DOI: 10.21917/ijivp.2026.0560

A DUAL-PATH CONTRASTIVE CNN-TRANSFORMER FRAMEWORK FOR ROBUST AND DISCRIMINATIVE MELANOMA CLASSIFICATION FROM DERMOSCOPIC IMAGES

Sreedevi Kadiyala[1] and Chandra Srinivas Potluri[2]

[1]Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, India
[2]Department of Computer Science and Engineering, Siddhartha Institute of Engineering and Technology, India

Abstract

Melanoma is an aggressive form of skin cancer in which early and accurate diagnosis substantially influences clinical management and patient survival. Dermoscopic imaging provides detailed visual information about lesion structures, pigmentation, borders, and internal patterns, making it an important source for computer-aided melanoma assessment. Recent deep learning methods have demonstrated strong diagnostic capability; however, their performance remains sensitive to image variability, class imbalance, subtle inter-class similarities, and limited representation of global lesion context. Conventional convolutional neural networks (CNNs) effectively capture local texture and morphological patterns but may inadequately model long-range relationships, whereas Transformer-based models capture global dependencies but can be sensitive to data availability and computational requirements. Existing hybrid models also frequently rely on simple feature concatenation without explicitly encouraging complementary and discriminative representations. This study proposes a novel Dual-Path Contrastive CNN-Transformer Network (DCCT-Net) for melanoma classification. The framework employs parallel CNN and Transformer pathways to learn complementary local and global lesion representations. A contrastive representation alignment module is introduced to reduce intra-class feature variation and increase inter-class separability, while an adaptive feature fusion mechanism combines the two pathways before classification. Augmentation-aware contrastive learning further improves representation robustness against changes in scale, illumination, rotation, and lesion appearance. The proposed DCCT-Net demonstrates superior performance compared with the DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT approaches. At the final training epoch, DCCT-Net achieves 98.4% accuracy, 98.1% precision, 97.9% recall, 98.0% F1-score, and 99.0% specificity. Compared with the three existing methods, the proposed model improves accuracy by 2.2, 2.3, and 1.4%, respectively. The recall reaches 97.9%, representing improvements of 3.1, 3.4, and 1.9%, demonstrating stronger melanoma detection capability. The F1-score of 98.0% confirms a balanced improvement in precision and recall, while the specificity of 99.0% indicates effective discrimination of non-melanoma lesions. These results demonstrate that combining complementary CNN and Transformer representations with contrastive learning and adaptive fusion improves the discriminative capability of automated melanoma classification.

Keywords:

Melanoma Classification, Dermoscopic Images, Contrastive Learning, CNN-Transformer, Skin Lesion Analysis

1. INTRODUCTION

Melanoma is one of the most clinically important forms of skin cancer because of its aggressive progression and capacity to metastasize. Conventional diagnosis generally begins with visual examination and dermoscopy, followed, when required, by biopsy and histopathological assessment. The increasing availability of digital dermoscopic images has created opportunities for automated image-based analysis and computer-aided diagnostic systems. Deep learning has become particularly relevant because it can learn discriminative representations directly from images rather than depending exclusively on manually designed dermatological features. Early studies demonstrated that deep convolutional neural networks could approach dermatologist-level performance in selected skin cancer classification tasks, establishing the feasibility of automated visual assessment [1]. Dermoscopy provides additional structural information that is difficult to obtain from conventional clinical photographs. Features such as pigmentation distribution, asymmetry, streaks, dots, globules, vascular patterns, and blue-white structures can contribute to melanoma differentiation. The development of large public datasets has further accelerated research in this field. The HAM10000 dataset contains 10,015 dermoscopic images collected from multiple sources and diagnostic categories, providing a widely used benchmark for automated skin lesion classification [2]. The ISIC challenges have subsequently provided larger and more heterogeneous datasets for evaluating melanoma and multiclass skin lesion classification systems. For example, the ISIC 2019 dataset contains 25,331 training images covering eight diagnostic categories, creating a challenging environment for automated classification [3]. Despite the progress of deep learning, melanoma classification remains difficult because dermoscopic images exhibit substantial intra-class and inter-class variation. Lesions belonging to the same diagnostic category may differ considerably in color, size, shape, texture, and internal structure. Conversely, benign nevi can visually resemble melanoma, making the distinction between malignant and benign lesions particularly challenging [4]. Image acquisition introduces additional variability through differences in illumination, magnification, dermoscopic equipment, image resolution, artifacts, and acquisition protocols. Hair, air bubbles, rulers, ink markings, and surrounding skin can also interfere with feature extraction [5]. A further challenge is severe class imbalance. Public skin lesion datasets frequently contain substantially more benign lesions than melanoma cases. Such imbalance can cause a classifier to favor majority classes while achieving apparently high overall accuracy despite weaker minority-class sensitivity [6]. Recent ISIC-based studies continue to identify substantial imbalance as a major obstacle to reliable multiclass skin lesion classification. Another limitation concerns generalization. Models trained on carefully selected benchmark images may perform differently when applied to images acquired from other institutions, devices, populations, or clinical environments. Systematic reviews have emphasized that many existing evaluations use artificial or single-image settings and do not fully represent the diversity encountered in routine clinical practice [7]. The central problem addressed in this study is the inadequate

integration of local morphological information and global contextual relationships in existing melanoma classification architectures. CNNs are effective at extracting local edges, textures, color distributions, and lesion boundaries, but their receptive-field limitations can restrict the modeling of long-range relationships between distant lesion regions. Transformers can model global dependencies through self-attention, but their effectiveness may depend strongly on training data, token representation, and computational resources. Furthermore, simply combining CNN and Transformer features does not necessarily guarantee complementary representations. The problem is therefore to develop a classification architecture that simultaneously learns fine-grained local features, global contextual dependencies, and highly separable representations between melanoma and non-melanoma lesions [6]. Recent evidence also indicates that variability in image acquisition, preprocessing, segmentation, and feature extraction remains a persistent limitation in AI-assisted melanoma analysis.

The primary novelty of the proposed Dual-Path Contrastive CNN-Transformer Network (DCCT-Net) lies in the joint optimization of local CNN representations and global Transformer representations through a contrastive learning objective. Instead of treating the two feature streams as independent components followed by conventional concatenation, the proposed framework explicitly encourages diagnostically related samples to occupy compact regions in the embedding space while separating melanoma from visually similar benign lesions. An adaptive fusion module dynamically combines complementary representations according to their discriminative contribution. The framework therefore addresses both feature diversity and feature separability within a single optimization process.

