Challenges in Information, Communication and Computing Technology – V. Sharmila et al. (Eds)

Demystifying heart disease prediction: The role of iris imaging and explainable AI

S. Geetha, B. Deepa, I. Sivaranjani, M. Sowmiya, P. Karthick & N.K. Priyadharshini
Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India

ABSTRACT: Iridology is a fairly new scientific discipline interested in the study of the beauty-contact areas of the body that are termed the iris. It can even assess problems with a variety of biological activities. Since the nervous system and brain connect the body's tissues and organs, the condition of each organ may be immediately revealed through the iris, and in this way, the cardia characteristics were retrieved through eye and feature expression and image pre-processing methods. In the system proposed, Explainable AI is now combined with machine learning techniques like K nearest neighbor and the random forest algorithm. If they unite, they will feel even stronger, more efficient, and self-reliant. With this component, estimating the probability of coronary artery disease becomes possible.

Keywords: Iris, Iridology, Coronary Artery disease, random forest, Image pre-processing & Explainable AI

1 INTRODUCTION

Each of the iris's most complicated tissue structures is the focus of iridology research. Through the analysis of iris patterns this method evaluates the state of the body' systems and organs. It can be applied to identify problems with many biological functions. The iris may be used to immediately assume the condition of each organ since the nervous system and brain connect the body's tissues and organs to the iris. To extract cardiovascular properties from the eye feature extraction and pre-processing were performed. In this system, Machine learning algorithm approaches Random Forest algorithm and KNN are used with Explainable AI.

2 LITERATURE REVIEW

One of the most prevalent illnesses in the modern world is cardiovascular disease (high-definition), that can be caused by a wide range of factors, such as overbearing blood sugar, type 2 diabetes,varying cholesterol, sleeplessness, and a number of others [Almustafa (2020)]. When it breaks down, our entire body may become polluted, and this illness typically results in death. Worldwide, here was a 41% increase in coronary artery disease from 1990 to 2013 [Awan *et al.* (2018)]. These factors include high blood pressure, gender, age, chest pain factor, and the Electrocardiogram (ECG), which measures heart function and is frequently utilized to build prediction models [Chen and Henigjinda (2021)]. High blood cholesterol levels that surpass optimal parameters a blood cholesterol level of less than 200 mg/dL is ideal can cause serious illnesses like strokes and heart attacks [Daniel *et al.* (2019)]. Image processing methods, particularly image segmentation, can be used to detect heart through the iris [Fadilla *et al.* (2022)]. This is because the most effective ways to lower the cardiovascular risk in

DOI: 10.1201/9781003559085-70

people with symptomatic PAD revascularization or medication management included—remain little recognized [Jones *et al.* (2017)]. The viewpoints of medical studies and gathering data are applied to identify various metabolic disorders [Lakshmanarao *et al.* (2019)]. The multidimensional perceptron (MLP for short) mapping look at offers better precision up to 6.62%, while multiple linear regression increases AGSV, as per the results [Ozbilgin and Kurnaz (2022)]. Tests that include tests for blood, electrocardiogram (ECG), effort tests, Holter tests, and cardiograms (the organization) are often used to identify CAD technology [Özbilgin *et al.* (2023)]. In medical fields, improved picture processing and data mining methods are employed as an useful disease diagnostic tool [Sarnant and Agarwal (2018)]. In order to determine cases of autonomic damage to nerves based on iridology, more research utilizing neural networks that are artificial has been conducted [Yohannes *et al.* (2020)].

3 RESEARCH METHOD

3.1 *Iridescence*

Iridescence is the study which utilizes the human iris as a diagnostic apparatus. The iris is utilized in medicine to measure how well the human organs are functioning. Each of the 60 segments that make up iridology's segmentation the organ's status or operation is depicted by the eyeball. The functioning of the body's left organs is indicated by the state of their respective left iris, while the health of the right organs is shown by the status of the right iris. Figure 1 shows the display the intended system's flow.

