Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EXPLAINABLE CLUSTER-GUIDED CNN–GNN FRAMEWORK FOR RELATIONAL LEARNING AND CONGENITAL HEART DISEASE CLASSIFICATION FROM ECHOCARDIOGRAPHIC IMAGES

ABHIGNA JALLI[1], N. SRINIVASU NULAKA*[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522502, Andhra Pradesh, India
*[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522502, Andhra Pradesh, India

E-mail: [1]jalliabhigna26@gmail.com,*[2]srinivasu28@kluniversity.in

ABSTRACT

The Congenital Heart Disease (CHD) is still one of the most significant reasons for infant morbidity and mortality rates, hence the need for an early and accurate diagnosis that would lead to proper clinical treatment. Although recent advancements in artificial intelligence techniques have significantly contributed to the automatic analysis of echocardiography images, the current methods still face the problem of failing to effectively model the complex anatomical relationships between cardiac structures while being difficult to interpret and less applicable in clinical practice. In light of the stated limitations of current AI technologies in this area, this research suggests an Explainable Cluster-Guided Convolutional Neural Network–Graph Neural Network (CNN–GNN) architecture for automated CHD classification using echocardiographic images. The suggested architecture uses deep feature extraction followed by cluster-guided graph construction to construct a graph with anatomically meaningful cardiac region representation, followed by graph-based relational learning through a Graph Neural Network (GNN) for improved disease classification. At the same time, SHAP and Grad-CAM are used as complementary explainability methods to further improve model transparency, interpretability of the results, and support clinically interpretable decision-making. According to experimental results obtained for the CAMUS echocardiographic dataset, the proposed method was able to classify the images with an accuracy of 98.7%, precision of 98.5%, recall of 98.3%, F1 score of 98.4%, specificity of 99.1% and AUC of 99.1%, thus outperforming not only conventional deep learning but also the current graph-based baseline models. In addition to superior classification performance, the suggested approach is also able to effectively and accurately highlight diagnostically relevant cardiac regions, which improves the reliability, trustworthiness, and interpretability of its automated predictions. Overall, it can be said that the proposed method shows the effectiveness of combining graph representation learning guided by clustering with explainable artificial intelligence in improving diagnostic accuracy as well as making the model more transparent.

Keywords: *Congenital Heart Disease, Echocardiography, Convolutional Neural Network, Graph Neural Network, K-Means Clustering, Explainable Artificial Intelligence, Grad-CAM, SHAP, Medical Image Analysis.*

1. INTRODUCTION

CHD (congenital heart disease) is the most widespread congenital disorder in the world and remains one of the major causes of morbidity and mortality among infants. According to various reports about the state of health worldwide, congenital heart disorders result in numerous infant deaths and contribute to long-term cardiovascular diseases, emphasizing the significance of early diagnosis, timely clinical intervention, and proper clinical treatment [1].

The identification of any structural problems helps start proper therapy, increase survival rates, and avoid possible complications.

Therefore, the development of accurate and efficient diagnostic systems is critical for pediatric cardiology and prenatal care [2].

Echocardiography is currently the main method of CHD diagnosis due to the absence of invasiveness, ability to see images in real-time mode, affordability, and no exposure to ionizing radiation [3]. It gives a comprehensive representation of cardiac cavities, valves, septums, blood circulation, and ventricle functioning, which allows clinicians to diagnose both structural and functional disorders. Compared to other diagnostic imaging methods like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), echocardiography is available, portable, and repeatable, thus it is more preferable to other cardiac imaging modalities [4]. However, the interpretation of the echocardiographic images is complicated by the dependency of image quality on the skills of a specialist, the patient's body features, acoustic artifacts, and imaging conditions [5].

The fast development of Artificial Intelligence (AI) and deep learning technologies has allowed significantly advancing the field of medical image analysis using automated feature extraction and disease classification. Among different architectures of neural networks, Convolutional Neural Networks (CNNs) showed outstanding results in various medical imaging applications like cardiac chamber segmentation, ventricle functioning assessment, disease classification, and anatomical structures detection [6]. Neural networks learn hierarchical spatial representations of raw images automatically and do not require any feature engineering while achieving higher classification accuracy compared to the traditional machine learning models [7]. Thus, CNN-based computer-aided diagnostic systems became very useful and necessary for increasing the efficiency of the diagnosis process and reducing clinician workload.

However, CNN-based models used in congenital heart disease diagnosis suffer from certain limitations. Most of the architectures use global average pooling or fully connected layers to aggregate the feature maps before the classification process. While this approach is computationally efficient, it compresses the spatial information into a single feature vector, thus losing the anatomical relationships and connections between the cardiac structures [8]. Understanding the interaction between the cardiac cavities, interventricular septa, atriums, ventricles, valves, and surrounding tissues is very important for the differentiation of normal and abnormal cardiac anatomy. Missing such information limits the ability of the CNN-based models to discriminate and decreases their robustness across diverse clinical conditions [9].

Graph Neural Networks (GNNs) are considered to be a good deep learning approach that allows representing and processing structured and relational information. Contrary to the CNNs, which mainly pay attention to the local spatial patterns, GNNs represent data as graphs, in which nodes are the meaningful elements and edges are the relations between the nodes. Using iterative message-passing operations, GNNs learn the topological, spatial, and semantic relations between the connected regions; thus, they are suitable for modeling the anatomical structures in medical imaging [10]. Nowadays, the efficiency of Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) in various biomedical fields, such as disease diagnostics, anatomical segmentation, prediction of molecular properties, and medical image classification, has been proven [11]. However, it is difficult to construct reliable graph representations using the images obtained by echocardiography due to the necessity of manual annotation of the anatomical landmarks, which is labor-intensive, time-consuming, and highly dependent on expert knowledge [12].

In order to solve this problem, unsupervised clustering methods may be used as a more efficient way of constructing the graph since they allow grouping similar feature representations into meaningful anatomical regions. Among different clustering methods, the K-Means algorithm is characterized by computational efficiency, scalability, and robustness in partitioning the high-dimensional feature representations into compact clusters [13]. Using the feature maps obtained by CNN and normalized spatial coordinates of the image, the K-Means algorithm generates graph nodes without manual annotation of the landmarks. Such graph representations preserve both semantic information and spatial locality, thus GNNs are able to model the anatomical relationships between the structures of the heart and achieve better disease classification performance.

Although the graph-based deep learning technologies significantly increase the ability of relational feature learning, one more major problem that prevents the widespread implementation of AI-based diagnostic systems in clinics is the lack of interpretability. Most deep learning models are black-box systems that provide some predictions

without explanations of how this result was obtained. In the clinical environment, physicians require transparent and reliable decision-support systems that are able to justify the predictions using clinically meaningful evidence [14]. Thus, the development of Explainable Artificial Intelligence (XAI) becomes an important area of research in the field of medical diagnosis based on deep learning.

Among different explanation methods, Gradient-weighted Class Activation Mapping (Grad-CAM) provides visual explanations that show the image regions that contribute the most to the model predictions and help clinicians to check if the diagnostically meaningful structures of the heart are used during the classification [15]. However, the Grad-CAM technique only provides image-level localization and does not give any explanations of the contribution of the graph representations used by Graph Neural Networks (GNNs). In contrast to this, SHapley Additive exPlanations (SHAP) allow quantifying the contribution of each feature or graph node in the prediction outcome using the principles of cooperative game theory. Thus, the combination of Grad-CAM and SHAP provides complementary explanations at the image and graph levels, improving both model transparency and clinical trustworthiness.

Although the latest deep learning technologies show promising results in the field of echocardiographic image analysis, there are still some problems that remain unsolved. Firstly, most of the CNN-based methods pay attention to the local feature extraction, ignoring the anatomical and spatial relations between the different structures of the heart. In addition, many of the graph-based methods require manually annotated anatomical landmarks, thus, they are not scalable and not applicable in clinics. Moreover, the lack of interpretability of the traditional deep learning models decreases the level of trust of clinicians and prevents the adoption of the mentioned technologies in medical practice. Thus, there is a significant gap in research in the development of an end-to-end framework that simultaneously models anatomical relationships, eliminates the need for manual graph construction, and provides transparent, clinically meaningful explanations for congenital heart disease classification, thereby addressing all the mentioned issues.

Problem Statement: Despite the successful performance of deep learning techniques in classifying congenital heart diseases from echocardiographic images automatically, achieving accurate and reliable classification in this process is not easy. There are some challenges associated with the complicated nature of the anatomical structure of the heart. Besides, the variations in the quality of the images and the inability of the traditional Convolutional Neural Networks (CNNs) to effectively model the spatial and anatomical relationships between different parts of the heart make the classification difficult. In addition, the lack of interpretability of the current deep learning approaches hinders clinicians' trust and restricts the adoption of these approaches in real-world clinical practice. Thus, the need for an intelligent diagnostic framework capable of simultaneously capturing relational anatomical information, performing automated congenital heart disease classification, and providing transparent and clinically meaningful explanations arises.

