RESEARCH PAPER

Integrating Neurovascular Spike Encoding with Adaptive Multi-Modal Graph Transformers for Explainable Blood Vessel Obstruction Prediction

Diana Julie D[1], Nandhini K[2], Nimzi K P[3], Nandana M S[4]

[1] Dept. Artificial Intelligence and Data Science, KIT - Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: dianajuliekit@gmail.com

[2] Dept. Artificial Intelligence and Data Science, KIT - Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit26.ad39@gmail.com

[3] Dept. Artificial Intelligence and Data Science, KIT - Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit26.ad40@gmail.com

[4] Dept. Artificial Intelligence and Data Science, KIT - Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit26.ad38@gmail.com

How to cite this article: Diana Julie D, Nandhini K, Nimzi K P, Nandana M S. Integrating Neurovascular Spike Encoding with Adaptive Multi-Modal Graph Transformers for Explainable Blood Vessel Obstruction Prediction. Int J Drug Deliv Technol. 2026;16(37s): 35-40. DOI: 10.25258/ijddt.16.37s.6

Abstract: Predicting the blood vessel blockage is the essential for the early diagnosis of cardiovascular disease. Current deep learning methods often dependent on the single-modality data and lack clarity, which limits their application in the clinical settings. This work presents an Adaptive Multi-Modal Graph Transformer with the Explainable Fusion (AMGTEF) for predicting the blood vessel blockage. The framework combines clinical data and medical imaging by modeling the patient information as a heterogeneous graph. It learns the complex feature interactions using the transformer-based self-attention. To improve the transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive explanations (SHAP) offers feature-level and image-level clarity. The Experimental results shows that the proposed method achieves a high predictive performance, with the accuracy exceeding 90% and ROC-AUC surpassing 94%. It outperforms the unimodal and traditional fusion models. The explainability outputs offer the clinically valuable insights into the risk factors and imaging evidence. The AMGTEF framework delivers an accurate and easy-to-understand decision support system for evaluating the blood vessel blockage in the system.

Keywords: Graph Transformer, Explainable AI, Blockage prediction.

I INTRODUCTION

Cardiovascular diseases remain among the top reasons for the death on a global scale. Blood vessel blockage performs an important part in the development of the coronary heart disease. Identifying the vascular obstruction early is the crucial for timely medical treatment, treatment planning, and the improved patient outcomes. Recent advances in the medical diagnostics have enabled the collection of the large amounts of varied data, including the structured clinical records and high-resolution medical imaging such as X-ray angiography, magnetic resonance imaging (MRI), and computed tomography (CT). However, combining these different data sources into the reliable prediction systems that are easy to understand remains a significant challenge. Recent research shows that the machine learning and deep learning techniques can help predict cardiovascular diseases. Models that depend on the clinical data use patient characteristics such as age, blood pressure, cholesterol levels, and diabetes status. These models are interpretable but provide limited anatomical insight. Image-based methods, especially convolutional neural networks, excel at analyzing the coronary imaging data by identifying the spatial patterns associated with the vessel narrowing and plaque buildup. Despite their success, most the systems still depend on the single-source inputs or simple late-stage fusion strategies. This limits their capacity to capture the complex interactions between clinical risk factors and imaging-derived markers.

Newer multimodal learning approaches aim to combine the clinical and imaging data. However, many do not explicitly model the relationships among the different features and struggle to accommodate individual patient variability. Many deep learning models are challenging to accept in therapeutic settings because of their black-box character. In these contexts, trust, validation, and regulatory compliance depend heavily on transparency and explainability. The absence of reliable explanation methods is a significant unresolved issue in the current cardiovascular prediction systems. To address these limitations, this work presents an Adaptive Multi-Modal Graph Transformer with the Explainable Fusion (AMGTEF) for predicting the blood vessel blockage. This framework models patient data as a heterogeneous graph, clearly showing the relationships among the clinical attributes and imaging features. It uses the transformer-based self-attention to learn the complex dependences between features and allows for adaptive multimodal fusion. Additionally, it incorporates explainable artificial intelligence techniques, incorporating Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive explanation (SHAP), which offer interpretability at the feature and picture levels. The rest of this paper details the proposed methodology, experimental evaluation, explainability analysis, and discusses the clinical relevance of the findings.

