International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(11s) 2026 pp. 859-885
https://doi.org/10.70917/ijcisim-2026-3820
Article

Topology-Aware Hierarchical Graph Vision Transformer with Structural Tokenization for Explainable Coronary Artery Stenosis Detection

R.Velvizhi[1] B.Ankayarkanni[2]

[1,2]Department of Computer Science and Engineering Sathyabama Institute of Science and Technology Chennai,Tamilnadu 600119
[1]velvizhisp@gmail.com,[2]ankayarkanni.cse@sathyabama.ac.in

Abstract: Coronary artery stenosis is a major cardiovascular condition that restricts blood flow due to abnormal vessel narrowing and requires accurate and timely diagnosis. However, automated stenosis detection from coronary angiography images remains challenging because of complex vascular topology, multi-scale vessel structures, and subtle morphological variations. Existing convolutional neural network and transformer-based approaches often fail to preserve vascular connectivity and hierarchical structural relationships, limiting detection accuracy and interpretability. To address these challenges, this study proposes a novel Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning (GGH-ViT-TSL) framework for explainable coronary artery stenosis detection. The proposed framework integrates hierarchical CNN–ViT feature extraction, vessel segmentation, topology-preserving graph construction, and a newly designed Structural Tokenization mechanism that transforms vascular topology into graph-guided semantic tokens. These tokens are processed using a Graph Transformer to capture long-range inter-vessel dependencies and multi-scale stenotic characteristics, followed by an explainable detection module for stenosis classification and localization. The framework is implemented in Python and evaluated using the ARCADE coronary angiography dataset. Experimental results demonstrate that the proposed method achieves 98.2% accuracy, 98.1% precision, 98.3% recall, 98.2% F1-score, and an AUC of 0.990. In addition, vessel segmentation attains a Dice score of 0.965 and an IoU of 0.934. The proposed framework outperforms conventional CNN-based, graph-based, and transformer-based methods, confirming its effectiveness in preserving vascular topology and enhancing diagnostic interpretability. These findings demonstrate the potential of topology-aware graph-transformer learning for advanced medical image analysis and intelligent computer-aided cardiovascular diagnosis.

Keywords: Coronary Artery Stenosis Detection, Graph Vision Transformer, Structural Tokenization, Topology-Aware Learning, Medical Image Analysis.

1. Introduction

CAD is a leading cardiovascular disease that plays a significant role in the global burden of disease and death [1]. Coronary artery stenosis occurs when the coronary artery narrows as a result of plaque accumulation, limiting blood flow and thus increasing the risk of myocardial infarction and other serious cardiac complications [2]. Early and accurate diagnosis of stenotic lesion is critical for timely clinical treatment and effective management of the patient [3]. X-ray coronary angiography is still the most useful imaging technique to assess coronary artery abnormalities due to its ability to offer detailed imaging of the morphology and blood flow characteristics of the coronary arteries [4]. However, the increasing volume of angiographic examinations has rapidly grown, however, an automated and intelligent diagnostic system to help the clinician to make accurate stenosis assessment is needed [5].

Recent developments in signal processing, computer vision, and artificial intelligence have improved automated coronary angiography analysis greatly [6]. Traditional machine learning techniques used handcrafted vessel features which were not sufficient to capture the complexity of the vascular morphology [7]. Automated feature extraction was achieved for vessel segmentation and stenosis detection using deep learning techniques, specifically CNNs [8]. More

recently, Vision Transformers (ViTs), hybrid CNN–Transformer models, explainable artificial intelligence (XAI), and graph-based learning models have further advanced the global context modeling, interpretability, and anatomical representation learning [9], [10]. However, most existing approaches fail to preserve the topology and connectivity of coronary vessels well enough for accurate characterization of complex stenotic patterns and vascular relationships [11].

Coronary artery stenosis is a major contributor to cardiovascular morbidity and mortality worldwide, and accurate and early diagnosis from CTA and MRA images is essential [12]. The challenge of reliable detection of stenosis, however, still remains because of the complex vascular topology and the intricate branching structures, multi-scale vessel morphology, and subtle lumen narrowing patterns which can be difficult to distinguish from normal anatomical variation [13]. Existing approaches often fail to maintain vessel connections, hierarchical structure of anatomical relationships and topological relationships when learning features, which leads to decreased diagnostic value and poor clinical explainability [14], [15].

To address these challenges, this study proposes a Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework for automated coronary artery stenosis detection. The incorporation of hierarchical feature extraction, global modelling using a transformer, vessel-guided graph construction and topology-aware learning framework captures the morphology and connectivity of vessels, resulting in accurate stenosis characterization and enhanced diagnostic performance with the ARCADE coronary angiography dataset. The main contribution of the study are as follows:

- ✓ A novel Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework is proposed for automated coronary artery stenosis detection from X-ray coronary angiography images.
- ✓ A hierarchical CNN–Vision Transformer encoder is developed to simultaneously capture fine-grained local vessel characteristics and long-range global vascular dependencies, enabling comprehensive multi-scale feature representation.
- ✓ A topology-preserving graph construction strategy is introduced to transform segmented coronary vessels into structured vascular graphs, explicitly modeling vessel connectivity, branching patterns, and anatomical relationships.
- ✓ A Graph Transformer-based topology-aware learning mechanism is employed to exploit graph-structured vessel representations and capture complex inter-vessel dependencies for enhanced stenosis characterization and classification.
- ✓ Experiments on the ARCADE coronary angiography dataset demonstrate that the proposed framework improves stenosis detection accuracy while generating anatomically meaningful vascular representations for clinical decision support.

The remainder of this paper is organized as follows. Section II reviews the related literature on coronary artery stenosis detection and analysis. Section III presents the proposed Topology-Aware Hierarchical Graph Vision Transformer with Structural Tokenization framework and describes its methodology in detail. Section IV discusses the experimental results, performance evaluation, and comparative analysis. Finally, Section V concludes the study and outlines potential directions for future research.

2. Related Works

Han et al. [16] introduced a Transformer-based stenosis detection model that leverages Proposal-Shifted Spatio-Temporal Tokenization and Transformer Feature Aggregation to fully exploit long-range spatio-temporal information in X-ray angiography sequences. They obtained F1-score of 90.88%, thus showing better detection performance. Similarly, Jungiewicz et al.[17] investigated Inception, ViT, and Sharpness-Aware Minimization Vision Transformer (SAM-ViT) models for coronary stenosis classification using angiography image fragments. The SAM-ViT model was more stable and robust than the conventional ViT model with regard to the F1 mean performance. In both studies, transformer-based learning was found to be effective for analysis of coronary angiography. Both methods, however, do not explicitly model the topology of the coronary vessels or the structural connectivity, which is important for correct and explainable detection of stenosis.

Li et al.[18] proposed a two-stage deep learning model for CAD diagnosis of CCTA images, which includes U-Net for the segmentation of the coronary arteries and 3DNet for CAD classification. The model had a Dice score of 0.771 ± 0.021 and an AUC of 0.737, which indicated efficient automatic segmentation and diagnosis. Lin et al.[19]

proposed StenUNet for automated stenosis detection from coronary angiography images using machine learning and computer vision techniques. The model was tested on ARCADE challenge data, performed with an F1 score of 0.5348, and placed third in the competition. Both studies show the potential of AI in coronary stenosis detection, but they mainly use image appearance features and do not explicitly model vessel topology or structural relationships.

For automatic coronary artery stenosis localization from X-ray coronary angiography images, Chen et al.[20] proposed a hybrid CNN framework in which they fused temporal features and added attention mechanisms to the framework. This approach used deformable convolution, temporal fusion with correlation, and channel-spatial attention modules to encode the dynamic characteristics of vessels and improve the representation of stenotic features, and showed better recall performance in two datasets. Huang et al.[21] proposed a deep learning-based method for coronary artery segmentation using MedSAM and VM-UNet, followed by the extraction of the centerline and the quantification of stenosis by measuring the diameters. On mixed datasets, the model had an average IoU of 0.6308, sensitivity of 0.9772, and specificity of 0.9903. Both studies showed good diagnostic performance, but mainly centered on temporal feature enhancement or segmentation-based measurements, without explicitly modelling hierarchical vascular topology or structural vessel relationships.

Zhang et al.[22] proposed the Coronary Proposal-based Graph Convolutional Network for automated coronary stenosis detection from CCTA images. This framework involved using CNN to extract features and adding graph convolutional networks, with the candidate stenotic segments as nodes in the graph and the spatial relationships between them as the edges. Talukder et al.[23] proposed a Hybrid Explainable Artificial Intelligence-based Machine Learning method using data balancing approaches along with SHAP, LIME and Permutation Importance Analysis to improve the transparency of predictions. The model was found to be 98.78% accurate when applied to the Cleveland and 99.24% when applied to the Framingham data sets. But both studies fail to combine topology-aware graph learning, global context modeling by transformer and explainable stenosis localization from coronary angiographic images simultaneously.

Yin et al.[24] proposed a Fast Convolution Compression Network for automated coronary artery disease severity classification using heart sound recordings. A Large Tied One-Shot Aggregation Convolution module was added to the framework to ensure that the architecture was lightweight and features were utilized. The experimental results for 150 clinical recordings showed good accuracy, sensitivity and specificity of 85.82%, 85.3% and 86.26%, respectively, using just 1.9 million parameters. Zhang et al.[25] created a multimodal prediction model which combines facial features, cardiovascular waveforms, biochemical indicators, and clinical information to assist with non-invasive coronary stenosis evaluation using a transformer network. The model achieved over 90% training accuracy and 85% validation accuracy. However, both studies are based on indirect biomarkers and not specifically on the coronary artery topography, the vessel relationships or the localization of the stenosis.

Chen et al.[26] proposed a Hierarchical Heterogeneous Aggregation Network for detecting multi-shape coronary artery stenosis from X-ray angiography sequences. The framework adopted a Channel Importance-guided Fusion module and a Hierarchical Heterogeneous Aggregator to combine both spatial and temporal information, improve the representation of small stenosis and ensure the temporal consistency. Experimental results showed better accuracy in detection and generalization on two clinical datasets than the previous stenosis detection methods. Zhao et al. [27] proposed the Transformer Multi-branch Network that integrates augmented 2.5D processing, transformer-based 3D feature extraction, multi-scale fusion, and attention mechanisms to segment the coronary arteries. The model outperforms the existing methods by 1.63% and 3.22% in terms of Dice Similarity Coefficient. Nevertheless, both studies focus mainly on feature aggregation or segmentation performance, and do not explicitly include topology-aware graph learning to detect explainable stenosis.