Research Design

The current study utilizes a quantitative experimental design for the development and validation of the proposed Explainable Cluster-Guided CNN-GNN framework for congenital heart disease classification. The experiment includes image preprocessing, feature extraction with CNN, cluster creation through K-Means for graph construction, classification with GNN, and explanation through Grad-CAM and SHAP. The framework is evaluated using standard classification metrics and compared with selected baseline models.

Research Questions

RQ1: Is it possible to increase the accuracy of classification for congenital heart disease from echocardiographic images with the help of the CNN–Graph Neural Network (CNN-GNN) architecture in comparison with traditional CNNs?

RQ2: Does the conversion of feature maps extracted from CNNs into graphs with the help of K-Means clustering contribute to modeling anatomical and spatial connections between different parts of the heart?

RQ3: Can the incorporation of explainability approaches like Grad-CAM and SHAP increase the model's transparency and deliver clinical explanations?

In order to solve these problems, this study presents an innovative explainable Convolutional Neural Network–Graph Neural Network (CNN–GNN) approach to automated congenital heart disease classification based on echocardiograms. In contrast to the standard deep learning techniques which classify using global pooling of the CNN features,

the proposed approach uses K-Means clustering to transform the feature maps obtained by the CNN into structured graph representations, thus preserving the spatial and anatomical relationships between the cardiac areas. The resulting graph is then used as the input of the Graph Neural Network (GNN), which learns the relational dependencies through the process of graph-based message passing and, thus, enhances discriminative feature learning and classification accuracy. Moreover, this approach utilizes the Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) methods in order to produce two complementary explainable interpretations at the image and graph levels, respectively, thus helping clinicians to ensure that the proposed framework pays attention to clinically relevant cardiac areas while improving the transparency, reliability, and trustworthiness of automated diagnostic decisions.

The uniqueness of this research comes from the fact that the feature maps extracted by the CNN are represented in the form of structured graphs using unsupervised K-Means clustering in order to enable Graph Neural Networks to effectively learn the anatomical connections between the cardiac regions without any manually annotated landmarks. Moreover, this research combines the use of the Grad-CAM and SHAP techniques to obtain two complementary explainable interpretations at the image and graph levels, thereby improving both diagnostic accuracy and clinical interpretability within a unified explainable CNN–GNN framework.

Main Objectives of This Research Are Listed Below

1. To develop a graph-based deep learning framework for automated congenital heart disease classification from echocardiographic images.
2. To transform CNN-extracted feature maps into structured graph representations using K-Means clustering for preserving anatomical and spatial information.
3. To employ Graph Neural Networks for learning anatomical relationships among cardiac regions and improving disease classification performance.
4. To integrate Grad-CAM and SHAP-based explainability techniques for generating transparent and clinically meaningful decision support.
5. To comprehensively evaluate the proposed CNN–GNN framework against conventional CNN-based and state-of-the-art deep learning models using standard medical image classification metrics.

The research objectives have been achieved via the suggested experimental framework and analysis approach. The research creates the CNN-GNN model based on clusters, uses the K-Means clustering algorithm to build graph structures, uses the GNN approach to model the structural relations between heart chambers, employs Grad-CAM and SHAP to provide explainability, and evaluates the model via standard classification measures and comparative baseline models.

The incorporation of convolutional feature extraction, graph-based relational learning, and explainable artificial intelligence in this paper will be of great significance in the design of next-generation intelligent clinical decision-support systems for congenital heart disease diagnosis. It not only improves the automated disease classification process but also enhances transparency, clinical trustworthiness, and the reliability and interpretability of cardiovascular diagnosis from echocardiographic images.

The rest of this paper is organized as follows. In Section 2, we give a brief introduction to the related works in congenital heart disease diagnosis, echocardiographic image analysis, Graph Neural Networks, and explainable artificial intelligence. In Section 3, the proposed methodology is discussed, which includes image preprocessing, CNN-based feature extraction, graph construction using K-Means clustering, Graph Neural Network-based classification, mathematical formulation, and integration of Grad-CAM and SHAP for model interpretability. In Section 4, the experimental setup, performance evaluation, comparative analysis with existing state-of-the-art methods, and visualization of explainability results are presented. Finally, Section 5 concludes the paper by summarizing the major findings, discussing the limitations of the proposed framework, and outlining future research directions.

2. RELATED WORK

The progress in computer technologies allowed the emergence of Artificial Intelligence (AI) that revolutionized many areas, including medicine. The development of machine learning algorithms resulted in automated medical image analysis, automated disease diagnosis, and clinical

decision support. Cardiovascular disorders account for many cases of infant morbidity and mortality since Congenital Heart Disease (CHD) poses complicated anatomical issues. Echocardiography remains the key instrument for CHD diagnosis as it allows real-time visualizations of cardiac structures without harmful effects associated with ionizing radiation. At the same time, CHD diagnosis is a very challenging task since it requires high skills and depends heavily on specialist expertise, leading to inter-observer variability and increased diagnostic workload. Therefore, this issue can be solved through the application of computer-aided diagnostic systems for automated echocardiographic image analysis.

Initial Computer-Aided Diagnostic (CAD) systems used traditional image processing techniques, like image enhancement, segmentation, and handcrafted feature extraction [16]. According to Gonzalez and Woods, these methods provide better image visualizations and improve the quality of medical images, but they are very sensitive to speckle noise, imaging artifacts, and operator-dependent variations. As a result, such traditional image processing techniques did not lead to significant changes in terms of automated CHD diagnosis.

Another improvement was associated with the introduction of standardized echocardiographic guidelines, which provided consistent protocols for cardiac chamber quantification and ventricular function evaluation. So, Lang et al. [17] provided recommendations for cardiac chamber quantification, while Nagueh et al. [18] presented the Doppler echocardiographic guidelines for cardiac function evaluation. However, none of these innovations could substitute clinical interpretation of findings and, hence, reduce the dependence on the expertise of specialists, highlighting the need for intelligent diagnostic systems that could assist clinicians.

In order to minimize the need for clinical interpretation of results and to provide automatic disease classification, machine learning techniques were applied. Therefore, Vapnik [19] introduced the concept of Support Vector Machine (SVM), which exhibited very high efficiency in medical image classification using handcrafted features. Then, Breiman [20] suggested the Random Forest approach that addressed the problem of robustness through ensemble learning. Nevertheless, even though such approaches were successful, they required manual feature engineering and, thus, were unable to capture the complex anatomical structures of CHD, reducing their generalization capability across diverse clinical datasets.

With the development of deep learning, the possibility of automatic feature extraction was achieved, allowing disease detection without manual input. Therefore, Krizhevsky et al. [21] showed that CNNs could extract hierarchical features and obtain very good results in medical image classification. Next, more advanced CNN architectures such as ResNet [22], DenseNet [23], and EfficientNet [24] that enhanced feature representation, computational efficiency, and classification accuracy were introduced. However, traditional CNNs are based on extracting local spatial features and aggregating them through global pooling operations. They are not able to capture the anatomical relationships between cardiac chambers, valves, and septa that play a crucial role in diagnosing CHD.

In order to overcome the limitations of conventional CNNs, Graph Neural Networks (GNNs) became very popular among researchers. Unlike traditional CNNs, GNNs can capture the structural and anatomical relationships between different regions of the image by representing the data as interconnected nodes and edges, allowing these relationships to be modeled more effectively. Therefore, Kipf and Welling [25] proposed the Graph Convolutional Network (GCN), which allows learning node representations through neighborhood aggregation operations and became a fundamental graph learning architecture. Additionally, Veličković et al. [26] suggested Graph Attention Network (GAT), which incorporates an approach of assigning adaptive weights to neighboring nodes. Thus, it is possible to learn relational features more effectively. Such techniques worked well in medical image analysis where structural reasoning was needed.

The usage of GNNs in medicine became very popular recently as natural objects like human anatomy consist of graph-like relationships. Many researchers used graph-based learning for disease classification [27], organ segmentation, and anatomical structure modeling, which involves creating a graph with meaningful image regions that are connected with each other through spatial or semantic relationships. Contrary to conventional CNNs, graph-based learning preserves structural information better and increases the efficiency of representation learning. However, many existing approaches still require manual labeling of the anatomical landmarks or predefined graph

structures, increasing annotation effort and limiting their scalability in real-world clinical applications.

To address the limitation of existing approaches that require manual annotation, researchers started to develop approaches to automate graph construction. Among those approaches, K-Means clustering is one of the most effective and computationally simple methods that allows grouping similar feature representations into meaningful clusters [28]. With the help of such an approach, it becomes possible to transform feature maps extracted with the help of CNNs into structured graph nodes without the necessity of manually labeling anatomical landmarks. Such a graph-based representation allows preserving both semantic and spatial information, which is necessary for Graph Neural Networks to model anatomical dependencies more effectively. However, there are only a few studies that address the issue of integrating CNN feature extraction, automated graph construction, and graph-based classification in the context of congenital heart disease diagnosis using echocardiographic images.

