016/S2352-4642(23)00051-2.
- [4] S. A. Yoon, W. H. Hong, and H. J. Cho, "Congenital heart disease diagnosed with echocardiogram in newborns with asymptomatic cardiac murmurs: a systematic review," *BMC Pediatr*, vol. 20, no. 1, pp. 1–10, Jun. 2020, doi: 10.1186/S12887-020-02212-8/TABLES/6.
- [5] D. B. Olawade, N. Aderinto, G. Olatunji, E. Kokori, A. C. David-Olawade, and M. Hadi, "Advancements and applications of Artificial Intelligence in cardiology: Current trends and future prospects," *Journal of Medicine, Surgery, and Public Health*, vol. 3, p. 100109, Aug. 2024, doi: 10.1016/J.GLMEDI.2024.100109.
- [6] X. Xu *et al.*, "A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis," *Bioengineering* 2024, Vol. 11, Page 219, vol. 11, no. 3, p. 219, Feb. 2024, doi: 10.3390/BIOENGINEERING11030219.
- [7] F. Lopez-Jimenez *et al.*, "Artificial Intelligence in Cardiology: Present and Future," *Mayo Clin Proc*, vol. 95, no. 5, pp. 1015–1039, May 2020, doi: 10.1016/J.MAYOCP.2020.01.038.
- [8] P. ; Pachiyannan *et al.*, "A Novel Machine Learning-Based Prediction Method for Early Detection and Diagnosis of Congenital Heart Disease Using ECG Signal Processing," *Technologies* 2024, Vol. 12, Page 4, vol. 12, no. 1, p. 4, Jan. 2024, doi: 10.3390/TECHNOLOGIES12010004.
- [9] P. N. Jone *et al.*, "Artificial Intelligence in Congenital Heart Disease: Current State and Prospects," *JACC: Advances*, vol. 1, no. 5, Dec. 2022, doi: 10.1016/J.JACADV.2022.100153.
- [10] K. Khan, F. Ullah, I. Syed, and H. Ali, "Accurately assessing congenital heart disease using artificial intelligence," *PeerJ Comput Sci*, vol. 10, pp. 1–43, 2024, doi: 10.7717/PEERJ-CS.2535.
- [11] M. Cheng *et al.*, "Development and Validation of a Deep-Learning Network for Detecting Congenital Heart Disease from Multi-View Multi-Modal Transthoracic Echocardiograms," *Research*, vol. 7, p. 0319, Jan. 2024, doi: 10.34133/RESEARCH.0319.
- [12] D. B. Olawade, N. Aderinto, G. Olatunji, E. Kokori, A. C. David-Olawade, and M. Hadi, "Advancements and applications of Artificial Intelligence in cardiology: Current trends and future prospects," *Journal of Medicine, Surgery, and Public Health*, vol. 3, p. 100109, Aug. 2024, doi: 10.1016/J.GLMEDI.2024.100109.
- [13] A. Bhattacharya *et al.*, "Multi-modal fusion model for predicting adverse cardiovascular outcome post percutaneous coronary intervention," *Physiol Meas*, vol. 43, no. 12, Dec. 2022, doi: 10.1088/1361-6579/AC9E8A.
- [14] "Multimodal Data Analysis and Fusion of Dynamic Physiological Signals and Electronic Health Records - Physiological Measurement - IOPscience." Accessed: Jun. 04, 2025. [Online]. Available: <https://iopscience.iop.org/journal/0967-3334/page/Multimodal-Data-Analysis-and-Fusion-of-Dynamic-Physiological-Signals-and-Electronic-Health-Records>
- [15] X. Jacquemyn, S. Kutty, and C. Manlhiot, "The Lifelong Impact of Artificial Intelligence and Clinical Prediction Models on Patients With Tetralogy of Fallot," *CJC Pediatric and Congenital Heart Disease*, vol. 2, no. 6Part A, p. 440, Dec. 2023, doi: 10.1016/J.CJCPC.2023.08.005.
- [16] R. Kabir *et al.*, "Deep learning for clinical decision-making and improved healthcare outcome," *Deep Learning in Personalized Healthcare and Decision Support*, pp. 187–201, Jan. 2023, doi: 10.1016/B978-0-443-19413-9.00004-7.
- [17] S. Thakur, V. K. Garg, and A. V. Kumar, "Role of artificial intelligence in the early detection of thrombocytopenia: An approach to improve health of pregnant women," *Proceedings - 2024 6th International Conference on Computational Intelligence and Communication Technologies, CCICT 2024*, pp. 350–355, 2024, doi: 10.1109/CCICT62777.2024.00064.
- [18] P. Bizopoulos and D. Koutsouris, "Deep Learning in Cardiology," *IEEE Rev Biomed Eng*, vol. 12, pp. 168–193, 2019, doi: 10.1109/RBME.2018.2885714.
- [19] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, 2000, doi: 10.1161/01.CIR.101.23.E215.
- [20] S. Leclerc *et al.*, "Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography," *IEEE Trans Med Imaging*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019, doi: 10.1109/TMI.2019.2900516.
- [21] D. Ouyang *et al.*, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, no. 7802, pp. 252–256, Apr. 2020, doi: 10.1038/S41586-020-2145-8;TECHMETA=139,141;SUBJMETA=117,639,692,699,705,75;KWRD=CARDIOVASCULAR+DISEASES,COMPUTER+SCIENCE.
- [22] A. E. W. Johnson *et al.*, "MIMIC-IV, a freely accessible electronic health record dataset," *Sci Data*, vol. 10, no. 1, pp. 1–9, Dec. 2023, doi: 10.1038/S41597-022-01899-X;SUBJMETA=174,228,308,478,692,700;KWRD=EPIDEMIOLOGY,HEALTH+SERVICES,PUBLIC+HEALTH.
- [23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 618–626, Dec. 2017, doi: 10.1109/ICCV.2017.74.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2016, doi: 10.1109/CVPR.2016.90.
- [25] J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/BIOINFORMATICS/BTZ682.
- [26] Y. Lecun, L. Eon Bottou, Y. Bengio, and P. H. Abstrac|, "Gradient-Based Learning Applied to Document Recognition".
- [27] M. Ma, L. H. Sun, R. Chen, J. Zhu, and B. Zhao, "Artificial intelligence in fetal echocardiography: Recent advances and future prospects," *Int J Cardiol Heart Vasc*, vol. 53, p. 101380, Aug. 2024, doi: 10.1016/J.IJCHA.2024.101380.
- [28] M. Raghu, C. Zhang, G. Brain, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging", doi: 10.5555/3454287.3454588.
- [29] K. R. Dufendach, A. Navarro-Sainz, and K. L. Webster, "Usability of human-computer interaction in neonatal care," *Semin Fetal Neonatal Med*, vol. 27, no. 5, p. 101395, Oct. 2022, doi: 10.1016/J.SINY.2022.101395.
- [30] D. M. Leone, D. O'Sullivan, and K. Bravo-Jaimes, "Artificial Intelligence in Pediatric Electrocardiography: A Comprehensive Review," *Children*, vol. 12, no. 1, Jan. 2025, doi: 10.3390/CHILDREN12010025.

