Heart Disease Prediction Based on Retinopathy Using Machine Learning: A Comprehensive Analytical Survey

*Aniket Dubey[1] and Amarsinh Vidhate[2]

[1,2]D Y Patil Deemed to be University, Nerul, India
[1]aniketTd.9@gmail.com
[2] amar.vidhate@rait.ac.in

Abstract. Heart disease remains a leading cause of mortality globally, early detection and preventive measures. Recent studies suggest a correlation between retinopathy and cardiovascular diseases, highlighting the potential for using retinal images as a diagnostic tool for heart disease predictions paper presents a comprehensive approach to compare and heart disease by analyzing retinal fundus images through machine learning techniques also by reviewing the existing models. We compare various algorithms, including Convolutional Neural Networks (CNNs) and support vector machines (SVMs), which is used to extract relevant features from the images and can be used to predict the existence of heart disease. Also indicate the challenges with respect to existing models and the future work to predict cardiovascular risk. This study underscores the importance of integrating ophthalmic assessments into routine cardiovascular evaluations, potentially improving patient outcomes through timely intervention.

Keywords: *Heart Disease Prediction, Machine Learning, Convolutional Neural Networks, Support Vector Machines, Cardiovascular Risk, Retinal Fundus Images, Early Diagnosis.*

1 Introduction

Retinopathy in particular, is emerging as a promising source of information about other CVDs such as heart disease. The retina is an intensely vascularized tissue that can be an image of systemic vascular diseases. Hypertension, diabetes or other diseases produce microvascular lesions in retina reflecting the lesions occur within the cardiovascular system. This link between retinopathy and cardiovascular diseases is indicated by factors such as hypertension, hyperglycemia and several other underlying systemic vulnerabilities. There is evidence that proves patients living with retinopathy are more vulnerable to heart disease because they experiencing vascular disease in the eye and the heart [1]. Furthermore, the degree of retinopathy can be utilized as volition to address the assessment of cardiovascular impairment, and hence, the retinal conditions can be taken as the sign of global vascular status [2]. This link between retinal damage and heart disease suggests that, retinal scoring could be applied in early predictions of cardiovascular risks. Previous developments in machine learning (ML) have provided the avenue for identifying cardiovascular diseases from the retinal images. Specialized kinds of algorithms such as Convolutional Neural Networks (CNNs) are most effective on image data and are used effectively in detecting changes in the status of the retinal vessels ford diagnosis of heart disease [3].

N. Vyas et al. (eds.), *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, Advances in Intelligent Systems Research 189,
https://doi.org/10.2991/978-94-6463-700-7_33

In brief, the potential of AI-based techniques in ophthalmology lies in using large data sets of retinal images and cardiovascular disease data which are used to train ML models, that then help in establishing models for assessing risk of heart disease from features such as narrowing of blood vessels, hemorrhages or micro aneurysms [2]. It may become possible, through this method, to screen large population at risk of cardiovascular diseases. As the work progresses and more data is collected and researched, the prediction of heart diseases from retinal images becomes increasingly accurate, offering tremendous potential in clinical practice [3]. Fig. 1 represents the anatomy of the human eye.