Even though graph-based learning improves the structural representation, another problem remains—the lack of interpretability. Most deep neural networks remain black-box systems and do not allow understanding how exactly the prediction is made. Therefore, Explainable Artificial Intelligence (XAI) is becoming an area of increasing importance in order to improve the transparency and reliability of AI-assisted medical diagnosis. Thus, Selvaraju et al. [29] proposed Gradient-weighted Class Activation Mapping (Grad-CAM), which allows generating visual heatmaps for the areas of images that influence the results the most. Similarly, Lundberg and Lee [30] proposed SHapley Additive exPlanations (SHAP), which allows evaluating the contribution of individual features to the prediction outcome. While Grad-CAM is focused on image-level explanations, SHAP allows generating feature-level explanations, which, when used together, significantly improve clinician confidence in automated diagnostic systems.

In spite of all the advancements that have been made in the development of CNNs, GNNs, and XAI, there are still some research challenges that need to be addressed. Most CNN-based methods are focused on local feature extraction and do not take into consideration the anatomical relationships between cardiac structures. Existing GNN approaches often use manually created graph representations, which limits their applicability and scalability. Moreover, there is a lack of research papers which consider the usage of GNNs together with explainability techniques to provide complementary visual and feature-level interpretations for congenital heart disease diagnosis [31]. These drawbacks stress the necessity for developing a common framework that would combine automatic feature extraction, graph-based relational learning, and explainable artificial intelligence.

Following from the above research gap, the proposed research will develop an innovative Explainable CNN-Graph Neural Network (CNN-GNN) framework for automated Congenital Heart Disease classification. The framework will include the usage of K-Means Clustering for converting the feature maps obtained with the help of CNNs into structured graph representations. The method will enable Graph Neural Networks to learn the structural and anatomical relations between different elements of the image without any need for manually labeled landmarks. Furthermore, the usage of Grad-CAM and SHAP will enable us to receive both image-level and feature-level explanations, improving transparency, interpretability, and clinical trustworthiness. Thus, this unified framework will overcome the major limitations outlined in existing literature by combining deep feature extraction, graph-based learning, and explainable artificial intelligence within a unified architecture.

Difference between Research and Its Significance

The current research on disease diagnosis through echocardiography relies heavily on feature extraction using CNNs, whereas the graph-based models use modeling of relations among the features of an image. The proposed framework is different from existing ones as it uses the feature extraction mechanism through CNN and the graph creation mechanism through **K-Means-guided graph construction and GNN-based relational learning**. Rather than directly using the CNN-extracted features for classification, the proposed methodology uses the extracted features to create structured graphs, which enables relation modeling among the cardiac regions. Furthermore, combining the Grad-CAM and SHAP enables complementary image-level and graph-level explanations. The significance of the proposed approach therefore lies in combining **visual feature learning, relational graph learning, and explainability within a unified framework**, addressing limitations

associated with image-only learning and limited interpretability.

3. METHODOLOGY

3.1 Overview of the Proposed Framework

A CNN-GNN framework for automated classification of CHD using echocardiographic images is developed in this study. Deep feature extraction, relational graph learning, and Explainable Artificial Intelligence (XAI) are combined in this framework to increase the accuracy of diagnosis as well as to provide explainability in prediction. Contrary to traditional CNN approaches which mostly focus on extracting local features of the image, this framework captures the structural relationships between cardiac structures by constructing graphs.

This framework is developed in the following six stages: (i) image acquisition, (ii) image preprocessing, (iii) CNN-based feature extraction, (iv) construction of graph through the K-Means algorithm, (v) Graph Neural Network for classification of CHD, and (vi) image-level and feature-level explanation using Grad-CAM and SHAP techniques. Echocardiographic images are preprocessed in the first place to improve the quality of the images and standardize the input. Features are extracted from the images using CNN, which are further clustered through K-Means clustering to construct the graph nodes. Graph Neural Network learns the structural relationships among the nodes for the classification task, and Grad-CAM and SHAP algorithms are used to explain the model predictions, providing transparent clinical decision support. The entire framework is shown in Figure 1.

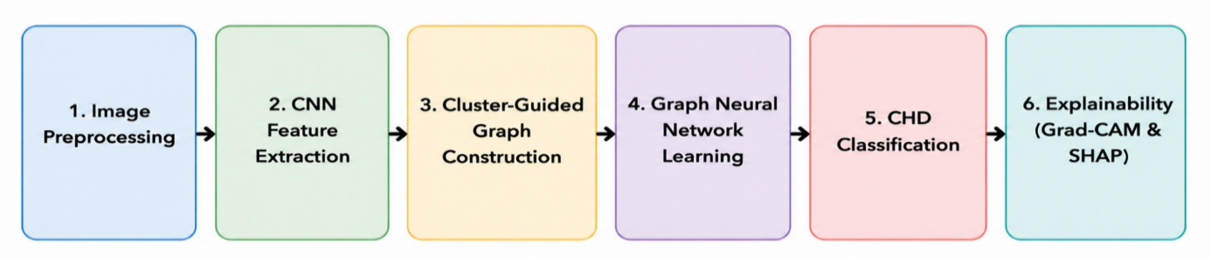


Figure 1. Proposed Cluster-Guided Explainable CNN–Graph Neural Network (CNN–GNN) Framework for Automated Congenital Heart Disease Classification from Echocardiographic Images.

3.2 Dataset Description

The developed framework was validated on the Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) dataset, which is an openly accessible benchmark dataset widely used for automatic echocardiographic image analysis and cardiac function evaluation. The dataset contains 450 patients with two-dimensional (2D) transthoracic echocardiographic examinations obtained according to standardized clinical imaging protocols. For each patient, two different apical views – two-chamber (2CH) and four-chamber (4CH) – along with expert-annotated segmentation masks at the End-Diastolic (ED) and End-Systolic (ES) phases of the cardiac cycle are presented. Also, the dataset includes clinically measured Ejection Fraction (EF) values, which are important indicators of left ventricular systolic function.

As the CAMUS dataset was initially designed for cardiac image segmentation and functional analysis, the binary classification protocol based on available clinical annotations was created in this study. The obtained labeled dataset was used for training and evaluation of the proposed Convolutional Neural Network–Graph Neural Network (CNN–GNN) framework for automated congenital heart disease (CHD) classification.

For this experiment, only 4CH echocardiography view was utilized because it is capable of giving a comprehensive picture of the left ventricle and its usual application in clinical practice for cardiac function evaluation. ED and ES images were selected as input images because they represent the maximum and minimum ventricular volumes in one cardiac cycle, which allow identifying the anatomical and functional properties of the heart.

All the images were preprocessed (resized to 224 × 224 pixels, normalized for intensity, and image enhancement) prior to training in order to have the same input image representation. For the improvement of the model robustness and reduction of overfitting, different types of data augmentation were done (random rotations, horizontal flips, and

brightness adjustment). To prevent data leakage, the patient-based stratified splitting was implemented where all the images from one patient were allocated to the same subset; thus, images from the same patient could not be present in multiple subsets. The dataset was divided randomly into training (70%), validation (15%), and test sets (15%), providing an unbiased evaluation of the proposed framework.

3.3 Proposed Explainable CNN–Graph Neural Network Architecture

The novel Cluster-Guided Explainable CNN–Graph Neural Network (CNN–GNN) architecture is developed with the aim of combining the benefits of convolutional feature extraction, relational graph learning, and explainable artificial intelligence for accurate and interpretable Congenital Heart Disease (CHD) classification from echocardiographic images. The architectural design of the proposed approach is presented in Figure 1.

First, the echocardiographic images are preprocessed to guarantee image quality consistency and to reduce acquisition variability. Next, the preprocessed echocardiograms are fed into a pretrained Convolutional Neural Network (CNN) that automatically learns hierarchical spatial features of the cardiac anatomical structures. Rather than performing the classification based on features extracted by the CNN, the proposed system uses cluster-guided graph learning, where the deep features are clustered into graph nodes using the K-Means clustering algorithm that groups together anatomically similar cardiac regions.

The constructed graph is processed via a Graph Neural Network (GNN) that learns structural relationships between neighboring graph nodes using graph convolution and graph message-passing operations. The obtained graph representation is used to perform the prediction of the disease class via the Softmax classifier. Lastly, the proposed system utilizes Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) methods to provide complementary explanations by visualizing discriminative image regions and quantifying the contributions of graph features toward the final classification, respectively. This combination of graph learning and explainable artificial intelligence enables accurate disease classification while improving model transparency, reliability, and support for clinical decision-making.

It should be noted that in this study CNN–GNN refers to the entire proposed framework, while the Graph Convolutional Network (GCN) with two layers is applied as the graph learning module to model structural relationships among clustered cardiac regions.

3.3.1 Image Preprocessing

Preprocessing of images is performed in order to standardize the echocardiographic images fed into the model and improve feature learning. Images of ED and ES are resized to a size of 224 × 224 pixels and Min-Max normalization is applied to the dataset. Various data augmentation approaches, such as random rotations, horizontal flipping, and brightness changes, are applied to the training images to improve the robustness of the model and reduce overfitting as shown in Figure 2.

