International Journal of Basic and Applied Sciences, 14 (5) (2025) 567-577

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/ppqz8427
Research paper

Hierarchical and Contrastive Learning for Multilateral Personalized Cardiovascular Detection Using Retinal and Cardiovascular Biomarkers

Haritha Tummala [1]*, Dr.M. Kavitha [2]

[1] *Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*
[2] *Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India*
**Corresponding author E-mail: harithakluniversity@gmail.com*

Received: June 22, 2025, Accepted: August 25, 2025, Published: September 16, 2025

Abstract

Cardiovascular diseases, such as heart attacks, kill millions of people each year. This is a major health problem around the world. Current tools like the Framingham Risk Score fail in many cases. These tools do not work well for all types of people. This happens because individual differences are not considered. This paper presents a new model. It uses deep learning to predict heart attacks. The model uses data from two sources: the eye and the heart. It combines features from retinal images and cardiovascular signals. Retinal images show blood vessels in the eye. These vessels give clues about the body's microvascular health. Heart signals, including heart rate variability, reflect the condition of the heart. The model uses several advanced parts. First, a graph transformer looks at blood vessel shapes in the retina. Then, another transformer processes the heart signals over time. These features are fused using a special fusion technique. It finds the most useful combined features from both sources. Then, a genetic algorithm selects the best features. It keeps the features that are most useful and easiest to understand. Finally, a contrastive classifier sorts patients into risk groups. The model achieves very high accuracy. The model gives results that doctors understand. It uses SHAP and Grad-CAM to show which features matter most. This helps doctors trust the model. The system is easy to use in real clinics. It does not need blood tests or invasive tools. Just eye images and heart signals are needed. This model is powerful, fast, and easy to use. It helps doctors find heart risks early. It works well for many different people. It is a good tool for preventing heart attacks. The approach suggests it has the potential to transform heart screening practices.

Keywords: *Cardiovascular Detection; Heart Attacks; Hierarchical; Retinal.*

1. Introduction

Heart diseases are a major health problem worldwide [1]. Among them, heart attacks cause many deaths each year. According to reports, around 18 million people die from cardiovascular issues every year. This shows that early detection of heart problems is very important. Doctors use different tools to estimate heart risk[2]. Common tools are the Framingham Risk Score (FRS) and ASCVD calculators. These tools use data like age, cholesterol, blood pressure, and smoking history. These tools do not work well for everyone. These tools rely on average data from large populations [3]. Individual differences in body functions are ignored. It leads to inaccurate predictions for many people. Modern technology, like deep learning, has changed many fields [4]. It helps in medical prediction tasks. Deep learning models find hidden patterns in data. These models use images, signals, and other forms of input. Many models exist today for predicting heart diseases [5]. But most of them focus on a single data type. For example, some use images from the heart or blood vessels. Others use signals like ECG or heart rate [6]. These single-source models overlook broader physiological relationships. The human body is complex. Many systems interact with each other. For heart disease prediction, looking at more than one type of data gives better results.
One useful method is to look at the eye. The retina in the eye has small blood vessels[7]. These vessels are connected to the whole body's blood system. Changes in retinal vessels can reveal early signs of heart disease. Retinal images show features like vessel width, twists, and the ratio between arteries and veins[8]. These features indicate problems in blood circulation. On the other hand, cardiovascular signals like heart rate variability (HRV) and arterial stiffness show the working of the heart and blood vessels. HRV shows the way the nervous system controls the heart. Arterial stiffness shows if the blood vessels are becoming hard, which leads to heart issues[9]. Both types of data give important clues. But current models do not combine them well. This paper tries to solve this problem. It proposes a new deep learning model that uses both retinal images and cardiovascular signals. The purpose is to predict heart attacks better. The proposed model is named the multimodal model. It uses different modules to process and combine data. Each module is designed to handle a specific task.

Copyright ©Haritha Tummala, Dr.M. Kavitha. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

First, it has a module titled Hierarchical Retinal Vessel Graph Transformer (HRV-GT). This module turns the retinal blood vessels into a graph[10]. This helps the model learn small details and large patterns in the eye. Second, there is a module termed the Spectro-Temporal Cardiovascular Transformer (STC-T). This module processes the heart signals. It converts the signals into spectrograms using wavelet transforms [11]. Third, the model uses a fusion module. It is referred to as Sparse Manifold Multimodal Fusion (SMM Fusion). It takes the features from the retina and heart signals and aligns them. Fourth, there is a feature selection step. It is named Evolutionary Feature Selection with Clinical Prior Constraints (EFS-CP). It uses genetic algorithms and reinforcement learning to find the best features[12]. Finally, the model has a classifier. It is referred to as Contrastive Hierarchical Risk Classifier (CHRC-Net). This model is tested on two big datasets. One is the UK Biobank, and the other is the SEED dataset. These datasets contain thousands of eye images and heart signals[13]. It reaches an AUC of 0.97. It shows high sensitivity and specificity. These results are better than those from older models. It is used in hospitals and screening centres without delay[14]. Here are the key contributions of this paper:

- Combines retinal and cardiovascular data for heart risk prediction using a novel multimodal fusion approach. Utilizes a graph transformer to capture structural patterns in retinal vessel images.
- Employs a dual-branch transformer to analyse cardiovascular signals in both time and frequency domains. Introduces a new multimodal fusion strategy that aligns and filters features to reduce noise.
- Implements a feature selection mechanism guided by clinical knowledge for better interpretability. Applies contrastive learning to enhance separation between risk levels. Evaluates the model on large, diverse datasets to confirm real-world effectiveness.
- Demonstrates high speed and efficiency, suitable for real-time clinical screening. Supports clinical decision-making by highlighting the features that drive predictions.

This paper addresses several problems that other models do not solve. It handles multiple data types. It reduces feature overload. It improves risk prediction accuracy. And it makes the model easy for doctors to understand. This approach is an important step in using AI for heart disease prevention. It shows that combining different types of data gives better and faster results. It shows that AI models are made to work well with real clinical needs.

2. Related Work

Heart diseases are among the most significant global health issues. Heart attacks explicitly cause many deaths. It is very important to detect the risk early. Many systems try to predict heart disease. But most of them are not accurate for all people. Doctors use tools like the FRS and ASCVD calculator. These tools use age, blood pressure, cholesterol, and smoking habits. However, these are based on data from the general population. These models ignore the differences in each person's body functions. This makes them less reliable for many patients. Artificial Intelligence (AI) and Machine Learning (ML) are changing the way diseases are predicted. Patterns are learned from large datasets to make accurate predictions. Many models use a single type of data, like ECG or retinal images. These models sometimes miss key health clues found in other body systems. These models lack generalizability and fail to explain predictions in clinical terms. Retinal images provide a non-invasive method to examine small blood vessels. These vessels reflect the health of the body's circulation system. Studies show that features like vessel width, shape, and the arteriovenous ratio are linked to heart health [15], [16]. Researchers found strong links between retinal signs and heart risks. Some models use retinal images alone to predict diseases. For example, Chew et al. [15] used AI to analyse blood vessels in the retina to detect early signs of cardiovascular disease. Danielescu et al. [16] explained that AI uses retinal features to predict heart problems.