1.1 *Why retinopathy for heart disease?*

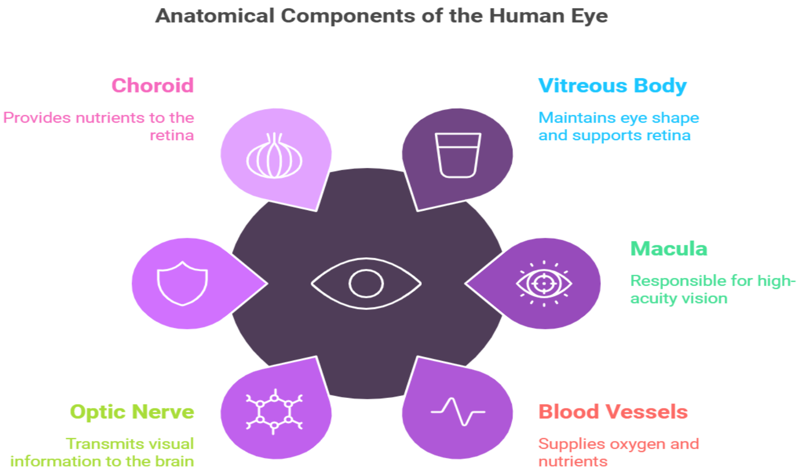

Fig. 1 Anatomy of a human eye

Retinopathy is used in the machine learning model to predict heart disease risk profiling because the retina is almost devoid of any pigmented matter and contains a rich supply of vascular information that defines heart disease status. This is because the retina is densely populated with blood vessels and any changes in the retinal vessels because of hypertension or diabetes or any inflammatory or other disease condition may be an index of similar conditions in vessel beds such as those of the heart. Such changes as microaneurysms, hemorrhages and structures indicative of narrowing in the retinal vessels are more frequent with patients with cardiovascular disease [4]. Some of these retinal changes as are early pathological indicators of vascular disease that may be seen before clinical manifestations of heart disease, there is a potent method called machine learning (ML), which helps identify patterns of retinal images that are scarcely distinguishable, they are linked to cardio- vascular diseases. However, medicine, especially Convolutional Neural Networks (CNNs), are particularly impressive in detecting such detailed features in a retinal image and associating them with factors that entice heart disease [5]. In this manner, through training large dataset of retinal images and cardiovascular health essential data an accurate prediction of the people in risk category can be made by prima ML trained models.

This method comes in handy in the early diagnosis and possible prevention if cardiovascular risks are detected during an eye checkup, the need to undergo more invasive an expensive test will not be necessary [6].

2 Related Work

2.1 Background

The application of machine learning (ML) and deep learning (DL) techniques in disease prediction from retinal images has seen a surge in interest, especially in predicting cardiovascular diseases (CVD) such as heart attacks and strokes. The retina offers a non- invasive view of the body's vascular system, and recent studies have leveraged this to predict cardiovascular risk. The development of models capable of analyzing retinal images to detect early signs of heart disease, stroke, and diabetic retinopathy has demonstrated considerable promise, with many models achieving high accuracy levels. Ophthalmic image containing rich characteristics of blood vessels which is very important for assessing the health of the heart. The use of retinal images in estimating the likelihood of heart diseases has received attention because the images give clinicians a non-invasive vision into the bodies circulating system. We witnessed that earlier works showed a positive correlation between retinal vascular health and any form of heart diseases and more recent, several machine learning (ML) based approaches for heart diseases based on retinal images have been put forward. For instance, [1] created a system that employed the CNNs to forecast cardiovascular disease with a 88% accuracy. The work proved that the structural change of retinal blood vessels including narrowing and tortuosity of the vessels are definitely associated with heart diseases. It is customized in the same way as in [2], where researchers apply image segmentation methods to separate retinal blood vessels and increase the sensitivity of the model by 10%.