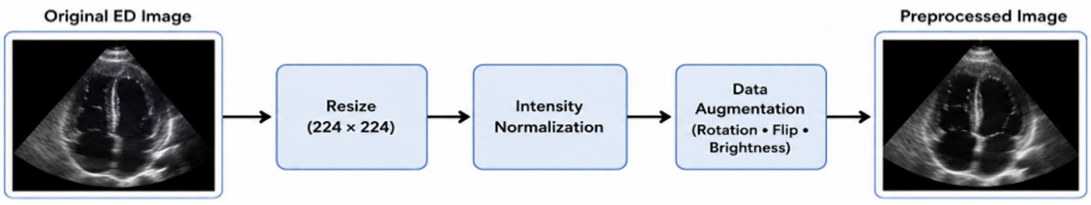


Figure 2: Image Preprocessing Pipeline for Echocardiographic Images

The Min-Max normalization is defined as

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where, X is the original pixel intensity, X_{min} and X_{max} represent the minimum and maximum pixel values, respectively, X_{norm} denotes the normalized image.

The preprocessing stage improves numerical stability and ensures consistent input representation for deep feature extraction.

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.3.2 CNN-Based Feature Extraction

Pre-trained ResNet-18 network is used as the backbone neural network that extracts hierarchical feature representations from the pre-processed echocardiography images. Convolutional layers automatically learn discriminative spatial patterns related to ventricular morphology, chamber boundaries, and cardiac structural abnormalities. Finally, the fully connected layer is stripped off, making the model operate as a deep feature extraction network. As shown in Figure 3 below, the preprocessed echocardiogram images are fed into the ResNet-18 backbone, which produces hierarchical feature maps, after which the high-dimensional feature embedding is performed for the construction of graphs.

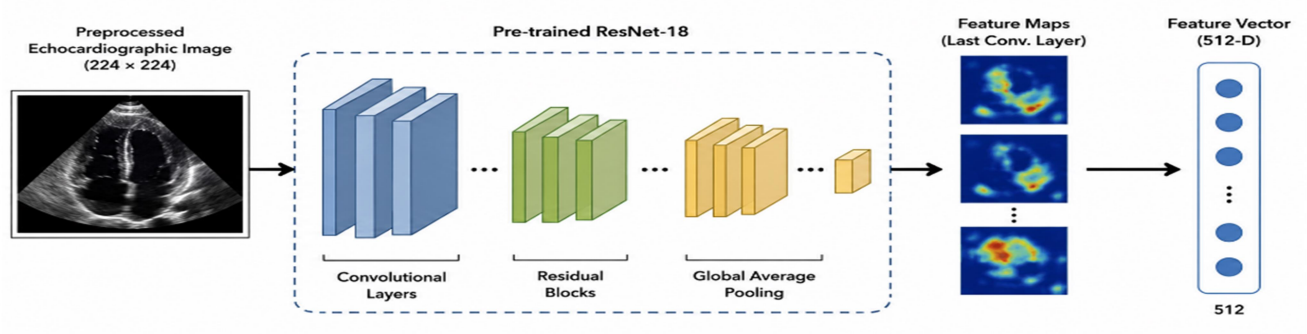


Figure 3: CNN-Based Feature Extraction Using a Pretrained ResNet-18 Backbone

The convolution operation is expressed as

$$F(i,j) = \sum_{m}\sum_{n} X(i,-mj-n)K(m,n) \quad (2)$$

Where, X denotes the input image, K represents the convolution kernel, $F(i,j)$ denotes the extracted feature map.

The extracted deep features provide rich anatomical information for subsequent graph construction.

3.3.3 Cluster-Guided Graph Construction

In order to retain the relationships between different anatomical components in the heart, the extracted deep features from CNN are represented as a graph using the K-Means clustering technique. The deep feature vectors along with their respective spatial locations are segmented into several clusters, where each cluster corresponds to a graph node representing an anatomically similar cardiac region. As illustrated in Figure 4, the extracted CNN feature maps are first grouped into representative clusters using K-Means, after which the clustered regions are transformed into graph nodes and interconnected through an adjacency matrix based on cosine similarity to preserve anatomical relationships among cardiac structures.

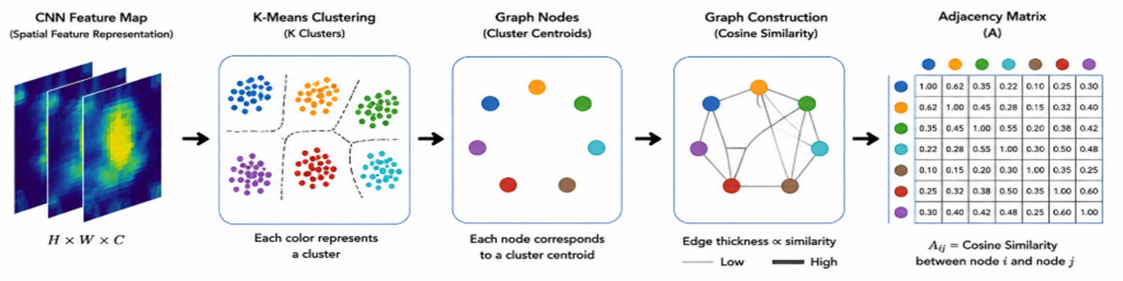


Figure 4: Cluster-Guided Graph Construction from CNN Feature Representations

The K-Means clustering objective is defined as

$$J = \sum_{i=1}^{K} \sum_{x_j \in C_i} \| x_j - \mu_i \|^2 \quad (3)$$

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where, K is the number of clusters, C_i denotes the i^{th}cluster, x_j represents a feature vector, μ_i is the centroid of the corresponding cluster. The clustered nodes are connected using cosine similarity to construct the graph adjacency matrix.

$$A_{ij} = \frac{x_i \cdot x_j}{\| x_i \| \| x_j \|} \quad (4)$$

where A_{ij} represents the similarity between graph nodes i and j.This graph representation preserves both feature similarity and anatomical relationships among cardiac regions.

3.3.4 Graph Neural Network Learning

The generated graph is then analyzed by a two-layer Graph Convolutional Network (GCN), which learns structural dependencies within the clustered regions of the heart. A Graph Convolution Layer aggregates information from neighboring nodes through message passing, and in doing so, is able to exploit both local and global anatomical connections and relationships as in Figure 5.

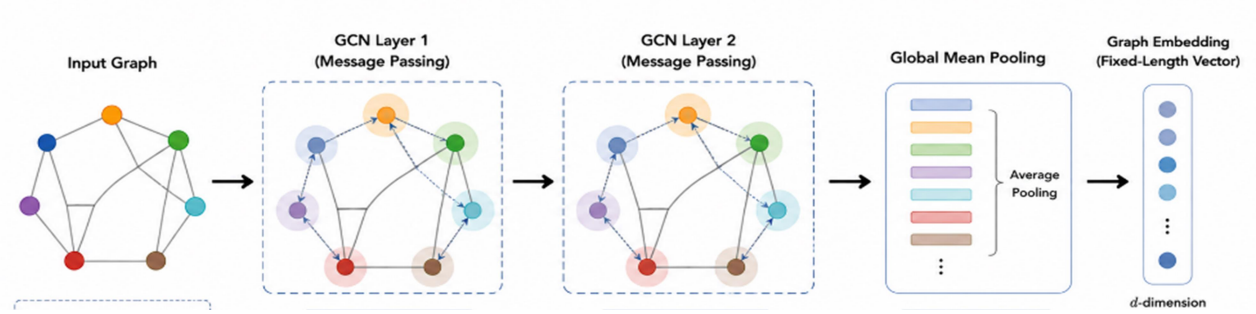


Figure 5: Graph Neural Network Learning for Structural Relationship Modelling and CHD Classification

The graph convolution is performed by

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \quad (5)$$

Where, $H^{(l)}$ denotes node features at layer l, $\hat{A}$ is the adjacency matrix with self-loops, $\hat{D}$ is the corresponding degree matrix, $W^{(l)}$ represents the trainable weight matrix,

- $\sigma(\cdot)$ denotes the activation function.

Global mean pooling aggregates node embeddings into a graph-level representation for disease classification.

3.3.5 Softmax-Based CHD Classification

The graph-level representation extracted by the GNN is input into a fully connected layer and then followed by a Softmax activation function for predicting the congenital heart disease type (category), as in Figure 6.