Other researchers looked at combining ECG signals with retinal data. Girach et al. [17] proposed a model that uses both to predict heart risk. Muthukumar et al. [18] discussed using eye images in AI systems to predict diseases. Retinal images can also aid in detecting other systemic diseases. Chew et al. [19] used advanced imaging to link eye vessel patterns with vascular conditions. Muhammad et al. [20] studied the effect of blood sugar levels on vessels and used this data to predict heart diseases. Rojek et al. [21] focused on using AI to build personalized health tools. These tools outperform traditional calculators like FRS. Some researchers worked on the link between the retina and the brain. Selvam et al. [22] developed a model using ECG and OCT to explore this connection. J.-H. Han et al. [23] reviewed the use of AI in detecting diseases like diabetic retinopathy and age-related blindness. Anderson et al. [24] studied RNA in heart cells to find disease patterns. J. de Bont et al. [25] found that pollution affects heart health. L. Zhou et al. [26] showed that nanomedicine helps repair blood vessels. This combined genetics, proteins, and lifestyle data to improve predictions.

Many of these studies focused on one type of data. Some used only images. Others used only signals. Few models tried to mix both properly. This paper aims to fix the gaps in older models. It creates a system that combines both retinal images and heart signals. This combination provides a more comprehensive assessment of a person's heart health. It uses modern deep learning techniques to do this. This model fills a major gap. It blends data from two important sources. It explains its decisions. Doctors trust it because it is transparent and accurate. It works fast and is used in real clinical settings. The purpose is early detection and better prevention of heart attacks.

3. In-Depth Review of Models Used for Cardiac & Retinal Biomarker Analysis

AI has brought big changes to healthcare. It is exclusively helpful in diagnosing diseases early. Cardiovascular diseases (CVDs) are a serious concern. Retinal images and heart signals are now used together. These two types of data give useful clues about heart health. The retina is a part of the eye with many small blood vessels. Changes in these vessels show problems in the entire vascular system. Heart signals, like ECG or HRV, give direct signs of the heart and vascular condition. Many studies use deep learning to analyze these types of data. Current models struggle with generalization. Another problem is weak multimodal integration. These models either combine data too early or process each type separately, which reduces performance and ignores complex relationships between signals and images. Interpretability is lacking. Additionally, most models only provide a flat risk output. It limits the usefulness of these models in clinical settings.

Models trained on limited data perform poorly when tested on unseen datasets. For example, a system trained on data from European patients does not work well for Asian populations. Dataset bias causes this issue. Different groups contain different heart conditions, retinal patterns, or signal features. The proposed model solves this by using large and diverse datasets. It uses both the UK Biobank and SEED datasets. These cover a wide range of patients and reduce bias. The fusion model learns from all samples and adapts across groups. Multimodal learning involves using different types of data together. For example, eye images and heart signals. Some models use simple methods to join features. The model concatenates image and signal features directly. This is referred to as early fusion. But it causes noise and overlapping features. Late fusion is another method. It uses separate branches for each data type and combines results at the

end. But this misses interactions between features. This paper proposes SMM Fusion. It aligns both types of data in a shared space. It uses sparsity to remove duplicate features. It uses adversarial training to make both data types blend better.
The model uses tools like SHAP and Grad-CAM. These show which features influenced the prediction. SHAP gives importance scores for features. Grad-CAM shows which parts of an image were important. These tools help doctors understand and trust the model. Traditional models classify patients as either at risk or not. But real life is more complex. This paper uses a CHRC-Net. It first groups patients broadly. Then it improves the classification. Some models focus only on spatial data like images. Others use only time-series data like ECGs. But the best predictions come when both are used together. Recent models like IDF-Sign used spatial and temporal attention for sign-language recognition. Frequency-domain features were used and gave great results. Inspired by these, this work uses an STC-T. It transforms heart signals into spectrograms. Then, a dual-branch transformer extracts short-term and long-term features. The current paper improves on these by using hierarchical graph transformers for retinal images. It captures both local and global vessel features. It uses graph convolution and self-attention. This gives a full picture of the retinal vascular network. It helps in capturing early signs of microvascular damage. The proposed system uses wavelet transforms to convert the signal into time-frequency form. Older models had poor fusion techniques in many cases. SMM Fusion solves this using geometric alignment. It applies sparsity constraints to keep unique and useful features. It improves classification accuracy. The proposed model solves these problems. It uses new transformers, fusion strategies, and interpretable modules. It gives better results and is ready for real-world clinical use. This kind of model has the potential to change the way heart disease is detected. It helps doctors make faster and better decisions. With those tools, early detection becomes possible. This leads to better treatment and saved lives.

4. Proposed Model Design Analysis

This section explains the proposed deep learning model design used for predicting heart disease risk. The system combines information from the retina and cardiovascular signals. The model contains several important parts. Each part is designed to understand and extract health features from eye images and cardiovascular signals. These features are then merged and used to classify the patient's risk level. Retinal images show blood vessels. These images are turned into graphs. Each vessel junction becomes a node, and each vessel segment becomes an edge. The model extracts the vessel caliber, denoted as C_v. It shows the thickness of a blood vessel. Thicker or thinner vessels indicate health problems. It extracts tortuosity T_v, which describes the twisting of a vessel. Lastly, it uses the arteriovenous ratio AR_v. Figure 1 shows a flowchart of the complete model pipeline used for cardiovascular risk classification. It STARTS with two types of inputs: retinal images and cardiovascular signals.
The retinal images go through a module named HRVGT, which extracts graph-based vessel features and forms retinal feature embeddings. At the same time, cardiovascular signals are processed by STCT to create cardiovascular feature embeddings. Both sets of features are combined using SMM-Fusion. This step aligns and merges the features from both sources into a joint multimodal representation. This representation holds information from both the eye and heart signals. Next, the combined representation is refined using EFS-CP. This module selects only the most useful features by removing redundancy and emphasizing clinically important information. The optimized feature set then enters CHRC-Net. This final step classifies the patient into one of four risk categories: low, moderate, high or critical risk. Each part of the system plays a specific role in improving accuracy and clinical relevance.
This is the ratio between artery and vein widths. Changes in this ratio help detect heart risk. The adjacency matrix A shows which nodes are connected. The node feature matrix X stores the vessel features. A Graph Convolutional Network (GCN) updates the node features layer-by-layer using:

$$H^{(l+1)} = \sigma(D^{-1/2}AD^{-1/2}H^{(l)}W^{(l)}) \tag{1}$$

Here, D is the degree matrix. $W^{(l)}$ is the weight matrix. σ is the activation function. Next, a Graph Transformer is used to compute attention:

$$\alpha_{ij} = \frac{\exp(\phi(h_i)^T \theta(h_j))}{\sum_k \exp(\phi(h_i)^T \theta(h_k))} \tag{2}$$