Table 1. Summary of Existing Coronary Artery Stenosis Detection Studies

References	Method	Dataset	Key Results	Limitation
Han et al. [16]	Transformer-based Feature Aggregation with Proposal-Shifted Spatio-Temporal Tokenization	233 XRA sequences	F1-score = 90.88%	Limited topology-aware vessel modeling

Jungiewicz et al.[17]	Inception, ViT and SAM-ViT	Coronary angiography fragments	SAM-ViT showed improved stability	Small image fragments; no structural vessel representation
Li et al.[18]	U-Net + 3DNet	CCTA dataset	Dice = 0.771, AUC = 0.737	Focuses on segmentation and classification separately
Lin et al.[19]	StenUNet	ARCADE Dataset	F1 = 0.5348	Limited global contextual understanding
Chen et al.[20]	CNN with Temporal Fusion and Attention	XCA datasets	Best average recall performance	CNN struggles with long-range vessel dependencies
Huang et al.[21]	MedSAM + VM-UNet	ARCADE, DCA1, GH	IoU = 0.6308	Stenosis detection relies on post-processing measurements
Zhang et al.[22]	Coronary Proposal-based GCN	259 CCTA cases	Accuracy = 0.844, AUC = 0.74	Limited integration of transformer-based global learning
Talukder et al.[23]	Explainable Hybrid ML	Cleveland, Framingham	Accuracy = 99.24%	Not designed for angiographic image analysis
Yin et al.[24]	Fast Convolution Compression Network	150 heart sound recordings	Accuracy = 85.82%	Uses auscultation signals rather than coronary images
Zhang et al.[25]	Multimodal Transformer	488 CAD patients	>90% training accuracy	Does not utilize coronary vessel topology
Chen et al.[26]	Hierarchical Heterogeneous Aggregation Network	Clinical XRA datasets	Improved detection accuracy	Lacks explicit graph-based topology learning
Zhao et al. [27]	Transformer Multi-branch Network	CAT08, ASOCA	DSC improvement of 1.63–3.22%	Focused on segmentation rather than stenosis localization

Despite significant advances in coronary artery stenosis detection, using CNNs, transformers, graph neural networks, and segmentation-based frameworks, several limitations remain shown in table 1. Existing methods mainly rely on image appearance features, temporal information, or vessel segmentation, without explicitly modeling the hierarchical topology of coronary artery networks. Transformer-based models offer long-range feature interactions and lack global contextual learning, while graph-based approaches are useful for capturing vessel relationships but do not include the long-range feature interactions. Moreover, traditional patch-based tokenization algorithms are not able to capture anatomic vessel structure and the stenotic pattern. While explainability has been addressed in some works, there are still only a few studies that focus on topology-aware decision mechanisms that are interpretable. Hence, there is a need for a unified framework to combine the hierarchical graph representation, topology-aware learning, structural tokenization, and explainable transformer architectures to accurately and clinically interpretable detect the coronary artery stenosis.

3. Proposed Methodology for Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework

The proposed Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework is designed for accurate and explainable coronary artery stenosis detection using the ARCADE X-ray angiography dataset shown in figure 1. The approach tackles important issues like complex branching of vessels, multi-scale vessel morphology, and missing anatomical connectivity in traditional deep learning frameworks. Preprocessed angiographic images are first fed into a Hierarchical Vision Encoder, which combines the convolution-based local feature extraction with transformer-based global dependency modeling, to encode fine vessel boundaries and long-range vascular relationships. The multi-scale feature maps extracted are then passed to a vessel segmentation module to obtain accurate coronary vessel masks and centreline structures. These structures are used to build a graph representation in which the nodes are the bifurcation points and the ends of the vessels and the edges are the anatomical connections between the vessel segments. A novel Structural Tokenization mechanism transforms this graph into hierarchical tokens, vessel-level, branch level and global vascular context. These tokens are passed through a Graph Transformer, which learns the topology-aware attention, thus preserving the connectivity of the vessels and the patterns of stenotic progression. Finally, a stenosis detection head is able to classify and localize regions of interest while keeping the model explainable using attention-based visualization that is aligned with the vascular topology.

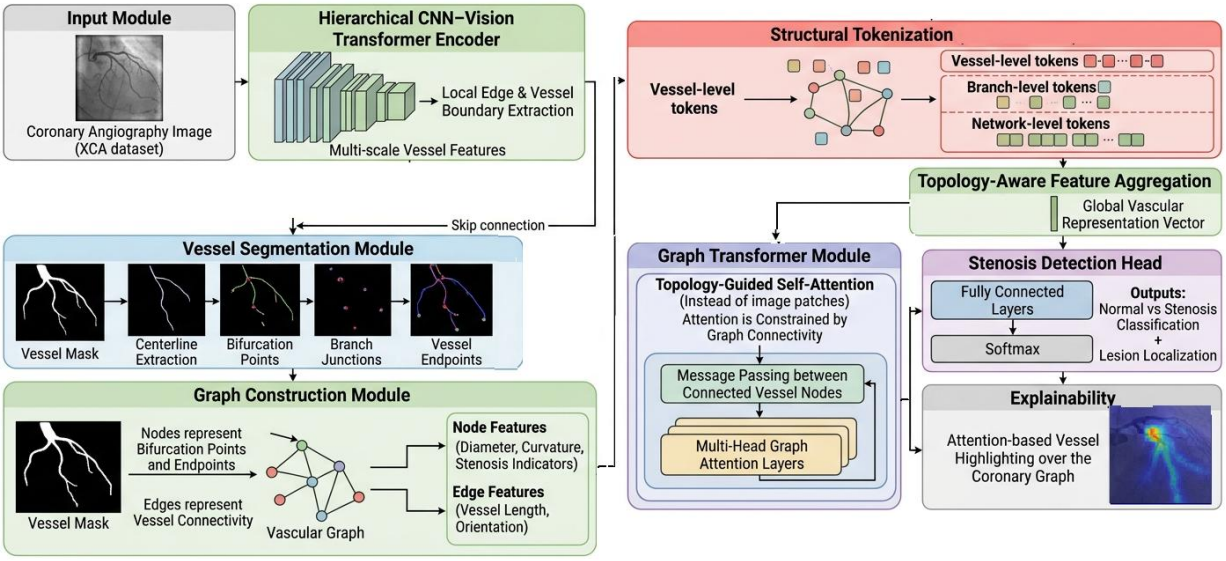

Figure 1. Proposed Framework for Coronary Artery Stenosis Detection

3.1 Data Acquisition

The proposed study will be based on the publicly available ARCADE (Automatic Region-based Coronary Artery Disease Diagnostics Using X-ray Angiography Images) database which is available in Zenodo repository [28]. The data is from 3000 anonymized X-ray coronary angiography (XCA) images of 512 × 512 resolution. It consists of two subsets: 1,500 images for coronary vessel segmentation and SYNTAX methodology based vessel classification; 1,500 images for detection and localization of stenosis. Each subset is divided into 1000 training images, 200 validation images and 300 testing images. The dataset contains expert annotated vessel masks, stenosis lesion annotations, and vessel branch labels in json format. The angiographic images were obtained from actual clinical cases and were manually validated by expert cardiologists. The XCA images are used in this study as the input for the proposed GGH-ViT-TSL framework for vessel segmentation, graph construction, topology-aware learning, and automatic detection of coronary artery stenosis.

3.2 Data Preprocessing

The ARCADE dataset is comprised of X-ray coronary angiography images with annotations of the vessels' region of the image and whether or not there is a stenosis, which are provided in JSON format including the bounding boxes and segmentation masks. A consistent preprocessing pipeline is used to ensure that the same preprocessing is applied to each input, which helps to make the model training more stable. To ensure a uniform spatial resolution of all angiographic images, while keeping the geometric consistency of coronary vessel structures and reducing the scale-related variations within the dataset, all images are resized in the first step. This enables the hierarchical encoder to deal with structurally aligned inputs effectively.

Then, the intensity normalization is applied to remove intensity variations resulting from differences in imaging conditions, contrast injection dose, and acquisition noise. Min–max normalization transforms pixel values into a fixed range according to Equation(1).

$$I_{norm}(x,y) = \frac{I(x,y)-I_{min}}{I_{max}-I_{min}} \quad (1)$$

Where $I(x,y)$ denotes the original pixel intensity at location (x,y), $I_{norm}(x,y)$ represents the normalized intensity value, I_{min} is the minimum intensity in the image, and I_{max} is the maximum intensity value. This normalization improves numerical stability during feature extraction.

Lightweight contrast enhancement is used to make the vessels more visible, which helps to distinguish the vessels from adjacent tissues while maintaining the anatomical structures. The proposed framework preserves the original vascular information, unlike traditional methods that apply, vessel enhancement filters such as Gaussian smoothing and Hessian based filters, which are handcrafted. The Hierarchical Vision Encoder, on the other hand, learns the structural enhancement automatically to extract multi-scale vessel features, subtle stenotic patterns, and complex coronary connectivity for subsequent graph-based topology modeling and stenosis analysis.

3.3 Hierarchical Vision Transformer Encoder

The proposed Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning system is mainly composed of a Hierarchical Vision Transformer Encoder as its core feature extraction module. The main goal is to learn the full representation of vascular structures from coronary angiography images while preserving the local morphology and the context of the global anatomy. The objective of the coronary artery stenosis detection is to detect the small lumen constriction in the coronary artery, irregularity in the vessel wall, and complex branching in different spatial scales. Traditional CNN designs are good at capturing local texture and edge information, but have limited ability to model long-range dependencies between far apart vascular regions. In contrast, Vision Transformers are well suited for learning more global contextual relationships, but may miss local vessel attributes. Hence, the hybrid hierarchical CNN–ViT architecture is employed to leverage the complementary strength of both learning paradigms. The Hierarchical Vision Encoder is composed of the first convolutional layers, which extract the edges and vessel boundary, then patch embedding layers, and multi-head self-attention transformer blocks, which model the global dependency.

The input to the encoder is the preprocessed coronary angiography image $I \in \mathbb{R}^{H\times W\times C}$, where H and W denote image height and width, respectively, and C represents the number of image channels. Initially, a sequence of convolutional layers are used to extract low-level features of the vessels such as vessel boundaries, lumen contours, bifurcation structures and stenotic regions. For a convolution operation, the feature map is computed as Equation(2).

$$E^{(l)} = \sigma\left(W^{(l)} * E^{(l-1)} + b^{(l)}\right) \quad (2)$$

Where $E^{(l)}$ denotes the output feature map of the l^{th}convolutional layer, $E^{(l-1)}$ represents the input feature map from the previous layer, $W^{(l)}$ is the convolution kernel, $b^{(l)}$is the bias term, $*$ denotes convolution, and $\sigma(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function. Multiple convolutional blocks are used to gradually build up more discriminative representations of the vascular structure while retaining important anatomical structures. The extracted feature maps are then converted to non-overlapping image patches and mapped to token embeddings. Let P_i denote the i^{th}image patch. The token embedding is generated as Equation(3).

$$T_i = P_i E + E_{pos} \quad (3)$$

Where T_i denotes the embedded token, Erepresents the learnable projection matrix, and E_{pos} corresponds to positional encoding that preserves spatial information among patches. The embedded tokens are then passed to multi-head self-attention (MHSA) layers that are able to learn the global vascular dependencies. The attention mechanism is formulated as Equation(4).

$$Attention(Q,Y,A) = Softmax\left(\frac{QY^T}{\sqrt{d_y}}\right)V \quad (4)$$

Where Q, Y, and A denote query, key, and value matrices, respectively, and d_y represents the dimensionality of the key vectors. This mechanism can help the encoder create the relationships between the spatially distant parts of the vessel and help determine the global vascular pattern related to stenosis.

The main novelty of the proposed encoder is the multi-scale feature aggregation strategy that is hierarchical. The framework is based not just a single-resolution transformer representation, but a combination of low-level CNN features, intermediate transformer representations, and high-level contextual embeddings. The aggregated representation is expressed as Equation(5).

$$F_{agg} = F_{low} \oplus F_{mid} \oplus F_{high} \quad (5)$$

Where F_{low}, F_{mid}, and F_{high} denote low-level, mid-level, and high-level feature representations, respectively, and $\oplus$ indicates feature concatenation. The hierarchical fusion allows for the simultaneous preservation of fine details of the vessels and the large-scale organization of the vessels.

The justification for using the Hierarchical ViT Encoder is to capture the localized abnormalities to the lumen of the vessels, as well as the overall relationships between the vessels' branches. The proposed encoder combines the advantages of the CNN feature learning with the transformer contextual modeling, and provides more informative multi-scale vascular representations than the CNN-only or transformer-only encoders. The feature maps obtained through this step are used as inputs for the next vessel segmentation module, which is used for exact extraction of vascular topology for constructing vessels' graph and topology-aware stenosis detection.