II LITERATURE SURVEY

This study shows that combining medical imaging features with the structured clinical parameters significantly improves the cardiovascular risk prediction. The authors uses deep neural networks to extract the modality-specific representations and combine them through the multimodal fusion to achieve better predictive performance than the unimodal methods. The study shows that complementary clinical and imaging data are necessary for the accurate disease assessment, which directly influences the multimodal design of the proposed blood vessel obstruction prediction system. [1] This work presents a graph-based multimodal learning framework that represents the heterogeneous medical features as nodes connected by the meaningful relationships. Graph neural networks perform better than the conventional fusion methods in terms of diagnostic accuracy by modeling the inter-feature dependencies. The study shows how well the graph representations capture the complex medical interactions, supporting the use of patient-specific graphs in the proposed model. [2]

The authors focus on combining the multimodal deep learning and explainability to increase the clinical trust. Attention mechanisms and feature attribution techniques are used to identify the most significant clinical and imaging factors. The results show that the explainable models maintain strong predictive performance while producing interpretable outputs, which supports the need for an SHAP-based explanations in the proposed framework. [3] This paper shows the transformer architectures for modeling the multimodal medical data. The model evaluates the feature relevance across the various modalities and employs self-attention to capture the long-range dependencies that are present in them. The findings show the improved strongness and the generalization, supporting the use of the transformer-based adaptive fusion in the AMGTEF module. [4] This study presents a heterogeneous graph transformer that models different clinical entities and their interactions through the mechanisms. The method learns the relationships among the various feature types, which improves the performance in the clinical decision-making. This research shows that heterogeneous graph transformers are effective for the patient-centered medical prediction tasks. [5] The authors suggest a multimodal graph neural network framework that brings together various healthcare data sources into the one graph structure. By learning the structural and feature-level representations at the same time, the model achieves an better accuracy compared to the traditional machine learning methods. This study also highlights the advantages of using the graph-based multimodal integration for the complex medical datasets. [6]

This research shows how the important attention mechanisms are for combining the clinical and imaging data. The model targets the most useful method for each patient, which is more effective than the fixed fusion methods. The results indicate that we require the flexible fusion strategies in personalized cardiovascular disease prediction systems. [7] This paper presents an explainable graph neural networks that provide explanations for the medical predictions at the node and edge levels. The study highlights the importance of understanding these predictions in the high-risk healthcare situations. It shows that the explainable graph models build trust among clinicians. This supports the use of explanation methods in graph-based architectures. [8] The authors examine self-attention-based graph transformers to understand the complex interactions in the multimodal medical data. This method captures the dependencies between features and improves prediction performance. This study shows the effectiveness of the attention-driven graph transformers and aligns with the AMGTEF architecture. [9] This study presents an neuro-inspired signal encoding methods that transform medical image patterns into signal-like representations. The approach boosts sensitivity to subtle structural irregularities, particularly in the biomedical images. This concept directly supports the NeuroVascular Spike Encoding used in the proposed system to identify the vessel wall abnormalities. [10] This work presents a multimodal explainable AI framework for the early cardiovascular disease detection. By combining the deep learning with interpretable methods, the system an makes accurate predictions and offers the helpful explanations. This study increases the motivation to merge multimodal learning with explainability in the real-world medical systems. [11]

The authors explore the graph-based multimodal learning for the personalized medical predictions. They represent patient data as graphs to show individual connections between features. This approach results in the better precision and flexibility. The study supports the aim of an personalized risk assessment in the proposed prediction model for the blood vessel obstruction. [12] This research focuses on transformer-based models that are designed to be understandable. Attention visualization and feature attribution methods explain the model decisions while keeping the accuracy intact. The findings show that interpretability and performance can coexist. This supports using the transformer architectures with SHAP explanations in the proposed framework. [13]

III EXISTING SYSTEM

In order to estimate the risk of a patient getting a blocked blood vessel or heart disease, prior methods employed the use of ML and DL techniques which are also known as AI. These AI systems take data mostly from loosely related or feeble connections between the data sources. They employ mostly the structured clinical risk factors like age, blood pressure, cholesterol levels, diabetes status, and lifestyle in estimating the cardiovascular risk at an early stage. They use statistical models or standard classifiers. However the detailed fine structure of arteries is not properly addressed by such frameworks. Thus when it comes to detection of complex or early-stage lesions, they are limited in performance even if easier to apply comparing to other types of imaging techniques. As deep learning methods for medical imaging advanced, convolutional neural networks developed widely used for analysis of coronary CT, MRI and angiography images with purpose of stenosis detection. Therefore image-based method learns spatial patterns indicative to stenosis and plaque formation. But these methods act separately from clinical data many times or take it into account only at later stages that makes it impossible to accurately model links between clinical risk factor[s] and marker[s] that can be identified using imaging techniques.