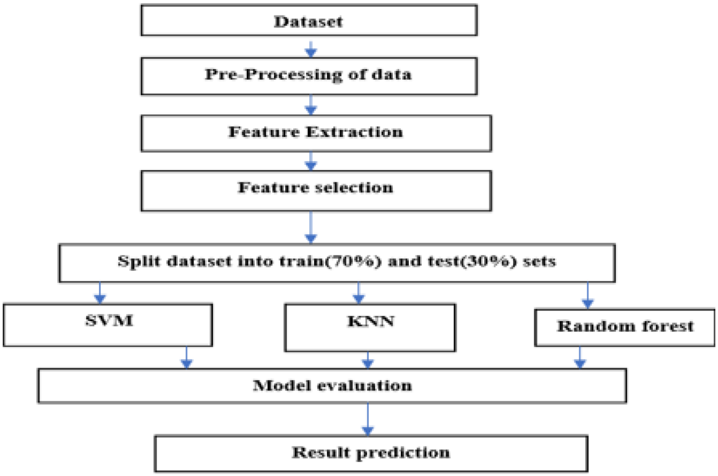

Figure 1. Display the intended system's flow.

3.2 *Load dataset*

Acquiring and arranging pertinent data for analysis is the process of loading the dataset in an iridology project. For the purpose of researching relationships between features of the eyes and medical disorders, the collection probably include iris photos or associated data. Essential stages involve monitoring data quality, cleansing the dataset, and accounting for factors including age, gender, and medical history. In the realm of iridology, this preliminary work establishes the framework for further investigation and pattern identification. Find an appropriate dataset's source. Typically, numerous types of analysis are conducted using distinct datasets for heart disease and Iris. Health related features are included in datasets for heart disease prediction, whereas the Iris dataset is frequently used for classification problems pertaining to iris emerge.

3.3 *Data preprocessing*

In ML loading dataset is the first step, then the data pre-processing refers to the methodical cleaning, modification, and arrangement of unprocessed data in order to get it ready for training a machine learning model. This is particularly relevant when predicting heart disease using the Iris dataset. Improving the quality of the data is the purpose of this process in order to make it suitable for efficient model training and prediction. In this pre-processing, managing missing values, encoding categorical variables, scaling features, and resolving any imbalance in the data are important processes. Using the features of the Iris dataset, the objective is to provide a new dataset that best supports the selected machine learning algorithm for precise and trustworthy heart disease prediction.

3.4 *Feature selection*

The goal of feature selection is to create graphs that, a machine learning algorithm used from the previously processed images. To use the Random Forest approach, features must be extracted from the pictures. Each vector value or phrase in the graph is counted on the quantity of times it appears. Feature selection can help the model perform better by reducing the number of dimensions in the iris.With the use of iris technology, certain characteristics pertaining to the iris's structure may be identified as potential heart disease indicators. Selecting iris features, common procedures involve determining which aspects are most connected with heart disease, removing less relevant ones, and assessing the statistical significance of each feature. This procedure contributes to the development of a more precise and effective model for iris-based heart disease prediction.

3.5 *Normalization*

The iris dataset can be used to predict heart disease, while normalization is necessary for machine learning models to produce effective predictions. The model's performance may be enhanced by ensuring that all features have a comparable scale. If iris technology is used to predict cardiac disease, the process of normalizing the iris data to a standard range is called normalization. Due merely to their greater size, certain variables are thus prevented from controlling the prediction model. On the other hand, it ensures that the various elements, such patterns or dimensions, gathered from iris scans are consistent in size.

3.6 *Region of interest*

To diagnose the heart through the iris, some elements are required; therefore, processing the complete image is not required. By using iridescence, the left-hand portion of the iris's cardiac projection can be observed in the area between 02.00 and 03:15. In the Region of Interest (ROI) technique, only a portion of the photograph needs to be processed. The image of the heart reflected by the left eye's iris stays after all of these parts have been eliminated used to support labelled examples into the model, allowing it to learn patterns and relationships (Figure 2).

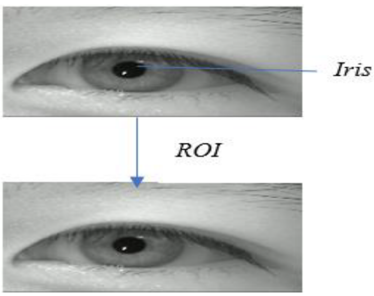

Figure 2. ROI of the iris.

3.7 *Training and testing*

On average, the iris dataset is utilized for developing a machine learning model which utilizes training and testing data to predict cardiovascular disease. The training data is unseen circumstances will be predicted in order to assess the model's effectiveness and see how well it generalizes to new data.

4 PROPOSED METHODOLOGY MACHINE LEARNING MODELS

Making use of many different machine learning approaches, such as randomized forest computing and k least cousins. Train the models on the pre-processed dataset to learn patterns indicative of heart disease. Common split ratios are 80-20 or 70-30 for training and testing.