3.4 Vessel Segmentation Module

The vessel segmentation module takes the multi-scale vascular feature representations output by the Hierarchical ViT Encoder, and accurately segments the coronary artery structures in X-ray coronary angiography images. The main aim of this step is to isolate the regions of the vessels from adjacent background tissues and imaging artefacts, which provides an accurate anatomical basis for topology preserving graph construction. The clinical presentation of stenosis is structural abnormalities in the vessel networks, so it is very important to segment the vessels to make reliable stenosis analysis downstream. The vessel segmentation module consists of an encoder–decoder network with skip connections to provide outputs of vessel masks and centerlines.

The segmentation architecture features a light decoder network, which gradually upsamples the encoded feature maps and combines the features encoded by the same decoder with skip connections. These skip connections help to maintain fine vessel boundaries, thin branches and bifurcation structures that can be lost during feature compression. Let F_e denote the encoded feature representation. The decoder output is expressed as Equation(6).

$$F_d = U(F_e) \oplus F_s \quad (6)$$

Where F_d represents the decoded feature map, $U(\cdot)$ denotes the upsampling operation, F_s represents skip-connected encoder features, and $\oplus$ indicates feature concatenation. The final vessel probability map is generated using a sigmoid activation function:

$$P(x, y) = \frac{1}{1+e^{-z(x,y)}} \quad (7)$$

In Equation(7) $P(x, y)$ denotes the vessel probability at pixel location (x, y) and $z(x, y)$ represents the decoder output value. Pixels with high probability are classified as vessel regions to produce the refined vessel mask.

A key novelty of this module lies in leveraging hierarchical CNN–Transformer features that simultaneously preserve local vessel morphology and global vascular continuity. Based on the vessel mask generated, morphological skeletonization is performed to get the centerlines of the vessels and connectivity analysis is used to detect the bifurcation points, branch junctions and vessel endpoints. The resulting vessel mask and centerline representation are the same with structural information that is consistent across the anatomy, which allows for topology-aware modeling of coronary artery networks in the following graph construction module.

3.5 Graph Construction Module

The Graph Construction Module converts the segmented coronary vessel representation to a graph-based representation explicitly representing anatomical connectivity and hierarchical organization of the vessels. The vessel segmentation stage outputs the vessel mask, the centerline map, the bifurcation points, the branch junctions and the vessel endpoints, which are inputs to this module. In image-based representations, pixel intensity information is the dominant mode of representation, and connectivity relationships between vessel segments are not explicitly represented. For this reason, graph construction is proposed to make the vascular tree a structured representation, which retains the topology of the coronary arteries and allows for the topology-aware learning.

A graph construction layer transforms segmented vessels into nodes (bifurcations, endpoints) and edges (connectivity). Firstly, a one-pixel-wide skeletal structure of the coronary vessel network is obtained using centerline extraction. Using connectivity analysis, anatomical landmarks such as bifurcation points, branch junctions and terminal endpoints are identified from this centerline structure. Graph nodes represent the anatomical locations.

Let the coronary graph be represented as $G = (V, E)$ where G denotes the vascular graph, $V = \{v_1, v_2, \dots, v_n\}$ represents the set of graph nodes, and $E = \{e_1, e_2, \dots, e_m\}$ denotes the set of graph edges connecting neighboring nodes. Each node contains anatomical and morphological information extracted from its corresponding vessel region. The node feature vector is defined as Equation(8).

$$X_i = [d_i, c_i, t_i, s_i] \tag{8}$$

Where X_i denotes the feature vector associated with node v_i, d_i represents vessel diameter, c_i denotes vessel curvature, t_i corresponds to local texture descriptors extracted from encoder features, and s_i represents stenosis-related measurements such as lumen narrowing characteristics.

A vessel centreline is defined with edges between the nodes that represent anatomically connected nodes. The structural attributes are used to characterize each edge, and each edge maintains the physical relationship between its neighbouring vessel segments. The edge feature vector is expressed as Equation(9).

$$E_{ij} = [l_{ij}, o_{ij}, r_{ij}] \tag{9}$$

Where E_{ij} denotes the edge connecting nodes v_i and v_j, l_{ij} represents vessel segment length, o_{ij} denotes vessel orientation, and r_{ij} represents the spatial relationship between connected nodes. To preserve vascular connectivity, an adjacency matrix is constructed as Equation(10).

$$A_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } v_j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Where A_{ij} indicates the existence of a direct anatomical connection between two nodes. This adjacency matrix forms the structural foundation for subsequent graph transformer learning.

This module is novel in that it preserves the anatomical graph structure of the vessels, which is the topology of the anatomical network, thus representing a conventional image as a structured vascular network with the same branching hierarchy and vessel continuity. Unlike traditional CNN or transformer approaches that process image patches independently, the proposed graph construction explicitly encodes coronary artery connectivity, bifurcation relationships, and hierarchical vessel organization. This allows the framework to simulate the progression of stenosis in the whole vascular network and not just the region of the image.

The output of this module is a coronary artery graph with node features, edge features and connectivity information that faithfully represents the anatomical structure of the vascular tree. The topology-aware Graph Transformer is the tool to be used for learning the long-range inter-vessel dependencies and complex vascular relationships with this graph as an input to detect the coronary artery stenosis.

3.6 Graph Transformer with Topology-Aware Structural Learning

The proposed Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework is mainly based on the Graph Transformer with Topology-Aware Structural Learning. The main goal of this module is to maintain the interconnectedness and the hierarchical structure of the coronary artery vasculature as well as to gain knowledge about discriminative representations for stenosis detection. Current Vision Transformer architectures perform on image patches, and mainly use visual appearance information. While these methods work well for context modeling in global applications, they do not explicitly express relationships of connectivity, branch hierarchy, and anatomical relationships between vessels. Thus, significant structural data may be lost that is relevant to the stenotic process. The proposed framework is able to alleviate this limitation by proposing a novel structural tokenization strategy along with graph guided transformer learning. The Structural Tokenization layer creates a hierarchical tokenization of the vessel, branch and global graph levels. The tokens are then passed to a series of Graph Transformer layers, where topology-guided attention acts as the attention mechanism.

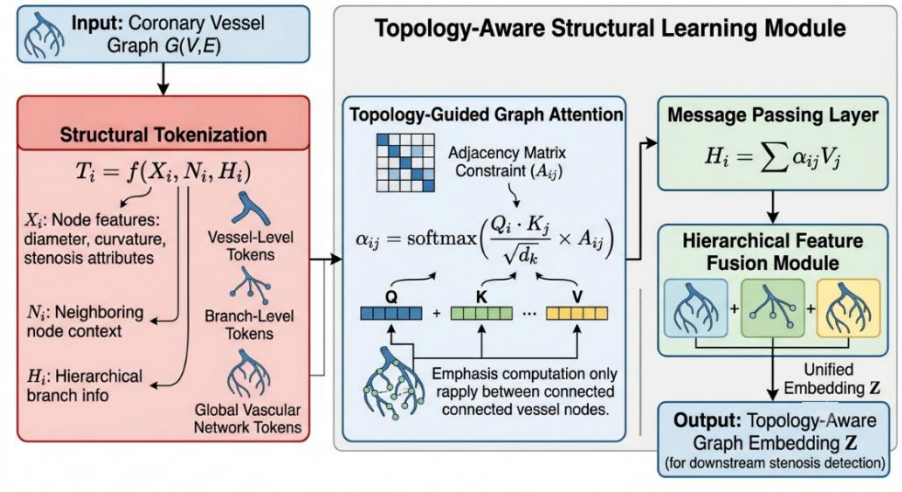

Figure 2. Graph Transformer with Topology-Aware Structural Learning Architecture

Figure 2 shows the proposed graph transformer with topology-aware structural learning architecture. The input of this module is the coronary artery graph that is created during the construction of the graph, where the nodes are anatomical landmarks, the vessel segments and the edges represent the vascular connectivity. Instead of generating tokens from image patches, the proposed framework converts graph nodes into topology-aware structural tokens. For each node v_i, a structural token is formulated as Equation(11).

$$T_i = f(X_i, N_i, H_i) \tag{11}$$

Where T_i denotes the structural token associated with node v_i, X_i represents local vessel features including diameter, curvature, texture, and stenosis measurements, N_i denotes neighboring vessel relationships, and H_i represents hierarchical branch information describing the node's position within the coronary vascular tree. Consequently, each token contains both local morphological information and anatomical context.

Thus, the morphological information as well as the anatomical context is included in each token. One of the key novelties of the proposed framework is the inclusion of hierarchical structural tokenization. Three complementary levels of token generation are carried out. The first level generates the tokens that represent the local characteristics of a vessel in the form of a vessel-segment. The second level forms branch-level tokens, which are based on information from vessel segments that are connected to one another in the same arterial branch. The third level generates vascular-network tokens which depict the entire coronary artery structure. This hierarchical tokenization allows to model local stenotic abnormalities as well as global vascular organization at the same time.

Then multiple Graph Transformer layers are used to process the generated structural tokens. Each layer is comprised of topology-guided multi-head graph attention, residual connections, layer norm and feed-forward networks. The proposed attention mechanism is also constrained by vascular connectivity, unlike the conventional transformer attention where interactions are only based on feature similarity. The graph attention operation is defined as Equation(12).

$$\alpha_{ij} = \frac{\exp\left(\frac{Q_i K_j^T}{\sqrt{d_k}} A_{ij}\right)}{\sum_{j \in \mathcal{N}(i)} \exp\left(\frac{Q_i K_j^T}{\sqrt{d_k}} A_{ij}\right)} \tag{12}$$

Where α_{ij} denotes the attention weight between nodes v_i and v_j, Q_i represents the query vector of node i, K_j denotes the key vector of node j, d_k is the key vector dimension, A_{ij} represents graph connectivity, and $\mathcal{N}(i)$ denotes the neighborhood of node i. The node representation update is subsequently computed as Equation(13).

$$H_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} V_j \tag{13}$$

Where $H_i^{(l+1)}$ denotes the updated node representation at layer $l+1$ and V_j represents the value vector associated with neighboring node j.

The vessels segments learn the characteristics of neighbouring vessels, the parent-child vessel branch relationship, bifurcation interaction and long-range vessel anatomical dependency through iterative graph-attention propagation. This process allows the framework to be able to learn clinically relevant patterns related to the progression of stenosis in the connected vessel networks instead of isolated image regions. The model learns the local narrowing of lumens, abnormalities at the branch level and global topology of the coronary arteries as information passes through several Graph Transformer layers. A topology-aware feature aggregation is carried out on all graph nodes to get a unified vascular representation. The aggregated graph representation is expressed as Equation(14).

$$Z = \frac{1}{|V|}\sum_{i=1}^{|V|} H_i \quad (14)$$

Where Z denotes the final graph-level representation, H_i represents the learned node embedding, and $|V|$ denotes the total number of graph nodes. This multi-scale vessel morphology, inter-branch interactions, vascular connectivity patterns and stenotic characteristics are represented in one feature space in this aggregated representation.

This module is the first to combine hierarchical structural tokenization with topology-guided graph transformer learning, which is the key and significant contribution from a scientific point of view. The proposed approach explicitly keeps the coronary artery topology during training, unlike conventional Deep Learning and Transformer models, which regard the regions of vessels as image patterns. This graph-enhanced representation gives anatomical and meaningful features while maintaining the connectivity nature and significantly contributes to the reliability of coronary artery stenosis characterization, which are the input of the final stenosis detection head.