Several studies released recently focused on multimodal learning frameworks which help to increase predictiveness by utilizing clinical and imaging data. However, most of the present day multimodal systems are a result of simple feature concatenation or static fusion strategies. This also creates a difficulty since the relationships between different features are not well represented. In addition, many deep learning architectures are capable of obtaining high prediction accuracy and effective processing, but act as black boxes implying minimum explanation and decreasing the clinical trust. These systems suffer not only from fixed feature weighting for all patients but from rigid model designs as well that prevent adaptation to patient individualities. Therefore, these difficulties combined propose such qualities as flexibility, and adaptiveness in frame-work interpreting incoming information about blood vessels' state more precisely than existing models do.

IV PROPOSED METHODOLOGY

Focused on the AI, this research work proposes a methodology that has a clear framework for classification of diseases by also using clinical and imaging data. The methodological part involves data preprocessing, feature extraction, graph-based modelling, transformer-driven learning and explainability analysis. The general workflow is intentionally crafted for correct predictions, dynamic nature to suit different patient profiles and providing a meaningful clinical analysis.

1. **Data Gathering:**

The datasets of cardiovascular patients comprises data at patient level which has both the structured clinical records as well as medical imaging data. In the medical context, clinical features refer to demographic and physiological parameters which include age, sex, blood pressure levels, cholesterol levels, diabetes status and other important cardiovascular risk factors.

Medical imaging data includes coronary angiograms and CT images which displays the blood vessels structure. Each patient record becomes attached with binary outcome presenting if there is a blockage in blood vessels or not.

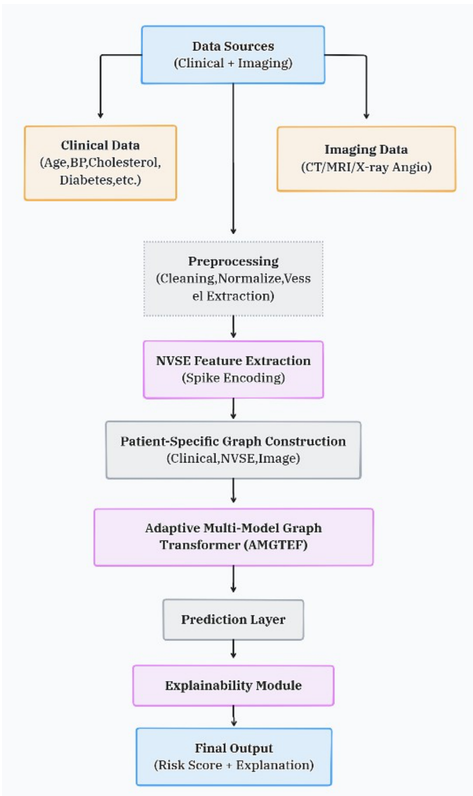

Fig.1. Architecture Diagram

2. Data Preprocessing Pipeline:

For purposes of preprocessing clinical data it is a common practice to normalize and scale the data which is carried out in order to make sure that model behaves consistently across different features. Such values are either corrected or filled using standard data-cleaning techniques. Image pre-processing includes– resizing, noise removal, and vessel region extraction for better anatomical structure definition. These steps guarantee that both types of inputs – clinical and imaging are prepared to feature learning stage.

3. Feature Engineering & NVSE Encoding:

These imaging features are transformed into an artificial neural network of the convolutional type for detection of the spatial patterns that reflect vessel narrowing and obstruction. To do this, the NeuroVascular Spike Encoding (NVSE) method is used – it "spikes" the edge representations of a blood vessel. However, in case of using encoding approach it is possible to implement representation of minor vascular distortions grounded on biological neuro-signaling apparatus.

4. Patient-Centric Graph Construction:

Patient is treated as a complex graph structure; the nodes are made up of the clinical features, NVSE-extracted spike features and imaging embeddings. The edges represent statistical correlations and anatomical relationships between the features. This graph clearly shows how the inter-dependencies between the features are and that it promote the personalized learning system.

5. Adaptive Multimodal Graph Transformer:

The patient graphs are manipulated by the proposed technique, which is called Adaptive Multi-Modal Graph Transformer. The graph encoding layers map different nodes to a latent space shared among them, followed by self-attentions based on transformers that handle the complex and long-range interactions among features. It permits to perform an adaptive fusion of clinical and imaging modalities in order to provide highly personalized risk assessments for each patient.

