DOI: 10.22146/jnteti.v14i1.13658

Analysis of Facial Areas to Identify CHD Risks Based on Facial Textures

Budi Sunarko[1], Agung Adi Firdaus[1], Yudha Andriano Rismawan[1], Anan Nugroho[1]

[1] Faculty of Engineering, Universitas Negeri Semarang, Semarang, Jawa Tengah 50229, Indonesia

[Received: 25 June 2024, Revised: 17 January 2025, Accepted: 6 February 2025]
Corresponding Author: Budi Sunarko (email: budi.sunarko@mail.unnes.ac.id)

ABSTRACT — Early screening for coronary heart disease (CHD) remains insufficiently addressed, underscoring the need for a more effective screening tool. Previous studies have reported a classification accuracy of only 72.73%, which is inadequate. This study aimed to develop and evaluate a machine learning model or diagnose CHD using facial texture features and to compare the performance across different facial regions to provide recommendations for improvement. The research involved constructing a machine learning model that extracted texture features from six facial regions of interest (ROIs) using the gray level co-occurrence matrix (GLCM) and employed an artificial neural network (ANN) algorithm. The datasets were full-face images of CHD patients (positive) and healthy people (negative). The face parts identified were the right crow's feet, right canthus, nose bridge, forehead, left canthus, and left crow's feet. A total of 132 (72 positive and 60 negative CHD) datasets were divided into 80% (n = 106) training data and 20% (n = 26) testing data. The developed model achieved a notable accuracy of 76.9%. The findings revealed that two facial regions—canthus and forehead—demonstrated excellent accuracy of 80.97% and 90%, respectively. Meanwhile, the crow's feet and nose bridge regions showed good accuracies at 73.50% and 65%, respectively. Based on the results, this research has proven to be able to become a model for early CHD screening with good accuracy and faster execution.

KEYWORDS — Coronary Heart Disease, Facial Texture Feature, Artificial Neural Network, Region of Interest.

I. INTRODUCTION

Coronary heart disease (CHD) is a prevalent cardiovascular condition and remains one of the leading causes of mortality worldwide [1], [2]. Despite its severity, there is a significant gap in public awareness and early screening efforts for CHD. Many individuals are reluctant to participate in routine health check-ups due to a lack of understanding, insufficient awareness, and financial barriers. In rural areas, medical resources are often scarce, and even in urban centers, there are substantial challenges, such as high costs, extended waiting times, and other obstacles [3]. As a result, heart disease may advance unnoticed until it reaches a critical stage, potentially leading to severe complications. Therefore, preventive strategies, including early diagnosis, are crucial.

Existing diagnostic techniques for CHD include coronary angiography, electrocardiograms (ECG), blood tests, treadmill tests, computed tomography (CT) scans, and magnetic resonance imaging (MRI) scans [4]. However, many of these methods are costly and less efficient, and some are even invasive [5]. These factors make it challenging for many individuals to access early cardiac screenings, particularly due to economic constraints and the limitations of current technology. There is a pressing need for a more efficient and cost-effective CHD screening tool to enhance public accessibility and awareness of early detection.

Machine learning program offers a consistent, rapid, and accurate method for diagnosing CHD [5]. By harnessing expert capabilities, this system aids in assessing CHD risk and offering effective follow-up recommendations with high precision. This technology, achievable through artificial intelligence, ensures efficient and accurate diagnosis. Existing artificial intelligence research for the prediction of heart disease mostly uses text data in the form of medical records as a parameter for determining decisions. The disadvantage of such methods is their low efficiency. Using facial images can be one solution to this problem [6], [7]. The diagnostic process can be significantly expedited by using facial images, making it a highly efficient tool for early CHD screening.

Each individual's face is a unique bioidentity identifier for humans. It gives information about age, gender, race, consciousness, mood, and health state [8]. The leading theory of this study is the correlation between facial texture and disease status, as evidenced in quantitative and qualitative-based studies [3], [6]–[9]. In the previous study, visualization tests have shown that the forehead and nose are more effective as the region of interest (ROI) for detecting CHD risk than other facial areas [6]. In contrast, the left and right eye regions yielded inconsistent results, and the ear was the least effective. Several studies have shown a strong association between high plasma total or low-density lipoprotein (LDL) cholesterol levels and various conditions such as coronary heart disease (CHD), insulin resistance, diabetes mellitus (DM), hypertension, stroke, dyslipidemia, obesity, and hyperuricemia, especially in individuals with xanthelasma [10], [11]. LDL functions as the main transporter of cholesterol, triglycerides, and other lipids throughout the body [12]. Xanthelasma palpebrarum (XP), which is characterized by black plaques at the inner corner of the eyelids, especially on the upper eyelid, consists of xanthoma cells—foamy histiocytes with intracellular lipid deposits in the upper reticular dermis [10], [12]. Xanthelasma is known as a marker of atherosclerosis, a major cause of cardiovascular disease [12]. In addition, patients with CHD often exhibit certain facial features such as forehead wrinkles, typical balding hair patterns, thickening of the upper eyelids, and ear creases [11], [13]. Other studies have also identified additional facial indicators in CHD patients, including preauricular folds, corneal arches, and acrochordons [14].

Several studies, particularly in China, have explored the application of facial imaging to identify CHD [6], [7]. The related study still used eight parts of the face such as the outer corner of the eye (crow's feet) right and left, the inner corner of the eye (canthus) right and left, the nose, forehead, right auricle, dan left auricle [6]. The performance of the classification model in that study was also classified as less high, at around 72.73%. ROIs with bad performance contributions, such as auricles, were still used [6]. A recent study [4] used six areas of the frontal face to develop several classification models using the outer corner of the eye (crow's feet) right and left, the inner corner of the eye (canthus) right and left, nose, and forehead without auricle. The best model performance was achieved at around 92.8% of the area under curve (AUC). Based on these findings, improvisation is needed to increase the overall performance of the system through ROI selection. According to previous relevant research, not all parts of the face represent or correlate with cardiovascular health conditions, so only a few ROIs correlating positively were selected. Selecting the right ROIs based on facial skin texture can better discriminate CHD patients from healthy individuals [6].