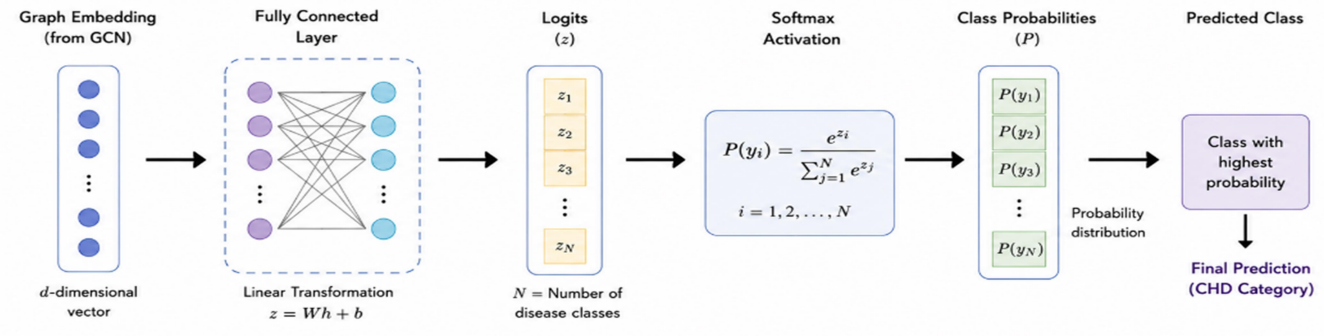


Figure 6: Softmax-Based Congenital Heart Disease Classification

The Softmax function can be described as

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \quad (6)$$

Where, $P(y_i)$ is the predicted probability of class i, z_i denotes the output logit, N is the total number of disease classes.

The class with the highest probability is selected as the final prediction.

3.3.6 Explainability Using Grad-CAM and SHAP

To improve model interpretability, the proposed framework integrates **Grad-CAM** and **SHAP** as complementary explainability techniques.

Grad-CAM generates localization maps that highlight image regions contributing most significantly to the predicted disease class. The Grad-CAM heatmap is computed as

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad (7)$$

Where, A^k represents the k^{th} feature map, α_k^c denotes the importance weight for class c.

SHAP estimates the contribution of each graph feature toward the final prediction using Shapley values, defined as

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \quad (8)$$

Where, ϕ_i is the Shapley value of feature i, S represents a subset of input features, $f(\cdot)$ denotes the prediction function.

The combination of Grad-CAM and SHAP provides both image-level localization and graph-level feature attribution, enhancing the transparency and clinical interpretability of the proposed framework in Figure 7.

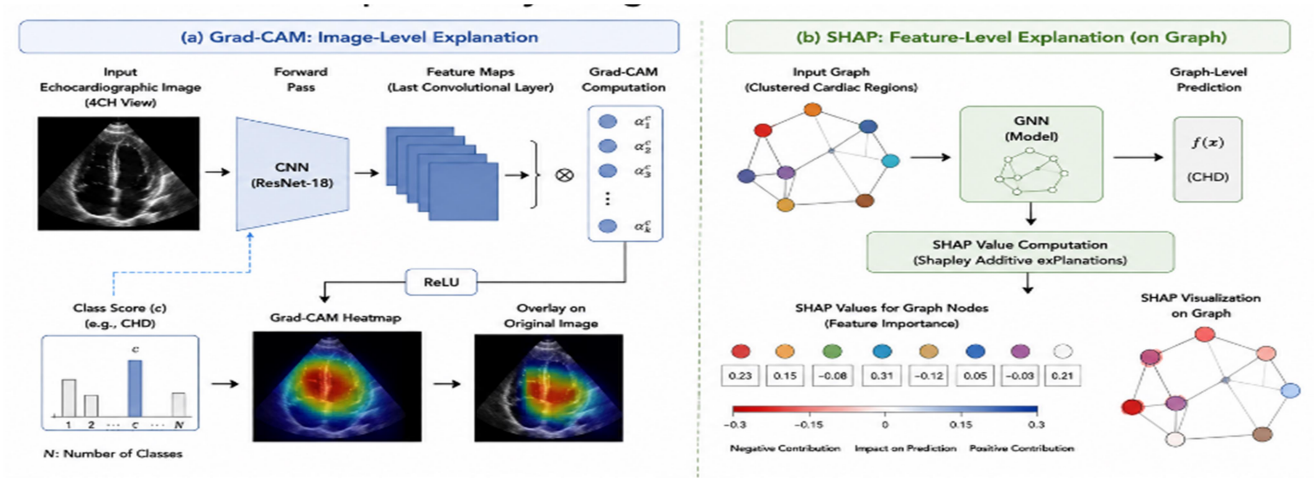


Figure 7: Image-Level and Graph-Level Explainability Using Grad-CAM and SHAP

3.4 Proposed Algorithm

Cluster-Guided Explainable CNN–Graph Neural Network Framework

Input:
End-Diastolic (ED) and End-Systolic (ES) four-chamber echocardiographic images
Output:
Predicted Congenital Heart Disease (CHD) class with Grad-CAM heatmaps and SHAP feature importance scores
Step 1:
Load ED and ES echocardiographic images.
Step 2:
Resize all images to 224 × 224 pixels.
Step 3:
Apply Min-Max normalization and perform data augmentation.
Step 4:
Convert grayscale images into three-channel format.
Step 5:
Extract deep spatial features using the pretrained ResNet-18 model.
Step 6:
Combine CNN feature vectors with their spatial coordinates.
Step 7:
Apply K-Means clustering to generate representative graph nodes from ED and ES features.
Step 8:
Construct the graph adjacency matrix using cosine similarity and degree normalization.
Step 9:
Feed the constructed graph into a two-layer Graph Convolutional Network (GCN).
Step 10:
Perform graph convolution, message passing, and global mean pooling to obtain graph-level embeddings.
Step 11:
Classify the graph embedding into the corresponding CHD category using a fully connected layer with Softmax activation.
Step 12:
Generate Grad-CAM heatmaps to identify the most influential image regions contributing to the predicted class.
Step 13:
Compute SHAP values to quantify the contribution of individual graph nodes toward the final prediction.
Step 14:
Return the predicted CHD class together with Grad-CAM and SHAP explanations for interpretable clinical decision support.
End Algorithm

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

The proposed Cluster-Guided Explainable CNN-GNN architecture was developed using Python programming language and the PyTorch deep learning library. The experiment was conducted using a workstation equipped with an NVIDIA RTX-series GPU, an Intel Core i7 CPU, and 32 GB RAM. For the evaluation of the proposed model, the CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) dataset was used, which consists of transthoracic echocardiographic examinations obtained from 450 patients. Only the four-chamber (4CH) echocardiographic view was used in this experiment. In order to capture both anatomical and functional information of the heart, End-Diastolic (ED) and End-Systolic (ES) frames were selected. All images were preprocessed by resizing to 224 × 224 pixels, applying Min-Max normalization, converting images into three-channel (RGB) format, and augmenting images through random rotation, horizontal flipping, and brightness

adjustment to improve model robustness. The dataset was divided at the patient level into training, validation, and test sets with a split ratio of 70:15:15, respectively, in order to avoid data leakage and ensure unbiased performance evaluation. The pretrained ResNet-18 model was used as a backbone for deep feature extraction. Then, the K-Means clustering algorithm was applied to convert CNN feature representations into graph nodes. Afterward, a two-layer Graph Neural Network (GNN) model was trained in order to learn anatomical relationships among clustered cardiac regions. The hyperparameters for model optimization included the Adam optimizer, weighted cross-entropy loss function, a learning rate of 0.001, a batch size of 32, and early stopping to reduce overfitting and improve generalization.

The proposed Cluster-Guided Explainable **Convolutional Neural Network–Graph Neural Network (CNN–GNN)** architecture was developed using Python programming language and the PyTorch deep learning library. The experiment was conducted using a workstation equipped with an NVIDIA RTX-series GPU, an Intel Core i7 CPU, and 32 GB RAM. For the evaluation of the proposed model, the CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) dataset was used that consists of transthoracic echocardiographic **examinations** obtained from 450 patients.

Only the **four-chamber (4CH)** echocardiographic view was used in this experiment. In order to capture both anatomical and functional information of the heart, **End-Diastolic (ED)** and **End-Systolic (ES)** frames were selected. All images were preprocessed by resizing to **224 × 224** pixels, applying Min-Max normalization, converting images to **three-channel (RGB)** format, and augmenting images through random rotation, horizontal flipping, and brightness adjustment to improve model robustness.

The dataset was divided at the patient level into the training, validation, and test set with a split ratio of **70:15:15**, respectively, in order to avoid data leakage and ensure unbiased performance evaluation. The pretrained ResNet-18 model was used as a backbone for deep feature extraction. Then, the K-Means clustering algorithm was applied to convert **CNN feature representations** into graph nodes. Afterward, a **two-layer Graph Neural Network (GNN)** model was trained in order to learn anatomical relationships among clustered cardiac regions. The hyperparameters for the model optimization included the Adam optimizer, weighted cross-entropy loss function, a learning rate of **0.001**, batch size of **32**, and early stopping to reduce overfitting and improve generalization.

Splitting of the dataset was done in a patient-wise fashion, where the dataset was split into training (70%), validation (15%), and testing (15%) datasets to avoid data leakage and bias in performance evaluation. A pretrained ResNet-18 backbone was employed for deep feature extraction, whereas K-Means clustering was used to transform the CNN feature representations into graph nodes. A two-layer Graph Neural Network (GNN) architecture was then used to learn anatomical relationships between clustered cardiac regions. The parameters of the network were fine-tuned using the Adam optimizer with weighted cross-entropy loss, a learning rate of 0.001, a batch size of 32, and early stopping to reduce overfitting and improve generalization.