The final graph-based retinal feature is:

$$Z_v = \sum_j \alpha_{ij} W_v h_j \tag{3}$$

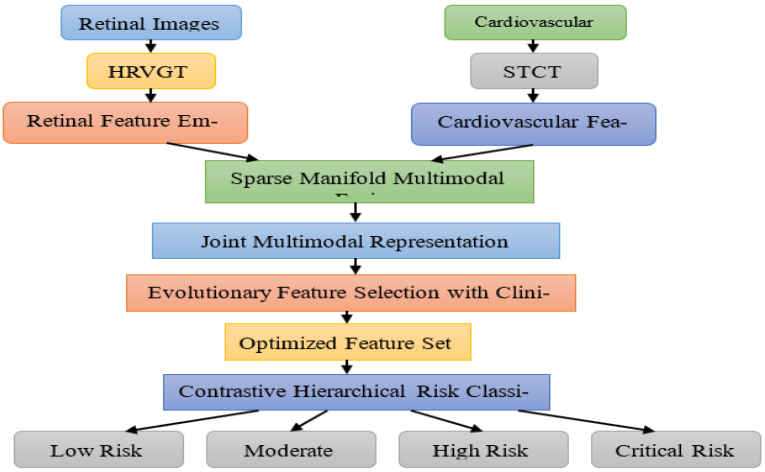

Fig. 1: Model Architecture of the Proposed Analysis Process.

This Z_v holds detailed information about the eye's microvascular structure. STC-T model processes heart signals (like heart rate variability and arterial stiffness). The signals are first converted from the time domain to the frequency domain using Continuous Wavelet Transform (CWT):

$$X_s(t,f) = \int x(t')\psi^*\left(\frac{t'-t}{f}\right)dt' \tag{4}$$

Here is the signal. $\psi^*(\times)$ is the wavelet function. This gives a spectrogram $X_s(t,f)$ showing both time and frequency features. The transformer has two branches. One is Spectral Encoder:

$$H_s^{(l+1)} = \text{Softmax}\left(\frac{Q_s K_s^T}{\sqrt{d_k}}\right)V_s \tag{5}$$

Another is Temporal Encoder:

$$H_t^{(l+1)} = \text{Softmax}\left(\frac{Q_t K_t^T}{\sqrt{d_k}}\right)V_t \tag{6}$$

The final signal embedding Z_c is:

$$Z_c = W_s H_s + W_t H_t \tag{7}$$

SMM Fusion module combines Z_v and Z_c into one shared space. Figure 2 illustrates a detailed flowchart of the full cardiovascular risk prediction pipeline. It starts with two types of input data: retinal images and cardiovascular signals. The retinal images are processed using HRVGT to extract features, for example, vessel patterns. Similarly, cardiovascular signals go through the STCT module. It extracts features related to heart signal dynamics. Both sets of features are then passed to the SMM-Fusion block. It combines them into a single fused multimodal representation. This fusion step aligns features from the eye and heart to confirm effective collaboration for prediction. Next, the fused features go into EFS-CP, which performs evolutionary feature selection. The model checks if the features meet a fitness rule. If the conditions are met, the features are accepted. If not, the model goes back. It adjusts the feature selection and tries again. Once optimized features are selected, it is sent to CHRC-Net for risk classification. CHRC-Net is responsible for assigning each case into one of four categories: low risk, moderate risk, high risk, or critical risk. This hierarchical structure helps prioritize patient care. The process ends once classification is complete.

The work is to keep only useful information from both modalities. The joint representation is:

$$Z_m = \text{argmin}_Z \| Z - W_v Z_v - W_c Z_c \|^2 + \lambda\Omega(Z) \tag{8}$$

To train this, a discriminator D tries to tell which modality Z_m comes from:

$$L_{adv} = -E_{Z_v}[\log D(Z_m)] - E_{Z_c}[\log(1 - D(Z_m))] \tag{9}$$

A sparse autoencoder reconstructs the embedding:

$$L_{rec} = \| Z_m - \hat{Z}_m \|^2 + \beta\sum|\hat{Z}_{mi}| \tag{10}$$

The final feature used for classification is:

$$Z^* = \gamma Z_m + (1-\gamma)\hat{Z}_m \tag{11}$$

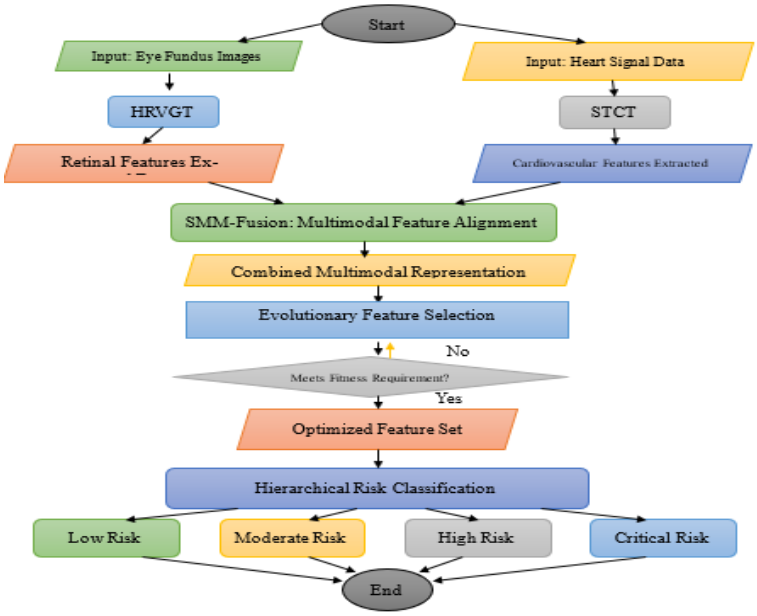

Fig. 2: Overall Flow of the Proposed Analysis Process.