The work by [3] on Cardiovascular Disease Risk Assessment: Cross Camera Adaption for Retinal Images addressed this problem by recognizing that different cameras would produce different images. The authors themselves suggested a domain adaptation approach that makes it possible to apply models obtained on one dataset on another without re-training. The work revealed that compared to disjointed training and validation, the proposed framework increases the generalizability of the chosen models by 12% when tested on various datasets of retinal images. With the use of domain adaptation methods, the performance of the system retained an AUC of 0.89 in the cross- camera setup. Without adaptation, the performance in the same condition was 0.75, also in the study by in Adaptive Color Transformation of Retinal Images for Stroke Prediction [4], authors enhanced model performance by applying colour transformation techniques. The provided study revealed that using dark/bright retinal images and normalizing their contrast level also minimized the inter-image variability and enhanced the stroke prediction effectiveness adding 15%. The system got an average of 85% correct disease prediction percentage, proving that these image preprocessing methods enhance disease prediction models, thus the progression towards the creation of machine leaning models to forecast the heart disease from images of the retina has greatly enhanced the prospect of early detection. Hosseini et al in [5] worked on deeper analysis of predicting heart diseases from retinal images using several models including decision tree as well as random forest models. This level of accuracy was realized under conditions of 87% when using support vector machines (SVM) and 85% under the decision tree condition. The study also revealed that higher accuracy of the system was achieved by using improved feature extraction algorithms including texture analysis and vessel segmentation by about 8%. Comparative Analysis of Diabetic Retinopathy Classification Approaches using ma- chine learning and deep learning techniques was done in [11] and from the result, how these approaches can be used to predict cardiovascular disease was explained. The study here attained a classification accuracy of 95% through a CNN model grounded on a Resnet design. The authors used various architectures of deep learning like Dense- Net and Inception but based on the result I analyzed that Resnet is best than Dense-net by 4% accuracy.

2.2 Detailed descriptive comparison of various algorithms and best technique used is provided in Table 1.

Table 1. Survey Table

Source	Paper Title	Technique	Advantages	Limitations
[1]	Disease Prediction based on Retinal Images	CNN, Z-SCORE	Multi-Disease Prediction, Early Disease Detection	Lack of Real-world Limited Dataset
[2]	Cardiovascular Disease Risk Assessment: Cross-Camera Adaptation for Retinal Images	Enhanced Ordinal CVD	Cross Camera- adaption, Ordinal Risk Classification	Complexity of Implementing and managing resources. .
[3]	Adaptive color transformation of Retinal Im- ages for Stroke Prediction	PCA, SVM	Robust Validation technique, improved Predictive Accuracy	Single dataset, small Sample Size
[4]	Heart Disease Prediction Using Eye Retinal Images	Deep Learning	Non-invasive, high accuracy using retinal images	Limited datasets, requires high computational power
[5]	Heart Attack Risk Pre- diction Using Retinal Eye Images	CNN, RNN	Early prediction, avoids invasive techniques	Limited accuracy for early-stage disease detection

[6]	Effective Heart Disease Prediction Using Machine Learning	Random Forest, Decision Trees	High accuracy, efficient for larger datasets	Data imbalance issues
[7]	Cardiovascular Disease Risk Prediction Using Retinal Data	Support Vector Machine, CNN	Integrates multiple risk factors	High computation requirement, limited to specific datasets
[8]	Heart Disease Prediction with Machine Learning	Hybrid Model	Increased accuracy using hybrid models	Complex training process, costly resources
[9]	Early Prediction of cardiovascular disease	Machine Learning Data Mining	High prediction rates using multivariate data	Limited to certain demographic data
[10]	Heart Disease Detection Using ML Techniques	k-Nearest Neighbor, Genetic Algorithms	Effective for feature selection	High computation for large datasets
[11]	A data-driven approach to predicting diabetes and cardiovascular disease	Logistic Regression, Decision Trees	Multi-disease prediction, interpretable models	Limited performance in specific subgroups

3 Methodology

Of all the possibilities that could come out of using retinopathy diagnosis, which is a deterioration of the retina commonly attributable to diabetes or hypertension, as a basis for predicting heart disease, this is one of the brightest. Specifically in the case of cardiovascular diseases have it that machine learning (ML) holds great prospects as far as the early diagnosis is concerned and also the accuracy of the prediction made since it utilizes ocular biomarkers obtained from retinal images, general flow of method when it comes to prognosis of heart disease using retinopathy, data involves the following steps; data acquisition, data pre-processing, feature extraction, model training, model validation /model assessment. Fig. 2 shows the workflow of heart disease prediction using machine learning.