4.1 *KNN (K-Nearest Neighbor)*

Regression and classification issues are handled utilizing a k-nearest-neighbors model (KNN), a supervised machine learning technique. When presented with a fresh input for classification, KNN finds the K training set data points that are closest to it using a selected distance metric (usually Euclidean distance). The new instance's class is determined by the majority class among these neighbors KNN (Figure 3) to estimates the target value in regression by averages, and assuming a weighted mean of the values of the closest K neighbors. The selection of K and the distance metric influence the algorithm's performance and vulnerability to exceptions.

$$\text{Accuracy} = \frac{\text{Total Number of Predictions}}{\textit{Number of Correct Prediction} \times 100\%} \quad (1)$$

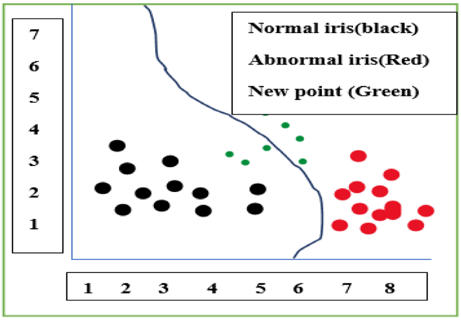

Figure 3. K-Nearest Neighbor.

4.2 *Random forest algorithm*

Random woodland learning is a prominent machine learning technique used in the controlled learning method. It can be employed in both the classification and regression challenges in machine learning. Integrating a number of classifiers to improve the model's accuracy and answer a complex issue has been referred to as combined learning, and it forms the core of this method. Between the ensemble learning techniques that are frequently used to predict heart disease using the iris dataset is the random forest algorithm. The approach is widely recognized for being able to process complex data and generate accurate forecasts. For each choice tree in the forest, a randomly selected portion of the features and data is used for training, and the mean of the dataset's inputs yields the final prediction. Applying the randomized forest method, this precision score contributes in our approximation of cardiopulmonary diagnosis. This accuracy track helps us expect the heart diseases by roughly 0.81 while we use a random forest proximity (Figure 4).

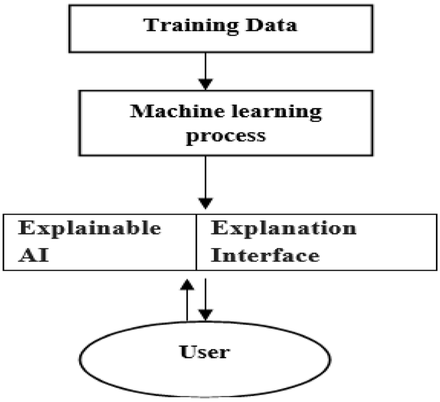

Figure 4. Explainable AI.

4.3 *Model evaluation*

Various measurements are utilized to assess the efficacy of the KNN model, including accuracy, precision, or recall. Enhancing the model's performance can be achieved through modifying the model's hyper parameters or employing a different type of feature selection.

$$F1score = \frac{2 \times Precision + Recall}{Precision \times Recall} \tag{2}$$

4.4 *XAI*

Enhancing the transparency and clarity of the model's decision-making process refers to Explainable AI. Both the Random Forest and k-Nearest Neighbors (KNN) algorithms possess some level of explainability. KNN, a non-parametric with slow learning capabilities categorizes items based on the majority opinion of their closest neighbors, assigning them to the most prevalent group. The explainable aspect of KNN stems from its straightforward approach, which bases a data point's outcome on the classifications of its nearest neighbors. Therefore, to facilitate decision-making in the classification process, it is essential to identify and exhibit the closest neighbors along with their associated labels. Therefore, to initiate the decision-making phase in classification, it is crucial to recognize and showcase the nearest neighbors along with their associated labels. To illustrate the Random Forest decision-making process, one must consider the role of each feature in the classification procedure. Typically, the reduction of impurity is proportionate to the contribution of a specific feature. While both Random Forest and KNN provide some level of explanation, the manner in which their decisions are clarified varies. Random Forest relies on assessing feature importance, whereas KNN's explanation is more straightforward, relying on the classification of nearest neighbors.