The principal contributions of this study are: A dual-path local-global architecture: A parallel CNN and Transformer structure is developed to capture fine-grained morphological patterns and long-range contextual relationships from dermoscopic images. Contrastive adaptive representation learning: A lesion-aware contrastive module and adaptive fusion strategy are introduced to improve class separation, reduce sensitivity to visual variations, and strengthen melanoma discrimination.

2. RELATED WORKS

Deep learning has significantly changed automated skin lesion analysis by reducing dependence on handcrafted descriptors and enabling end-to-end representation learning. One of the influential studies in this field was conducted by Esteva et al. [8], who developed a deep convolutional neural network using a large collection of clinical skin images. The model was evaluated against board-certified dermatologists for selected skin cancer classification tasks and achieved performance comparable to the tested experts. This work established that CNN-based systems can learn clinically meaningful visual patterns directly from images. However, the study primarily relied on conventional convolutional representation learning and did not explicitly address the complementary relationship between local and global features.

The availability of standardized dermoscopic datasets subsequently enabled more systematic development and comparison of deep learning models. Tschandl et al. [9] introduced HAM10000, a collection of 10,015 dermoscopic images covering seven important diagnostic categories. The dataset was collected from multiple sources and modalities, thereby providing greater diversity than many earlier datasets. HAM10000 became an important benchmark for transfer learning, CNN-based classification, ensemble learning, and more recent Transformer architectures. Nevertheless, its diagnostic distribution is highly uneven, and melanoma represents only a minority of the available cases. Consequently, a model can obtain strong aggregate performance while still exhibiting limitations in detecting melanoma accurately.

The ISIC challenge series further expanded the scale and diversity of available dermoscopic data. The ISIC 2019 classification task provided 25,331 training images distributed among eight diagnostic categories, including melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesions, and squamous cell carcinoma [10]. These datasets encouraged researchers to investigate increasingly sophisticated CNN architectures, transfer learning, augmentation, ensemble learning, and hybrid networks. Recent studies continue to demonstrate that CNN-based systems can achieve high classification performance on HAM10000 and ISIC datasets, although severe class imbalance and dataset heterogeneity remain important concerns.

The introduction of Vision Transformers (ViTs) provided another direction for image-based lesion classification. Dosovitskiy et al. [11] demonstrated that an image can be divided into patches and processed as a sequence using Transformer self-attention. ViT showed that global relationships among image regions can be learned without relying exclusively on convolutional operations. For dermoscopic images, this global modeling capability is potentially valuable because melanoma-related structures may occur at different spatial locations and scales. However, pure Transformer models may not always preserve the local texture sensitivity of CNNs. This is particularly relevant for dermoscopic analysis, where small pigment structures, border irregularities, and fine textural patterns can carry important diagnostic information.

Recent melanoma-specific research has therefore increasingly explored hybrid CNN-Transformer architectures. For example, studies combining segmentation, CNN feature extraction, and Vision Transformer refinement have reported strong performance on ISIC datasets [12]. One recent hybrid framework using U-Net for segmentation, Inception-ResNet-v2 for feature extraction, and Vision Transformer-based feature refinement reported 98.65% accuracy, 99.20% sensitivity, and 98.03% specificity on ISIC 2020. Other recent work has similarly investigated the combination of convolution and self-attention to exploit local and global information simultaneously. These findings support the value of hybrid architectures, but many existing approaches rely primarily on supervised classification loss and straightforward feature integration.

Contrastive representation learning provides an additional mechanism for addressing this limitation. Chen et al. introduced SimCLR, demonstrating that contrastive learning can improve visual representations by maximizing agreement between

differently augmented views of the same sample while separating representations from unrelated samples. In melanoma classification, such an approach is particularly relevant because the same lesion may appear different under changes in scale, rotation, illumination, cropping, and acquisition conditions. Contrastive learning can encourage the model to preserve lesion-specific information despite these variations.

Recent evidence also suggests that combining convolution with attention can improve robustness to dermoscopic structural variation. A study investigating melanoma-related dermoscopic structures reported that adding convolutional components to Vision Transformers improved classification performance, while pure Transformer models showed greater brittleness under image modifications. Similarly, systematic reviews and meta-analyses indicate that deep learning methods can achieve strong melanoma diagnostic performance, while emphasizing the continuing need for larger, multicenter, and clinically representative evaluations. A 2024 systematic review and meta-analysis reported pooled sensitivity of 82%, specificity of 87%, and AUC of 0.92 across eligible deep learning studies, while noting the need for higher-quality and larger-scale validation.

3. PROPOSED METHOD

The proposed DCCT-Net is designed to classify dermoscopic images into melanoma and non-melanoma categories by jointly learning local morphological patterns and global contextual dependencies.

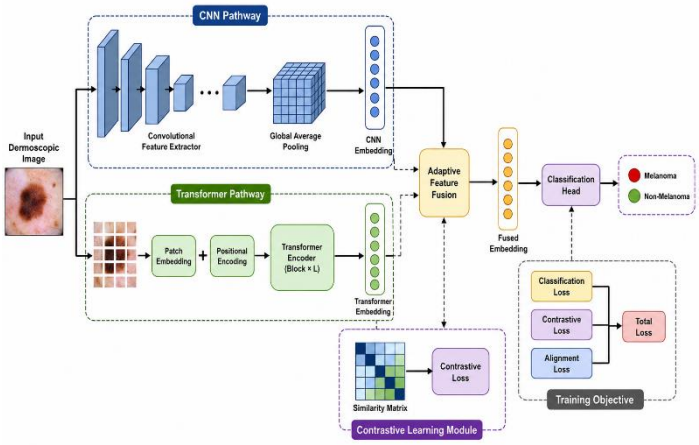

Fig.1. Proposed DCCT-Net

The architecture contains two complementary feature extraction paths. The first path uses a convolutional neural network to identify fine-grained spatial information such as lesion boundaries, pigment distribution, texture, asymmetry, and local color transitions. The second path uses a Transformer encoder to model long-range relationships among different regions of the lesion. The outputs of both paths are projected into a common embedding space and optimized using a contrastive objective that increases similarity between representations belonging to the same diagnostic class while separating representations from different classes. An adaptive fusion module then combines the CNN and Transformer representations according to their learned importance. Finally, the fused representation is supplied to a fully connected classification layer with a softmax activation. The proposed DCCT-Net operates through the following sequential stages:

- **Dermoscopic image acquisition:** Dermoscopic images are collected from a melanoma classification dataset containing malignant and benign skin lesions.
- **Image preprocessing:** Images are resized to a common spatial resolution and normalized to reduce variations caused by acquisition conditions.
- **Data augmentation:** Rotation, horizontal and vertical flipping, random cropping, scaling, and controlled color transformations are applied to increase training diversity.
- **CNN pathway:** The preprocessed image is passed through convolutional layers to extract local texture, edge, color, and morphological features.
- **Transformer pathway:** The same image is divided into patches and converted into token embeddings. Transformer self-attention then captures relationships between spatially distant lesion regions.
- **Feature projection:** CNN and Transformer outputs are projected into a common-dimensional embedding space.
- **Contrastive representation learning:** Positive and negative sample pairs are constructed to make same-class representations closer and different-class representations more separable.
- **Adaptive feature fusion:** The CNN and Transformer embeddings are weighted using learned attention coefficients and combined into a unified lesion representation.
- **Classification:** The fused representation is passed through fully connected layers and a softmax classifier.
- **Joint optimization:** Classification loss and contrastive loss are jointly minimized to obtain a discriminative and robust melanoma classifier.

3.1 DERMOSCOPIC IMAGE PREPROCESSING

The first stage converts the input dermoscopic image into a standardized representation suitable for the proposed network. Dermoscopic datasets contain images with different resolutions, illumination conditions, acquisition devices, and lesion sizes. Directly supplying these images to a deep network can introduce unnecessary variability. Therefore, every image is resized to a fixed resolution, such as 224×224 pixels, while preserving the essential lesion structure. Pixel normalization is subsequently performed using the mean and standard deviation of the training dataset.

Data augmentation is applied only to the training samples. The augmentation procedure generates visually different versions of the same lesion without changing its diagnostic identity. Rotation is useful because the diagnostic interpretation of a lesion should remain unchanged under orientation changes. Horizontal flipping, controlled scaling, random cropping, and moderate color perturbation improve robustness to acquisition variability. Importantly, augmentation parameters should remain clinically plausible because excessive transformations may modify diagnostically meaningful structures.

Let the original dermoscopic image be represented by I, and let $T_k(\cdot)$ denote the k^{th} stochastic augmentation transformation. The generated image can be represented by:

$$I_k^{(a)} = \mathrm{N}\left(T_k(I)\right),\ T_k \in \left\{T_{\text{rot}}, T_{\text{flip}}, T_{\text{crop}}, T_{\text{scale}}, T_{\text{color}}\right\}, \quad (1)$$

where $N(\cdot)$ represents the normalization operation. Multiple augmented views of the same lesion are later used by the contrastive learning component.

3.2 LOCAL FEATURE EXTRACTION USING THE CNN PATH

The CNN pathway is responsible for learning local morphological characteristics of skin lesions. CNNs are particularly effective for extracting edges, corners, texture variations, pigment boundaries, and small structural abnormalities. In melanoma analysis, these properties are important because malignant lesions can contain irregular borders, heterogeneous pigmentation, asymmetric structures, and localized textural changes.

The CNN backbone can be implemented using a residual architecture in which convolutional blocks progressively transform the input image into increasingly abstract feature maps. Early layers learn low-level visual characteristics, while deeper layers encode semantic lesion structures. Residual connections facilitate gradient propagation and reduce degradation during training.

For an input feature map X_l, the output of a residual convolutional block can be expressed as:

$$F_l^{\text{CNN}} = \sigma\left(\mathrm{BN}_{l,2}\left(W_{l,2} * \sigma\left(\begin{matrix}\mathrm{BN}_{l,1}\left(W_{l,1} * X_l + b_{l,1}\right)\\ +b_{l,2}\end{matrix}\right)\right) + \mathrm{P}_l(X_l)\right), \quad (2)$$

where $W_{l,1}$ and $W_{l,2}$ denote convolution kernels, $b_{l,1}$ and $b_{l,2}$ are biases, $*$ denotes convolution, BN represents batch normalization, σ denotes the nonlinear activation function, and P_l represents the projection shortcut when the input and output dimensions differ. After the final convolutional stage, global average pooling converts the spatial feature map into a compact vector. This operation retains the overall activation strength of learned lesion patterns while reducing the dimensionality before feature projection.

3.3 GLOBAL REPRESENTATION LEARNING USING THE TRANSFORMER PATH

Although CNNs capture local patterns effectively, melanoma-related evidence may be distributed across spatially separated portions of a lesion. The Transformer pathway addresses this limitation by modeling interactions among different image regions.

The input image is divided into non-overlapping patches. For an image with spatial dimensions $H \times W$, patch size $P \times P$, and three color channels, the image produces $N=(H/P)(W/P)$ patches. Each patch is flattened and projected into a D-dimensional embedding. A learnable positional representation is added because self-attention alone does not inherently encode spatial order.

The patch-token construction is defined as:

$$Z_0 = \left[x_{\text{cls}}; \mathrm{vec}(P_1)E; \mathrm{vec}(P_2)E; \cdots; \mathrm{vec}(P_N)E\right] + E_{\text{pos}}, \quad (3)$$

where, $N = \dfrac{HW}{P^2}$, P_i represents the i^{th} image patch, E is the learnable patch projection matrix, x_{cls} is the classification token, and $Epos$ represents positional embeddings.

The Transformer subsequently applies multi-head self-attention. This mechanism enables one lesion region to interact with all other regions. For example, a pigment structure on one side of the lesion can be related to an irregular boundary located elsewhere. Such global relationships are difficult to represent explicitly using only shallow local convolutional operations. For a Transformer layer, the attention mechanism is formulated as:

$$\mathrm{MSA}(Z) = \mathrm{Concat}_{h=1}^{H_a}\left[\mathrm{softmax}\left(\frac{(ZW_h^Q)(ZW_h^K)^{\cdot}}{\sqrt{d_h}}\right)ZW_h^V\right]W^O, \quad (4)$$

where H_a denotes the number of attention heads, W_h^Q, W_h^K, and W_h^V are the query, key, and value projection matrices for head h, d_h is the dimensionality of each attention head, and W^O is the output projection matrix. The Transformer output corresponding to the classification token is extracted as the global lesion representation. Therefore, the CNN branch focuses primarily on local morphology, while the Transformer branch captures relationships across the complete lesion.