Based on this background, an analysis of the performance of each ROI is required. It is one of the basic theories for enhancing the accuracy of this technology, as the inclusion of components with low accuracies will deteriorate diagnostic accuracies. This research is essential because the selection of parts of the area with high accuracy will optimize the accuracy of the diagnosis system. The novelty given in this study is a modification of the use of ROIs with six locations, which was done by eliminating the ear part. That way, beside the main goal of improving performance, this research can augment the system efficiency. Therefore, the study of facial part selection analysis for basic diagnostic purposes is an effort to improve the performance and efficiency of CHD diagnostic methods by utilizing various facial textures.

This study sought to create and evaluate a machine learning model for diagnosing CHD by utilizing facial images, specifically focusing on the texture features of facial skin. Additionally, it seeks to compare the performance of different facial regions to provide recommendations for enhancing the efficiency and effectiveness of CHD diagnosis applications that rely on facial images. Through this research, a machine learning model can be produced to develop early screening technology for CHD that is very efficient in increasing access to and awareness about CHD in the community.

II. RELATED WORKS

Researchers in China have recently conducted studies on disease detection with artificial intelligence technology, particularly in CHD diagnosis by facial imagery. CHD detection has been developed with a method utilizing eight facial features, encompassing the right side and left side areas of the eyes, nose bridge, forehead, and the two earlobes [6]. This study used the gray level co-occurrence matrix (GLCM) for the features extraction algorithm, then classified them into two classes (positive or negative) using a decision tree and random forest algorithm. However, this method still did not yet achieve a high accuracy, with the highest accuracy recorded at approximately 72.73% when utilizing the random forest algorithm. Higher accuracy is needed to create ready-made medical technology in society. Additionally, the fact that some groups of people consider earlobes to be private renders them ineffective.

Another research is about the feasibility of a deep learning model to classify CHD using the facial image [7]. This study has proven that facial images can help in the prediction of CHD using the deep convolutional neural network and Diamond-Forrester classification algorithm. The ROIs used in this study were the cheeks, forehead, nose, around the eyes, mouth, earlobes, and chin. The highest performance result of this study was around 73% of the AUC for the deep learning method. Further studies are needed to improve prediction performance so that technology is more useful and applicable in the medical world.

Another study developed machine learning models to evaluate the risk of CHD by analyzing facial texture features, utilizing the GLCM algorithm [4]. The study focused on specific facial ROIs, including the outer eye corners (crow's feet), the inner eye corners (canthus), the nose, and the forehead. These features were then classified using support vector machine (SVM), decision tree, and artificial neural network (ANN) algorithms. Six GLCM features—contrast, dissimilarity, homogeneity, energy, correlation, and angular second moment (ASM)—were extracted from the ROI images. The ANN model outperformed the others, achieving the highest performance with an AUC score of 92.8%. Recent research indicates that optimizing a CHD prediction model based on facial analysis requires further investigation to identify the most relevant facial areas with the highest accuracy.

III. METHODOLOGY

This research was conducted at Universitas Negeri Semarang, Indonesia commencing on June 1, 2021, and concluding in June 2022. The datasets were mainly collected from Dr. Kariadi Hospital and Diponegoro National Hospital in Semarang, Indonesia. In addition, datasets were also taken from residents domiciled in Jawa Tengah who matched the inclusion requirements. The tools used in this research included the ASUS Vivobook personal computer with Core i7 and 12 GB RAM processor, iPhone 8 Plus smartphone with a resolution of 12 Megapixels camera, and Fuji XA3 camera. While the software was Anaconda Navigator application version 3, Spyder (Python 3.9), Orange Python 3, Microsoft Excel 2016, and OpenCV. This research methodology consisted of data acquisition, preprocessing, feature extraction, training and testing model, and classification analysis.

A. DATASETS

The inclusion criteria for positive samples comprised Indonesian men and women aged between 20 and 59 years who were positively diagnosed with CHD by cardiologist and were available to support medical documentation. The facial images were taken using devices with a minimum resolution of 13 MP for DSLR or 48 MP for smartphone rear cameras, showing full face views from the neck upwards (front, right, left, and top). The exclusion criteria for positive samples comprised individuals with serious facial diseases or conditions that were visibly apparent, as well as those with blurry or unclear facial images where the facial texture was not discernible. Similarly, for negative samples, the inclusion criteria comprised Indonesian men and women aged between 20 and 59 years, who were negatively diagnosed with CHD by a cardiologist, along with the same image resolution and quality requirements. Exclusion criteria for negative samples are similar to that for

positive samples, disqualifying individuals with visible serious facial conditions or unclear facial photos.

The datasets collected are primary data from researchers in the form of full-face images of CHD patients (positive) and healthy people (negative). The sampling technique used was accidental sampling. Images were selected through inclusion and exclusion criteria formulated with heart and blood vessel specialists so that the data collected were valid. Inclusion criteria are criteria by which the subject of the study can be used as a research sample because it qualifies as a sample. Meanwhile, the exclusion criteria are criteria when the research sample cannot be a sample because it does not qualify as a research sample.