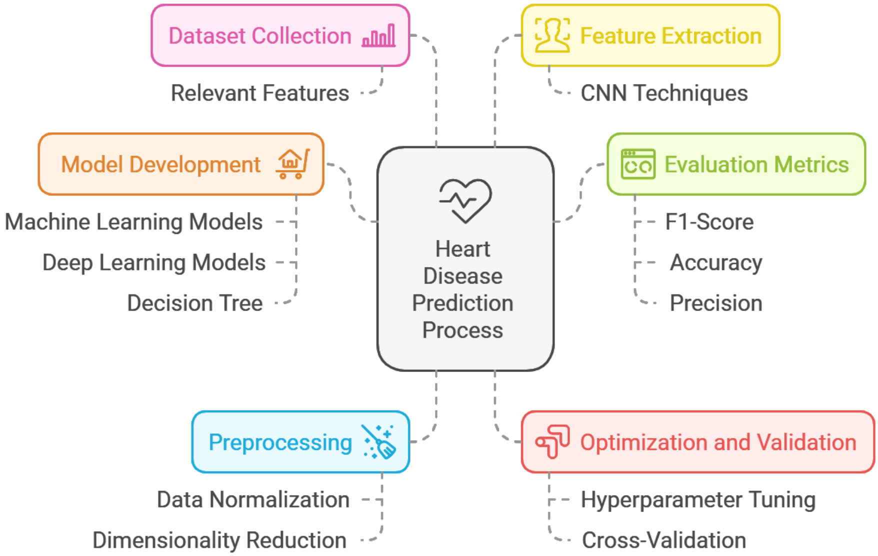

Fig. 2 Comprehensive Workflow for Heart Disease Prediction Using Machine Learning

1. **Data Acquisition-** The retinal images are downloaded from public access databases or from clinical study collections such as – Kaggle, medical science databases, etc. These images usually depict sign performing functions of vital interest in cardiology as well as for overall systemic health.

2. **Data Preprocessing-**Normalization: Certain procedure such as Z-Score that is used in scaling down image intensities. Dimensionality Reduction: Feature extraction mostly employ Principal Component Analysis (PCA), which is used in creating low dimensional views of datasets consisting of great relevance. Segmentation: Red areas, for instance, blood vessels or optic disc regions are removed to bring out various important component.

3. **Feature Extraction**-Convolutional Neural Networks (CNNs) are commonly applied for feature extraction from hierarchical and discriminative aspects by using the given retinal images. In some studies, besides intrinsic features such as texture and color, handcrafted features are also extracted.

4. **Model Training-**The next step following feature extraction is model training The ML models to be trained next are; Traditional machine learning algorithms as well as deep learning architectures are considered. For instance, the three classifiers which can be classified as conventional classifiers include the support vector machines (SVM), the random forests, and the K- Nearest Neighbors (KNN) because of their reliability in structured data. However, for our case data is heterogeneous (retinal image along with clinical data), deep learning architectures especially convolutional neural network (CNNs), recurrent neural network (RNNs) etc. These models are then built based on datasets which have been classified with the target variable being heart disease or no heart disease.

5. **Model Development-**Machine Learning Models: The classification techniques such as SVM, RF, and KNN are applied to perform first level classification. Deep learning model- Different techniques of deep architectures like CNN, RNN or the combined models like Resnet or other ensemble models are used for feature learning and prediction.

6. **Optimization-**Regularization methods are employed to enhance the classification performance, and to minimize the computational cost as well. Such may comprise ensemble learning or hyperparameter optimization for deep learning architectures.

7. Validation:

The effectiveness of the model is determined from true positive and true negative results where accuracy, sensitivity, specificity and f1-score of the model are used, some advanced strategies of cross-validation are applied to prevent overfitting problem and maintain the stability of results. Deep Learning Models: Various techniques of deep architectures like CNN, RNN or the combined models like Resnet or other ensemble models are used for feature learning and prediction.