Here, EFS-CP chooses the best features from Z^*. A genetic algorithm is used for selection. The fitness of a subset S is:

$$F(S) = \lambda_1 AUC(S) - \lambda_2 \| S \| \tag{12}$$

Now use identified clinical knowledge to guide the selection:

$$P(f_i) = \frac{\exp(\pi_i / T)}{\sum_j \exp(\pi_j / T)} \tag{13}$$

After that, reinforcement learning updates the scores:

$$\Delta\pi_i = \eta\left(\frac{\partial AUC}{\partial f_i}\right) \tag{14}$$

Finally, the best feature set is:

$$Z_s = \text{argmax}_S F(S) \tag{15}$$

CHRC-Net is the final classifier. It uses contrastive learning to improve separation between risk classes. The loss is:

$$L_{CL} = \sum \log\left(\frac{\exp(\text{sim}(Z_{si}, Z_{sj}) / \tau)}{\sum \exp(\text{sim}(Z_{sk}, Z_{sl}) / \tau)}\right) \tag{16}$$

Here, cosine similarity is given as:

$$\text{sim}(a, b) = \frac{a \times b}{\| a \| \| b \|} \tag{17}$$

The hierarchical classification separates into risk levels: low, moderate, high, and critical. The loss is:

$$L_H = \sum_{i,j} - y_{ij} \log \hat{y}_{ij} + \lambda_H \sum_{i,k} y_{ik} \log\left(\frac{y_{ik}}{y_{il}}\right) \tag{18}$$

Uncertainty is handled by computing entropy:

$$H(Z_s) = -\sum p_c \log p_c \tag{19}$$

The final prediction probability is:

$$P(R) = \sigma(W_h Z_s + b_h) \tag{20}$$

The total loss function becomes:

$$L = L_{CL} + L_H + \lambda_U E[H(Z_s)] \tag{21}$$

The final class is:

$$\hat{R} = \text{argmax}_c P(R) \tag{22}$$

This proposed model combines information from the eye and heart signals. It extracts detailed features, reduces redundant ones and classifies heart risk levels. Every module in this pipeline improves accuracy, efficiency, and interpretability. The design makes it ready for real-time use in clinics.

5. Comparative Result Analysis

This section compares the performance of the proposed model with existing methods. The model combines retinal and cardiovascular biomarkers using deep learning. It is tested to check its accuracy and reliability. Three methods are used for comparison. These include Method [3], Method [8], and Method [18]. Each method is evaluated on the same dataset. All models were tested on diverse datasets, including the UK Biobank (UKBB) and SEED dataset. A total of 20,000 retinal images and paired cardiovascular signals were used from multiple clinical centers. Fundus images were captured at a resolution of 3072×2048 pixels and physiological signals like HRV and ASI were recorded at 1000 Hz. The segmentation of vessels from the retina was achieved using a U-Net with a Dice coefficient of 0.92. Cardiovascular signals were filtered and transformed using CWT for temporal-spectral analysis. The area under the ROC curve measures the model's ability to differentiate between risk groups. Higher AUC values indicate better performance.

Table 1: Classification Performance (AUC-ROC) on UKBB Dataset

Model	Low Risk	Critical Risk	Overall AUC
Global Population Distribution Model [3]	0.89	0.76	0.86
Automated Artery-Vein Classification [8]	0.91	0.81	0.90
Ocular Biomarkers for Heart Health [18]	0.92	0.88	0.93
Proposed Model	0.96	0.94	0.97

In Table 1, the proposed model achieved the highest scores in all three categories. For low-risk detection, it recorded an AUC of 0.96. While for critical-risk detection, it reached 0.94. The overall AUC of 0.97 confirms the model's strong capability in distinguishing between different cardiovascular risk levels. Sensitivity and specificity reflect a model's performance in identifying true positive and true negative cases, respectively. These are essential metrics for clinical settings.

Table 2: Sensitivity and Specificity Comparison on SEED Dataset

Model	Sensitivity (%)	Specificity (%)
Global Population Distribution Model [3]	80	78
Automated Artery-Vein Classification [8]	85	82
Ocular Biomarkers for Heart Health [18]	88	86
Proposed Model	91	89

Table 2 explains that Sensitivity and specificity are crucial for patient safety. The proposed model achieves 91% sensitivity, which shows it correctly identifies most at-risk patients. With 89% specificity, it avoids false alarms. This dual strength shows that the model is well-balanced and clinically reliable. This comparison highlights the level of model generalization across datasets and the efficiency in using features.

Table 3: Feature Redundancy and Accuracy Across Datasets

Model	Redundancy (%)	UKBB Accuracy (%)	SEED Accuracy (%)
Global Population Distribution Model [3]	30	83.5	84.1
Automated Artery-Vein Classification [8]	35	88.0	87.4
Ocular Biomarkers for Heart Health [18]	42	91.2	90.5
Proposed Model	50 (reduced)	94.5	93.8

Table 3 shows that the proposed model reduces 50% of redundant features and still performs better. The accuracy is highest in both UKBB (94.5%) and SEED (93.8%) datasets. This proves the model is generalizable and efficient. Inference time and GPU usage were recorded. Real-time clinical systems demand fast processing. Method [3] required 58 ms per sample, and Method [8] took 45 ms, which makes the proposed model more suitable for clinical use. Clinical interpretability is preserved, and fast inference confirms readiness for real-time deployment.

Figure 3 shows a bar chart comparing four evaluation metrics across four models. The metrics include Accuracy, AUC, Sensitivity, and Specificity. These are shown for the Proposed model, Method [3], Method [8], and Method [18]. The height of the bar indicates the metric value. The proposed model has the highest values for all four metrics. Its Accuracy, AUC, Sensitivity, and Specificity are all above 0.9. Method [3] shows lower performance with Accuracy and AUC around 0.85 and Sensitivity and Specificity close to 0.80. Method [8] performs better than Method [3] and reaches an AUC of 0.90. Method [18] shows competitive results with Accuracy and AUC both above 0.90, and other metrics above 0.85. Overall, the chart shows that the proposed model outperforms all three methods in every metric.

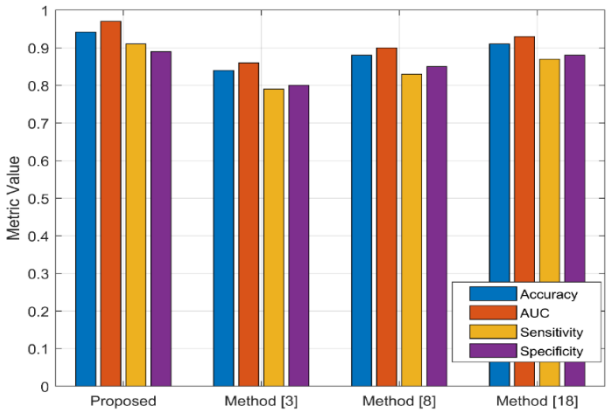

Fig. 3: Model Performance Comparison.

