**IEEE CS Bangalore Chapter Internship and Mentorship Program - 2025**

**Duration: 1st April 2025 to 30th September 2025**

**Monthly Progress Report Template**

**Project ID:**P64

**Name of the Students:**

1. Shreyas R Katti
2. Samarth D
3. Veeresh P T

**University/ College Name:** S G Balekundri Institute of Technology,Belagavi,Karnataka

**Name of the Mentor:** Dr. Ashok Kumar A R

**Mentor’s Organization:** R V College of Engineering,Bengaluru

**Title of the Project:** Retinal AI- A Deep Learning Framework for Early Prediciton of Cardiovascular Diseases via Retinal Imaging

**Progress of the Project**

1.  Problem Definition

2.  Literature Review

3.  Existing System

4.  Proposed System

5.  Knowledge gained - Tools, Technology, Courses etc.

6.  Architectural Framework

7.  Project Implementation

8.  Results

9.  Conclusion and Future Work

10. Research article preparation (if applicable)

**PROBLEM DEFIINTION**

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, with millions of lives lost every year. One of the biggest problems in treating CVDs is that they are often detected too late. Most current methods—like blood tests, ECGs, or stress tests—are either invasive, expensive, or only used after symptoms appear. This makes early detection difficult, especially in rural or low-income areas.

In recent years, researchers have started looking at the **retina**—the back of the eye—as a way to detect signs of CVD early. The retina has many tiny blood vessels that reflect the health of the body’s overall blood circulation. Changes in these vessels—such as narrowing, twisting, or changes in their structure—have been linked to heart disease, stroke, high blood pressure, and other health problems. Because retinal imaging is quick, non-invasive, and widely available, it could become a powerful tool for early diagnosis.

Many research papers and conference presentations have already shown that it’s possible to use artificial intelligence (AI), especially machine learning and deep learning, to analyze retinal images and predict the risk of CVD. However, there are still some challenges. The accuracy of these models varies a lot depending on the algorithm used, the quality of images, the dataset, and the population it was trained on. Some models work well in one study but fail to perform in real-world settings.

**LITERATURE REVIEW**

Retinal imaging has emerged as a promising non-invasive method for detecting cardiovascular diseases (CVDs). Various research studies have explored the use of deep learning models to extract and analyze features from retinal images for CVD prediction.

1. **H. G. T. S et al. (2024)**  
   In the paper *“Prediction of Cardiovascular Disease from Retinal Images using Deep Learning”* [AISP 2024], the authors proposed a deep learning model based on U-Net and CNN architectures to detect early signs of CVD from retinal images. They focused on segmenting blood vessels and extracting key retinal features. The study highlighted the potential of non-invasive diagnostics but acknowledged the need for improvement in model robustness and accuracy.  
   **DOI:** 10.1109/AISP61711.2024.10870647
2. **V. Willis, B. Zhou, and Q. Liu (2024)**  
   In *“Detection and Prediction of Cardiovascular Disease Using Fundus Images with Deep Learning”* [ICNC-FSKD 2024], the authors utilized CNNs and transfer learning to classify fundus images for heart disease risk. They integrated fuzzy systems and AI to boost predictive performance. This study demonstrated promising accuracy but also raised concerns about dataset variability.  
   **DOI:** 10.1109/ICNC-FSKD64080.2024.10702244
3. **N. D. Bisna, P. Sona, and A. James (2025)**  
   In their IEEE Access paper *“Retinal Image Analysis for Heart Disease Risk Prediction: A Deep Learning Approach”*, the authors introduced a deep learning framework using EfficientNet-B3 and multimodal data fusion. The use of Grad-CAM helped visualize important retinal regions influencing predictions. This work emphasizes explainability and high accuracy.  
   **DOI:** 10.1109/ACCESS.2025.3562433
4. **S. Shaikh et al. (2023)**  
   The study *“Heart Disease Prediction Using Eye Retinal Images”* [ICAST 2023] implemented deep learning models to classify CVD risk from retinal features. Their approach combined machine learning with feature extraction techniques. While results were promising, the study noted limitations in real-time application and generalization.  
   **DOI:** 10.1109/ICAST59062.2023.10454943
5. **N. S and A. T (2024)**  
   In *“Enhancing Cardiovascular Risk Prediction with Vision Transformer and Retinal Image Analysis”* [ICACRS 2024], the authors applied vision transformers along with recurrent neural networks for retinal image analysis. This approach improved performance on noisy or complex datasets and introduced a novel combination of attention-based modeling and CVD diagnostics.  
   **DOI:** 10.1109/ICACRS62842.2024.10841560
6. **P. D et al. (2024)**  
   The paper *“Diagnosis Of Cardiovascular Diseases Using Retinal Fundus Scans Via Deep Learning”* [AMATHE 2024] explored CNN-based classification of fundus scans for heart disease detection. Their focus was on increasing model precision and integrating diagnostic tools. However, the paper also pointed out a lack of publicly available Indian datasets for validation.  
   **DOI:** 10.1109/AMATHE61652.2024.10582150