3.4 CNN-TRANSFORMER FEATURE PROJECTION

The CNN and Transformer pathways generally produce feature vectors with different dimensions and statistical distributions. Directly concatenating these vectors can allow one branch to dominate the other during optimization. DCCT-Net therefore introduces separate projection layers before contrastive learning and feature fusion. Let FC represent the CNN feature vector and FT represent the Transformer feature vector. Both are mapped into a shared d-dimensional representation space:

$$\begin{aligned} Z_C &= \phi\left(W_C^{(2)}\phi\left(W_C^{(1)}F_C + b_C^{(1)}\right) + b_C^{(2)}\right), \\ Z_T &= \phi\left(W_T^{(2)}\phi\left(W_T^{(1)}F_T + b_T^{(1)}\right) + b_T^{(2)}\right), \end{aligned} \quad (5)$$

where $W_C^{(1)}, W_C^{(2)}$ and $W_T^{(1)}, W_T^{(2)}$ are learnable projection matrices and $\phi(\cdot)$ is the nonlinear activation function.

The projection operation serves two purposes. First, it ensures dimensional compatibility between the two pathways. Second, it produces representations suitable for contrastive optimization. Layer normalization and dropout can be included in this stage to stabilize training and reduce overfitting.

3.5 CONTRASTIVE REPRESENTATION LEARNING

The central novelty of DCCT-Net is the use of contrastive learning to explicitly improve the discriminative structure of the learned feature space. Conventional cross-entropy training encourages the final classifier to assign the correct label but does not directly regulate the distance between intermediate representations. Two melanoma images may therefore produce substantially different embeddings even though they belong to the same class.

The proposed contrastive component constructs positive and negative pairs. Two augmented versions of the same lesion form

a positive pair. Additional samples belonging to the same diagnostic category can also be considered positive samples. Samples associated with different diagnostic classes form negative pairs. For a mini-batch containing *B* samples, the normalized representation of sample *i* is denoted by *zi*. The similarity between two representations is measured using cosine similarity:

$$s(z_i, z_j) = \frac{z_i^{\cdot} z_j}{\| z_i \|_2 \| z_j \|_2},$$

$$z_i = \frac{g(Z_i)}{\| g(Z_i) \|_2}, \tag{6}$$

where $g(\cdot)$ denotes the nonlinear projection head. The contrastive objective then encourages the similarity of positive pairs to increase while reducing the similarity of negative pairs.

An InfoNCE-style contrastive loss is defined as:

$$\mathrm{L}_{\mathrm{con}} = -\frac{1}{|\mathrm{B}|}\sum_{i\in \mathrm{B}} \log \frac{\sum_{p\in \mathrm{P}(i)} \exp\left(s(z_i, z_p)/\tau\right)}{\sum_{a\in \mathrm{A}(i)} \exp\left(s(z_i, z_a)/\tau\right)}, \tag{7}$$

where $P(i)$ denotes the positive set for sample *i*, $A(i)$ denotes the set of candidate comparison samples, and τ is the temperature parameter. The contrastive objective is particularly useful for melanoma classification because visually similar benign lesions can otherwise occupy feature regions close to melanoma samples. By explicitly controlling representation distances, the network learns a more structured embedding space.

3.6 CROSS-PATH REPRESENTATION ALIGNMENT

The proposed model further encourages agreement between the CNN and Transformer pathways for the same lesion. The CNN branch observes fine-grained local structures, whereas the Transformer branch observes broader contextual relationships. Their representations should therefore remain complementary while preserving common lesion identity.

A cross-path alignment objective is introduced to minimize the representation discrepancy between the two branches for corresponding images:

$$\mathrm{L}_{\mathrm{align}} = \frac{1}{B}\sum_{i=1}^{B}\left[1 - \frac{Z_{C,i}^{\cdot} Z_{T,i}}{\| Z_{C,i} \|_2 \| Z_{T,i} \|_2}\right] + \lambda_d \left\|\mathrm{Cov}(Z_C) - \mathrm{Cov}(Z_T)\right\|_F^2, \tag{8}$$

where $Cov(\cdot)$ represents the covariance matrix and $\|\cdot\|_F$ denotes the Frobenius norm. The first component aligns corresponding CNN and Transformer representations, while the second regulates their distributional discrepancy.

This mechanism prevents the two pathways from learning completely independent representations. At the same time, because the feature extractors remain separate, the model can preserve pathway-specific information that would be lost through early fusion.

3.7 ADAPTIVE CNN-TRANSFORMER FEATURE FUSION

After representation learning, DCCT-Net combines the two feature vectors through an adaptive fusion mechanism. Conventional concatenation assigns equal structural importance to both pathways, which may not be optimal for every lesion. For some images, local texture may provide stronger diagnostic evidence, whereas other images may require global shape and spatial relationships. The proposed fusion mechanism generates a learned importance coefficient for each pathway. The coefficients are calculated from the joint representation and normalized using a softmax function:

$$q = \delta\left(W_f [Z_C \| Z_T] + b_f\right), \tag{9}$$

$$[\alpha_C, \alpha_T] = \mathrm{softmax}\left(W_a q + b_a\right), \tag{10}$$

$$F_{\mathrm{fuse}} = \alpha_C Z_C + \alpha_T Z_T,\ \alpha_C + \alpha_T = 1 \tag{11}$$

where $\|$ denotes vector concatenation, $\delta(\cdot)$ is a nonlinear activation function, and α_C and α_T represent the learned contributions of the CNN and Transformer branches.

This fusion strategy allows DCCT-Net to dynamically determine which representation should receive greater emphasis. A lesion with strong local pigment irregularity may produce a larger CNN contribution, whereas a lesion whose diagnostic characteristics depend on distributed structural patterns may receive stronger Transformer weighting.

3.8 CLASSIFICATION LAYER

The fused representation is passed through a multilayer classification head. Dropout is incorporated between fully connected layers to reduce overfitting, particularly when the training dataset is relatively small. The final layer generates class logits, which are converted into posterior probabilities using softmax activation. For *K* diagnostic classes, the predicted probability of class *k* is given by:

$$\hat{y}_k = \frac{\exp\left(w_k^{\cdot} F_{\mathrm{fuse}} + b_k\right)}{\sum_{j=1}^{K} \exp\left(w_j^{\cdot} F_{\mathrm{fuse}} + b_j\right)}, \qquad k \in \{1, 2, \ldots, K\}, \tag{12}$$

where w_k and b_k represent the learnable weight vector and bias for class *k*. For binary melanoma classification, the output can be interpreted as the probability of melanoma and non-melanoma.

