

FCM-Encoder: a novel interpretable framework with vision–language reasoning for coronary artery disease and solitary pulmonary nodule classification

Received: 14 February 2026

Accepted: 3 July 2026

Published online: 06 August 2026

Cite this article as:

Feleki A., Papageorgiou E., Apostolopoulos I.D. *et al.* FCM-Encoder: a novel interpretable framework with vision–language reasoning for coronary artery disease and solitary pulmonary nodule classification. *BMC Med Inform Decis Mak* (2026). <https://doi.org/10.1186/s12911-026-03695-8>

Anna Feleki, Elpiniki Papageorgiou, Ioannis D. Apostolopoulos, Dimitris J. Apostolopoulos, Nikolaos Papathanasiou & Nikolaos I. Papandrianos

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

FCM-Encoder: A Novel Interpretable Framework with Vision–Language Reasoning for Coronary Artery Disease and Solitary Pulmonary Nodule Classification

Anna Feleki¹, Elpiniki Papageorgiou¹, Ioannis D. Apostolopoulos¹, Dimitris J. Apostolopoulos², Nikolaos Papathanasiou², and Nikolaos I. Papandrianos^{1,*}

1. *Department of Energy Systems, University of Thessaly, Gaiopolis Campus, 41500 Larisa Greece, annafele1@uth.gr, elpinikipapageorgiou@uth.gr, japostol@iceht.forth.gr, npapandrianos@uth.gr*
2. *Department of Nuclear Medicine, University Hospital of Patras, Rio, 26504, Greece, dimap@med.upatras.gr, nikopapath@upatras.gr*

* Author to whom correspondence should be addressed

Abstract

Although modern machine-learning models attain strong predictive performance, their clinical adoption remains limited by a lack of transparency. This concern is particularly relevant in multimodal diagnostic pipelines, where the contribution of each modality to the final prediction is more difficult to discern. Relatively few studies integrate heterogeneous clinical and imaging data, where they typically employ decision- or feature-level fusion, and provide limited insight into the cross-modal interactions that underline the diagnostic decision. We propose a Fuzzy Cognitive Map Encoder (FCM-Encoder) framework for multimodal explainable classification. Imaging features extracted by a Convolutional Neural Network (CNN) are mapped into low-dimensional concepts and combined with normalised clinical concept activations inside a unified FCM. The connections to the diagnostic output are initialised from the Spearman rank correlation of each concept with the label, and all interconnections are then refined by a particle swarm optimisation strategy enhanced with Lévy-flight exploration. Under 10-fold cross-validation, the framework reached $81.82\% \pm 4.51\%$ accuracy for Coronary Artery Disease (CAD) from clinical variables and polar maps, and $93.39\% \pm 3.09\%$ for Solitary Pulmonary Nodule (SPN) classification in Non-Small Cell Lung Cancer (NSCLC) from CT, PET, and clinical data. The ablation analysis confirmed that each component contributes to the overall performance, the image-encoder comparison showed that the selected encoder matches or exceeds substantially larger pretrained backbones, and the graph-baseline comparison demonstrated that the FCM-based reasoning surpasses graph neural networks. FCM-Encoder further links imaging concepts to clinical labels and generates clinically aligned narratives through a Vision–Language module.

Keywords: Fuzzy Cognitive Maps, Deep Learning, Classification, FCM-Encoder, Qwen, NLG

1. Introduction

1.1. Prologue

Cardiovascular and oncological diseases remain the top causes of death worldwide, challenging early, accurate diagnostic solutions that are also interpretable for guiding treatment decisions. Coronary Artery Disease (CAD) represents one of the most significant challenges in contemporary cardiovascular medicine and has a substantial impact on patient health and quality of life worldwide. Its progression is therefore a critical public health priority [1]. The primary characteristic is the gradual development of atherosclerotic plaques in coronary arteries that reduce myocardial perfusion and are known to cause ischemic events [1, 2]. Timely and appropriate detection of CAD is essential to enhance patients' outcomes and guide clinical decision-making. [2, 3].