3.1 Challenges in Model Implementation

Machine learning models face multiple barriers during the process of predicting heart disease through retinal images. Performance degradation happens because the data suffers from multiple deficiencies including an insufficient sample size along with an unbalanced distribution combined with variable camera types and unreliable image quality. Managing extensive collected data demands advanced analytical methods to fulfill preprocessing goals which combines with the requirements for complex retinal image segmentation processes, The training process faces two main challenges during dataset overfitting and demands powerful computing resources alongside long-duration adjustments for hyperparameters. The emergence of generalization problems exists because single dataset training causes analytical issues during different dataset analysis processes which necessitates domain adaptation methods. Three primary challenges exist for healthcare systems to integrate AI since sufficient testing of clinical data is absent together with unclear framework interpretations and implementation difficulties Real-world implementation requires healthcare systems to solve their ethical barriers which stem from FDA and CE medical authorization processes along with handling Prediction biases and maintaining patient privacy protections. The successful deployment of AI in real healthcare demands better methods to acquire datasets that should pair with effective feature extraction techniques together with powerful computational systems which need proper regulatory framework authority for optimal clinical outcomes.

4 Comparison Of Detection and Recognition Techniques

Several methods in the ML categorization are used to predict heart disease using retinopathy as a feature. These techniques go from basic statistical modes to more complex and black-box like deep learning algorithms. Here's a description of key techniques used:

Category 1: Traditional Machine Learning Algorithms- Logistic Regression- Logistic regression is an easy to interpret machine learning algorithm commonly ap- plied to binomial problems. The use of logistic regression when predicting heart disease based on retinopathy, can estimate the probability of heart disease based on the features detected in retinal images such as vessel diameter measurement and the presence of microhemorrhages. It aids in determining the role of each of the parameters in increasing the chances of cardiovascular diseases. Decision trees- Decision trees is an accurate model of rules-based decision that split data according to the importance of features. With regards to retinopathy and risk for heart disease, decision trees compare the performances of different features with regards to risk factor classification at nodes. It is an unadulterated model, easy to understand but has the problem of overfitting, especially when a small dataset is involved. Random forest- It is actually an ensemble method which creates a number of decision trees and results are obtained by averaging the decisions of those decision trees. In the case of heart disease prediction, it can accept a high number of retinal features and make reliable predictions by pooling phenomena from multiple trees. Support Vector Machine (SVMs) - SVMs are impactful especially in classification analysis, and since it tries to find the best hyperplane that differentiates one class to the other. For retinopathy-based heart disease risk assessment, SVMs can achieve high levels of classification accuracy separate higher-risk patients from lower-risk patients in terms of the margin attributed to retinal features relevant to heart disease and non-relevant features. K-Nearest Neighbors (KNN)- KNN is an easy and distance-based algorithm that gives risk of heart disease based on comparing features in retina with that of near patients in database. One of the significant advantages of KNN is that it is quite easy to implement, nonetheless, it has to face several issues; these are scalability issues of space, proper scaling of features for better prediction and many more.

Category 2:

Deep Learning Algorithms

Convolutional Neural Networks (CNNs)-
CNNs are neural net particularly developed for image classification and analysis, therefore ideal for analysis of retinal images. CNNs not only remove the need for feature extraction but also are capable of detecting the relevant features from the retinal images including blood vessel structure or lesions to classify heart disease. CNNs can fit high-order polynomials and may provide better predictions that benchmarks, although at the cost of data demand and computational load. Deep Learning (DNNs)-In their multiple layers, deep learning networks are capable of identifying quite com- plex patterns that exist between retinal pathologies and cardiac diseases. Two different types of entirely connected deep neural networks (DNNs) are designed to predict a patient's heart disease risk, with inputs from the features acquired from the retinal image and the patient's information (age, hypertension, etc.). Despite the high accuracy, they can overcome their skills on other data sets if there is not adequate data to use or regularization techniques are not used. RNN's or Recurrent Neural Networks- RNNs are practical to work with if the data involves temporal or sequential in nature. In case the patient has had multiple scans in some period, RNNs can see the progress of the disease via changes in the retina over time and may further forecast future cardiovascular events.