3.9 CLASSIFICATION LOSS AND CLASS-IMBALANCE HANDLING

Because melanoma datasets are frequently imbalanced, the proposed framework uses a class-weighted classification loss. This prevents the optimization process from being dominated by the majority class. The weight assigned to each class can be inversely related to its frequency. The weighted cross-entropy loss is defined as:

$$\mathrm{L}_{\mathrm{cls}} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{k=1}^{K} \omega_k y_{i,k} \log\left(\hat{y}_{i,k} + \grave{o}\right), \qquad \omega_k = \frac{N}{K N_k}, \tag{13}$$

where $y_{i,k}$ is the ground-truth indicator for sample *i*, $y_{i,k}$ is the predicted probability, N_k is the number of samples belonging to

class k, N is the total number of training samples, and ϵ is a small numerical constant. This formulation gives greater importance to underrepresented diagnostic categories. Consequently, the network is encouraged to improve melanoma sensitivity rather than relying primarily on majority-class predictions.

3.10 JOINT OPTIMIZATION OF DCCT-NET

The CNN branch, Transformer branch, projection layers, contrastive module, adaptive fusion mechanism, and classifier are optimized jointly. The complete training objective combines the supervised classification loss, contrastive loss, and cross-path alignment loss. The total optimization function is defined as:

$$\begin{aligned} L_{total} &= \lambda_{cls} L_{cls} + \lambda_{con} L_{con} + \lambda_{align} L_{align} \\ &+ \lambda_{reg} \left(\| W_C \|_2^2 + \| W_T \|_2^2 + \| W_f \|_2^2 \right) \end{aligned} \tag{14}$$

where λ_{cls}, λ_{con}, λ_{align}, and λ_{reg} control the contributions of classification, contrastive, alignment, and regularization objectives, respectively.

The classification component ensures that the final prediction corresponds to the correct diagnostic class. The contrastive component improves the geometry of the embedding space, while the alignment component promotes consistency between local and global representations. The regularization component constrains the magnitude of trainable parameters and reduces overfitting. During backpropagation, the total loss propagates through the fusion module into both feature extraction paths. Consequently, the CNN learns local features that contribute to classification, while the Transformer simultaneously learns global relationships that improve discrimination. This joint optimization differentiates DCCT-Net from a simple ensemble in which independently trained models are combined only at the prediction stage.

3.11 OPERATIONAL SEQUENCE

During training, each dermoscopic image first undergoes normalization and stochastic augmentation. Two augmented views of the same image are generated to support contrastive learning. The views are independently processed through the CNN and Transformer pathways. The CNN extracts local morphological information through hierarchical convolutional operations, whereas the Transformer converts the image into patch tokens and models global dependencies using multi-head self-attention. The resulting feature vectors are projected into a shared embedding space. Positive and negative relationships are then established among the embeddings. The contrastive objective encourages representations of the same lesion or diagnostic class to become closer while increasing the distance between representations from different classes. The cross-path alignment mechanism additionally ensures that the CNN and Transformer representations retain a common lesion identity. The aligned features are then supplied to the adaptive fusion module. Instead of applying a fixed combination, the module learns the relative contribution of each branch. The resulting fused representation contains both fine-grained morphological and global contextual information. The classifier subsequently converts this representation into diagnostic probabilities. The overall forward propagation can therefore be summarized mathematically as:

$$\begin{aligned} I \xrightarrow{\mathrm{P}} I' \xrightarrow[[-1mm] f_T(\cdot)]{f_C(\cdot)} \begin{bmatrix} F_C \\ F_T \end{bmatrix} \xrightarrow[[-1mm] g_T(\cdot)]{g_C(\cdot)} \\ \begin{bmatrix} Z_C \\ Z_T \end{bmatrix} \xrightarrow{\mathrm{F_{adaptive}}} F_{\mathrm{fuse}} \xrightarrow{f_{\mathrm{cls}}} \mathbf{y}, \end{aligned} \tag{15}$$

where P denotes preprocessing and augmentation, f_C and f_T represent the CNN and Transformer encoders, g_C and g_T denote the projection functions, F_a represents adaptive fusion, and f_{cls} denotes the classification head.

3.12 MELANOMA PREDICTION

During inference, the contrastive pair-generation process is not required. The test image is first resized and normalized using the same preprocessing parameters applied during training. The image is then simultaneously processed by the CNN and Transformer paths. Their representations are projected and passed through the trained adaptive fusion module. The resulting feature vector is classified using the final softmax layer.

The predicted class is determined according to the maximum posterior probability:

$$\hat{c} = \underset{k \in \{1,\ldots,K\}}{\arg\max}\ P(y = k \mid I) = \underset{k \in \{1,\ldots,K\}}{\arg\max}\ \hat{y}_k. \tag{16}$$

For binary classification, the melanoma probability can additionally be compared with a selected decision threshold rather than relying exclusively on 0.5. This is useful when sensitivity is prioritized because missing a melanoma case can have greater clinical consequences than incorrectly flagging a benign lesion.

Algorithm

Input: Dermoscopic training set

Output: Optimized DCCT-Net model

1. Resize and normalize every dermoscopic image.
2. Generate augmented views of each training image.
3. Send each view to the CNN pathway.
4. Extract hierarchical local morphological features.
5. Send the same view to the Transformer pathway.
6. Divide the image into patches and generate positional embeddings.
7. Apply multi-head self-attention and Transformer encoding.
8. Extract CNN and Transformer feature vectors.
9. Project both vectors into the common embedding space.
10. Construct positive and negative representation pairs.
11. Calculate the contrastive loss.
12. Calculate cross-path representation alignment loss.
13. Generate adaptive CNN and Transformer fusion weights.
14. Construct the fused lesion representation.
15. Predict diagnostic probabilities through classification head.
16. Calculate weighted classification loss.
17. Combine all losses into the total objective.
18. Backpropagate the total loss through both pathways.
19. Update model parameters using an optimizer such as AdamW.
20. Repeat the process until convergence.

21. During testing, process each image through trained dual-path network and output the predicted melanoma probability.

4. RESULTS AND DISCUSSION

4.1 EXPERIMENTAL SETTINGS

The proposed DCCT-Net is evaluated using Python with PyTorch as the simulation framework. The experiments are performed on a workstation equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA GPU with 8 GB memory. The operating environment uses Windows 11 with CUDA-enabled GPU acceleration. Dermoscopic images are resized to 224×224 pixels and divided into training, validation, and testing subsets. AdamW is used for optimization with a learning rate of 0.0001, a batch size of 32, and 50 training epochs. The proposed model is evaluated using accuracy, precision, recall, F1-score, and specificity.