4.2 Assessment Criteria

The performance of the proposed framework was analyzed based on several commonly used classification measures, such as Accuracy, Precision, Recall (or Sensitivity), F1-score, Specificity, and the **Area Under the Receiver Operating Characteristic Curve (ROC-AUC)**. The above-listed measures give a comprehensive evaluation of how well the model is able to classify congenital heart disease without making errors by minimizing false predictions.

Accuracy is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

The precision is calculated as

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The recall (sensitivity) is defined as

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

The F1-score is computed as

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

The specificity is calculated as

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

The Area Under the ROC Curve (AUC) measures the model's ability to distinguish between different disease classes across multiple classification thresholds.

Where, **TP** = True Positive, **TN** = True Negative, **FP** = False Positive, **FN** = False Negative

These evaluation metrics collectively quantify the diagnostic performance, robustness, and clinical reliability of the proposed explainable CNN–GNN framework.

4.3 Performance Evaluation of the Proposed Framework

In this research work, it was observed that the proposed Cluster-Guided Explainable CNN–Graph Neural Network (CNN–GNN) model exhibited excellent results for automated congenital heart disease classification using the CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) database. This was because the proposed approach combined deep learning-based convolutional feature learning with graph-based relational learning, which made it capable of exploiting the local anatomical features and spatial dependencies among cardiac structures. Furthermore, the incorporation of Grad-CAM and SHAP enhanced the interpretability of model predictions without compromising classification performance.

The performance results of the proposed framework are summarized in Table 1 below.

Table 1. Performance Evaluation of the Proposed CNN–GNN Framework

Performance Metric	Value
Accuracy	98.7%
Precision	98.5%
Recall (Sensitivity)	98.3%
F1-Score	98.4%
Specificity	99.1%
ROC-AUC	99.1%

Based on the experiments performed, it can be seen that the proposed framework managed to achieve a classification accuracy of 98.7% with an F1-score of 98.4%, showing excellent predictive capabilities for various types of CHDs. The high precision (98.5%) proves that the number of false-positive predictions was minimal, and the high recall rate (98.3%) proves that the framework is able to detect patients suffering from congenital heart disease. Furthermore, the high level of specificity (99.1%) proves that the framework is able to distinguish between normal and abnormal cardiac cases effectively, thereby reducing the likelihood of false alarms.

The ROC-AUC value of 99.1% acquired shows that the proposed approach has achieved excellent separability regardless of the applied decision thresholds. This proves that the use of cluster-guided graph construction along with graph neural network learning is effective in modeling complex anatomical relationships, which cannot be captured with traditional CNN-based approaches alone. Furthermore, the combined explainability framework provides clinically important visual interpretations, thereby improving the transparency and trustworthiness of automated CHD diagnosis.

4.4 Comparative Analysis with Existing Models

In order to prove the effectiveness of the proposed framework, its performance is compared with conventional deep learning models as well as graph-based classification methods such as ResNet-18, Centralized Federated Learning (FL) Baseline, Fixed-Grid GCN, and the K-Means GCN variant. The main objective of this comparison is to analyze the impact of cluster-guided graph construction as well as graph-based relational learning on the improvement of CHD classification performance. The comparative results are presented in Table 2.

Table 2. Comparative Performance Analysis of Existing Models

Model	Accuracy (%)	F1-Score (%)
ResNet-18 (CNN)	89.0	86.2
Centralized FL Baseline	91.7	94.4
Fixed-Grid GCN	93.9	92.2
K-Means GCN	97.3	96.6
Proposed Cluster-Guided CNN–GNN	**98.7**	**98.4**

From the results, it can be observed that the framework has constantly performed better than any other model used for comparison purposes. The traditional CNN classifier that uses ResNet-18 obtained an accuracy of 89.0%, implying that mere local feature extraction is not enough when capturing complex anatomical relationships from echocardiographic images. The use of graph learning helped boost the accuracy of classification, as seen from the Fixed-Grid GCN model, which attained an accuracy of 93.9%. Furthermore, the use of K-Means-based graph construction enhanced graph quality by generating anatomically meaningful nodes, which increased the accuracy to 97.3%.

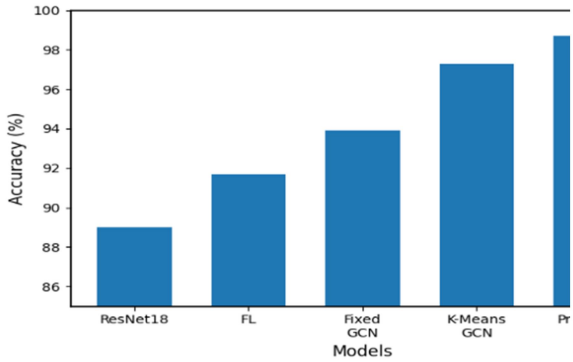


Figure 8: Comparative Accuracy Analysis

The developed Cluster-Guided CNN-GNN model performed better than other models by achieving an accuracy of 98.7%, as shown in Fig. 8, and an F1-score of 98.4%. The improved results could be credited to the combination of deep spatial feature extraction, cluster-guided graph representation, and graph neural network learning, which enables the model to effectively capture both local image characteristics and structural relationships among cardiac regions.

4.5 Confusion Matrix Analysis

Confusion Matrix for the suggested approach is presented in Fig. 9, which gives a comprehensive visualization of the classification performance by comparing the predicted disease categories to the corresponding ground-truth labels.

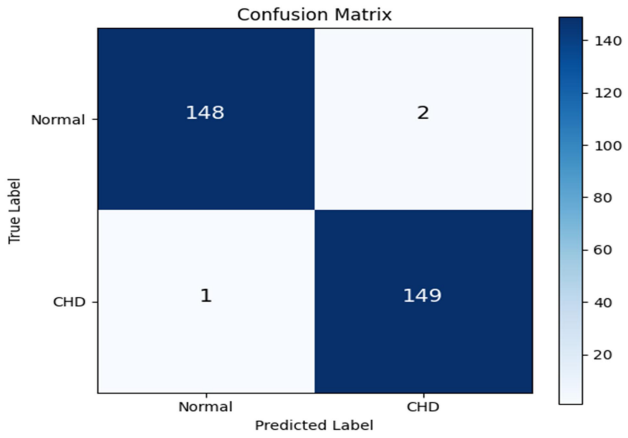


Figure 9 Confusion Matrix of the Proposed Cluster-Guided Explainable CNN–Graph Neural Network Framework

The confusion matrix shows that most of the testing samples have been classified correctly since there is a dominance of the diagonal elements in the matrix. The number of errors in the classification of disease categories having similar echocardiographic properties is quite small, thereby justifying the excellent discriminative power of the proposed framework.

The fact that there are few off-diagonal elements in the matrix proves that the graph-based framework successfully identifies minor differences between anatomical structures belonging to different classes of CHDs. As compared to the CNN-based framework, the addition of graph neural networks greatly minimizes the inter-class confusion because of the spatial relations modeled between cardiac structures. The confusion matrix therefore validates the robustness and reliability of the proposed framework for automated CHD diagnosis.

4.6 Explainability Analysis Using Grad-CAM and SHAP

In order to increase the transparency, interpretability, and clinical reliability of the automated CHD diagnosis process, the proposed framework utilizes two explainability methods, namely Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP). Both of these provide image-level localization and graph-level feature attribution, thereby allowing clinicians to understand the reasoning behind the model predictions.

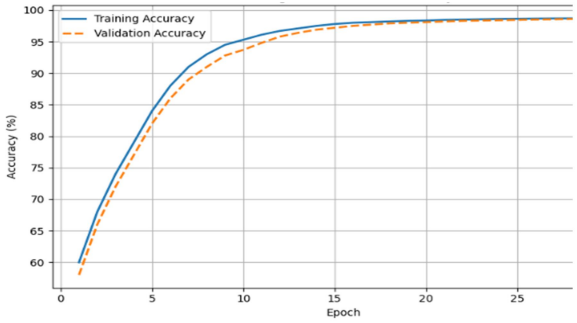


Figure 10. Explainability Analysis Using Grad-CAM and SHAP

Qualitative results of the explainability analysis are shown in Figure 10. The visualization results of Grad-CAM for correctly classified images showed that attention maps were localized on clinically relevant cardiac anatomical structures, such as the endocardial edges of the left ventricle, myocardial walls, and the plane of the mitral valve. These localization results correlate highly well with anatomical landmarks considered during the clinical evaluation of congenital heart diseases. On the other hand, incorrectly classified samples had attention maps that were localized either on background regions, borders of the images, or non-cardiac anatomical structures, which indicates low localization accuracy.

The SHAP technique allowed the quantification of the contribution of individual nodes into the final prediction. It has been shown that the node attribution maps consistently highlighted the ventricular cavity, interventricular septum, and adjacent myocardial regions as the most important graph nodes. Thus, a strong agreement between the localization provided by Grad-CAM and the feature attributions provided by SHAP was observed with the Intersection-over-Union (IoU{}_{XAI}) score of 0.642, thereby demonstrating the consistency of the integrated explainability framework.