Lung cancer is the leading cause of cancer-related deaths worldwide. Based on the size of cancer cells, lung cancer is divided into small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC makes up about 85% of lung cancer cases, including adenocarcinoma (ADC) at around 50%, squamous cell carcinoma (SCC) at about 40%, and large cell lung cancer [5]. The chances of survival and the selection of the appropriate treatment improve significantly with the early detection of lung cancer.

For CAD diagnosis, several reports have shown the growing utilization of artificial intelligence [3, 6–8]. Similarly, for SPN diagnosis, several studies have been reported [5, 9–13], focusing on the detection and characterization of SPNs. These methods aim at lowering the number of consecutive examinations and at

reducing the medical workload by ruling out obvious cases, as seen in [14, 15] for CAD studies and in [16, 17] for SPN characterization. Additional independent clinical factors, such as age and histology, may have a substantial role in making personalized survival predictions, with reference to CAD studies [8, 18] and for SPN [19, 20]. It has been proven beneficial to employ diagnostic assistance tools in such analyses, as it minimizes the need for extensive interventions, as demonstrated in [21, 22] for CAD and in [23, 24] for SPN classification. These studies have achieved high performance metrics, but they lack explainability. Consequently, nuclear experts are hesitant to fully incorporate these systems into their decision-making workflows due to the lack of transparency and interpretability.

On the other hand, Fuzzy Cognitive Maps (FCMs), which aim to support clinical decision making by a well-structured representation of the relationship among patient-specific factors, allow for a comprehensive and transparent view of the state of the patient [25]. FCMs, introduced by Kosko [26] as an enhanced version of cognitive maps, represent variables and system states with nodes, where relationships in the form of weighted values illustrate the influence among them. These values lie within the range of $[-1, 1]$. The interconnection values within FCMs signify the strength of relationships between nodes. More information about FCMs can be found in [27] [28] [26] [29] [30]. FCMs have exhibited noteworthy outcomes in medical diagnosis [28–32] and are proficient at establishing relationships among clinical variables, which are transformed to concepts of the FCM system, and enhancing the applicability and reliability of their outcomes.

There has been an increasing research interest in recent years in hybrid and multimodal approaches to medical diagnosis that combine clinical and imaging data, aiming to improve prediction accuracy and interpretability. Among them, concept-level methods have drawn attention because they explicitly model the reasoning process, offering a transparent alternative to fully data-driven deep learning models. These methods are presented in the following section.

1.2. Literature studies

This section demonstrates the related literature in two subsections: firstly, it reviews external literature on multimodal fusion approaches for CAD and SPN, and secondly it summarizes our earlier work on Deep Fuzzy Cognitive Maps (DeepFCM) and its learning techniques for CAD and SPN classification diagnosis. For a more analytical comparison, representative literature studies are presented in the Appendix, organised by input modalities, model architecture, evaluated instances, performance metrics, interpretability mechanisms, and reported limitations.

1.2.1. Multimodal studies for CAD diagnosis

A multimodal CAD classification framework was introduced in [33], by Apostolopoulos et al., to overcome the limitations of single-modality diagnostic approaches. The method combined myocardial perfusion imaging (MPI) SPECT polar maps with structured clinical variables, demonstrating that CNN-based imaging models alone inadequately capture patient-specific risk heterogeneity, while clinical-only models lack spatial ischemic representation. Deep CNN-derived polar-map features were fused with machine-learning classifiers trained on clinical data. Performance was reported in terms of accuracy, sensitivity, specificity, and AUC, achieving approximately 82% accuracy and an AUC of 0.90. However, fusion was performed at the decision level using independent learners, limiting cross-modal interaction. The technique was also still data-driven, lacking any mechanistic or interpretability aspect, and validation was limited to a single retrospective cohort.