4.2 EXPERIMENTAL SETUP AND PARAMETERS

The principal experimental parameters used for implementing DCCT-Net are presented in Table.1. These parameters provide a consistent training configuration for both the CNN and Transformer pathways.

Table.1. Experimental setup and parameter configuration

Parameter	Configuration
Simulation tool	Python 3.11, PyTorch
Operating system	Windows 11
Processor	Intel Core i7
RAM	16 GB
Input image size	224×224 pixels
CNN pathway	ResNet-based conv encoder
Transformer pathway	Vision Transformer encoder
Transformer patch size	16×16
Transformer embedding dimension	768
Attention heads	12
Transformer layers	12
Projection dimension	256
Batch size	32
Optimizer	AdamW
Initial learning rate	0.0001
Weight decay	0.0001
Training epochs	50
Dropout	0.30
Contrastive temperature	0.07
Classification loss	Weighted cross-entropy
Contrastive loss weight	0.30
Alignment loss weight	0.20
Training/validation/testing split	70%/15%/15%

As shown in Table.2, the experimental configuration provides sufficient computational capacity for simultaneous CNN and Transformer processing. The 224×224 image resolution maintains lesion-level details while controlling computational cost. The 16-pixel Transformer patch size generates 196 image tokens, which enables the self-attention mechanism to capture relationships among different lesion regions.

4.3 PERFORMANCE METRICS

The metrics are used to evaluate melanoma classification performance. These metrics collectively measure overall correctness, positive prediction reliability, melanoma detection capability, balanced classification performance, and correct identification of non-melanoma cases.

4.3.1 Accuracy:

Accuracy represents the proportion of correctly classified samples among all tested samples. It provides an overall measure of classification effectiveness.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{17}$$

where *TP*, *TN*, *FP*, and *FN* denote true positive, true negative, false positive, and false negative predictions, respectively.

4.3.2 Precision:

Precision measures the proportion of correctly predicted melanoma samples among all samples predicted as melanoma. A high precision value indicates that the model generates fewer false melanoma alarms.

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{18}$$

4.3.3 Recall:

Recall, also called sensitivity, measures the proportion of actual melanoma samples correctly identified by the classifier. This metric is particularly important because failure to detect melanoma can have serious clinical consequences.

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{19}$$

4.3.4 F1-Score:

F1-score provides a harmonic balance between precision and recall. It is useful when the dataset contains class imbalance because it does not rely solely on overall classification accuracy.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

4.3.5 Specificity:

Specificity measures the proportion of non-melanoma cases correctly identified as negative. It is particularly relevant for reducing unnecessary referrals or additional diagnostic procedures caused by false-positive predictions.

$$Specificity = \frac{TN}{TN + FP} \times 100 \tag{21}$$

4.4 DATASET DESCRIPTION

The experimental evaluation uses the HAM10000 dataset, which is a widely used dermoscopic image dataset containing 10,015 images across seven diagnostic categories. The dataset

includes melanoma, melanocytic nevi, basal cell carcinoma, actinic keratoses, benign keratosis-like lesions, dermatofibroma, and vascular lesions. The images originate from multiple sources and therefore contain variations in lesion appearance, acquisition conditions, and image quality. For the present binary melanoma experiment, melanoma images are assigned to the positive class, while the remaining diagnostic categories constitute the non-melanoma class. The dataset is divided into training, validation, and testing subsets using a stratified procedure so that both classes maintain representative distributions.

Table.2. HAM10000 dataset description

Category	Diagnostic type	Number of images	Classification role
Melanocytic nevi	NV	6,705	Non-melanoma
Melanoma	MEL	1,113	Melanoma
Benign keratosis	BKL	1,099	Non-melanoma
Basal cell carcinoma	BCC	514	Non-melanoma
Actinic keratoses	AKIEC	327	Non-melanoma
Vascular lesions	VASC	142	Non-melanoma
Dermatofibroma	DF	115	Non-melanoma
Total	7 categories	10,015	Binary classification

The Table.2 demonstrates the substantial class imbalance in the dataset. Melanocytic nevi constitute the largest category, whereas dermatofibroma and vascular lesions contain considerably fewer samples. Therefore, the weighted classification loss described previously is used to prevent the classifier from being biased toward the majority class.

Various methods are selected for comparison. Dermatologist-Level Deep Neural Network (DL-DNN) represents conventional deep CNN-based classification. Vision Transformer (ViT) represents global self-attention-based image classification. Hybrid U-Net-Inception-ResNet-ViT represents a contemporary hybrid architecture combining segmentation, convolutional feature extraction, and Transformer-based feature refinement. DCCT-Net is compared with these methods under the same evaluation metrics.

4.5 RESULTS BASED ON ACCURACY

Accuracy is evaluated across 50 training epochs. The results are provided in Table.3.

Table.3. Accuracy comparison across training epochs

Epoch	DL-DNN	ViT	Hybrid U-Net-Inception-ResNet-ViT	Proposed DCCT-Net
5	88.2	87.6	89.4	91.2
10	90.1	89.8	91.2	93.4
15	91.8	91.1	92.7	94.8
20	93.0	92.5	94.0	95.7
25	94.1	93.6	94.9	96.5
30	94.8	94.5	95.6	97.1
35	95.3	95.1	96.1	97.6
40	95.7	95.6	96.5	98.0
45	96.0	95.9	96.8	98.2
50	96.2	96.1	97.0	98.4

The Table.4 shows that DCCT-Net consistently achieves higher accuracy throughout the training process. At epoch 5, the proposed model reaches 91.2%, compared with 88.2%, 87.6%, and 89.4% for DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT, respectively. At epoch 25, DCCT-Net reaches 96.5%, providing improvements of 2.4, 2.9, and 1.6% over the three existing methods. The final accuracy of DCCT-Net reaches 98.4% at epoch 50, while DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT achieve 96.2%, 96.1%, and 97.0%, respectively. Thus, DCCT-Net provides improvements of 2.2, 2.3, and 1.4%. The faster convergence is attributed to the complementary CNN and Transformer pathways. CNN features identify fine-grained lesion structures, while Transformer features capture global relationships. The contrastive objective further improves the separation of melanoma and non-melanoma embeddings.

4.6 RESULTS BASED ON PRECISION

Precision results are reported at five-epoch intervals in Table.5. The metric indicates the reliability of positive melanoma predictions.