ROIs are areas of the image that will be further processed. Therefore, this study used ROI_m, ROI before preprocessing derived from manual cropping, and ROI_t, ROI after preprocessing derived from segmentation results. Six ROI_m were taken manually for each respondent. The facial parts identified were the outer corner of the left and right eyes (crow's feet), the inner corner of the right and left eyes (canthus), nose, and forehead. The ROIs were in a resolution of 400 × 400 pixels. Figure 1 shows the ROIs used in this research. The total datasets used were 132 images of 60 negatives and 72 positives. The datasets were then divided randomly into 80% (n = 106) as training data and the remaining 20% (n = 32) as testing data. The characteristics of the datasets used are shown in Table I.

TABLE I
CHARACTERISTICS OF DATASETS

Class	Average	Methods of Diagnosis		Gender	
		Angiography (%)	Electrocardiogram (%)	Man (%)	Women (%)
CHD	21-62 (42.5) years	83.3	16.67	36.37	18.19
Healthy	22-36 (30) years	0	100	31.81	13.63

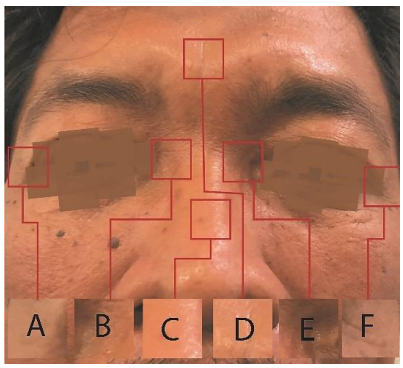

Figure 1. ROIs: (a) right crow's feet, (b) right canthus, (c) nose bridge, (d) forehead, (e) left canthus, (f) left crow's feet.

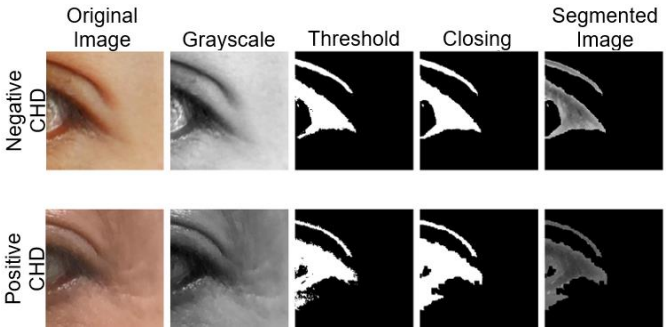

Figure 2. Segmentation of facial texture between healthy respondents and respondents with CHD.

B. PREPROCESSING

Images that were originally in RGB color were converted into grey levels. Starting from the thresholding of each face block. Thresholding used the threshold value of T as a benchmark to determine whether pixel on the image was converted to black or white. The thresholding was done with a 3 × 3 kernel, as delineated in (1) [15].

$$T_h(x) = \begin{cases} x & if\ |x| \geq \lambda \\ 0 & otherwise \end{cases} \quad (1)$$

In addition to skewness and color spacing, grayscale images are also segmented by binary threshold [16]. Visual inspection of the distribution shows that the threshold for the feature can provide a better classification. This research employed hard thresholding by selecting coefficients that exceeded a specified threshold λ and setting all other coefficients to zero. The next operation was the morphological operation of the image through a closing process, which was a combination of dilation and continued erosion. This process was used to cover or remove small holes in the surface. The closing process is useful for smoothing the contours and removing small holes so that the texture of an object is easier to recognize. The closing process is defined in terms of Minkowski addition ($\oplus$) and subtraction ($\ominus$) [17]. After the closing operation, black and white images were used as segmentation areas for grayscale images to obtain ROIs. The distinction between the textures of healthy and CFD respondents may be subtle to the unaided eyes; however, this classification is more distinguishable through quantitative methods. The characteristic focused on and obtained in this research was the texture of the skin. The texture feature is the regularity of certain patterns from the arrangement of pixels on the image. Segmentation of facial texture between healthy respondents and respondents with CHD is shown in Figure 2.

$$Closing\ (A, \sigma) = ((A \oplus \sigma) \ominus \sigma). \quad (2)$$

C. TEXTURE FEATURES EXTRACTION

Texture extraction was performed using the GLCM algorithm, initially introduced by Haralick in 1973 for feature extraction from terrain images [18]. GLCM analyses spatial patterns within an image by examining the relationship between pairs of pixels, considering their grayscale intensity, distance, and directions [18]. At angle directions of 0°, 45°, 90°, and 135°, the relationship between pixels was examined using this approach. Therefore, all features in (3) to (8) were calculated in all directions. The distance parameter refers to the distance between these pixel pairs in a particular direction. In terms of smooth images: smaller distance, smoother image capture or local texture patterns and larger distance, rougher image capture or more global texture patterns. In this research, a 5-pixel distance was used. For 400 × 400-pixel images, this distance is considered relatively small but not negligible because GLCM will capture local texture patterns but spread over a small region. It can provide a sense of texture details without reaching a truly global scale. This matrix is created by

counting the number of times a pixel with gray level value *i* appears in a particular spatial relationship with another pixel with gray level value *j*. Each element (*i*, *j*) of the matrix represents the number of instances of this pixel pair, which describes the image's spatial distribution of gray levels. Figure 3 is an example of a GLCM matrix for a pair of pixels in angle direction of 0° with a distance of 1 pixel [19].