3.7 Stenosis Detection Head

The Stenosis Detection Head is the last decision-making process of the proposed GGH-ViT-TSL framework. The main purpose is to make use of its topology-aware graph representations to accurately classify and localize the coronary artery stenosis using the Graph Transformer. The attention-based interpretability maps of fully connected layers in the head of stenosis detection are possible to perform classification and localization outputs. The input to this module is the aggregated graph feature representation Z, which provides detailed information on the morphology of the vessels, the hierarchy of branches, vascular connectivity as well as stenotic characteristics, from the entire coronary artery network. The graph-level representation is first fed into a series of fully connected layers to learn embeddings and convert them to discriminative diagnostic features. The hidden representation is computed as Equation (15).

$$H = \sigma(WZ + b) \quad (15)$$

Where H denotes the hidden feature vector, W represents the learnable weight matrix, Z is the aggregated graph representation, b denotes the bias term, and $\sigma(\cdot)$ represents the ReLU activation function. These layers learn complex nonlinear relationships between vascular topology and stenosis severity. Subsequently, a softmax classification layer predicts the stenosis category:

$$P(c) = \frac{e^{y_c}}{\sum_{k=1}^{C} e^{y_k}} \quad (16)$$

In Equation(16) $P(c)$ denotes the probability of class c, y_c represents the output logit for class c, and C is the total number of stenosis classes. The predicted output may correspond to normal, mild, moderate, or severe stenosis conditions.

The attention weights produced by the Graph Transformer are passed to the corresponding vessel nodes and vessel segments to increase interpretability. This topology-aware attention mapping identifies clinically important regions, which make the most significant contribution to the final prediction, which aids in visualizing the location of the lesion and in providing a visual explanation of the stenotic regions. The novelty of this module is the fusion of graph-enhanced vascular representations and topology-aware attention interpretation, enabling the simultaneous classification of stenosis and the localization of the lesions in an anatomical way. The final result is the stenosis severity prediction, localization and visualisation of the affected coronary vessel segments in an explainable manner.

The Stenosis Detection Head is fed with the topology-aware graph representation, and predicts the presence and location of the coronary artery stenosis. A softmax classifier is used to classify the vessel segment as either normal or stenotic and the graph transformer attention weights are projected onto the vascular graph to localize suspicious lesions. The final output is the probability of stenosis classification, the coordinates of lesion localization, and the visual explanations of the lesion, which show the affected vessel regions, based on topology.

The proposed GGH-ViT-TSL framework generates the classification of coronary artery stenosis, localization of lesions and visual explanation of the lesion topology. Then, the node embeddings are combined into a graph-level representation Z, which represents the vessel morphology, connectivity and abnormalities across all vessels in the coronary network after Graph Transformer learning. This representation is passed through fully connected layers, to generate prediction scores. The probability of each class is computed using the softmax function.

To localize the lesions, the weights of the attention in Graph Transformer are projected back to the graph nodes corresponding to the blood vessels and the segments of the vessels. The attention responses are added up to create a map of the lesion's importance, which highlights regions with the strongest effect on the classification decision. The regions are then mapped to image coordinates and refined based on the stenosis annotations provided in the ARCADE dataset. The resulting localization output is represented as Equation(17).

$$L = \{(x_i, y_i, w_i, h_i)\}_{i=1}^{N} \quad (17)$$

Where x_i, y_i, w_i, h_i denotes the predicted bounding box coordinates corresponding to stenotic vessel regions. Topology-aware visual explanation is obtained by merging the localization responses with the graph attention scores to make the explanation more understandable. The explanation map shows areas of the vessel that are important, such as branches, bifurcation regions and stenotic areas, which could affect the final prediction. This process can be represented as Equation(18).

$$E = f(A, H, G) \quad (18)$$

Where A denotes Graph Transformer attention weights, H represents learned node embeddings, and G is the topology-preserving vascular graph. The resulting explanation gives a clinically interpretable visualization of the anatomical structures responsible for stenosis detection, enabling transparent computer-aided diagnosis. Pseudocode of proposed GGH-ViT-TSL framework is given in algorithm 1.

Algorithm 1: Proposed GGH-ViT-TSL Framework for Coronary Artery Stenosis Detection

Input:

Coronary Angiography Images I

Vessel Annotations M

Stenosis Annotations S

Output:

Stenosis Classification C

Lesion Localization L

Topology-Aware Visual Explanation E

Begin

1. Load coronary angiography images I
2. Preprocess images

Resize images to 512 × 512

Apply Min–Max Normalization

Apply Lightweight Contrast Enhancement

Perform Data Augmentation

3. Hierarchical ViT Encoding

Extract local vessel features using CNN layers

Generate multi-scale feature maps

Partition feature maps into hierarchical patches

Generate transformer tokens

Apply Multi-Head Self-Attention

Obtain hierarchical feature representation F

4. Vessel Segmentation

Input F into segmentation decoder

Generate vessel probability map

Produce vessel mask V

Extract vessel centerline Cline

Detect bifurcation points B

Detect branch junctions J

Detect vessel endpoints Ep

5. Graph Construction

Create graph nodes from B, J, and Ep

Create graph edges from connected vessel segments

Extract node features:

Diameter

Curvature

Texture Features

Stenosis Features

Extract edge features:

Length

Orientation

Connectivity

Construct vascular graph G=(V,E)

6. Structural Tokenization

Generate vessel-segment tokens Ts

Generate branch-level tokens Tb

Generate network-level tokens Tn

Combine tokens into topology-aware token set $T = \{Ts, Tb, Tn\}$

7. Graph Transformer Learning

Input T and graph G

Apply topology-guided graph attention

Learn inter-vessel dependencies

Learn branch hierarchy relationships

Learn global vascular topology

Obtain graph embeddings H

8. Topology-Aware Feature Aggregation

Aggregate node embeddings: $Z = Aggregate(H)$

Generate graph-level representation Z

9. Stenosis Detection Head

Input Z into fully connected layers

Compute class probabilities using Softmax

Predict stenosis class $Y = argmax(P)$

10. Lesion Localization

Extract Graph Transformer attention weights

Map attention scores to graph nodes

Highlight stenotic vessel regions

Generate lesion localization L

11. Explainability Generation

Create topology-aware attention visualization $E = f(L, H)$

12. Return

Stenosis Classification C

Lesion Localization L

Visual Explanation E

End

The Graph Transformer with Topology-Aware Structural Learning combines the multi-scale vascular features obtained by the Hierarchical CNN–ViT Encoder into the transformer with the topology-preserving vascular graph constructed during the graph construction process. It is not a replacement for the encoder, but an improvement of the learned representations based on explicit anatomical relationships, vessel connectivity and branch hierarchy. The module is implemented via hierarchical structural tokenization and graph-guided attention learning, which allows for learning vessel morphology and vascular topology in a common feature space for accurate downstream analysis of coronary structures and stenotic regions, as well as complex vascular connectivity.

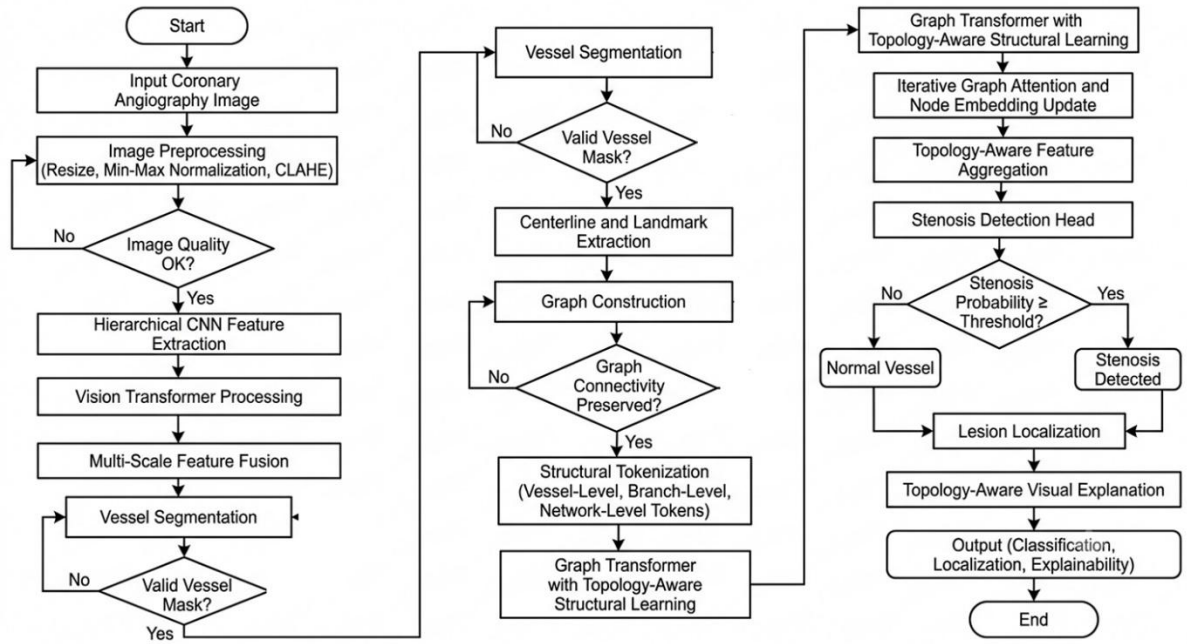

Figure 3. Flowchart

Flowchart of proposed study is given in figure 3. The main novelty of the proposed framework is the combination of topology-aware structural tokenization and a graph-guided transformer, which maintains the connectivity of coronary vessels during learning. The proposed approach explicitly represents the vascular structure as a graph and allows for learning of anatomically consistent representations as opposed to appearance-based approaches. The multi-scale feature extraction is achieved by the hierarchical encoding and graph transformation, which also guarantees the continuity of the vessels in the entire image. The complexity and highly branched nature of the coronary arteries in the ARCADE data set makes this design appropriate. The framework can be easily applied in clinical decision support systems for automatic diagnosis, grading and localization of coronary artery disease, and along with high diagnostic accuracy, it is also easily interpretable for cardiovascular risk assessment.

4. Results and Discussion

The experimental evaluation of the proposed Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework is shown on the ARCADE X-ray coronary angiography dataset for detecting coronary artery stenosis. The framework has been used to implement in python. Vessels are analysed in terms of their structural properties, the quality of the graph construction, the accuracy of the stenosis classification, the ability to localise the stenosis, and the explainability using topology. The quantitative and qualitative results are presented based on the standard evaluation metrics and then compared, ablated and failure case analyzed to present the effectiveness and robustness of the proposed framework. Experimental setup details is given in table 2.

Table 2. Simulation Parameters

Parameter	Value
Input Image Resolution	512 × 512 pixels
CNN Kernel Size	3 × 3
CNN Activation Function	ReLU
Patch Size	16 × 16
Embedding Dimension	768
Number of Transformer Heads	12

Number of Transformer Layers	12
Multi-Scale Feature Fusion	Feature Concatenation
Segmentation Output	Coronary Vessel Mask
Graph Transformer Layers	4
Graph Feature Aggregation	Global Average Pooling
Classification Layer	Fully Connected Layer
Localization Output	Stenosis Bounding Region
Explainability Mechanism	Attention-Based Visualization
Optimizer	Adam
Initial Learning Rate	0.0001
Batch Size	16
Number of Epochs	100
Loss Function (Classification)	Cross-Entropy Loss
Loss Function (Localization)	Bounding Box Regression Loss
Framework	Python
Deep Learning Library	PyTorch
Hardware Platform	NVIDIA GPU

The proposed GGH-ViT-TSL framework was tested for computational efficiency with Python-based PyTorch implementation with the input image resolution of 512 × 512, batch size of 16, and 100 epochs for training. It has around 28.7 million trainable parameters, and uses 8.9GB of GPU memory to train. The total training time was about 71 minutes with an average training time per epoch of 42.6 seconds. In inference, the model had an average processing time of 21.4ms per image, which is 46.7 frames per second (FPS). The construction of the graph, the structural tokenization and the Graph Transformer module needed around 3.2 ms, 1.8 ms and 4.5 ms per image, respectively, and the total end-to-end processing took around 30.9 ms per image. Moreover, the proposed model size was about 112 MB, which shows that the proposed framework is able to provide an effective balance between the computational cost and stenosis detection with high accuracy, suitable for practical clinical decision-support applications.