**EXISTING SYSTEM**

Most existing systems that aim to predict cardiovascular diseases (CVDs) using retinal images rely heavily on traditional convolutional neural networks (CNNs). These models are often limited to 10 to 17 layers and primarily use pre-trained architectures such as VGGNet, ResNet, or EfficientNet. While transfer learning offers decent performance, it restricts the model’s ability to fully understand retinal-specific features, especially when trained on unrelated image datasets like ImageNet.

These systems have shown promise in identifying general patterns in the retina—such as arteriolar narrowing or vessel tortuosity—but often lack robustness when applied to diverse or real-world datasets. Many of them also neglect fine-grained vascular features due to the depth limitations of CNNs. In addition, their accuracy is further limited by small datasets or poor generalization across populations.

**PROPOSED SYSTEM**

In our proposed system, we aim to improve the accuracy and generalization of CVD risk prediction using retinal fundus images by leveraging **Vision Transformers (ViTs)** instead of CNNs. Unlike CNNs, ViTs are capable of capturing long-range dependencies in an image and are well-suited for understanding global structural and vascular patterns in the retina.

We plan to **train a Vision Transformer model from scratch** rather than relying on pre-trained models. This will allow the model to learn features directly relevant to retinal image data instead of adapting from unrelated image domains. To further improve performance, we will explore the use of **traditional machine learning classifiers** (such as Random Forests or SVMs) on the extracted ViT features to enhance classification performance.

Our dataset will primarily include publicly available retinal fundus datasets such as **DRIVE**, **STARE**, and potentially **UK Biobank** (if access is granted). Data preprocessing will include normalization, contrast enhancement, and vessel segmentation to ensure high-quality input. Throughout the process, we aim to document the training pipeline, challenges, and insights gained from using ViTs in a medical imaging setting.

**Architectural Framework**

The architectural framework for our system is designed to process retinal fundus images and predict cardiovascular disease (CVD) risk using a Vision Transformer (ViT) integrated with machine learning classifiers. The system comprises the following modules:

1. **Input Acquisition**  
   Retinal fundus images are sourced from publicly available datasets like DRIVE and STARE. Additional datasets such as UK Biobank may be used upon request.
2. **Preprocessing**  
   Images undergo several enhancement techniques:
   * **Vessel Segmentation** to isolate vascular structures
   * **Gamma Correction** to improve image contrast
   * **Normalization and Resizing** to match the input size expected by the ViT
3. **Feature Extraction via Vision Transformer**  
   A Vision Transformer (ViT) model will be trained from scratch using the transformers library in Python with TensorFlow. The model learns high-level features from the retinal images by capturing both local and global dependencies in the visual data.
4. **Classification**  
   Extracted features are fed into different classifiers:
   * **Hybrid Random Forest with Linear Model (HRFLM)**
   * **Contextual Transformer Classifier**  
     These classifiers aim to enhance accuracy and reduce overfitting by leveraging ensemble learning and attention mechanisms.
5. **Model Evaluation**  
   The system will be evaluated using:
   * **5- to 10-fold Cross-Validation**
   * **Train-Test Split (75%-25%)**
   * Metrics such as accuracy, precision, recall, and F1-score
6. **Deployment (Future Scope)**  
   A lightweight version of the model could be considered for deployment in clinical settings or telemedicine platforms for real-time CVD risk assessment.