A higher contrast value indicates a greater variability in the image. Dissimilarity quantifies the distance between pixel pairs within a specific image region. Homogeneity, also known as inverse different moment (IDM), looks at how uniformly distributed the elements are throughout an image area. In the GLCM, energy is calculated as the sum of the squared elements with values between 0 and 1, where a higher value indicates a more uniform texture. With values ranging from 1 to -1, correlation evaluates the relationship between a pixel and its neighbor throughout the image. Finally, angular second moment (ASM) measures the uniformity of the image's gray levels; higher ASM values indicate a greater pixel similarity. These features' mathematical equations are described in (3) to (8) [4], [20].

$$Contrast = \sum_i \sum_j |i-j|^2 p(i,j) \quad (3)$$

$$Dissimilarity = \sum_i \sum_j p_{i,j} |i-j| \quad (4)$$

$$Homogenity = \sum_i \sum_j \frac{p_{i,j}}{1+(i-j)^2} \quad (5)$$

$$Energy = \sum_{i,j} {p_{i,j}}^2 \quad (6)$$

$$Correlation = \sum_i \sum_j \frac{(i-\mu_i)(j-\mu)p(i,j)}{\sigma_i \sigma_j} \quad (7)$$

$$Angular\ Second\ Moment = \sum_i \sum_j p_{i,j}^2 \quad (8)$$

where *P* denotes the probability matrix in GLCM, *i* denotes the reference *pixel*, *j* denotes the neighborhood pixel, and μ_i, μ_j, σ_i, σ_j denotes the mean and standard deviation of $P_{i,j}$.

D. TRAINING AND TESTING MODEL

The dataset split test was divided into 80% (n = 106) of the dataset for the training set and the rest 20% (n = 26) was used for the testing set. The split method was chosen because it has a higher level of data testing accuracy during the simulation. The dataset comprised six parts obtained from the facial image. The training and testing data distribution was done randomly and automatically using Orange Python 3 software. The training set, also called model development, is a grouping based on the unique features obtained from the features extraction process.

Based on the characteristics of feature values from several feature matrices above, the model learned to make classification rules. ANN algorithm was used to make classification rules when the model was trained. In the testing phase, the classification rule becomes the basis for decision-making for prediction. ANN functions by adjusting the weights (*w*) and biases (*b*) of the connections to ensure that the outputs align with the inputs. The general equations used in this process are shown in (9) until (12) [21].

$$y^1 = f^1(w^1 p + b^1) \quad (9)$$

$$y^2 = f^2(w^2 y^1 + b^2) \quad (10)$$

$$\vdots \qquad \vdots$$

$$O = y^n = f^n(w^n y^{n-1} + b^n), \quad (11)$$

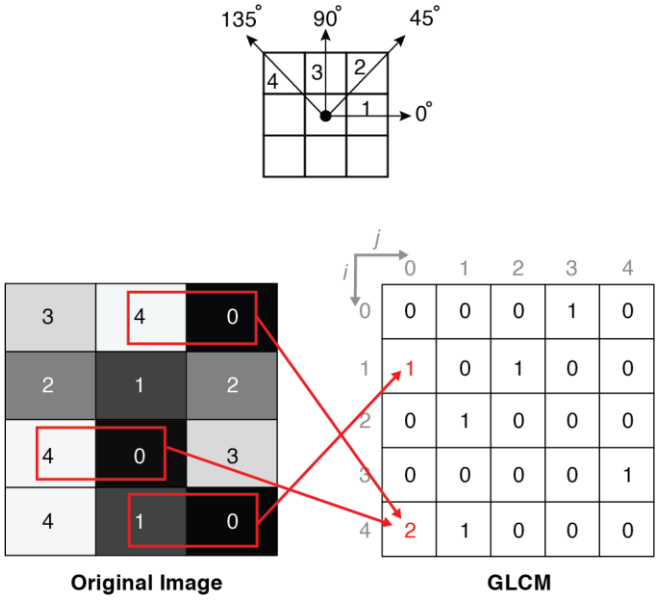

Figure 3. GLCM matrix for angle direction of 0° and a distance of 1 pixel.

$$\text{Where } p = \begin{bmatrix} p_1 \\ \vdots \\ p_r \end{bmatrix}, b^1 = \begin{bmatrix} b_1^i \\ \vdots \\ b_r^i \end{bmatrix}, and\ w^i = \begin{bmatrix} w_{11}^i & \cdots & w_{1r}^i \\ \vdots & \ddots & \vdots \\ w_{s1}^i & \cdots & w_{sr}^i \end{bmatrix}. \quad (12)$$

The evaluation process involves testing the classification model on a fresh dataset using the Orange 3 application to gauge its effectiveness. For this assessment, 20% of the dataset, equivalent to 26 samples, was reserved for testing. During this stage, visual tools like the confusion matrix and the receiver operating characteristic (ROC) curve were utilized to present the outcomes. These visualizations help assess the accuracy of each ROI and guide the selection of the most effective ROIs. The classification outcomes were then analyzed, sorted, and interpreted to draw meaningful conclusions. The confusion matrix records the comparison between predicted and actual labels, illustrating the four potential scenarios: a) true positive (TP) represents the number of CHD patients predicted correctly as CHD sufferers; b) true negative (TN) represents the number of healthy patients predicted correctly as healthy patients; c) false positive (FP) represents the number of healthy patients predicted incorrectly as CHD sufferers; d) false negative (FN) represents the number of CHD patients predicted incorrectly as healthy patients. These classifications are essential for evaluating the accuracy and reliability of the predictive model.

The model's performance is typically evaluated using several key metrics, including the area under the ROC curve (AUC), classification accuracy (CA), precision, recall, and the F1 score. The AUC indicates the model's effectiveness in distinguishing between positive and negative instances. The true positive rate (TPR), also known as recall, represents the proportion of actual positive cases that the model correctly identifies. On the other hand, the false positive rate (FPR) quantifies the fraction of negative cases that are mistakenly classified as positive. Classification accuracy measures the percentage of correct predictions, both positive and negative, across the entire dataset. Precision assesses the accuracy of positive predictions by comparing the number of true positives to the total positive predictions made. The F1 score then

synthesizes recall and precision into a single metric, offering a balanced overview of the model's predictive performance. Accuracy providing an overall view of the model's performance. However, imbalanced datasets can appear misleading accuracy, as it may inflate performance if one class dominates. Precision, on the other hand, calculates the proportion of true positive predictions out of all positive predictions made, making it useful when minimizing false positives is critical. It is valuable in applications such as rare disease diagnosis, reducing the number of incorrect positive diagnoses is important. Recall assesses the model's ability to identify all actual positive cases, which is crucial when minimizing false negatives is important. Recall is commonly prioritized in fields like fraud or disease detection, where identifying as many true positive cases as possible is essential. F1 score is the harmonic mean of precision and recall, offering a balanced measure when there is a trade-off between these two metrics. Detailed formulas for these metrics are provided below.