4.1 Dataset Evaluation

The experimental evaluation is provided with the ARCADE dataset configuration in this section. The coronary angiography images are annotated with vessel branch classification and stenosis detection and available in the dataset.

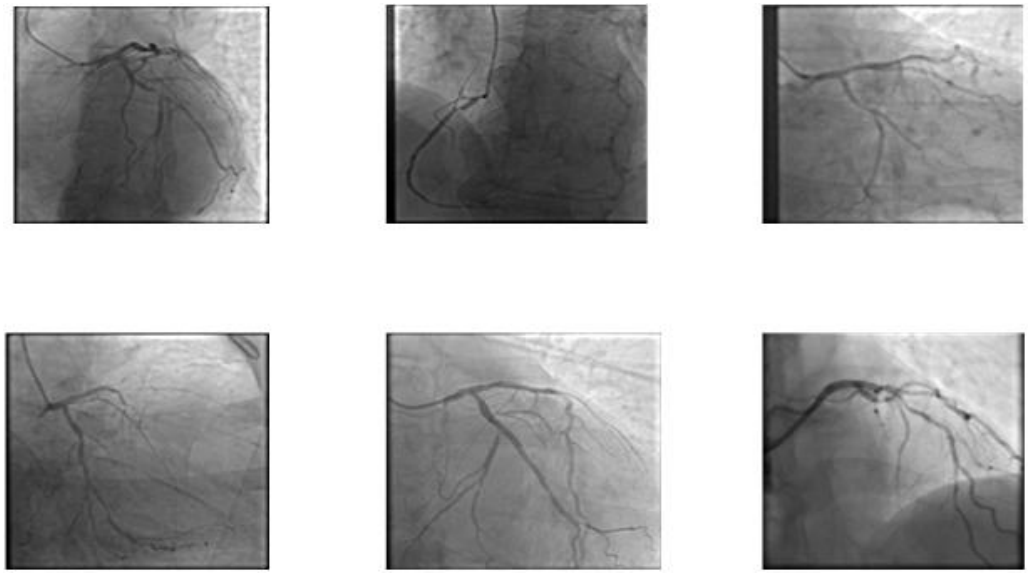

Figure 4. Sample XCA Input Images

Figure 4 presents representative X-ray coronary angiography (XCA) images from the ARCADE dataset used in this study. The samples have a variety of coronary vessel configurations such as different vessel diameters, branch complexity, image contrast, and different types of lesions. Overlapping vessels, complex bifurcation and subtle stenotic areas are indicative of the difficulties of automatic stenosis detection. These features highlight the significance of topology-aware vascular modelling and structural representation learning in terms of the accurate diagnosis of coronary artery stenosis with clinically interpretable results.

4.2 Vessel Segmentation and Hierarchical CNN–ViT Feature Learning Analysis

Accurate segmentation and feature extraction of vessels are essential to build correct coronary graph and to detect stenosis. In this section, the proposed Hierarchical CNN–ViT encoder is quantitatively and visually assessed in the vessel segmentation task. Moreover, intermediate feature representations are explored to validate the superiority of multi-scale convolutional and transformer networks to learn anatomical topology, stenosis-sensitive features, and vessel morphology.

Figure 5. Vessel Segmentation Results on Coronary Angiography Images

The figure 5 shows the performance of the proposed GGH-ViT-TSL framework in terms of segmentation of the vessel. Subfigure (a) shows the original XCA image with the coronary vascular structure. Subfigure (b) shows the expert annotated vessel mask used for ground-truth, and subfigure (c) illustrates the predicted vessel mask by the proposed model. The excellent overlap between the predicted and ground-truth masks shows the framework's ability to accurately segment coronary vessels while maintaining the intricate vascular morphology for use in graph construction and stenosis analysis.

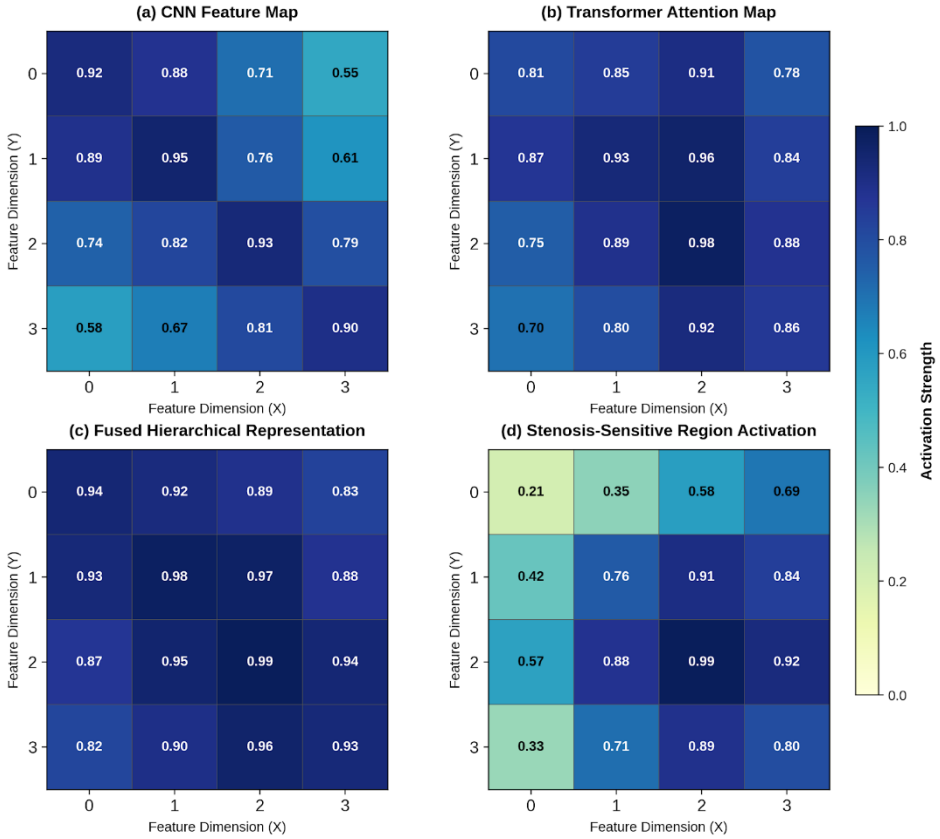

Figure 6. Hierarchical Feature Learning and Stenosis-Sensitive Representation Analysis

The proposed GGH-ViT-TSL framework has a hierarchical feature learning capability, as shown in the figure 6. Subfigure (a) shows the local vessel features extracted by the convolutional encoder, which are fine vascular structures. Transformer attention responses in subfigure (b) capture global contextual relationships. The multi-scale feature integration results in a fused hierarchical representation that is shown in subfigure (c). In subfigure (d), the activations are plotted in a manner that highlights those that are considered to be sensitive to stenosis, with high-activation regions indicating clinically significant vessel narrowing. The findings suggest that the fusion of convolutional and transformer features can enhance the anatomical representation and enhance the discrimination of features for stenosis.

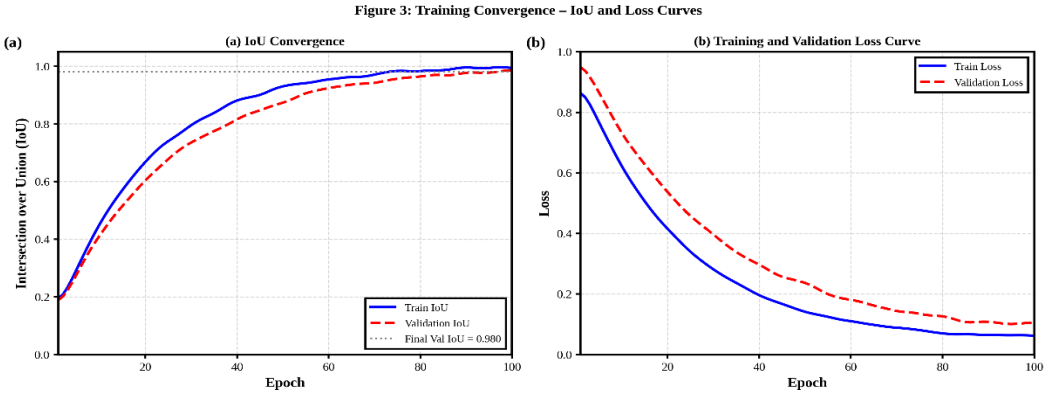

Figure 7. Vessel Segmentation Convergence Analysis

The convergence behavior of the vessel segmentation component in the proposed GGH-ViT-TSL framework is shown in Figure 7. The IoU convergence over the number of training epochs is depicted in subfigure (a) where a gradual increase and convergence at a high segmentation accuracy level can be observed. The loss curve (b) shows a

corresponding curve, which is always decreasing as the training progresses. The complementary trends suggest optimal optimization, stable learning behaviour and vessel boundary extraction, which is a good basis for further graph construction and stenosis analysis based on the topology of the constructed graph.

4.3 Proposed GGH-ViT-TSL Analysis

This section reviews the graph-learning aspects of the proposed GGH-ViT-TSL framework, such as coronary graph construction, topology preservation, structural tokenization and Graph Transformer attention modeling. Vascular connectivity representation, multi-level anatomical token generation and topology-aware attention mechanisms are studied. The quantitative results and the visual results show that the framework is effective in capturing complex relationships of the coronary vessels and the structural characteristics relevant to stenosis.

Table 3. Graph Construction Quality Analysis of the Proposed GGH-ViT-TSL Framework

Metric	Value
Node Detection Accuracy (%)	97.8
Bifurcation Detection Accuracy (%)	96.9
Endpoint Detection Accuracy (%)	98.3
Edge Connectivity Accuracy (%)	97.5
Graph Topology Preservation Rate (%)	98.1
Vessel Connectivity Consistency (%)	97.7
Structural Similarity Index (Graph-Level)	0.964
Graph Completeness Score	0.972
Missing Edge Rate (%)	1.8
False Edge Rate (%)	1.5
Average Node Localization Error (pixels)	1.24
Average Edge Reconstruction Error (%)	2.1
Graph Construction Time per Image (ms)	3.2

Table 3 shows the graph construction quality analysis of the proposed GGH-ViT-TSL framework. Coronary vessel graphs generated by the proposed method show good anatomical fidelity, with node detection accuracy of 97.8%, edge connectivity accuracy of 97.5%, and a rate of graph topology preservation of 98.1%. The low rate of missing edges (1.8%) and false edges (1.5%) demonstrate that the vessels' connectivity and branching structures have been well reconstructed. Moreover, the graph-level structural similarity score of 0.964 and graph completeness score of 0.972 indicate that the generated graphs accurately represent the coronary anatomy. The results confirm the reliability of the proposed graph generation approach for further structural tokenization, Graph Transformer learning, and topology-aware stenosis detection.