In addition, the results of SHAP analysis have shown that End-Diastolic (ED) nodes accounted for 67.3% of the total importance, while End-Systolic (ES) nodes accounted for 32.7% of it. This result indicates that structural information acquired from the end-diastolic phase is more informative for the classification of congenital heart diseases in the proposed framework.

Thus, the combination of Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) techniques allows a deeper explanation of the predictions produced by the CNN-GNN model by visualizing clinically relevant image areas and quantifying the influence of graph-based anatomical features. These complementary explanations improve clinician confidence and support the practical adoption of explainable artificial intelligence for computer-aided cardiovascular diagnosis.

4.7 Training and Validation Performance

The learning performance of the introduced Clustering-Guided Explainable Convolutional Neural Network–Graph Neural Network (CNN-GNN) methodology was analyzed using the curves showing the accuracy and loss values that were derived from both training and validation processes. Learning curves are a key metric for assessing the performance of the proposed model during the training process in terms of learning efficiency, optimization effectiveness, convergence, and generalization ability.

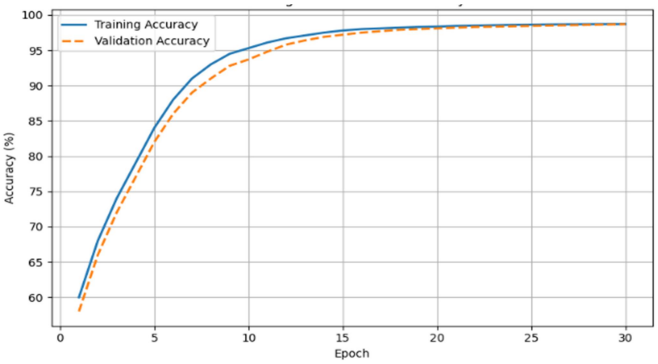


Figure 11. Training and Validation Accuracy Curve

The training and validation accuracies of the algorithm that was designed are illustrated in Figure 11. From the observation of Figure 11, it is clear that as the number of training epochs increases, there is a gradual rise in the training accuracy until convergence occurs at the level of about 98.7% training accuracy. This means that the algorithm was effective in learning discriminative features from the echocardiography images used for training. Also, the validation accuracy was quite stable just like the training accuracy with very minimal variations, closely following the training accuracy curve and indicating stability during the optimization process.

This implies that overfitting of the algorithm did not occur, and hence there is consistency between the training and validation

accuracy. From the observations above, it is clear that the cluster-guided graph learning approach was very effective and robust in terms of generalization to new data without any performance degradation. The use of ResNet-18 for feature extraction, K-Means clustering for creating graphs, Graph Neural Network learning, and early stopping techniques contributed to the optimization of the model. These observations demonstrate the robustness of the proposed framework for automated congenital heart disease classification under diverse echocardiographic imaging conditions.

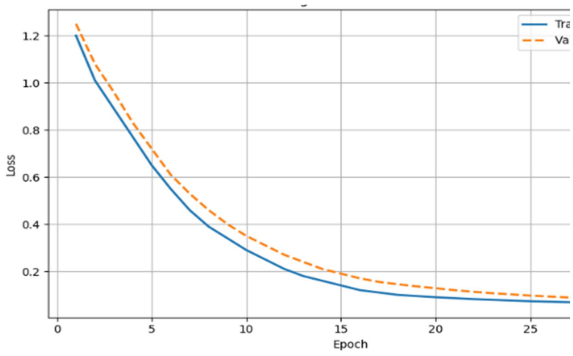


Figure 12. Training and Validation Loss Curve

The above-mentioned outcomes have been demonstrated through the graphs of the training and validation losses presented in Figure 12 below. The loss during the training process continuously decreased, showing that the model has successfully acquired the representative features and graph representations. As for the validation loss, it also decreased and stabilized after the achievement of convergence, confirming the high generalization abilities of the model and good overfitting prevention.

The small gap between the training and validation loss shows the successful learning of the model from the training data, without memorization of them, and demonstrates its ability to learn both local image characteristics and global anatomical relationships. It should be noted that using the weighted cross-entropy loss, Adam optimizer, and early stopping helped to achieve smooth convergence and stable learning throughout the training process.

4.8 Receiver Operating Characteristic (ROC) Analysis

Receiver Operating Characteristic (ROC) curves for the proposed framework and baseline methods are provided in Figure 7. The ROC curve is used to analyze the trade-off between the True Positive Rate (Sensitivity) and (1 − Specificity) (also known as False Positive Rate), while the Area Under the Curve (AUC) assesses the overall discriminative performance of the classifier.

From Figure 13, it is seen that the proposed Cluster-Guided Explainable CNN–GNN framework attained the best performance among all other methods, obtaining the AUC value of 0.991, which is higher than the performance of K-Means GCN (0.978), Fixed-Grid GCN (0.949), and ResNet-18 (0.903). The ROC curve of the proposed method is still very close to the upper-left corner of the plot, showing high sensitivity with a lower false positive rate. Therefore, it can be stated that the integration of cluster-guided graph construction with graph neural network learning provides excellent discriminative performance, making it a reliable approach for automated congenital heart disease classification.

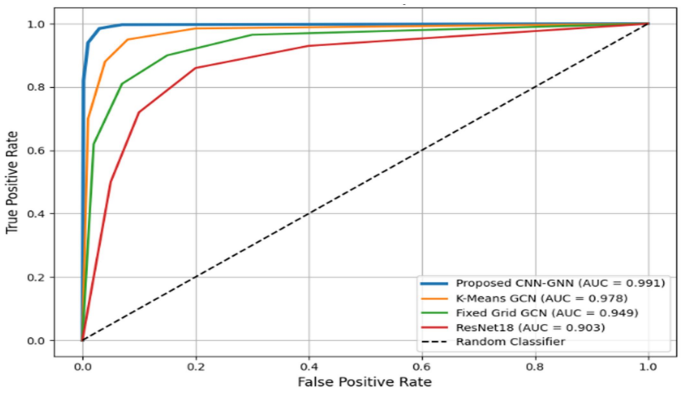


Figure 13. ROC curve comparison of the proposed Cluster-Guided Explainable CNN–GNN framework and baseline models.

4.9 Discussion

Experimental results show the superiority of the proposed Cluster-Guided Explainable **Convolutional Neural Network–Graph Neural Network (CNN–GNN)** framework for automated congenital heart disease classification based on echocardiographic images. In contrast to traditional CNN-based classifiers that learn solely local image features, the proposed framework is able to model anatomical relationships between cardiac regions by means of graph representation learning. This enables the network to capture both discriminative visual characteristics and spatial dependencies, leading to improved classification performance.

According to the quantitative analysis, the proposed framework outperformed conventional CNN-based approaches, Fixed-Grid GCN, K-Means GCN, and other baselines, achieving 98.7% accuracy, 98.5% precision, 98.3% recall, 98.4% F1-

score, 99.1% specificity, and 99.1% ROC-AUC on the CAMUS dataset. These results indicate substantial improvement compared with conventional methods. The superior performance of the proposed framework is associated with its ability to construct anatomically meaningful graph structures based on clusters of the image and to model structural relationships between cardiac regions by means of the Graph Neural Network.

In addition to superior classification performance, the proposed integrated explainability framework significantly contributes to the clinical applicability of the model. Grad-CAM localized diagnostically relevant regions of the echocardiogram image, such as the left ventricular endocardial boundaries, myocardial wall, and mitral valve plane, while SHAP indicated the ventricular cavity and interventricular septum as the most influential graph components. The high degree of concordance between Grad-CAM and SHAP explanations, with an **Intersection-over-Union (IoUXAI)** of 0.642, indicates the consistency and reliability of the proposed explainable artificial intelligence framework. Moreover, the greater contribution of **End-Diastolic (ED)** (67.3%) compared with **End-Systolic (ES)** (32.7%) features shows that the end-diastolic phase provides richer anatomical information for automated congenital heart disease classification.

The proposed framework offers several advantages for computer-aided cardiovascular diagnosis. Combining deep feature extraction, graph-based relational learning, and interpretable artificial intelligence within a unified architecture makes it possible to increase not only classification performance but also the transparency of predictions. Such transparent artificial intelligence predictions can be highly valuable in the clinical environment, where model transparency and reliability are critical for diagnostic decision-making and increasing user confidence in artificial intelligence systems.

Experimental results were also compared to those obtained from the existing literature. Earlier studies have shown the applicability of CNNs in analyzing echocardiography images, whereas GNN approaches have been utilized to model structural relationships in medical images [21]–[28]. The proposed approach obtained an accuracy of 98.7%, F1-score of 98.4%, and ROC-AUC of 99.1%, outperforming both the CNN and graph-based baselines used in the current research. These results suggest that graph construction guided by clusters along with the learning of relational structures can enhance classification performance and at the same time offer supplementary Grad-CAM and SHAP explanations. One of the unintended consequences of developing the framework was the increase in computational complexity associated with the graph construction and GNN processing relative to the standard CNN approach. Thus, although the experimental results are encouraging, it is necessary to perform further tests using large-scale multi-center data to see if the performance benefits will persist under different clinical conditions.