$$Recall = \frac{TP}{(FN+TP)} \quad (13)$$

$$False\ Positive\ Rate = \frac{FP}{(TN+FP)} \quad (14)$$

$$Classification\ Accuracy\ = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (15)$$

$$Precission = \frac{TP}{(TP+FP)} \quad (16)$$

$$F1\ Score\ = \frac{2\ (Recall\ x\ Precission)}{(Recall\ +\ Precision)}. \quad (17)$$

The ROC curve is generated from values obtained through calculations involving a confusion matrix, depicting the relationship between the FPR and the TPR. This graph is used to evaluate the effectiveness of prediction results. In classification models, the ROC curve serves as a visualization tool that helps assess and select classification performance on a binary scale ranging from 0 to 1. A diagonal ROC curve, which aligns with the 45° line or baseline, signifies poor model performance. Conversely, a curve above the baseline indicates a more effective classification model. Thus, a larger AUC indicates better performance. An AUC of 0.5 signifies no ability to discriminate between classes, while an AUC of 1.0 denotes perfect classification accuracy. Figure 4 illustrates the concept of the AUC [22].

E. CLASSIFICATION ACCURACY ANALYSIS

The analysis was carried out through data tabling and visualized with a bar chart. Data analysis was done by calculating the accuracy of each ROI and categorizing accuracy values based on the interval scale that had been designed based on [23]. A score ranging from 80 to 100% is classified as very good, indicating excellent performance. Scores between 71 and 80% are categorized as good, reflecting above-average performance. A percentage of 61 to 70% is considered fair, suggesting moderate performance. Scores in the range of 51 to 60% fall into the poor category, indicating below-average results. Finally, percentages from 0 to 50% are deemed very poor, signifying inadequate performance. This scale helps assess and classify the quality or success of an evaluation. The score of the prediction results in relation to the total number of predictions that have been made is called accuracy. When recommending the ROI for future research, categorization is used as a selection process.

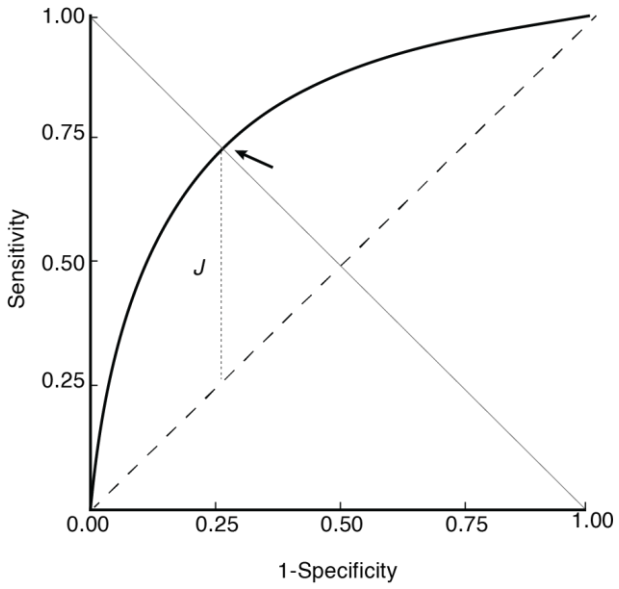

Figure 4. Illustration of area under curve (AUC).

IV. RESULTS AND DISCUSSION

A. TRAINING AND TESTING MODEL

The model was built by extracting texture feature scores from the training data to create classification guidelines. The ANN was trained using these guidelines to predict the likelihood CHD. The model divided potential outcomes into two risk categories: low risk, which is negative, and high risk, which is positive.

The training process involved labeling the dataset with positive and negative indicators of CHD. Following this step, the Keras library was imported. Keras was utilized to implement the neural network algorithm on top of TensorFlow for creating machine learning models. The next step involved configuring the hidden layers of the model, specifically by adding two hidden layers. Two activation functions were assigned: rectified linear unit (ReLu) for the hidden layers and sigmoid for the output layer. The fitting function was subsequently utilized to compile the ANN model, which was trained over 200 epochs; this denotes the number of iterations the algorithm executed over the entire training dataset. Finally, metrics like classification accuracy, AUC, precision, recall, and the F1-score were used to evaluate the model's performance.

The amount of data in the datasets and the number of epochs entered into the program determined the duration of the learning process. The more datasets and the number of epochs, the longer the learning process will take. The time for model execution in this study showed relatively fast results, with results of 20.13 s. The learning process displayed loss and accuracy values, meaning that if the loss value is large, the accuracy is small, and vice versa. After the learning process was complete, the program continued directly by testing the model with the existing test dataset and then displaying the results for each data.