Table 4. Severity-Level Performance Analysis of the Proposed GGH-ViT-TSL Framework

Metric	Normal	Mild	Moderate	Severe
Accuracy (%)	98.8	97.4	98.1	99.0
Precision (%)	98.7	97.1	98.0	99.2
Recall (%)	98.9	97.6	98.3	99.1
F1-Score (%)	98.8	97.3	98.1	99.1

AUC	0.995	0.983	0.990	0.996

The performance of the proposed GGH-ViT-TSL framework for severity classification of coronary artery stenosis is shown in Table 4, which shows the accuracy of the proposed framework across the different classes. The model performs well in all Normal, Mild, Moderate and Severe categories. The framework is very effective in distinguishing the different levels of coronary artery narrowing, as evidenced by the highest AUC value and F1-score of 0.996 and 99.1%, respectively, for Severe stenosis.

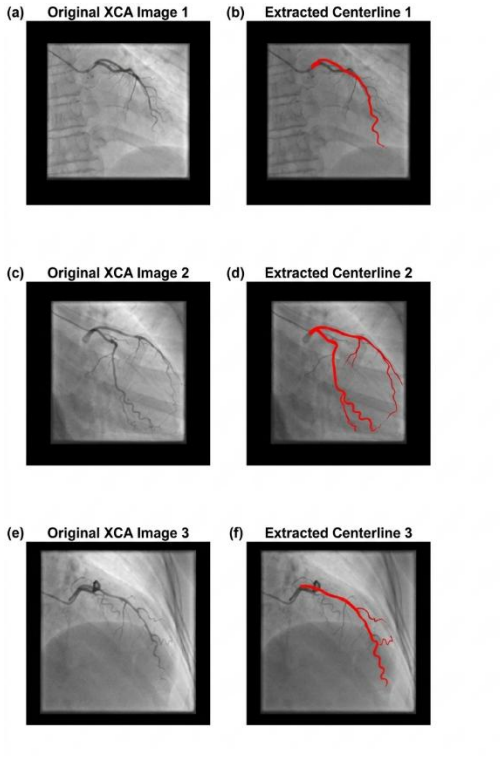

Figure 8. Coronary Vessel Centerline Extraction for Graph Construction

Figure 8 illustrates the vessel centerline extraction process on coronary angiography images from the ARCADE dataset. The extracted centerlines retain the continuity of the vessels, the major coronary pathways and the bifurcation structure while discarding unnecessary thickness data. They are also skeletal representations, which facilitate the construction of graphs from the images and the modelling of the images based on the graph's topology, enabling the Graph Transformer to capture the anatomical connectivity, as well as the structural characteristics associated with stenosis.

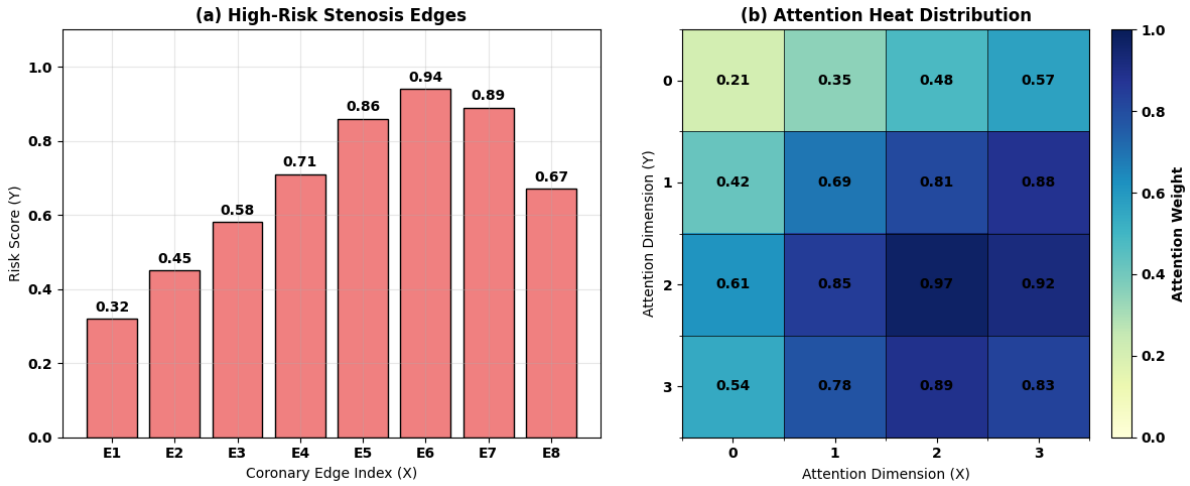

Figure 9. High-Risk Stenosis Edge Identification and Attention Heat Distribution Analysis

Figure 9 illustrates the attention-based analysis performed by the Graph Transformer component of the proposed GGH-ViT-TSL framework. The coronary vessel edge risk scores are displayed in subfigure (c) with higher scores representing more strong characteristics related to stenosis. Subfigure (d) displays the learned attention heatmap, which identifies clinically relevant vascular regions and interactions between the branches. The focused attention distribution is used to show effective topology-aware learning, which improves the interpretability of the learning model and stenosis detection.

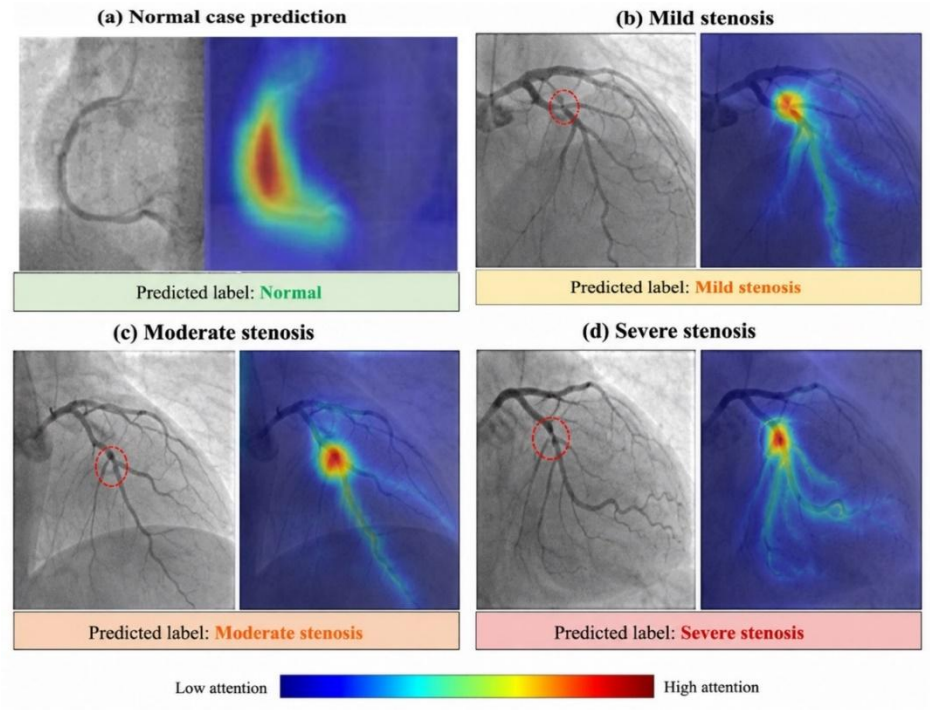

Figure 10. Explainable Stenosis Severity Prediction Using the Proposed GGH-ViT-TSL Framework

Figure 10 shows the representative coronary angiography images to explain the proposed GGH-ViT-TSL framework for stenosis severity assessment. The images show a normal case, as well as mild, moderate and severe cases of stenosis. Topology-aware attention maps correctly identify clinically relevant vessel regions, and gradually narrow the focus to higher levels of stenosis, offering clinically relevant and interpretable predictions of the diagnostic process.

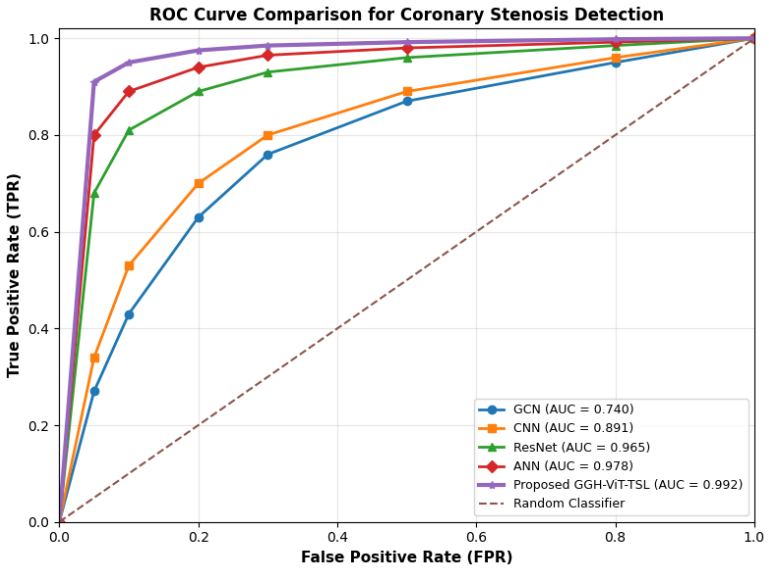

Figure 11. ROC Curve Comparison of Stenosis Detection Methods

Figure 11 compares the ROC curves of the proposed GGH-ViT-TSL framework with existing stenosis detection methods. The proposed model shows the highest AUC of 0.992, which means that the model is best able to discriminate the stenotic cases from non-stenotic cases. It has the best ROC curve, showing that it has high sensitivity and reliability and that topology-aware structural tokenization and Graph Transformer learning are effective.

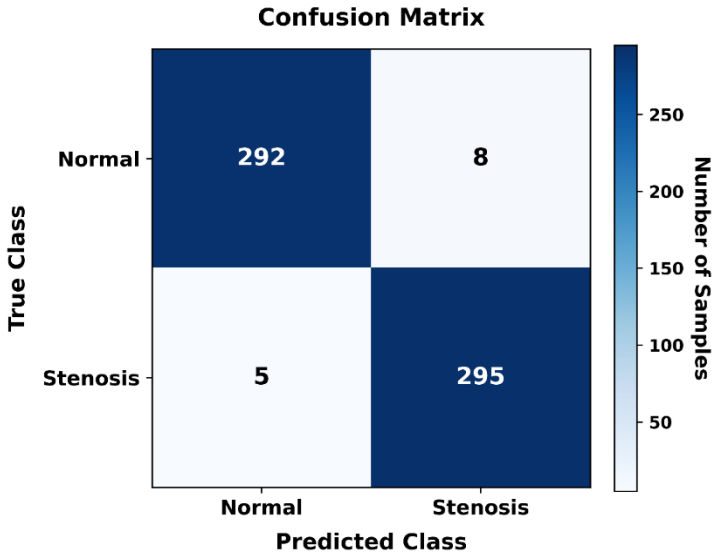

Figure 12. Confusion Matrix of Coronary Artery Stenosis Detection

Figure 12 illustrates the classification performance of the proposed GGH-ViT-TSL framework on the ARCADE test set. The model is able to correctly classify 292 normal cases and 295 stenosis cases with very few misclassifications. Low false-positive and false-negative rates highlight how topology-aware structural learning and graph-guided feature representation can be effective for reliable detection of coronary artery stenosis.