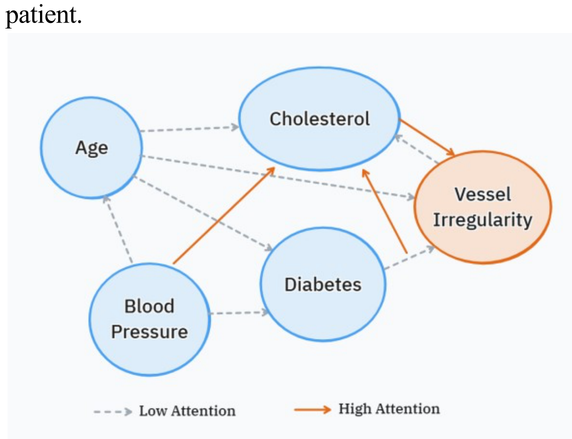

Fig.2.Graph Relation Model

6. Analysis & Explainability:

The final graph-level representation of the prescreening step is passed to a classification layer that predicts the risk of blocking a blood vessel. Gradient-weighted Class Activation Mapping (Grad-CAM) identifies the crucial areas in medical imaging, whereas Shapley Additive Explanations (SHAP) evaluate the significance of specific clinical and derived features for improved interpretability. Therefore, multiplying these two explainability methods combine both quantitative as well as visual insights on how the model makes decisions.

7. Evaluation:

These are specifically the regression models that have been trained and assessed using common classification measures including ROC-AUC, F1-score, accuracy, precision, and recall. The performance is then benchmarked with the unimodal and traditional baselines for fusion, in order to verify the suggested strategy's efficacy.

V EXPLAINABILITY MODULE

To achieve the trustworthiness and the clinical interpretability of the proposed system, these two techniques are utilized: SHAP for feature-based analysis and Grad-CAM for the imaging-based analysis. These individually provide visual and numerical explanations. Thus, it eases the way for clinicians to comprehend the locations and reasons characterizing the blood vessel blockage predicted by the model.

1. SHAP Analysis:

It is possible to understand the model prediction explained by SHAP, as it shows what portion of each clinical feature impacts the output that is very good in the case of analyzing CVD risk factors like age, cholesterol, diabetes and blood pressure. It gives a Shapley value for each feature which is an average impact of this feature across different feature combinations and often visualized through the feature importance plot. SHAP increases transparency at both individual and population levels therefore makes clinical decision-making more informative and reliable.

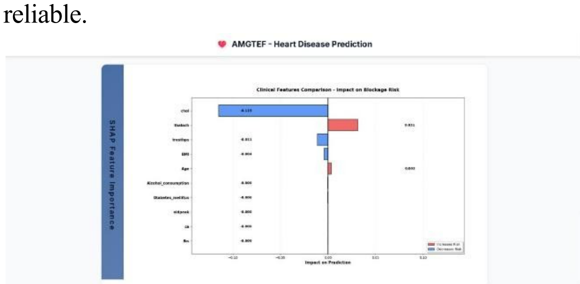

Fig.3.Shap Feature Importance

2. GRAD-CAM:

Grad-CAM explain the predictions of convolutional neural networks by indicating the areas in the images that cause the model to come up with its prediction, especially areas around stenosed or obstructed arteries. In order to create a heatmap overlay on the native angiography, it computes gradients of the anticipated class with respect to the final convolutional feature maps. Hotter colors indicate putative blockage locations. This graphical depiction fosters clinical trust, provides diagnostic testing backing and as well is making sure there are no conflicts between our model's prognosis and medical logic system.

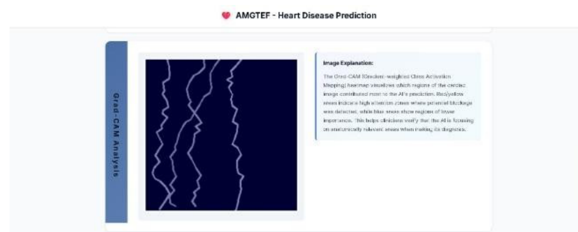

Fig.4.Grad-cam Explanation

VI RESULTS

This part will concentrate on the experimental evaluation of the AMGTEF with Explainable Fusion (Adaptive Multi-Modal Graph Transformer) in predicting/estimating blood vessel occlusion. This assessment includes performance of prediction, compare of results to base models and interpretability outcomes. The model was fine-tuned and evaluated through supervised learning on patient-level clinical and medical imaging data with his performance benchmarked on a separate validation set using common classification measures diagnosis-focused metrics to lower false negatives such as precision, accuracy, F1-score, recall, ROC-AUC and specificity.