B. MODEL EVALUATION

The confusion matrix obtained from the testing process on models using the ANN algorithm is described in Figure 5, obtained from Orange3 application. The testing results on the model with the application of the ANN algorithm, as shown in the Figure 5, indicated the results from 26 testing data, comprising 8 correct predictions for data with negative labels

Jurnal Nasional Teknik Elektro dan Teknologi Informasi
Volume 14 Number 1 February 2025

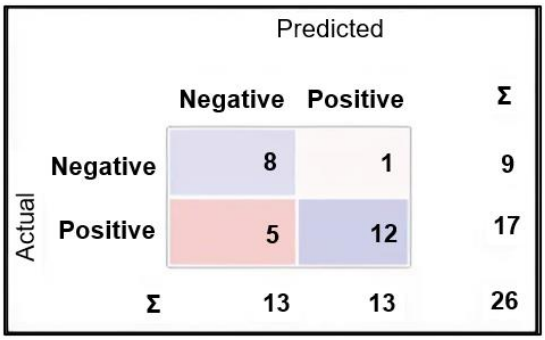

Figure 5. Output in the confusion matrix.

and 12 correct predictions for data with positive labels. In addition, 1 wrong prediction for data with positive labels and 5 false predictions for data with negative labels were found. In total, the model could make 20 correct predictions and 6 wrong predictions, with an accuracy percentage of 76.9%.

Apart from the classification accuracy, model evaluation can also be done by looking at several other quantities of data testing results, namely AUC, precision, recall, and F1 score. A binary value was used to measure this amount, with the lowest value was 0 and the highest score was 1. The complete evaluation's outcomes with weighted average over classes obtained from Orange3 application are depicted in the bar chart in Figure 6.

C. ROI PERFORMANCE ANALYSIS

Each testing data used had a label or actual condition from the data, which was between positive and negative related to CHD conditions. Predictions are values that represent the results of the classification task by the model created. Predictions are concluded based on the comparison of negative and positive values. A higher value between negative and positive determines the result of the classification prediction tends to have a positive or negative conclusion.

The model's ROC curve is depicted in Figure 7. The conformity between the data label (actual condition) and the model's prediction is the basis for drawing the prediction results' conclusion. When the labels of the data and the predictions match, the predictions are correct. The data of each ROI was tested by selecting 26 datasets at random using a split method with six ROIs to eliminate the possibility of the application system taking an unbalanced number of ROIs. This study chose to go into detail on 6 left crow's feet, 7 left canthus, 2 right canthus, 4 forehead, 4 right crow's feet, and 3 nose bridge data.

The result of data testing was subsequently calculated as the average or the mean for each ROI to determine the prediction performance on each part of the face image used. The average accuracy of each ROI showed that from the six parts analyzed through computational experiments, two ROIs had prediction accuracy with excellent categories: the left crow's feet, right canthus, and forehead. At the same time, the other three ROIs, including the left canthus, right crow's feet, and nose bridge were categorized as having good prediction accuracy. Right canthus was the ROI with the highest prediction accuracy at 95.50%, while right crow's feet and nose bridge were the ROIs with the lowest prediction accuracy at 65%. Although the three ROIs were classified into the category of good accuracy, the accuracy score obtained was still not high enough, around 65%–66.43%. At the same time, the other three ROIs with excellent accuracy were the left crow's feet, right canthus, and

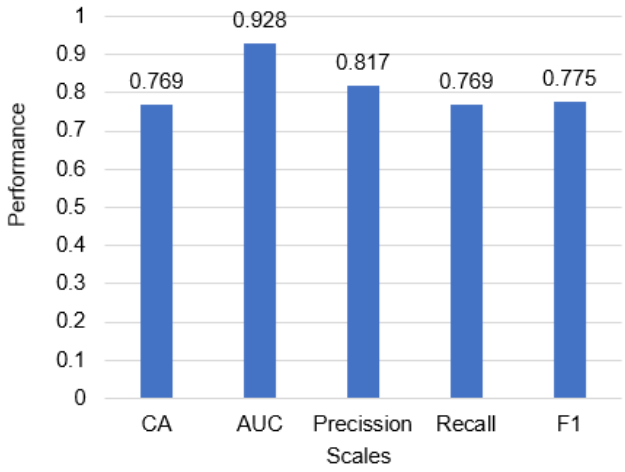

Figure 6. Performance output of the confusion matrix.

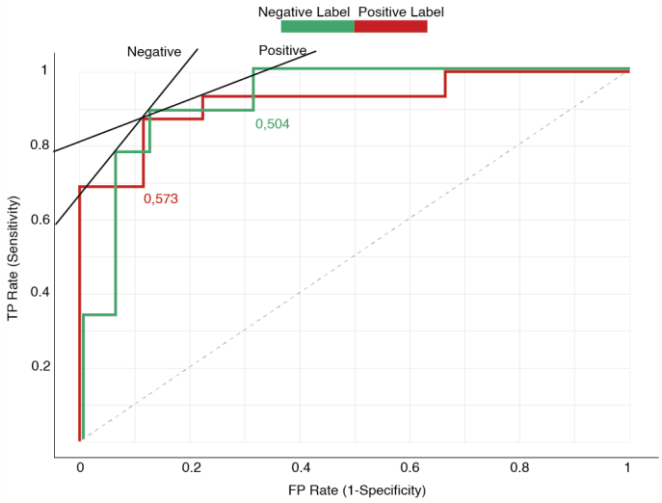

Figure 7. ROC curve of this model.

forehead, which had prediction accuracy of 82%, 95.50%, and 90%, respectively. The accuracy of each ROI of these values is shown in Figure 8.

The names of the facial regions were used to group the results of each ROI based on the findings of the earlier research. Since ROIs on crow's feet and canthus has two sides—right and left—the average accuracy is shown for ROI performance analysis. Table II presents the outcomes.

The results of the research showed that there were two facial regions with very high accuracy, namely the canthus and forehead, achieving accuracies of 80.97% and 90%, respectively. Meanwhile, two other facial regions were categorized as having high accuracy, namely the crow's feet and nose bridge, with accuracies of 73.50% and 65%, respectively. Compared to previous research, there are modifications or differences in the system developed in this study. The differences lie in the dataset, classification algorithm, and the selection of ROIs used. The previous study used a total of 1,528 ROI datasets from the Chinese ethnicity, whereas this study used 132 ROI datasets from the full Indonesian ethnicity. In terms of the classification algorithm, this study employed ANN, while the previous research used random forest and decision tree algorithms. Regarding the selection of ROIs, the previous study used 8 ROIs, whereas this study used 6 ROIs. The results of the average percentage of accuracy for each ROI are visualized in Figure 9.