**Project Implementation**

The project will be implemented using the following tools and steps:

* **Programming Language**: Python
* **Frameworks**: TensorFlow, Transformers (Hugging Face), OpenCV, scikit-learn
* **Datasets**: DRIVE, STARE (publicly available), UK Biobank (if approved)

**Implementation Steps:**

1. **Preprocessing**
   * Apply gamma correction to enhance contrast.
   * Perform blood vessel segmentation.
   * Normalize pixel values and resize images for input to the ViT model.
2. **Vision Transformer Setup**
   * Use Hugging Face's Transformers library to create a ViT architecture from scratch.
   * Train the ViT model from the ground up without relying on pretrained weights.
   * Use default hyperparameters initially (learning rate, optimizer), which may be fine-tuned later.
3. **Feature Extraction and Classification**
   * Pass the image embeddings from the ViT to external classifiers.
   * Implement and compare the performance of HRFLM and Contextual Transformer classifiers using scikit-learn or custom implementations.
4. **Model Training and Validation**
   * Use 5–10-fold cross-validation to ensure model robustness.
   * Apply a 75-25 train-test split to benchmark performance.
   * Evaluate using classification metrics.
5. **Result Tracking and Iteration**
   * Log training performance and model metrics.
   * Iterate on architecture and preprocessing steps based on observed accuracy and generalization.

**Results (Expected/Planned)**

The proposed Vision Transformer (ViT)-based model will be trained from scratch using segmented and preprocessed retinal fundus images. The model’s performance will be evaluated using multiple metrics: **Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R² score, F1-score, Precision, Recall**, and **Area Under the Curve (AUC)**.

The ViT model's results will be compared against several established pre-trained CNN-based models (such as EfficientNet, ResNet, and VGG) commonly used in the literature. We expect the ViT to **outperform CNNs in AUC and F1-score**, due to its superior ability to model long-range dependencies and spatial hierarchies in fundus images.

Cross-validation (5 to 10-fold) and a 75:25 train-test split will be used to ensure generalization and avoid overfitting.

**Conclusion and future work**

Retinal imaging has emerged as a powerful non-invasive tool for early detection of cardiovascular disease (CVD), offering the potential to identify systemic microvascular changes. Existing research primarily relies on CNN-based architectures with transfer learning, which, while effective, are limited by fixed pre-trained weights and shallow depth. Our study introduces a Vision Transformer trained from scratch, specifically tailored for fundus image analysis.

Although the project is ongoing, we aim to demonstrate that transformer-based architectures can capture global retinal features more effectively, leading to improved predictive performance. This could represent a significant advancement in AI-driven retinal diagnostics for cardiovascular risk assessment.

* **Multimodal Integration**: Incorporating clinical data (e.g., age, cholesterol, blood pressure) alongside fundus images for more robust prediction.
* **Explainability**: Implementing Grad-CAM or SHAP to interpret which retinal features contribute most to predictions.
* **Data Expansion**: Requesting access to larger datasets like **UK Biobank** to improve model robustness.
* **Real-world Testing**: Applying the model in clinical settings to validate its practicality and effectiveness.
* **Model Optimization**: Experimenting with different transformer architectures and hyperparameter tuning to enhance accuracy.

**Signature of the Mentor**

**Date:**