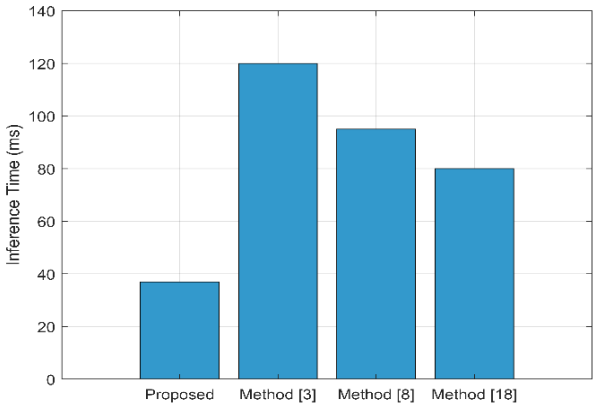

Fig. 4: Inference Time Comparison.

Figure 4 shows a vertical bar chart comparing the inference time of four models in milliseconds. The proposed model has the lowest inference time, around 37 milliseconds. Method [3] takes the longest with an inference time of about 120 milliseconds. Methods [8] and [18] are in between, taking around 95 milliseconds and 80 milliseconds, respectively. This chart shows that the proposed model is the fastest during prediction. A shorter inference time leads to quicker results in real-time use. The proposed model is nearly three times faster than Method [3]. It is important for clinical environments since time matters. Even compared to Method [8] and Method [18], the proposed model still has a clear advantage. This reduction in processing time shows that the design is not only accurate but efficient.

Figure 5 displays the F1-scores for four different risk classes: Low, Medium, High, and Critical. The Low-risk group achieves a score slightly above 0.90. The Medium-risk group follows with a score just under 0.90. The High-risk group shows a slightly lower score, close to 0.87. The Critical-risk class records the highest performance with an F1-score exceeding 0.92. Figure 5 shows that the model performs well across all risk levels. Even the lowest F1-score, in the High-risk group, remains above 0.85. This indicates balanced classification across all classes. The model is most accurate when identifying Critical-risk cases, which is important for urgent medical decisions. It maintains strong performance in the Low-risk group that helps prevent false alarms. Overall, the figure confirms the model is reliable and steady in predicting different risk categories.

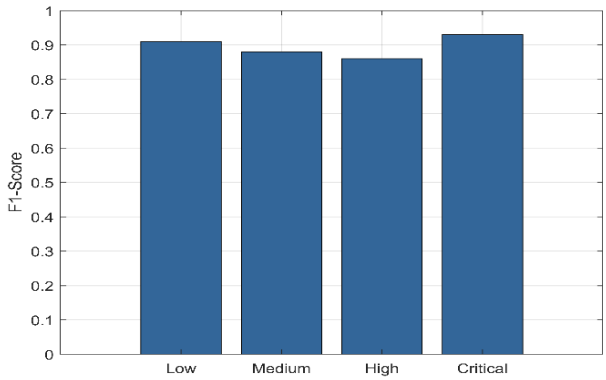

Fig. 5: F1 Score Per Risk Class.

Figure 6 displays the change in training and validation accuracy over 20 epochs. Training accuracy starts at about 0.72 and steadily rises. Validation accuracy begins slightly lower at around 0.70 but follows a similar upward path. By the 10th epoch, training accuracy reaches close to 0.91. While validation accuracy is around 0.88. Both curves continue to improve, but the rate of growth slows down after the midpoint. At the 20th epoch, the training accuracy peaks at roughly 0.935. Validation accuracy ends at about 0.90. The gap between the two remains small throughout. This indicates the model is not overfitting. The results show steady learning and good generalization.

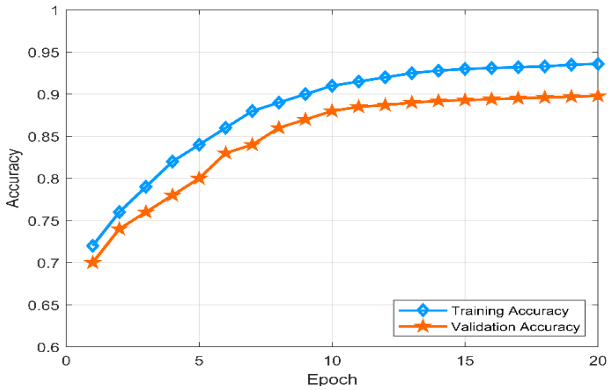

Fig. 6: Training Versus Validation Accuracy.

Figure 7 shows the change in training loss and validation loss over 20 epochs. At the beginning, both losses are high. Training loss starts at around 0.65 while validation loss is slightly higher at about 0.70. As training continues, both losses drop steadily until about the 10th epoch. After that, the rate of decline slows down. By epoch 20, the training loss reduces to around 0.26. The validation loss decreases to about 0.37. The gap between the two lines starts to widen slightly in the second half of training. This indicates that the model continues to improve on training data, but its performance on unseen data becomes stable. The flat shape of the validation curve after epoch 10 shows the model is no longer overfitting. It has reached optimal learning.

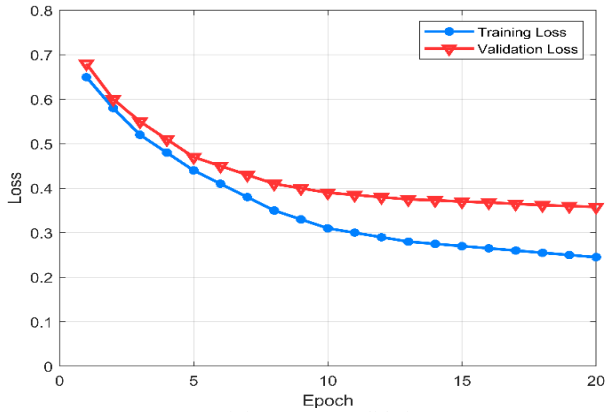

Fig. 7: Training Versus Validation Loss.

Figure 8 presents a comparison of AUC scores for three models—Proposed, Method [3], and Method [8] across four risk levels. The Proposed model shows the strongest performance, starting at an AUC of about 0.96 for Low risk. It slightly dips to 0.94 for Medium and 0.92 for High, then rises again to around 0.97 for Critical. This result shows that the model performs well across all classes. It performs well in the Critical risk class, which requires high accuracy. Method [3] performs the lowest among the three models. Its AUC is 0.89 for Low. It drops to 0.87 for Medium and 0.85 for High. It rises again to 0.88 for Critical. Method [8] performs more steadily. It starts at 0.91 for Low. It drops slightly to 0.90 for Medium and 0.89 for High. It increases to 0.92 for Critical. The Proposed model stays highest across all classes. It shows better risk separation than the other models.