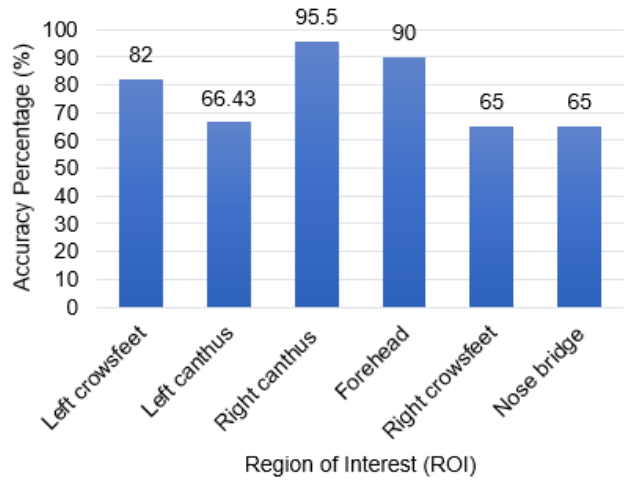

Figure 8. Bar chart of average accuracy of each ROI.

TABLE II
AVERAGES ACCURACY OF ROIS

ROIs	Average Accuracy (%)
Crow's feet	73.50
Canthus	80.97
Forehead	90
Nose bridge	65

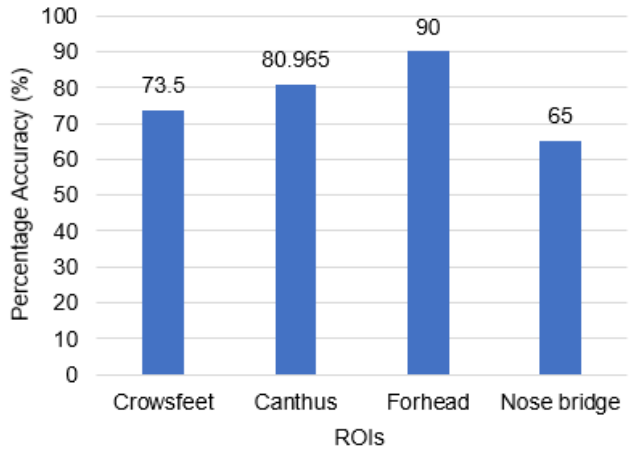

Figure 9. Average percentage of ROI accuracy.

The results also show differences in relation to the proposed method modification. This study had a slightly higher average ROI. Additionally, each ROI used in this research produced accuracy with a significant difference. This aspect needs to be examined through further research by increasing the quantity and diversity of the dataset to see if there are any differences or influences related to the amount of the dataset on accuracy. Along with the development of the algorithm, it is essential to conduct testing that incorporates modified deep-learning algorithms to observe the differences.

V. CONCLUSION

The research findings indicate that the machine learning model designed for diagnosing CHD, using texture features from six facial regions and employs the ANN algorithm, demonstrates strong performance, attaining an accuracy rate of 76.9%. The research results indicated that two parts of the face had excellent accuracy, namely the canthus and forehead, with 80.97% and 90% accuracy, respectively. The other two parts of the face were categorized into good accuracy, namely crow's feet and nose bridge, with accuracy of 73.50% and 65%, respectively. Based on the research results, the developed machine learning model is capable of diagnosing CHD only through a facial photo uploaded and processed in a short time of about 20.13 s. Therefore, the target of designing a model that is easy to use and fast in execution compared to clinical methods has been realized. Although it is easier and faster, the current CHD diagnosis model cannot replace the existing clinical method, as it still needs further development before it can be deemed suitable for implementation. The benefit of this research is to create a better path of discovery in future research. In addition, this program has proven to be able to become a model for early screening of CHD with good accuracy. It is recommended that further research increases the number of datasets and conduct trials of implementing deep-learning algorithms that have the potential to get better model accuracy.

CONFLICTS OF INTEREST

The authors explicitly declare that in their study titled "Analysis of Facial Areas to Identify CHD Risk Based on Facial Textures," there are no conflicts of interest to report. They assert that all aspects of the research, including the design, methodology, data analysis, and interpretation of results, were conducted independently and without any external influences or biases. The authors further confirm that no financial, personal, or professional relationships could have influenced the findings of this research, ensuring the integrity and objectivity of their work.

AUTHORS' CONTRIBUTIONS

Conceptualization, Yudha Andriano Rismawan, Budi Sunarko; methodology, Agung Adi Firdaus and Budi Sunarko; data analysis, Agung Adi Firdaus and Budi Sunarko; resources, Yudha Andriano Rismawan and Anan Nugroho; data acquisition, Agung Adi Firdaus, Anan Nugroho, and Yudha Andriano Rismawan; writing—original draft preparation, Agung Adi Firdaus; writing—reviewing and editing, Yudha Andriano Rismawan; supervision, Budi Sunarko.

ACKNOWLEDGMENT

The authors extend our gratitude to the Faculty of Engineering at Universitas Negeri Semarang, Dr. Kariadi Hospital, and Diponegoro National Hospital for their invaluable support in conducting this research. Their contributions and assistance were essential to the success of this study, and the authors deeply appreciate their collaboration and commitment to advancing our work.

REFERENCES

[1] S.I. Ayon, M.M. Islam, and M.R. Hossain, "Coronary artery heart disease prediction: A comparative study of computational intelligence techniques," *IETE J. Res.*, vol. 68, no. 4, pp. 2488–2507, Jul./Aug. 2022, doi: 10.1080/03772063.2020.1713916.