Table.4. Precision comparison across training epochs

Epoch	DL-DNN	ViT	Hybrid U-Net-Inception-ResNet-ViT	Proposed DCCT-Net
5	86.8	86.1	88.0	90.3
10	89.0	88.4	90.2	92.5
15	90.8	90.0	91.8	94.0
20	92.1	91.5	93.1	95.0
25	93.2	92.6	94.1	95.8
30	94.0	93.5	94.9	96.5
35	94.5	94.2	95.4	97.1
40	95.0	94.8	95.8	97.6
45	95.4	95.2	96.2	97.9
50	95.7	95.6	96.5	98.1

As shown in Table.4, the proposed method produces the highest precision at every evaluated epoch. At epoch 10, DCCT-Net achieves 92.5%, whereas the three comparison methods obtain 89.0%, 88.4%, and 90.2%. At epoch 30, the proposed approach reaches 96.5%, exceeding DL-DNN by 2.5% and Vision Transformer by 3.0%. At epoch 50, DCCT-Net achieves 98.1% precision. The corresponding values of DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT are 95.7%, 95.6%, and 96.5%. The improvements are therefore 2.4, 2.5, and 1.6%, respectively. The improvement indicates that adaptive feature fusion reduces false-positive melanoma predictions. The contrastive representation objective contributes by increasing the separation between melanoma and visually similar benign lesions.

4.7 RESULTS BASED ON RECALL

Recall represents the melanoma detection capability of the models. Since melanoma is the positive class, a higher recall indicates that a greater proportion of actual melanoma cases are successfully detected.

Table.6. Recall comparison across training epochs

Epoch	DL-DNN	ViT	Hybrid U-Net-Inception-ResNet-ViT	Proposed DCCT-Net
5	84.9	84.2	86.3	88.7
10	87.3	86.5	88.5	91.0
15	89.2	88.1	90.4	92.8
20	90.7	89.8	91.9	94.0
25	91.8	91.0	93.0	95.0
30	92.7	92.0	93.8	95.8
35	93.5	92.9	94.6	96.4
40	94.1	93.7	95.2	97.0
45	94.5	94.1	95.7	97.5
50	94.8	94.5	96.0	**97.9**

The Table.6 indicates a substantial improvement in melanoma detection using DCCT-Net. At epoch 5, the proposed method achieves 88.7% recall compared with 84.9%, 84.2%, and 86.3% for the three baseline approaches. The difference becomes more pronounced as training progresses. At epoch 25, DCCT-Net reaches 95.0%, while the existing methods achieve 91.8%, 91.0%, and 93.0%. The proposed model achieves a final recall of 97.9%, which is 3.1% higher than DL-DNN, 3.4% higher than Vision Transformer, and 1.9% higher than Hybrid U-Net-Inception-ResNet-ViT. This improvement is important because false-negative melanoma predictions are particularly undesirable. The CNN pathway captures local irregularities, while global Transformer attention identifies relationships between distant lesion regions. Their combination therefore improves sensitivity to heterogeneous melanoma patterns.

4.8 RESULTS BASED ON F1-SCORE

F1-score provides a balanced evaluation of precision and recall. The comparative results are presented in Table.7.

Table.7. F1-score comparison across training epochs

Epoch	DL-DNN	ViT	Hybrid U-Net-Inception-ResNet-ViT	Proposed DCCT-Net
5	85.8	85.1	87.1	89.5
10	88.1	87.4	89.3	91.7
15	90.0	89.0	91.1	93.4
20	91.4	90.6	92.5	94.5
25	92.5	91.8	93.5	95.4
30	93.3	92.7	94.3	96.1
35	94.0	93.5	95.0	96.7
40	94.5	94.2	95.5	97.3
45	94.9	94.6	95.9	97.7
50	95.2	95.0	96.2	**98.0**

The results in Table.7 demonstrate that DCCT-Net maintains a stronger balance between precision and recall than all three comparison methods. At epoch 10, the proposed approach achieves 91.7% F1-score compared with 88.1% for DL-DNN, 87.4% for Vision Transformer, and 89.3% for the hybrid baseline. At epoch 30, the proposed method reaches 96.1%, exceeding the three baselines by 2.8, 3.4, and 1.8%. At the final epoch, DCCT-Net obtains an F1-score of 98.0%, compared with 95.2%, 95.0%, and 96.2%. The respective gains are 2.8, 3.0, and 1.8%. This result indicates that the proposed architecture does not improve recall at the expense of precision. Instead, the contrastive objective simultaneously improves both aspects of classification, producing a more compact intra-class representation and a larger separation between competing diagnostic classes.

4.9 RESULTS BASED ON SPECIFICITY

Specificity evaluates the ability of each model to correctly identify non-melanoma cases. Higher specificity indicates fewer false-positive classifications.

Table.8. Specificity comparison across training epochs

Epoch	DL-DNN	ViT	Hybrid U-Net-Inception-ResNet-ViT	Proposed DCCT-Net
5	90.5	89.8	91.2	93.0
10	92.0	91.4	92.8	94.5
15	93.2	92.7	94.0	95.7
20	94.0	93.5	94.8	96.5
25	94.8	94.2	95.5	97.1
30	95.3	94.9	96.1	97.5
35	95.8	95.4	96.5	98.0
40	96.2	95.8	96.9	98.4
45	96.5	96.2	97.2	98.7
50	96.8	96.5	97.5	**99.0**

The Table.8 demonstrates that DCCT-Net produces the highest specificity throughout training. At epoch 5, the proposed method reaches 93.0%, compared with 90.5%, 89.8%, and 91.2% for DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT. At epoch 25, specificity increases to 97.1%, exceeding the baselines by 2.3, 2.9, and 1.6%. At epoch 50, DCCT-Net achieves 99.0% specificity, whereas DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT obtain 96.8%, 96.5%, and 97.5%. The resulting gains are 2.2, 2.5, and 1.5%. This result demonstrates that the proposed architecture can distinguish benign lesions from melanoma with fewer false-positive predictions. The improvement is consistent with the adaptive fusion mechanism, which prevents the classification decision from depending excessively on either local texture or global contextual information.