5 RESULT AND ANALYSIS

In the iris dataset, the random forest method and K-nearest neighbors (KNN) are applied to predict heart disease. The performance of the models, which includes accuracy, precision, and recall, is evaluated based on the results. The random forest algorithm involves a collection of selected trees that yield accurate forecasts, while KNN depends on the nearby majority class. The analysis consists of evaluating the strengths and weaknesses of the models, identifying significant features that impact predictions, and comparing the effectiveness of random forest and KNN in predicting heart disease using the iris dataset. This is achieved through the use of Explainable AI for clear interpretation.

6 CONCLUSION AND FUTURE WORK

To sum up, the use of random forest and K-nearest neighbors (KNN) alongside the iris dataset offers valuable insights into predicting cardiac disease. Both models demonstrate strengths in heart disease prediction, with random forest providing robustness through ensembles and KNN relying on proximity for classification. The selection between the two models depends on specific needs, and further adjustments could improve predictive accuracy. Furthermore, the use of Explainable AI methods enhances model interpretability, highlighting the crucial features influencing predictions. This study sets a strong foundation for informed decision-making when applying machine learning methods to forecast heart disease. Future research exploring various organs, such as brain tumors and kidney deficiencies, holds significant potential for advancements.

REFERENCES

Almustafa, K. M. *et al.* (2020). Prediction of heart disease and classifiers' sensitivity analysis; *BMC bioinformatics*, 21:1–18.

Awan, S. M., Riaz, M. U., and Khan, A. G. *et al.* (2018). Prediction of heart disease using artificial neural network; *VFAST Transactions on Software Engineering*, 6(1):51–61.

Chen, J. I. Z., and Hengjinda, P. *et al.* (2021). Early prediction of coronary artery disease (CAD) by machine learning method-a comparative study; *Journal of Artificial Intelligence*, 3(01):17–33.

Daniel, M., Raharjo, J., and Usman, K. *et al.* (2019). Iris-based image processing for cholesterol level detection using gray level co-occurrence matrix and support vector machine; *Engineering Journal*, 24(5).

Fadilla, N. A., Sakti, M. L. P., Setyaningsih, N., Zhafran, N., Pramudyawardhana, T. A., Shelomita, V. A., Fahmi, A. *et al.* (2022). Cholesterol detection through iris using Daugman's and GLCM based on K-means clustering; *Paper presented at the 2022 International Seminar on Application for Technology of Information and Communication* (iSemantic).

Jones, W. S., Baumgartner, I., Hiatt, W. R., Heizer, G., Conte, M. S., White, C. J., Mahaffey and colleagues (2017) conducted a study comparing ticagrelor with clopidogrel in patients who had undergone lower extremity revascularization in peripheral artery disease. *The study was published in Circulation*, *135*, *Issue 3*, pages 241–250.

Lakshmanarao, A., Swathi, Y., and Sundareswar, P. S. S. *et al.* (2019). Machine learning techniques for heart disease prediction; *Forest*, 95(99): 97.

Ozbilgin, F., and Kurnaz, C. *et al.* (2022). An alternative approach for determining the cholesterol level: Iris analysis; *International Journal of Imaging Systems and Technology*, 32(4):1159–1171.

Özbilgin, F., Kurnaz, Ç., and Aydın, E. *et al.* (2023). Prediction of coronary artery disease using machine learning techniques with iris analysis; *Diagnostics*, 13(6):1081.

Pan, Y., Fu, M., Cheng, B., Tao, X., and Guo, J. *et al.* (2020). Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform; *IEEE Access*, (8):189503–189512.

Prakash, K., and Saradha, S. *et al.* (2023). Efficient prediction and classification for cirrhosis disease using LBP, GLCM and SVM from MRI images; *Materials Today: Proceedings*, (81):383–388.

Samant, P., and Agarwal, R. *et al.* (2018). Machine learning techniques for medical diagnosis of diabetes using iris images, *Computer Methods and Programs in Biomedicine*, 157:121–128.

Yadav, D. C., and Pal, S. *et al.* (2020). Prediction of heart disease using feature selection and random forest ensemble method; *International Journal of Pharmaceutical Research*, 12(4):56–66.

Yogeshwaran, R., and Vishal, S. *et al.* (2023). Disease detection based on iris recognition. *Paper Presented at the 2023 International Conference on Energy: Materials and Communication Engineering* (ICEMCE).

Yohannes, C., Nurtanio, I., and Halim, K. *et al.* (2020). *The Potential of Identifying Heart Disease Through Iridology. Presentation Delivered at the IOP Conference Series: Materials Science and Engineering*