[2] C. Shao, J. Wang, J. Tian, and Y. Tang, "Coronary artery disease: From mechanism to clinical practice," in *Coronary Artery Disease: Therapeutics and Drug Discovery*, M. Wang, Ed., Singapore, Singapore: Springer, 2020, pp. 1–36.

[3] B. Jin, L. Cruz, and N. Gonçalves, "Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis," *IEEE Access*, vol. 8, pp. 123649–123661, Jun. 2020, doi: 10.1109/ACCESS.2020.3005687.

[4] Y.A. Rismawan *et al.*, "Development of coronary heart disease diagnosis system based on facial imagery," in *Proc. 3rd Conf. Fundam. Appl. Sci. Adv. Technol. 2022*, 2022, p. 020003, doi: 10.1063/5.0180182.

[5] J. Wang *et al.*, "A stacking-based model for non-invasive detection of coronary heart disease," *IEEE Access*, vol. 8, pp. 37124–37133, Feb. 2020, doi: 10.1109/ACCESS.2020.2975377.

[6] S. Lin *et al.*, "Face analysis for coronary heart disease diagnosis," in *2019 12th Int. Congr. Image Signal Process. BioMed. Eng. Inform. (CISP-BMEI)*, 2019, pp. 1–5, doi: 10.1109/CISP-BMEI48845.2019.8966020.

[7] S. Lin *et al.*, "Feasibility of using deep learning to detect coronary artery disease based on facial photo," *Eur. Heart J.*, vol. 41, no. 46, pp. 4400–4411, Dec. 2020, doi: 10.1093/eurheartj/ehaa640.

[8] J. Qiang *et al.*, "Review on facial-recognition-based applications in disease diagnosis," *Bioengineering*, vol. 9, no. 7, pp. 1–16, Jul. 2022, doi: 10.3390/bioengineering9070273.

[9] U. Thirunavukkarasu, S. Umapathy, K. Janardhanan, and R. Thirunavukkarasu, "A computer aided diagnostic method for the evaluation of type II diabetes mellitus in facial thermograms," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 871–888, Sep. 2020, doi: 10.1007/s13246-020-00886-z.

[10] S. Gondane, A. Maherda, and R. Kothiwala, "To study the prevalence of metabolic syndrome and dyslipidemia in patients of xanthelasma palpebrarum at a tertiary care hospital," *Asian J. Diabetol.*, vol. 21, no. 3, pp. 10–14, Oct. 2020.

[11] H.-C. Chang, C.-W. Sung, and M.-H. Lin, "Serum lipids and risk of atherosclerosis in xanthelasma palpebrarum: A systematic review and meta-analysis," *J. Am. Acad. Dermatol.*, vol. 82, no. 3, pp. 596–605, Mar. 2020, doi: 10.1016/j.jaad.2019.08.082.

[12] P. Kampar, Q. Anum, and S. Lestari, "The correlation between lipid profile and xanthelasma," *Berk. Ilmu Kesehat. Kulit Kelamin*, vol. 32, no. 2, pp. 119–125, Aug. 2020, doi: 10.20473/bikk.V32.2.2020.119-125.

[13] A.K.I. Suman *et al.*, "Association of xanthelasma palpebrarum (XP) with cardiovascular disease (CVD) risk factors," *Asian J. Med. Biol. Res.*, vol. 5, no. 4, pp. 324–329, Dec. 2019, doi: 10.3329/ajmbr.v5i4.45271.

[14] B.I. Fitrasanti, A. Onggo, W. Sugirman, and B. Ciptawan, "Prevalence of cutaneous markers in coronary artery disease cases," *Bali Med. J.*, vol. 10, no. 2, pp. 877–880, Aug. 2021, doi: 10.15562/bmj.v10i2.2531.

[15] E.R. Dougherty, *Digital Image Processing Methods*. New York, NY, USA: CRC Press, 1994.

[16] S. Ibrahim *et al.*, "Automated platelet counter with detection using k-means clustering," *Ann. Emerg. Technol. Comput. (AETiC)*, vol. 7, no. 5, pp. 39–49, Oct. 2023, doi: 10.33166/AETIC.2023.05.004.

[17] S. Sellán, J. Kesten, A.Y. Sheng, and A. Jacobson, "Opening and closing surfaces," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–13, Dec. 2020, doi: 10.1145/3414685.3417778.

[18] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[19] B.M. Jebin, M.A. Rejula, and G. Eberlein, "Neonatal Seizure detection using GLCM feature extraction & AlexNet classification," *Multimed. Tools Appl.*, Oct. 2024, doi: 10.1007/s11042-024-18779-8.

[20] N. Iqbal, R. Mumtaz, U. Shafi, and S.M.H. Zaidi, "Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms," *PeerJ Comput. Sci.*, vol. 7, pp. 1–26, May 2021, doi: 10.7717/PEERJ-CS.536.

[21] M. Madanan, *Neural Network and Deep Learning*. Uttar Pradesh, India: BlueRose Publishers, 2022.

[22] S. Parodi, D. Verda, F. Bagnasco, and M. Muselli, "The clinical meaning of the area under a receiver operating characteristic curve for the evaluation of the performance of disease markers," *Epidemiol. Health*, vol. 44, pp. 1–10, Oct. 2022, doi: 10.4178/epih.e2022088.

[23] M.G. Nugraha, S. Utari, D. Saepuzaman, and F. Nugraha, "Redesign of students' worksheet on basic physics experiment based on students' scientific process skills analysis in Melde's law," in *4th Int. Semin. Math. Sci. Comput. Sci. Educ.*, 2018, pp. 1–8, doi: 10.1088/1742-6596/1013/1/012038.