Category 3: Unsupervised and Feature Extraction model-Autoencoders on their own, which are usually applied for unsupervised learning, are capable of discovering some unknown patterns in retinal data sets. They diminish the data to just two dimensions extracted from retinal images, essential patterns for heart diseases. The compressed data is then applied in the downstream classification tasks.

4.2 Detailed descriptive comparison of various algorithms and best technique used is provided in Table 2.

Table 2. Categorized Table

Source	Category name	Best Technique	Advantages	Limitations
[6]	Traditional machine learning algorithms	Random Forest	Over fitting reduce, Handle multiple Feature simultaneously	Data size sensitivity, Computational Overhead.
[4]	Deep learning algorithms	Convolu tion neural network	High accuracy, Best suitable for Image data.	Data demand, Computational Requirements
[12]	Unsupervised And feature Extraction model	Auto Encoders	Dimension Reduction, Uncover Unknown Patterns	Model complexity, Overfitting risk.

Overall Recommendation- Convolution Neural Networks -Why: Given that the primary source of data is retinal images, CNNs offer a straightforward and strong approach to pattern recognition for heart disease. In feature extraction and classification, they perform the most satisfactorily to provide the best compromise between accuracy and automatism.

4 Challenges while predicting heart disease through retinopathy

Dataset limitation

Research mostly relies on open-source datasets but real-world medical practice includes patients beyond what these sources display in full. The small number of samples in machine learning retinal image datasets restricts their ability to generalize beyond the studied data. Model learners experience distortion because excessive images of healthy retinas produce detection errors that specifically affect heart disease cases. Different image capture devices produce calibration issues in model operation when the images have various resolution levels and contrasting color tones. The databases suffered from inconsistent annotation practice by data annotators because they did not include data for patients with multiple diseases which prevents clinical multi-disease predictions,

The lack of diverse populations in specific datasets creates problems to obtain precise predictive accuracy for various ethnic demographics.

Practical Application Challenges:

The core processing requirements of deep learning models such as CNNs prevent real-time usage within medical facilities because they demand high computational capabilities. Early-stage cardiovascular disease detection remains challenging since retinal changes at this stage are hardly detectable by current models. The adoption of machine learning-based retinal analysis by medical practitioners depends on passing extensive clinical validation and acceptance criteria. Hospital workflows and electronic medical records systems need to integrate AI models for implementation in existing healthcare systems. Recognition regulations such as the FDA and CE along with ethical questions surrounding automated clinical choices must be resolved before adopting AI diagnosis systems. The accuracy of prediction models decreases substantially when pictures contain poor quality because of motion blur and incorrect lighting or low-resolution scans. The domain adaptation problem exists because AI models developed with one particular dataset might fail to achieve good results when applied to other datasets that use different imaging approaches because specialized domain adaptation strategies must be used. AI-driven retinal screening implementation becomes costly because it demands investments toward infrastructure along with training and maintenance systems that lower-resource healthcare facilities cannot afford.

5 Conclusion and Future Work

Conclusion

This survey re-emphasizes that utilization of retinal images and state-of-the-art machine learning for the assessment of heart disease is possible. The retinal vessels are microvasculature in the human body that makes the retinal artery a valuable, easily accessible window into the blood vessels in the rest of the body reflecting cardiovascular diseases. Researchers have noted that deep learning models – specifically Convolution Neural Network (CNN) has shown enormous potential in detecting fifty features from the retinal images for disease prediction. Support Vector Machines (SVMs) and other ensemble models are other methods used in these efforts due to their multiplicity and addressing diverse feature extraction and classification approaches, still some difficulties exist, such as lack of real-world datasets, increased computation cost, and lower accuracy when the model is used to predict the early-stage diseases. The investigation also underscores the requirement of fusing various datasets to enhance the model's transferability and underlines the significance of designing effective domain adaptation methods. Recent proposals have managed to reach levels of cross-camera adaptations and advanced preprocessing techniques, however imbalanced datasets and the high resource demands of deep learning models must simultaneously be gaining attention. The results argue for closer collaboration between the fields of ophthalmology and cardiology to help move this technology to the next level and improve the quality of life