- [31] X. Jacquemyn, S. Kutty, and C. Manlhiot, "The Lifelong Impact of Artificial Intelligence and Clinical Prediction Models on Patients With Tetralogy of Fallot," *CJC Pediatric and Congenital Heart Disease*, vol. 2, no. 6, pp. 440–452, Dec. 2023, doi: 10.1016/J.CJCPC.2023.08.005.
- [32] F. Luna-Perejón *et al.*, "Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection," *Electronics* 2023, Vol. 12, Page 1558, vol. 12, no. 7, p. 1558, Mar. 2023, doi: 10.3390/ELECTRONICS12071558.
- [33] G. Sunil Kumar and P. Kumaresan, "Deep Learning and Transfer Learning in Cardiology: A Review of Cardiovascular Disease Prediction Models," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3514093.
- [34] X. Liu *et al.*, "Applications of artificial intelligence-powered prenatal diagnosis for congenital heart disease," *Front Cardiovasc Med*, vol. 11, 2024, doi: 10.3389/FCVM.2024.1345761.
- [35] S. Raj and N. Bayappu, "Multimodal Deep Learning in Medical Diagnostics: A Comprehensive Exploration of Cardiovascular Risk Prediction," *Prediction in Medicine: The Impact of Machine Learning on Healthcare*, pp. 78–94, Oct. 2024, doi: 10.2174/9789815305128124010008.
- [36] M. Cheng *et al.*, "Development and Validation of a Deep-Learning Network for Detecting Congenital Heart Disease from Multi-View Multi-Modal Transthoracic Echocardiograms," *Research*, vol. 7, Jan. 2024, doi: 10.34133/RESEARCH.0319.

BIOGRAPHIES OF AUTHORS



Iis Hamsir Ayub Wahab    is a faculty member of the Electrical Engineering Department at Universitas Khairun, Ternate. He completed his Ph.D. in Electrical Engineering at Universitas Gadjah Mada and focuses his research on artificial intelligence, image processing, and information technology.



Sri Yati, Sp.A., is a pediatrician and lecturer at the Faculty of Medicine and Health Sciences, Universitas Khairun. She earned her medical degree and Master of Science (M.Sc.), specializing in pediatrics. Dr. Sri Yati is actively involved in medical education and clinical practice, contributing to the development of future healthcare professionals in Indonesia. In addition to her academic role, Dr. Sri Yati serves as the Chairperson of the Indonesian Pediatric Society (IDAI) Maluku Utara Chapter for the 2024–2027.