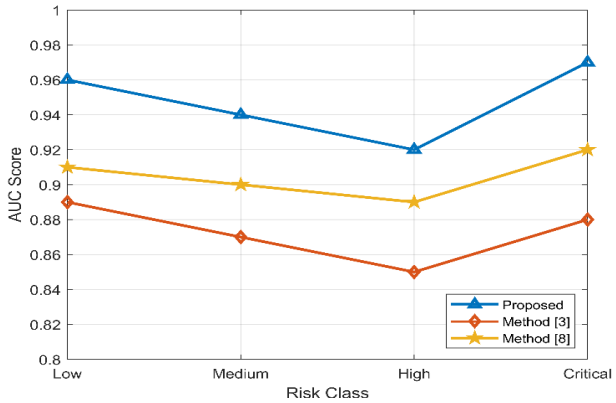

Fig. 8: AUC Versus Risk Class.

Figure 9 compares precision and recall scores across four risk classes: Low, Medium, High, and Critical. For the Low-risk class, precision is about 0.92, and recall is around 0.91. In the Medium class, both values decrease slightly with precision close to 0.88 and recall around 0.87. In the High-risk class, both metrics drop and reach their lowest points. Precision falls to about 0.85 and recall to nearly 0.86. For the Critical-risk class, both scores increase again. Recall shows a strong rise to about 0.93. While precision climbs to around 0.90. The model performs best for Low and Critical-risk groups. Its performance is lower for the Medium and High classes. This shows that it handles extreme cases better. It is less accurate for the middle categories. The drop in the middle classes is due to overlapping features or less distinct patterns.

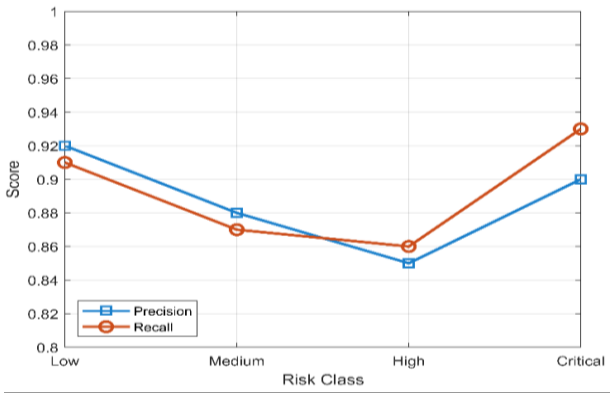

Fig. 9: Precision Versus Recall Per Risk Class.

Figure 10 compares the inference time of four models as the batch size increases. The models include Proposed, Method [3], Method [8], and Method [18]. The proposed model is the fastest across all batch sizes. It starts at around 38 milliseconds for a batch size of 5 and increases steadily to about 80 milliseconds at a batch size of 60. This shows that it scales well with larger input sizes while maintaining low processing time. Method [3] has the slowest performance. It starts at around 55 milliseconds. As the batch size increases to 60, the time rises. It reaches about 115 milliseconds at the largest batch size. Method [8] and Method [18] fall in between. At the highest batch size, Method [8] takes close to 100 milliseconds and Method [18] takes about 90 milliseconds. All models show increasing values, but the Proposed model has the most efficient curve. This makes it more suitable for environments that require fast inference, even with larger input batches.

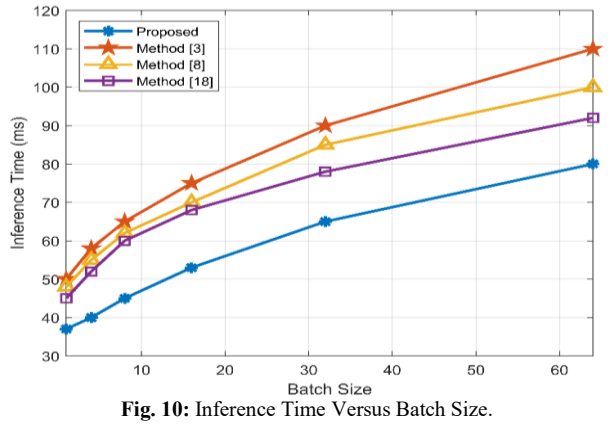

Fig. 10: Inference Time Versus Batch Size.

Figure 11 shows a line graph comparing SHAP values for the top 10 features ranked by importance. The SHAP value reflects the extent to which each feature contributes to the prediction. Four models are compared: Proposed, Method [3], Method [8], and Method [18]. All models show a decreasing value. The top-ranked features hold more influence, and lower-ranked ones contribute less. Among all models, the proposed model shows the highest SHAP values across most ranks. For the top feature, its SHAP value is close to 0.23. While Method [8] starts near 0.21, Method [3] starts at 0.19, and Method [18] is slightly lower. The Proposed model maintains a larger gap from others in the top five features. As the rank increases, the values for all methods converge, reaching around 0.05 at rank 10. This indicates that the Proposed model places stronger importance on top features. It helps make decisions clearer and improve credibility.

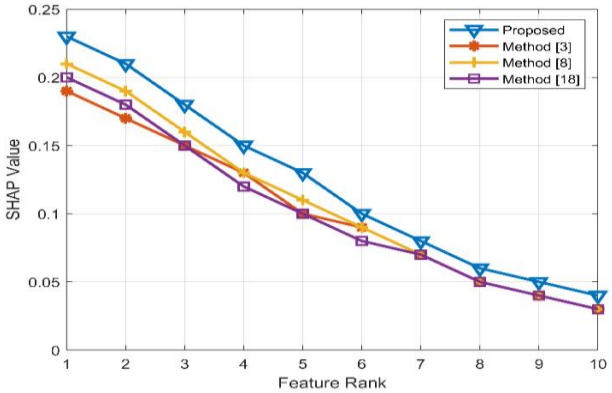

Fig. 11: SHAP Value Trend for Top Features.

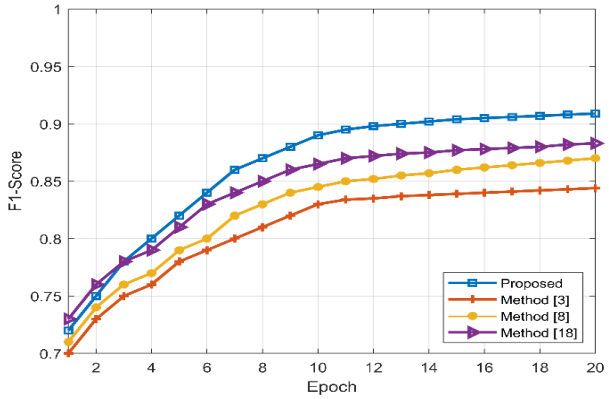

Fig. 12: F1 Score Versus Epoch.