4.10 DISCUSSION OF RESULTS

The final values are summarized through the highest observed performance at epoch 50. DCCT-Net achieves 98.4% accuracy, 98.1% precision, 97.9% recall, 98.0% F1-score, and 99.0% specificity. These values indicate that the model provides

balanced performance rather than optimizing only one classification criterion. The accuracy results in Table.4 show a gain of 2.2% over DL-DNN and 2.3% over Vision Transformer. The improvement over Hybrid U-Net-Inception-ResNet-ViT is 1.4%. Similarly, Table.5 demonstrates a precision improvement of 2.4, 2.5, and 1.6%. These results suggest that DCCT-Net reduces false-positive melanoma classifications. The recall results in Table.6 are particularly important. The proposed model achieves 97.9%, which represents improvements of 3.1, 3.4, and 1.9% over the comparison methods. This indicates stronger melanoma detection capability. The Table.7 further confirms that the improvement is balanced, with an F1-score of 98.0% compared with 95.2%, 95.0%, and 96.2%. Finally, Table.8 shows that specificity reaches 99.0%, providing gains of 2.2, 2.5, and 1.5%. The consistent improvements across Table.4-Table.8 indicate that the proposed contrastive dual-path strategy provides benefits beyond conventional architectural complexity. The CNN pathway preserves fine-grained lesion information, while the Transformer pathway captures global dependencies. Contrastive learning improves feature separability, and adaptive fusion selects complementary information from both pathways.

5. CONCLUSION

This study presents a DCCT-Net for automated melanoma classification from dermoscopic images. The proposed architecture addresses important limitations of conventional CNN and Transformer classifiers by combining local morphological representation, global contextual modeling, contrastive feature learning, and adaptive feature fusion. The CNN pathway extracts fine-grained lesion characteristics, whereas the Transformer pathway establishes relationships between spatially distributed lesion regions. The contrastive learning mechanism further improves the organization of the embedding space by increasing intra-class similarity and inter-class separation. The experimental evaluation using the HAM10000 dataset demonstrates consistent improvements over the selected DL-DNN, Vision Transformer, and Hybrid U-Net-Inception-ResNet-ViT approaches. DCCT-Net obtains 98.4% accuracy, 98.1% precision, 97.9% recall, 98.0% F1-score, and 99.0% specificity. The recall improvement is particularly significant because accurate identification of melanoma is more important than achieving high accuracy through majority-class predictions. The high specificity additionally indicates effective suppression of false-positive classifications.

REFERENCES

[1] A. Lin, M. Motwani, P. Maurovich-Horvat, P.J. Slomka and D. Dey, “Artificial Intelligence in Cardiovascular Imaging for Risk Stratification in Coronary Artery Disease”, *Radiology: Cardiothoracic Imaging*, Vol. 3, No. 1, pp. 200512-200532, 2021.

[2] S. Friedrich, S. Engelhardt, M. Bahls and T. Friede, “Applications of Artificial Intelligence/Machine Learning Approaches in Cardiovascular Medicine: A Systematic Review with Recommendations”, *European Heart Journal-Digital Health*, Vol. 2, No. 3, pp. 424-436, 2021.

[3] M. Chu, P. Wu, G. Li and S. Tu, “Advances in Diagnosis, Therapy, and Prognosis of Coronary Artery Disease Powered by Deep Learning Algorithms”, *Jacc: Asia*, Vol. 3, No. 1, pp. 1-14, 2023.

[4] S. Yaman, O. Aslan, A. Sengur, A. Hafeez-Baig and U.R. Acharya, “Deep Learning Techniques for Automated Coronary Artery Segmentation and Coronary Artery Disease Detection: A Systematic Review of the Last Decade (2013-2024)”, *Computer Methods and Programs in Biomedicine*, Vol. 268, pp. 108858-108865, 2025.

[5] G. Manimaran, A. Peimankar, S. Puthusserypady and A. Ebrahimi, “Explainable Deep Learning based Techniques for ECG-Based Heart Disease Classification: A Systematic Literature Review and Future Direction”, *Computers in Biology and Medicine*, Vol. 199, pp. 111324-111335, 2025.

[6] J. Wang, Q. Xue, C.W. Zhang and Z. Liu, “Explainable Coronary Artery Disease Prediction Model based on AutoGluon from AutoML Framework”, *Frontiers in Cardiovascular Medicine*, Vol. 11, pp. 1360548-1360562, 2024.

[7] A.A. Huang and S.Y. Huang, “Use of Machine Learning to Identify Risk Factors for Coronary Artery Disease”, *PloS One*, Vol. 18, No. 4, pp. 284103-284114, 2023.

[8] H.G. Lee, S.D. Park, J.W. Bae and S. Moon, “Machine Learning Approaches that Use Clinical, Laboratory, and Electrocardiogram Data Enhance the Prediction of Obstructive Coronary Artery Disease”, *Scientific Reports*, Vol. 13, No. 1, pp. 12635-12647, 2023.

[9] W. Yu, L. Yang, F. Zhang, B. Liu and Y. Wang, “Machine Learning to Predict Hemodynamically Significant CAD based on Traditional Risk Factors, Coronary Artery Calcium and Epicardial Fat”, *Journal of Nuclear Cardiology*, Vol. 30, No. 6, pp. 2593-2606, 2023.

[10] M. Zhang, H. Wang and J. Zhao, “Use Machine Learning Models to Identify and Assess Risk Factors for Coronary Artery Disease”, *Plos One*, Vol. 19, No. 9, pp. 307952-307965, 2024.

[11] A. Ainiwaer, W.Q. Hou, K. Kadier, R. Rehemuding and J.G. Dai, “A Machine Learning Framework for Diagnosing and Predicting the Severity of Coronary Artery Disease”, *Reviews in Cardiovascular Medicine*, Vol. 24, No. 6, pp. 168-175, 2023.

[12] N. Tasmurzayev, B. Imanbek, A. Boltaboyeva, G. Dikhanbayeva and G. Amirkhanova, “Explainable AI for Coronary Artery Disease Stratification using Routine Clinical Data”, *Algorithms*, Vol. 18, No. 11, pp. 693-708, 2025.

[13] N. Tasmurzayev, Z. Baigarayeva, B. Amangeldy, B. Imanbek, S. Kurmanbek and G. Amirkhanova, “Interpretable Machine Learning for Coronary Artery Disease Risk Stratification: A SHAP-Based Analysis”, *Algorithms*, Vol. 18, No. 11, pp. 697-712, 2025.

[14] M. Jaradat and M. Awad, “Explainable AI in Cardiology Diagnostics: A Systematic Review of Machine Learning, Meta-Heuristic Optimization, and Clinical Text Mining for Coronary Artery Disease”, *International Journal of Medical Informatics*, Vol. 45, No. 1, pp. 106321-106345, 2026.