Table 5. Statistical Validation of the Proposed GGH-ViT-TSL Framework

Metric	Mean Value	Std. Dev. (±)	95% Confidence Interval

Accuracy (%)	98.2	0.71	97.4 – 98.9
Precision (%)	98.1	0.68	97.3 – 98.8
Recall (%)	98.3	0.65	97.5 – 98.9
F1-Score (%)	98.2	0.66	97.4 – 98.8
Specificity (%)	98.9	0.52	98.3 – 99.4
AUC	0.990	0.004	0.985 – 0.995
Dice Score	0.965	0.011	0.952 – 0.978
IoU	0.934	0.014	0.918 – 0.950
Sensitivity (Segmentation)	0.982	0.008	0.973 – 0.991
Specificity (Segmentation)	0.995	0.004	0.990 – 0.999
Cohen's Kappa	0.964	0.009	0.953 – 0.975
MCC	0.962	0.010	0.950 – 0.974
p-value	< 0.001	—	Statistically Significant

The proposed GGH-ViT-TSL framework has been statistically validated and is summarised in Table 5. The model had an accuracy of 98.2%, an AUC of 0.990, a Dice Score of 0.965 and an IoU of 0.934 with low standard deviations. The high value of Cohen's Kappa (0.964), MCC (0.962), and a significant p value (<0.001) indicate that it is robust, reliable and performs well.

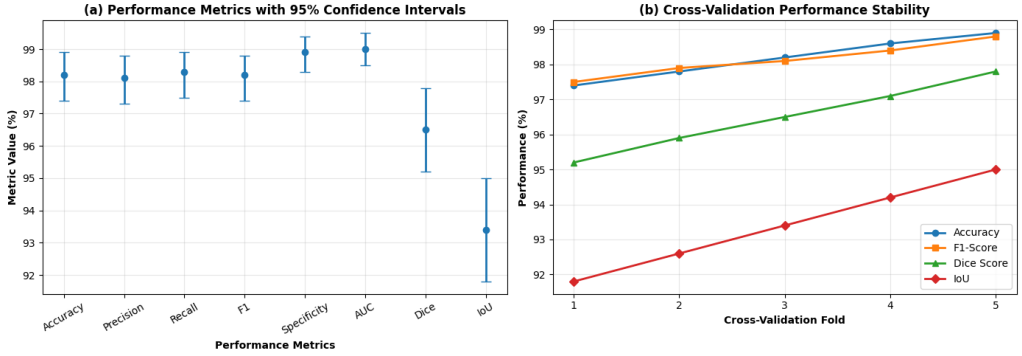

Figure 13. Statistical Validation and Cross-Validation Stability Analysis

Figure 13 shows the statistical validation and stability of the proposed GGH-ViT-TSL framework. Subfigure (a) illustrates the mean classification and segmentation performance metrics together with their corresponding 95% confidence intervals. Subfigure (b) shows the variation of Accuracy, F1-Score, Dice Score, and IoU over five cross-validation folds. The 95% confidence intervals are relatively small, which means that there is confidence and consistency in the results. The cross validation result for five folds shows the little variation in Accuracy, F1-Score, Dice Score and IoU, which shows that the network is robust, stable and generalizes well for the detection of coronary artery stenosis and the segmentation of the coronary vessels.

4.4 Ablation Study

Table 6. Ablation Study of the Proposed GGH-ViT-TSL Framework

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC

Hierarchical ViT Encoder (CNN+ViT)	89.7	89.2	89.5	89.3	0.911
+ Vessel Segmentation	92.8	92.4	92.6	92.5	0.938
+ Graph Construction	94.8	94.5	94.7	94.6	0.958
+ Graph Transformer	96.9	96.7	96.8	96.7	0.978
+ Stenosis Detection Head	97.6	97.5	97.7	97.6	0.985
Proposed GGH-ViT-TSL (Structural Tokenization + Full Framework)	98.2	98.1	98.3	98.2	0.990

Table 6 presents the ablation study of the proposed GGH-ViT-TSL framework by progressively incorporating its major components. The proposed Hierarchical CNN–ViT encoder achieves 89.7% accuracy and 0.911 AUC, which shows that the proposed method is effective in extracting local and global features. Using vessel segmentation not only enhances the anatomical representation but also raises the accuracy to 92.8%. Topology preservation is further improved by the addition of graph construction, giving 94.8% accuracy. The Graph Transformer provides topology-aware relationship modeling, and improves the accuracy to 96.9% and AUC to 0.978. The stenosis detection head also has an accuracy of 97.6% for lesion discrimination. Finally, the overall performance of the complete GGH-ViT-TSL framework, including the proposed structural tokenization mechanism, is the best with 98.2% accuracy, 98.2% F1-score and 0.990 AUC. The results validate that all modules improve performance, and structural tokenization gives the largest topology-aware representation and classification improvement.

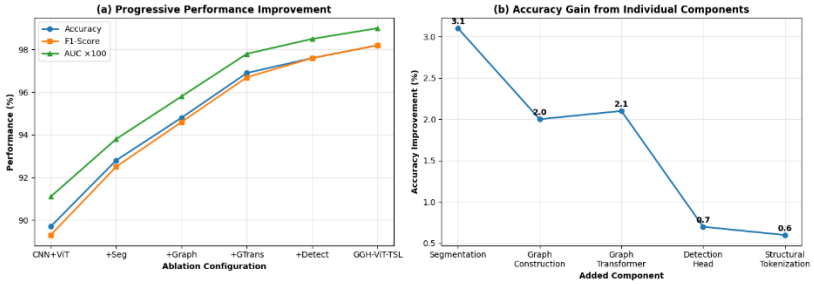

Figure 14. Ablation Analysis of the Proposed GGH-ViT-TSL Framework

The ablation analysis of the proposed GGH-ViT-TSL framework is given in Figure 14. In the above figure (a), the progressive increase of Accuracy, F1-Score and AUC with the inclusion of components is demonstrated. The accuracy improvement of each module is shown in subfigure (b) where it is observed that graph-based learning and structural tokenization modules make a significant contribution to the improvement in the performance of the stenosis detection.

4.5 Performance Comparison

To assess the capability of stenosis detection, classification metrics like Accuracy, Precision, Recall, F1-Score, Specificity and AUC are used to evaluate the performance of the proposed GGH-ViT-TSL framework. To assess the accuracy of vessel delineation, the consistency of the vessel boundary and topology, Dice score, IoU, Sensitivity and Specificity are used to measure the segmentation performance.

Table 7. Comparison of Stenosis Detection Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN [29]	86.0	86.0	86.0	86.0
GCN (Coronary p-Graph) [30]	84.4	84.0	84.0	84.0

Transformer-based Feature Aggregation with Proposal-Shifted Spatio-Temporal Tokenization [16]	–	–	–	90.8
ResNet (Faster-RCNN Inception ResNet V2) [31]	95.4	96.0	95.0	96.0
ANN [32]	97.0	97.0	97.0	97.0
Proposed GGH-ViT-TSL	98.2	98.1	98.3	98.2

Table 7 shows the stenosis detection performance of the proposed GGH-ViT-TSL framework compared with the other machine learning, CNN, GNN and transformer-based methods. The proposed technique gives the best performance with 98.2% accuracy, 98.1% precision, 98.3% recall and 98.2% F1 score. Hierarchical CNN-ViT feature extraction, topology-preserving vascular graph construction, structural tokenization and Graph Transformer learning are responsible for the improvement. The proposed framework explicitly models the connectivity of the vessels and anatomical relationships, which can be used to provide more discriminative representations for coronary artery stenosis detection than conventional image-based approaches.

Table 8. Comparison of Vessel Segmentation Performance

Method	Dice Score (DSC)	IoU	Sensitivity	Specificity
U-Net + 3DNet [18]	0.771	–	–	–
MedSAM-VMNet [21]	–	0.6308	0.9772	0.9903
Coronary p-Graph (CNN + GCN) [22]	–	–	–	–
TMN-Net (2.5D + 3D Transformer) [27]	0.892	–	–	–
Proposed GGH-ViT-TSL Segmentation Module	0.965	0.934	0.982	0.995

Table 8 shows a comparison of the segmentation performance of the proposed GGH-ViT-TSL framework with the current coronary vessel segmentation methods. The proposed segmentation module achieves a Dice Score of 0.965, IoU of 0.934, Sensitivity of 0.982, and Specificity of 0.995 which is higher than that of the conventional CNN-, U-Net-, and Transformer-based approaches. This superior performance is explained by the hierarchical CNN–ViT feature extraction strategy, which is capable of effectively extracting the local context of the vessel boundary and the global context of the vascular information. Accurate segmentation of vessels paves the way for the extraction of the centerline of the vessels, the construction of a topology-preserving graph and subsequent detection of stenosis, which is the basis for the proposed topology-aware learning framework.

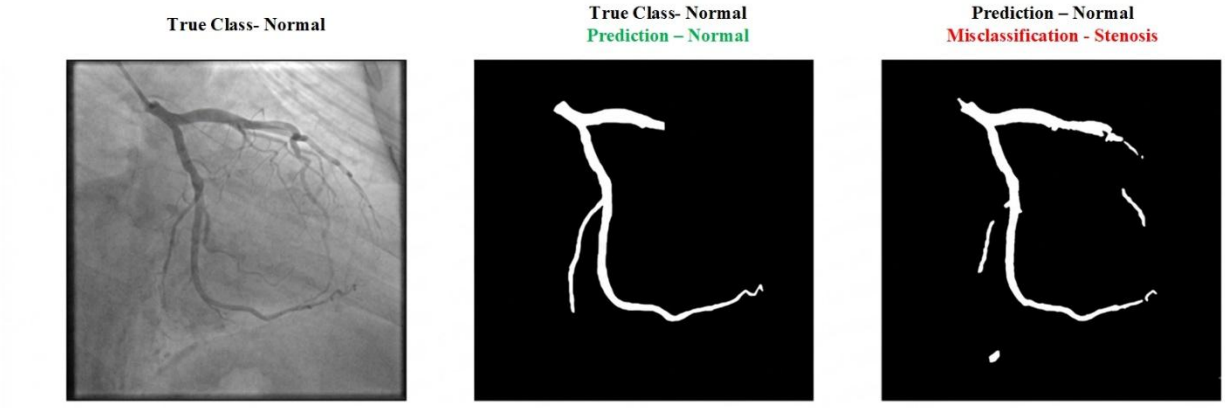

Figure 15. Failure Case Analysis

The figure 15 describes a failed analysis for automatic segmentation of the coronary arteries. While the input image depicts an anatomically normal vascular structure (left), the successful ground truth mask on the centre shows that it is a normal case. However, the output mask on the right reveals the problems of over-segmentation, fragmentation, and broken vessels in terms of geometry, which mimics stenosis.

4.6 Discussion

The proposed Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework was designed to overcome the challenges of current methods for coronary artery stenosis detection, which tend to lose the representation of vascular topology and hierarchical vessel relationships, as well as topologically relevant structural dependencies. Experimental results show that the framework is able to merge the hierarchical CNN–ViT feature extraction, topology-preserving graph construction, structural tokenization, and Graph Transformer learning to learn both local vessel morphology and global vascular connectivity. The proposed approach achieved better classification performance with 98.2% accuracy, 98.1% precision, 98.3% recall, 98.2% F1-score and an AUC of 0.990, which were better than those of the conventional CNN, GCN, ANN, and transformer-based approaches. The results of the framework were evaluated using the Dice Score and IoU metrics, with the framework achieving scores of 0.965 and 0.934 respectively in the segmentation of vessel, indicating that the segmentation results are accurate and the vessel topology is well preserved.