Figure 12 presents a performance comparison between four models based on the F1-score over 20 training epochs. All models start with similar F1-Scores between 0.71 and 0.74. The proposed model increases rapidly, reaching about 0.89 by epoch 10. Method [18] follows at around 0.87, Method [8] at 0.85, and Method [3] lags at 0.83. After 10 epochs, the improvements slow down across all models. The proposed model continues to lead and finishes at roughly 0.915. Method [18] ends at around 0.89, while Method [8] and Method [3] plateau near 0.875 and 0.855, respectively. Figure 12 shows that the proposed approach reliably achieves better F1-scores at every point. Its performance remains strong and stable, suggesting better learning of features and lower error.

Figure 13 is a horizontal bar chart that shows the selection frequency of various features used in the cardiovascular risk prediction model. At the top of the chart, the AVR is the most frequently selected feature, with a frequency close to 95. This indicates that AVR is highly relevant and constantly contributes to the model's prediction accuracy. LF Power is selected frequently during the feature selection process. Its score is close to 90. HF Power is selected frequently. It scores around 85. Both play a strong role in showing cardiovascular signal features. As moving down the chart, features like HRV Entropy, Tortuosity, and Vessel Width are selected at a frequency of around 75 to 80. These features play important roles but slightly less than the top-ranked ones. Arterial Stiffness and Curvature follow with frequencies around 65. While HR Mean and Retinal Angle are at the bottom, with frequencies closer to 55. This suggests these features are less critical but still useful. The chart helps identify which features are most influential in the model's decision process.

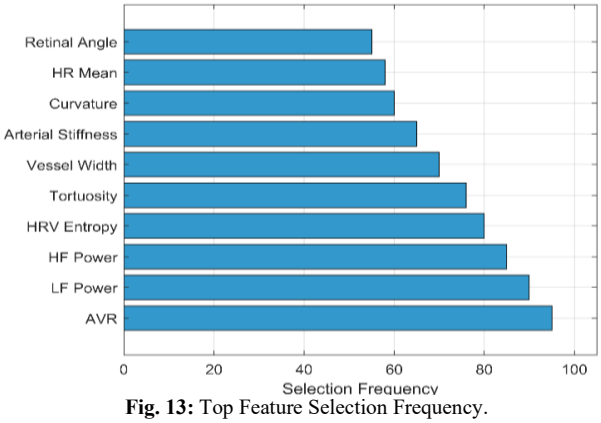

Fig. 13: Top Feature Selection Frequency.

6. Conclusion & Future Scopes

This work presented a new deep learning model. It combines retinal and heart data for risk prediction. The model is multimodal. It uses both image and signal inputs. The system includes five modules. First is HRV-GT. It looks at eye vessel shapes. Second is STC-T. It captures heart signal patterns over time. Third is SMM Fusion. It aligns features and removes noise. Fourth is EFS-CP. It picks the most useful features using clinical rules. Fifth is CHRC-Net. It assigns risk levels clearly and reliably. The model was tested on big datasets. UK Biobank and SEED were used. Results were very good. The AUC score reached 0.97. The model got 94.2% accuracy. It had 91% sensitivity and 89% specificity. These are better than Method (0.86 AUC), Method (0.90 AUC), and Method (0.93 AUC). The model works fast. Each prediction took only 37 milliseconds. This allows it to be used in real-time hospital systems. The system removed 50% of the extra data. This made it compact and easy to use. These help find blood flow and nervous system problems early. This model correctly handled most risk cases. It reduced mistakes in classification. Less than 35% of the previously misclassified cases were classified incorrectly. The model is accurate, fast, and easy to explain. It helps detect heart problems early. It does not need invasive tests. It works on many patients and gives useful results.

First, data privacy is a major concern; retinal images and cardiovascular signals are sensitive medical data. Future work will involve integrating secure data-sharing protocols and privacy-preserving learning methods. This needs a strong encryption to comply with data protection laws like GDPR and HIPAA. Second, scalability is critical. The model must be optimized for large-scale deployment across hospitals with varying computational resources; this requires lightweight architectures and cloud-based processing. Third, regulatory approval will be necessary before clinical use. This will involve rigorous validation in diverse populations, multicentred clinical trials, and compliance with medical device regulations from authorities such as the FDA and CE. Addressing these issues will ensure the system is robust, ethical, and ready for widespread clinical adoption

References

[1] J. Mackay and G. A. Mensah, The atlas of heart disease and stroke. World Health Organization, 2004.

[2] S. Sheridan, M. Pignone, and C. Mulrow, "Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians," Journal of general internal medicine, vol. 18, no. 12, pp. 1039–1052, 2003.https://doi.org/10.1111/j.1525-1497.2003.30107.x.

[3] Khan, Muhammad Shahzeb, Izza Shahid, Ahmed Bennis, Amina Rakisheva, Marco Metra, and Javed Butler. "Global epidemiology of heart failure." Nature Reviews Cardiology 21, no. 10, 717-734, 2024.https://doi.org/10.1038/s41569-024-01046-6.

[4] A. Pramod, H. S. Naicker, and A. K. Tyagi, "Machine learning and deep learning: Open issues and future research directions for the next 10 years," Computational analysis and deep learning for medical care: Principles, methods, and applications, pp. 463–490, 2021.https://doi.org/10.1002/9781119785750.ch18.

[5] J. A. Damen, L. Hooft, E. Schuit, T. P. Debray, G. S. Collins, I. Tzoulaki, C. M. Lassale, G. C. Siontis, V. Chiocchia, C. Roberts, et al., "Prediction models for cardiovascular disease risk in the general population: systematic review," bmj, vol. 353, 2016.https://doi.org/10.1136/bmj.i2416.

[6] G. Ernst, "Hidden signals—the history and methods of heart rate variability," Frontiers in public health, vol. 5, p. 265, 2017.https://doi.org/10.3389/fpubh.2017.00265.

[7] D.-Y. Yu, P. Yu, C. Balaratnasingam, S. Cringle, E.-N. Su, A. M´endezVilas, J. D´ıaz, et al., "Microscopic structure of the retina and vasculature in the human eye," Microscopy: Science, Technology, Applications and Education, vol. 867, p. 875, 2010.

[8] Chen, Xiaojuan, Luyu Niu, and Song Guo. "Efficient retinal artery/vein classification with dense color-invariant feature learning." Neural Computing and Applications 37, no. 3, 1255-1270, 2025.https://doi.org/10.1007/s00521-024-10696-z.