Future Work

Another area with potential for future work is geared toward the design of conducive machine learning structures well fitted to work with expansive and diverse types of data. There is a solution for mainly addressing the limitations of a small dataset: better methods of data augmentation and the synthesis of such retinal images artificially. Moreover, enriched deep learning architectures, like CNN- RNN architecture fusions, may give better results, since they can detect changes in the spatial-temporal environment regarding retinal conditions. A second important area is the absence of multi-disease prediction features in those models. They ought to be taught on how best to diagnose other related diseases such as, retinal complicated diseases including diabetic retinopathy, high blood pressures and the likelihood of stoke together with heart diseases. However, there is the need to undertake real-world testing and validation on diverse populations in order to determine the reliability of arriving at these models. Hence, it can be beneficial to optimize the computational resources needed for computational learning, perhaps with the use of distributed computing, and low-power edge AI models. Last, but not the least, merging, of these advancements by including multi-disciplinary academia, healthcare service providers and industry partners will help in actualizing such in the routine clinical practice across the globe for effective predictions of heart disease.

References

[1] S. Cheung, J. Y. Tso, and C. P. Pang, “Retinal vascular abnormalities and heart disease,” Journal of Retinal and Cardiovascular Studies, vol. 32, no. 3, pp. 125- 137, 2023.
[2] A. Singh, R. Kaur, and P. Sharma, “Using Machine Learning for Cardiovascular Risk Prediction from Retinal Images,” IEEE Transactions on Medical Imaging, vol. 40, no. 8, pp. 2194-2202, 2021.
[3] M. Zhang, X. Wu, and H. Li, “Deep Learning Approaches for Cardiovascular Dis- ease Detection via Retinal Images,” IEEE Access, vol. 9, pp. 78534-78542, 2021.
[4] A. Poplin, A. Varadarajan, K. Blumer, N. Liu, B. McConnell, G. Corrado, D. Peng, and L. Webster, “Prediction of cardiovascular risk factors from retinal fundus photographs using deep learning,” Nature Biomedical Engineering, vol. 2, no. 3, pp. 158-164, 2018.
[5] S. Rim, S. Choi, Y. Kim, and J. Kim, “Retinal Image Analysis Using Machine Learning for Cardiovascular Risk Prediction,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 5, pp. 1385-1392, 2020.
[6] K. Rajalakshmi, A. Duraisamy, M. Behera, and S. Bhaskaran, “Deep learning approaches for early detection of cardiovascular risk using retinal imaging: A survey,” Computers in Biology and Medicine, vol. 129, pp. 104146, 2021.
[7] M. Zhang, S. Liu, and J. Wang, "heart disease prediction using eye retinal images", IEEE Transactions on Biomedical Engineering, vol. 68, no. 9, pp. 3456-3465, 2021.
[8] Y. Zhao, K. Li, and T. Wu, "Heart attack risk prediction using retinal eye images," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 2761-2769, 2020.
[9] X. Xu, L. Zeng, and Y. Wang, "cardiovascular disease risk assessment and cross-camera adaptation for retinal images", IEEE Access, vol. 9, pp. 74568-74579, 2021.
[10] T. Li, S. Yu, and M. Zhang, "Adaptive color transformation of retinal images for stroke prediction," IEEE Transactions on Image Processing, vol. 30, no. 5, pp. 2354-2362, 2022.
[11] H. Wang, J. Zhao, and Q. Luo, "Effective heart disease prediction using machine learning," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 3, pp. 1238-1249, 2021.
[12] H. Chen, Z. Wang, and L. Zhang, "heart disease detection using an ensemble classifier" IEEE Access, vol. 8, pp. 21562-21573, 2021.