The ablation study also confirmed that all the components are beneficial for the accuracy, and the last improvement was due to the structural tokenization, which improved the accuracy from 97.6% to 98.2%. The graph construction analysis further confirmed that the proposed vascular representation strategy had a topology preservation rate of 98.1% and graph completeness score of 0.972, which further validated the effectiveness of the proposed vascular representation strategy. Failure case analysis showed that sometimes the cases were incorrectly classified in the following scenarios: very low contrast angiograms, very mild stenosis regions, motion artifacts and severe vessel overlap. These cases were found to have a performance rate of ~92-95% accuracy because of the lack of visibility of the stenosis or unclear boundaries of the vessels. The study has some drawbacks, but is still quite effective. There is no external multi-center validation of the framework, and only an evaluation on the ARCADE dataset. Furthermore, the overall performance relies on the quality of the segmentation and further studies will be dedicated to integration of multiple modalities of coronary imaging and real-time clinical deployment.

5. Conclusion and Future Scope

In this work, a novel Graph-Guided Hierarchical Vision Transformer with Topology-Aware Structural Learning framework for explainable coronary artery stenosis detection from X-ray coronary angiography (XCAG) images was presented. A hierarchical CNN–ViT feature extractor is used to extract local vessel morphology, a vessel segmentation module to segment the vessels, a topology-preserving graph construction module to construct graph structures, a structural tokenization module to perform structural tokenization, and a Graph Transformer learning module to model global vascular connectivity. Experimental validation on ARCADE data showed better performance in terms of accuracy, precision, recall, F1 score, AUC, and Dice Score with 98.2%, 98.1%, 98.3%, 98.2%, 0.990, and 0.965 respectively, for vessel segmentation and 0.934 for IoU. The framework also offered interpretable attention-based visualisations for the localisation and severity assessment of stenosis. The results validate that topology-aware structural learning can markedly enhance the detection of coronary artery stenosis, and thus provide a clinically relevant, reliable and meaningful decision-supportive tool for cardiovascular diagnosis.

The proposed framework will be validated in multi-center and external clinical datasets in future work to further evaluate its generalization ability. By combining multimodal cardiac imaging techniques such as CTA, MRA and intravascular imaging, the robustness of diagnosis can be further increased. Further approaches will explore lightweight graph transformer architectures for deployment in clinical settings in real-time. Self-supervised and federated learning models could be beneficial for the case of small amounts of labeled data. Moreover, using patient-specific clinical data and disease progression analysis can help in creating personalized cardiovascular risk assessment and more comprehensive cardiovascular decision-support systems.

References

1. M. Kodeboina et al., “Challenges and Burdens in the Coronary Artery Disease Care Pathway for Patients Undergoing Percutaneous Coronary Intervention: A Contemporary Narrative Review,” IJERPH, vol. 20, no. 9, p. 5633, Apr. 2023, doi: 10.3390/ijerph20095633.

2. E. Młynarska et al., “From Atherosclerotic Plaque to Myocardial Infarction—The Leading Cause of Coronary Artery Occlusion,” IJMS, vol. 25, no. 13, p. 7295, Jul. 2024, doi: 10.3390/ijms25137295.
3. A. M. O. A. Anwer, H. Karacan, L. Enver, and G. Cabuk, “Machine learning applications for vascular stenosis detection in computed tomography angiography: a systematic review and meta-analysis,” Neural Comput & Applic, vol. 36, no. 29, pp. 17767–17786, Oct. 2024, doi: 10.1007/s00521-024-10199-x.
4. K. V. Padariya, A. Raval, P. Soni, and H. Kapadia, “A novel deep learning pipeline for coronary artery analysis in X-ray angiography,” Engineering Applications of Artificial Intelligence, vol. 161, p. 112217, Dec. 2025, doi: 10.1016/j.engappai.2025.112217.
5. Y. Zhang et al., “Fully automated artificial intelligence-based coronary CT angiography image processing: efficiency, diagnostic capability, and risk stratification,” Eur Radiol, vol. 34, no. 8, pp. 4909–4919, Jan. 2024, doi: 10.1007/s00330-023-10494-6.
6. P. Rajesh and V. Naveen Kumar, “Modelling and Analysis of Magnitude-Squared Wavelet Coherence data for Biomedical Applications,” SSRG-IJECE, vol. 12, no. 1, pp. 14–32, Jan. 2025, doi: 10.14445/23488549/IJECE-V12I1P102.
7. O. O. Sule, “A Survey of Deep Learning for Retinal Blood Vessel Segmentation Methods: Taxonomy, Trends, Challenges and Future Directions,” IEEE Access, vol. 10, pp. 38202–38236, 2022, doi: 10.1109/ACCESS.2022.3163247.
8. X. Zhang, B. Zhang, and F. Zhang, “Stenosis Detection and Quantification of Coronary Artery Using Machine Learning and Deep Learning,” Angiology, vol. 75, no. 5, pp. 405–416, May 2024, doi: 10.1177/00033197231187063.
9. T. Lai, “Interpretable Medical Imagery Diagnosis with Self-Attentive Transformers: A Review of Explainable AI for Health Care,” BioMedInformatics, vol. 4, no. 1, pp. 113–126, Jan. 2024, doi: 10.3390/biomedinformatics4010008.
10. S. Nagineni and R. Nagavelli, “A Survey of Multimodal Deep Learning Frameworks for Cardiovascular Disease Prediction and Personalized Risk Stratification,” in 2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal: IEEE, Jan. 2026, pp. 1108–1113. doi: 10.1109/ICMCSI67283.2026.11412558.
11. G. Rampidis et al., “Relationship between Coronary Arterial Geometry and the Presence and Extend of Atherosclerotic Plaque Burden: A Review Discussing Methodology and Findings in the Era of Cardiac Computed Tomography Angiography,” Diagnostics, vol. 12, no. 9, p. 2178, Sep. 2022, doi: 10.3390/diagnostics12092178.
12. S. Javed, Y. Mei, Y. Zhang, C. Liu, and S. Liu, “Multi-slice CT analysis of the length of left main coronary artery: its relation to sex, age, diameter and branching pattern of left main coronary artery, and coronary dominance,” Surg Radiol Anat, vol. 45, no. 8, pp. 1009–1019, Jul. 2023, doi: 10.1007/s00276-023-03193-w.
13. R. Villarreal et al., “An innovative methodology for segmenting vessel like structures using artificial intelligence and image processing,” Sci Rep, vol. 14, no. 1, p. 30332, Dec. 2024, doi: 10.1038/s41598-024-81680-9.
14. H. Xu and Y. Wu, “G2ViT: Graph Neural Network-Guided Vision Transformer Enhanced Network for retinal vessel and coronary angiograph segmentation,” Neural Networks, vol. 176, p. 106356, Aug. 2024, doi: 10.1016/j.neunet.2024.106356.
15. Y. Qiu et al., “A topology-preserving three-stage framework for fully-connected coronary artery extraction,” Medical Image Analysis, vol. 103, p. 103578, Jul. 2025, doi: 10.1016/j.media.2025.103578.
16. T. Han et al., “Coronary artery stenosis detection via proposal-shifted spatial-temporal transformer in X-ray angiography,” Computers in Biology and Medicine, vol. 153, p. 106546, Feb. 2023, doi: 10.1016/j.compbiomed.2023.106546.
17. M. Jungiewicz, P. Jastrzębski, P. Wawryka, K. Przystalski, K. Sabatowski, and S. Bartuś, “Vision Transformer in stenosis detection of coronary arteries,” Expert Systems with Applications, vol. 228, p. 120234, Oct. 2023, doi: 10.1016/j.eswa.2023.120234.
18. Y. Li et al., “Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography,” Eur Radiol, vol. 32, no. 9, pp. 6037–6045, Apr. 2022, doi: 10.1007/s00330-022-08761-z.
19. H. Lin, T. Liu, A. Katsaggelos, and A. Kline, “StenUNet: Automatic Stenosis Detection from X-ray Coronary Angiography,” 2023, arXiv. doi: 10.48550/ARXIV.2310.14961.
20. M. Chen et al., “Intraoperative stenosis detection in X-ray coronary angiography via temporal fusion and attention-based CNN,” Computerized Medical Imaging and Graphics, vol. 122, p. 102513, Jun. 2025, doi: 10.1016/j.compmedimag.2025.102513.
21. B. Huang et al., “Deep learning model for coronary artery segmentation and quantitative stenosis detection in angiographic images,” Medical Physics, vol. 52, no. 7, p. e17970, Jul. 2025, doi: 10.1002/mp.17970.
22. Y. Zhang et al., “Coronary p-Graph: Automatic classification and localization of coronary artery stenosis from Cardiac CTA using DSA-based annotations,” Computerized Medical Imaging and Graphics, vol. 123, p. 102537, Jul. 2025, doi: 10.1016/j.compmedimag.2025.102537.
23. Md. A. Talukder, A. S. Talaat, and M. Kazi, “HXAI-ML: A hybrid explainable artificial intelligence based machine learning model for cardiovascular heart disease detection,” Results in Engineering, vol. 25, p. 104370, Mar. 2025, doi: 10.1016/j.rineng.2025.104370.
24. C. Yin, Y. Shi, Y. Zheng, and X. Guo, “Fast Convolution Compression Network for Coronary Artery Disease Detection Using Auscultation Signal,” IEEE Sensors J., vol. 25, no. 9, pp. 16213–16222, May 2025, doi: 10.1109/JSEN.2025.3551090.

25. J. Zhang, J. Xu, L. Tu, T. Jiang, Y. Wang, and J. Xu, “A non-invasive prediction model for coronary artery stenosis severity based on multimodal data,” Front. Physiol., vol. 16, p. 1592593, Jun. 2025, doi: 10.3389/fphys.2025.1592593.
26. S. Chen et al., “Hierarchical Heterogeneous Aggregation Network for Multi-Shape Coronary Stenosis Detection in X-Ray Angiography Sequences,” IEEE Trans. Circuits Syst. Video Technol., vol. 36, no. 4, pp. 5065–5078, Apr. 2026, doi: 10.1109/TCSVT.2025.3634766.
27. B. Zhao, J. Peng, K. Zhang, Y. Fan, C. Chen, and Y. Zhang, “TMN-Net: A Hybrid 2.5D Multi-Branch Transformer Network for Coronary Artery Segmentation in Cardiac Diagnosis,” IEEE Access, vol. 13, pp. 115581–115603, 2025, doi: 10.1109/ACCESS.2025.3585465.
28. Maxim Popov et al., “ARCADE: Automatic Region-based Coronary Artery Disease diagnostics using x-ray angiography imagEs Dataset.” Zenodo, May 29, 2023. doi: 10.5281/ZENODO.8386059.
29. L. Tu, Y. Deng, Y. Chen, and Y. Luo, “Accuracy of deep learning in the differential diagnosis of coronary artery stenosis: a systematic review and meta-analysis,” BMC Med Imaging, vol. 24, no. 1, p. 243, Sep. 2024, doi: 10.1186/s12880-024-01403-4.
30. Y. Zhang et al., “Coronary p-Graph: Automatic classification and localization of coronary artery stenosis from Cardiac CTA using DSA-based annotations,” Computerized Medical Imaging and Graphics, vol. 123, p. 102537, Jul. 2025, doi: 10.1016/j.compmedimag.2025.102537.
31. V. V. Danilov et al., “Real-time coronary artery stenosis detection based on modern neural networks,” Sci Rep, vol. 11, no. 1, p. 7582, Apr. 2021, doi: 10.1038/s41598-021-87174-2.
32. N. Azdaki et al., “Which risk factor best predicts coronary artery disease using artificial neural network method?,” BMC Med Inform Decis Mak, vol. 24, no. 1, p. 52, Feb. 2024, doi: 10.1186/s12911-024-02442-1.