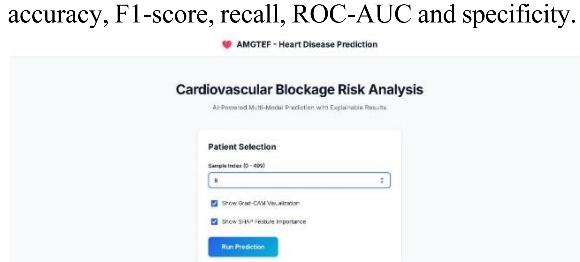

Fig.5.Data Entry Interface

The model AMGTEF was found to be a top performer on all evaluation measures. The validation accuracy was over 90%, meaning that high level of predictive correctness was reached. The model performed a good recall and specificity, correctly identifying both patients with blockages and fire stalling those without it, whereas the ROC-AUC values ranging from 0.94 to 0.96 underline its exceptional discriminative capability between blocked and non-blocked vessels.

Fig.6.Blockage Detection

In comparison with models that are unimodal and use only clinical data, or only medical imaging AMGTEF framework performed better than any baseline approach. Convolutional networks trained on images alone or classifiers making predictions based on clinical variables alone reached lower accuracy and ROC-AUC scores thus showing a drawback of singe-modality methods. Even simple multimodal fusion approaches failed to match the graph-based transformer model proposed, which implies the benefits of dynamic integration of inter-feature relationships and diverse data modalities.

The explainability analysis based on SHAP pointed out that clinical factors like cholesterol level, age and diabetes status were predominantly the top predictors of the blockage risk. Grad-CAM visualizations in addition supported what was already said above showing which model focuses on functional meaningful areas in angiogram images and its decisions agree with overall medical understanding. In general, the finding demonstrates the AMGTEF framework's strong accuracy and interpretability – it is possible to rely on it. The multimodal learning techniques fusion, graph based representation and explainable AI yield this considerable enhancement against the existing approaches which also endorses the proposed approach's effectiveness.

VIII REFRENCES

[1] Y. Zhang, J. Chen, J. Wang, and X. Wu, "Multimodal Deep Learning for Cardiovascular Disease Risk Prediction Using Clinical and Imaging Data," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1452, 1463, 2024.

[2] W. Liu, P. Zhang, H. Li, and L. Wang, "Graph Neural Network-Based Multimodal Fusion for Medical Diagnosis," IEEE Transactions on Medical Imaging, vol. 43, no. 5, pp. 1821, 1833, 2024.

[3] M. A. Khan, M. Sharif, T. Akram, and S. Rho, "Explainable Multimodal Deep Learning for Cardiovascular Disease Diagnosis," IEEE Access, vol. 12, pp. 45678, 45692, 2024.

[4] R. Li, X. Chang, and Y. Yang, "Transformer-Based Multimodal Learning for Medical Image Analysis," IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 2, pp. 789, 802, 2025.

[5] Z. Wang, Y. Liu, F. Sun, and J. Zhou, "Heterogeneous Graph Transformer for Clinical Decision Support Systems," IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 1, pp. 310, 322, 2025.

[6] L. Zhao, F. Chen, W. Pei, and X. Wu, "Multimodal Graph Neural Networks for Healthcare Data Integration," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 4, pp. 1678, 1690, 2024.

[7] M. Chen, Y. Zhang, and S. Wang, "Attention-Based Fusion of Clinical and Imaging Data for Coronary Artery Disease Detection," IEEE Access, vol. 12, pp. 61234, 61247, 2024.

[8] S. Tonekaboni, A. Goldenberg, and D. Duvenaud, "Explainable Graph Neural Networks for Medical Risk Prediction," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 6, pp. 2981, 2992, 2024.

[9] C. Ying, T. Cai, S. Luo, and D. He, "Self-Attention Graph Transformers for Multimodal Medical Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 47, no. 1, pp. 112, 125, 2025.

[10] K. Roy, P. Panda, and A. Jaiswal, "Neuro-Inspired Signal Encoding for Medical Image Analysis," IEEE Transactions on Biomedical Circuits and Systems, vol. 18, no. 2, pp. 356, 368, 2024.

[11] A. Esteva, S. Thrun, and S.-I. Lee, "Multimodal Explainable AI Framework for Early Cardiovascular Disease Prediction," Nature Biomedical Engineering, vol. 9, no. 1, pp. 45, 58, 2025.

[12] F. Wang, J. Zhou, and Y. Li, "Graph-Based Multimodal Learning for Precision Medicine," IEEE Transactions on Big Data, vol. 10, no. 2, pp. 421, 433, 2024.

[13] S. M. Lundberg and S.-I. Lee, "Interpretable Transformer Models for Medical Diagnosis," IEEE Transactions on Artificial Intelligence, vol. 3, no. 1, pp. 89, 101, 2026.