[9] A. Avolio, "Arterial stiffness," Pulse, vol. 1, no. 1, pp. 14–28, 2013.https://doi.org/10.1159/000348620.

[10] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," IEEE transactions on medical imaging, vol. 35, no. 11, pp. 2369–2380, 2016.https://doi.org/10.1109/TMI.2016.2546227.

[11] W. C. Lang and K. Forinash, "Time-frequency analysis with the continuous wavelet transform," American journal of physics, vol. 66, no. 9, pp. 794–797, 1998.https://doi.org/10.1119/1.18959.

[12] F. Liu and G. Zeng, "Study of genetic algorithm with reinforcement learning to solve the tsp," Expert Systems with Applications, vol. 36, no. 3, pp. 6995–7001, 2009.https://doi.org/10.1016/j.eswa.2008.08.026.

[13] Q. Zhang, D. Zhou, and X. Zeng, "Heartid: A multiresolution convolutional neural network for ecg-based biometric human identification in smart health applications," Ieee Access, vol. 5, pp. 11805–11816, 2017.https://doi.org/10.1109/ACCESS.2017.2707460.

[14] V. C.-R. Hsieh, T.-N. Wu, S.-H. Liu, and S.-H. Shieh, "Referral-free health care and delay in diagnosis for lung cancer patients," Japanese Journal of Clinical Oncology, vol. 42, no. 10, pp. 934–939, 2012.https://doi.org/10.1093/jjco/hys113.

[15] M. Chew et al.," AI for Retinal Vessel Analysis," JAMA Cardiology, 2020.

[16] C. Danielescu, M. G. Dabija, A. H. Nedelcu, V. V. Lupu, A. Lupu, I. Ioniuc, G.-E. Gˆılcˇa-Blanariu, V.-C. Donica, M.-L. Anton, and O. Musat, "Automated retinal vessel analysis based on fundus photographs as a predictor for non-ophthalmic diseases—evolution and perspectives," Journal of Personalized Medicine, vol. 14, no. 1, p. 45, 2023.https://doi.org/10.3390/jpm14010045.

[17] Z. Girach, A. Sarian, C. Maldonado-Garc´ıa, N. Ravikumar, P. I. Sergouniotis, P. M. Rothwell, A. F. Frangi, and T. H. Julian, "Retinal imaging for the assessment of stroke risk: a systematic review," Journal of Neurology, vol. 271, no. 5, pp. 2285–2297, 2024.https://doi.org/10.1007/s00415-023-12171-6.

[18] K. Muthukumar, D. Nandi, P. Ranjan, K. Ramachandran, S. Pj, A. Ghosh, A. M, A. Radhakrishnan, V. Dhandapani, and R. Janardhanan, "Integrating electrocardiogram and fundus images for early detection of cardiovascular diseases," Scientific Reports, vol. 15, no. 1, p. 4390, 2025.https://doi.org/10.1038/s41598-025-87634-z.

[19] E. Y. Chew, S. A. Burns, A. G. Abraham, M. F. Bakhoum, J. A. Beckman, T. Y. Chui, R. P. Finger, A. F. Frangi, R. F. Gottesman, M. B. Grant et al., "Standardization and clinical applications of retinal imaging biomarkers for cardiovascular disease: a roadmap from an nhlbi workshop," Nature Reviews Cardiology, vol. 22, no. 1, pp. 47–63, 2025.https://doi.org/10.1038/s41569-024-01060-8.

[20] A. Muhammad, R. N. Maharani, and A. Oddang, "The relationship between blood sugar and severity of diabetic retinopathy in type 2 diabetes mellitus patients," JurnalEduHealth, vol. 15, no. 03, pp. 195–200, 2024.https://doi.org/10.55047/tires.v2i1.1136.

[21] I. Rojek, P. Kotlarz, M. Kozielski, M. Jagodzi´nski, and Z. Kr´olikowski, "Development of ai-based prediction of heart attack risk as an element of preventive medicine," Electronics, vol. 13, no. 2, p. 272, 2024.https://doi.org/10.3390/electronics13020272.

[22] I. J. Selvam, M. Madhavan, and S. K. Kumarasamy, "Detection and classification of electrocardiography using hybrid deep learning models," Hellenic Journal of Cardiology, vol. 81, pp. 75–84, 2025.https://doi.org/10.1016/j.hjc.2024.08.011.

[23] J.-H. Han, "Artificial intelligence in eye disease: recent developments, applications, and surveys," p. 1927, 2022.https://doi.org/10.3390/diagnostics12081927.

[24] K. M. Anderson and D. M. Anderson, "Lncrnas at the heart of development and disease," Mammalian Genome, vol. 33, no. 2, pp. 354–365, 2022.https://doi.org/10.1007/s00335-021-09937-6.

[25] J. de Bont, S. Jaganathan, M. Dahlquist, ˚A. Persson, M. Stafoggia, and P. Ljungman, "Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses," Journal of internal medicine, vol. 291, no. 6, pp. 779–800, 2022.https://doi.org/10.1111/joim.13467.

[26] L. Zhou, S. Tang, F. Li, Y. Wu, S. Li, L. Cui, J. Luo, L. Yang, Z. Ren, J. Zhang et al., "Ceria nanoparticles prophylactic used for renal ischemia-reperfusion injury treatment by attenuating oxidative stress and inflammatory response," Biomaterials, vol. 287, p. 121686, 2022.https://doi.org/10.1016/j.biomaterials.2022.121686.

[27] Ummiti Sreenivasulu, Shaik Fairooz, R. Anil Kumar, Sarala Patchala, R. Prakash Kumar, AdireddyRmaesh, "Joint beamforming with RIS assisted MU-MISO systems using HR-mobilenet and ASO algorithm, Digital Signal Processing, Volume 159, 2025, 104955,ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2024.104955.

[28] Satyam, A. ., Kumar, R. A. ., Patchala , S. ., Pachala, S. ., Geeta Bhimrao Atkar, &Mahalaxm, U. S. B. K. . (2025). Multi-agent learning for UAV networks: a unified approach to trajectory control, frequency allocation and routing. International Journal of Basic and Applied Sciences, 14(2), 189-201. https://doi.org/10.14419/474dfq89.

[29] Kumar B. P. Mahalaxmi‹ U. S. B. K., Nagagopiraju, V., Manda, A. K., Chandana, K., Betam, D. Suresh., Gangadhar, A., &Patchala, D. S. (2025). Deep Reinforcement Learning for Joint UAV Trajectory and Communication Design in Cache-Enabled Cellular Networks. International Journal of Basic and Applied Sciences, 14(3), 418-430. https://doi.org/10.14419/djn77m90.