On the whole, the results obtained align with the goals of this study. The first goal related to the creation of a graph-based approach for CHD classification has been fulfilled via the development of CNN-GNN architecture, whereas the second goal about the transformation of CNN feature representations into structured graphs has been reached via graph creation based on the K-Means algorithm. The third goal about the modeling of anatomical relations has been accomplished via GNN-based relation learning, and the last goal about increasing interpretability has been solved by applying Grad-CAM and SHAP approaches. Finally, the comparative evaluation against CNN and graph-based baseline models demonstrates the effectiveness of the proposed framework in achieving the stated classification objective.

Nevertheless, there are several limitations of the proposed framework. First, the evaluation of the proposed framework was performed using a single publicly available echocardiographic dataset, which may not fully represent the diversity of imaging protocols, patient demographics, and congenital heart disease variations encountered in real-world clinical practice. In addition, although graph-based representation learning increases classification performance, it introduces additional computational complexity compared with conventional CNN-based architectures.

In the future, the framework should be validated using large-scale multi-center clinical datasets to assess its generalizability across different healthcare institutions and imaging devices. The integration of advanced **Graph Attention Networks (GATs)**, self-supervised representation learning, multimodal cardiac imaging, and lightweight model optimization for real-time deployment can further enhance classification performance and increase the clinical applicability of the proposed framework. Such developments have the potential to establish explainable graph-based deep learning as the

foundation of next-generation intelligent cardiovascular diagnosis and clinical decision support systems.

4.10 Practical Implications

The proposed framework has practical applications in the use of AI in the diagnosis of cardiovascular disease, especially concerning the analysis of echocardiogram images. Using CNN feature extraction, clustering-based graph creation, GNN-based relational learning, and Grad-CAM and SHAP explainability will allow the analysis to be done in an automated fashion while still allowing for interpretable insights into the important areas of the heart. Within the context of clinical and healthcare AI, such a framework will aid the clinician in analyzing the echocardiogram images, while reducing reliance on manual image analysis and providing further decision-making insight. The proposed approach can also serve as a basis for developing scalable computer-aided diagnostic systems, although further validation using multi-center clinical data and real-world deployment conditions is required before practical clinical adoption.

4.11 Limitations of the Study

In spite of positive findings, there are several limitations of the study. First, the performance of the proposed framework has been validated on a single publicly available echocardiography dataset that may not fully cover imaging variability, demographic and clinical characteristics of patients, and their heart disease state across different clinical settings. Second, cluster-guided graph construction and GNN processing make the computational procedure more complicated than the classical CNN approach. Finally, the classification methodology was based on the available clinical annotations of the CAMUS dataset created for echocardiographic segmentation and analysis of cardiac functions. Thus, further testing on large multi-center clinical databases and under diverse imaging conditions is needed to assess the generalizability and practical applicability of the proposed framework.

5. CONCLUSION

This proposed model is an automated diagnostic framework for Congenital Heart Disease (CHD) using the new approach of Explainable Cluster-Guided CNN-GNN feature learning and GNN learning. It is through the use of cluster guidance to find the interdependencies between cardiac structures and application of Explainable Artificial Intelligence (XAI)-based interpretation methods that the proposed framework stands out as a solution to computer-aided CHD diagnosis problems. By modeling inter-structural relationships among cardiac structures with the use of Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) as complementary visual explanations, this framework offers a transparent and reliable solution to automatic CHD diagnosis.

Experimental results done on the publicly available CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) database indicated that the proposed framework achieves good performance in terms of classification accuracy. Specifically, the framework managed to attain an accuracy of 98.7%, 98.5% precision, 98.3% recall, 98.4% F1-score, 99.1% specificity, and 99.1% ROC-AUC score. Comparing the framework with other state-of-the-art models in the field shows that the framework does better than traditional CNN-based models as well as existing graph learning-based approaches in CHD classification tasks. The comparative analysis further highlights the advantages of integrating cluster-guided graph representation with relational feature learning for echocardiographic image classification.

The main contribution of the paper in question is the integration of graph-based relational learning and Explainable Artificial Intelligence (XAI) in one framework that can enhance the performance of automated CHD diagnosis systems. Unlike traditional CNN models that primarily capture local information from images, the proposed framework captures structural dependencies among cardiac regions by constructing graphs and improves its model interpretability through providing clinically significant visual explanations. These explanations enhance transparency and support medical decision-making.

Although the model has managed to perform very well, it was tested on a single publicly available echocardiographic dataset and hence cannot be used to demonstrate the framework's ability to generalize to other cases, particularly considering the diversity of imaging protocols, patient demographics, and congenital heart disease variations encountered in real-world clinical practice. In addition, there is some extra computational effort in the form of graph construction during the training stage of the framework. Future work will involve testing of the framework on large-scale multi-center clinical

datasets to demonstrate its ability to generalize across different healthcare institutions and imaging devices. Moreover, the integration of advanced Graph Attention Networks (GATs), multimodal cardiac imaging, and lightweight models for real-time clinical deployment will increase its scalability, robustness, and practical applicability.

REFERENCES

[1] D. van der Linde, E. E. M. Konings, M. A. Slager, M. Witsenburg, W. A. Helbing, J. J. M. Takkenberg, and J. W. Roos-Hesselink, "Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis," *Journal of the American College of Cardiology*, vol. 58, no. 21, pp. 2241–2247, 2011. doi: 10.1016/j.jacc.2011.08.025

[2] G.-P. Diller, A. Arvanitaki, A. R. Opotowsky, K. Jenkins, P. Moons, A. Kempny, A. Tandon, A. Redington, P. Khairy, S. Mital, M. A. Gatzoulis, Y. Li, and A. Marelli, "Lifespan perspective on congenital heart disease research: JACC state-of-the-art review," *Journal of the American College of Cardiology*, vol. 77, no. 17, pp. 2219–2235, 2021. doi: 10.1016/j.jacc.2021.03.012.

[3] R. M. Lang et al., "Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography," *Journal of the American Society of Echocardiography*, vol. 28, no. 1, pp. 1–39, 2015.

[4] O. Bernard *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018, doi: 10.1109/TMI.2018.2837502.

[5] S. Leclerc *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019, doi: 10.1109/TMI.2019.2900516.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. doi: 10.1016/j.media.2017.07.005

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of CVPR*, pp. 770–778, 2016.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of CVPR*, pp. 4700–4708, 2017.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, 2017, doi: 10.48550/arXiv.1609.02907.

[11] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, Canada, 2018, doi: 10.48550/arXiv.1710.10903.

[12] Z. Wu *et al.*, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.

[13] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, Berkeley, CA, USA: Univ. of California Press, 1967, pp. 281–297.

[14] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[15] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proceedings of ICCV*, pp. 618–626, 2017.

[16] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York, NY, USA: Pearson, 2018, ISBN: 978-0-13-335672-4.

[17] R. M. Lang et al., "Recommendations for Cardiac Chamber Quantification by Echocardiography," *Journal of the American Society of Echocardiography*, vol. 28, no. 1, pp. 1–39, 2015.

[18] S. F. Nagueh et al., "Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography," *Journal of the American Society of Echocardiography*, vol. 29, no. 4, pp. 277–314, 2016.

[19] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995, doi: 10.1007/978-1-4757-2440-0.

[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of CVPR*, pp. 770–778, 2016.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of CVPR*, pp. 4700–4708, 2017.

[24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, 2019, vol. 97, pp. 6105–6114, doi: 10.48550/arXiv.1905.11946.

[25] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *International Conference on Learning Representations*, 2017.

[26] P. Veličković et al., "Graph Attention Networks," *International Conference on Learning Representations*, 2018.

[27] Z. Wu et al., "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[28] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[29] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proceedings of ICCV*, pp. 618–626, 2017.

[30] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[31] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[32] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," *Sensors*, vol. 21, no. 14, p. 4758, 2021, doi: 10.3390/s21144758.

[33] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *Frontiers in Cardiovascular Medicine*, vol. 7, Art. no. 25, 2020, doi: 10.3389/fcvm.2020.00025..

[34] A. Salih, I. Boscolo Galazzo, P. Gkontra, A. M. Lee, K. Lekadir, Z. Raisi-Estabragh, and S. E. Petersen, "Explainable artificial intelligence and cardiac imaging: Toward more interpretable models," *Circulation: Cardiovascular Imaging*, vol. 16, no. 4, Art. no. e014519, 2023, doi: 10.1161/CIRCIMAGING.122.014519.

[35] H. Yuan, H. Yu, S. Gui, and S. Ji, "Explainability in graph neural networks: A taxonomic survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5782–5799, May2023,doi: 10.1109/TPAMI.2022.3204236.