A support vector machine-based multimodal CAD classification framework was introduced in [34], by Arsanjani et al., to integrate heterogeneous clinical information with quantitative myocardial perfusion analysis. Rather than using raw images, the study relied on expert-engineered polar-map-derived perfusion and functional scores combined with clinical risk factors. Diagnostic performance was evaluated using accuracy and agreement with expert interpretation, achieving an accuracy of approximately 86% and an AUC of 0.92. While the framework demonstrated that structured multimodal fusion could outperform isolated quantitative imaging and expert assessment, reliance on handcrafted imaging descriptors limited the ability to capture complex imaging patterns, and the nonlinear SVM model offered limited interpretability of feature interactions. In addition, the absence of deep feature learning limited robustness across scanners and clinical populations.

Using a large multicenter cohort, Otaki et al. [35] evaluated an explainable deep learning framework (CAD-DL) for obstructive CAD detection that integrates SPECT myocardial perfusion imaging (MPI) polar maps with selected clinical parameters through early feature-level fusion. Raw stress perfusion, wall motion, and wall

thickening polar maps were processed through convolutional layers to extract imaging representations, while non-imaging clinical attributes (age, sex, and left ventricular end-diastolic and end-systolic volumes) were incorporated directly into the final fully connected layer of the network, jointly contributing to probability estimation rather than being combined at the decision level. The model was evaluated on 3,578 patients (10,734 vessels) using invasive coronary angiography (ICA) as reference. In repeated 10-fold validation, CAD-DL achieved a per-patient AUC of 0.84 (95% CI: 0.81–0.87), with sensitivity of 85.6% and specificity of 60.0%, outperforming stress total perfusion deficit (TPD) and expert visual interpretation. Per-vessel analysis yielded an AUC of 0.78 (95% CI: 0.76–0.80). External validation on 555 patients confirmed robustness (AUC = 0.80, 95% CI: 0.76–0.84). Interpretability was achieved through gradient-based attention maps showing important regions in the myocardium; however, interpretability was image-driven.

A deep learning-based framework for stress-only MPI CAD classification was proposed in [36], by Liu et al., to address inefficiencies in conventional stress–rest imaging workflows. Deep convolutional networks were used to extract features from stress SPECT polar maps, which were combined with selected clinical variables. The proposed model attained a reported accuracy of approximately 79% and AUC values around 0.80. The results showed that stress-only functional imaging augmented with clinical context can preserve diagnostic reliability. Nevertheless, the framework prioritised workflow efficiency over interpretability, remained largely data-driven, and was evaluated under a constrained acquisition protocol, limiting generalisability to more complex CAD presentations.

1.2.2. Multimodal studies for SPN diagnosis

In the following section, we present key studies that implement multimodal approaches for SPN malignancy characterization. Studies relying solely on single-modality models based on either imaging or clinical data are not considered, as these usually have limited diagnostic ability and inherently biased outcomes [9, 37–40].

In paper [41], Wang et al. tackled the problem of reducing false positives in CT-based lung cancer detection using explicit nodule morphology modeling in a deep learning framework. Their method combined a 3D convolutional neural network and a recurrent neural network for nodule detection, followed by CNN-based morphological feature extraction and XGBoost-based logistic regression for malignancy prediction. The experiments were carried out with the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset on radiologist-annotated nodule characteristics such as diameter, spiculation, lobulation, and internal composition. The proposed system achieved an accuracy of 82.39%, with diameter, spiculation, and lobulation identified as the most influential predictors. Despite improved false-positive control, this framework was still confined to CT imaging analysis and heavily relying on hand-crafted morphology features without explicit modeling over clinical/imaging concepts.

In study [42], He et al. addressed the need for highly sensitive, non-invasive early lung cancer diagnosis by integrating imaging, clinical variables, and cell-free DNA methylation biomarkers. Leveraging a large multicenter cohort of 1,380 patients across 24 hospitals to develop and extend the initial model with the addition of molecular biomarker information (PulmoSeek Plus), CIBM was developed. The combined framework achieved sensitivities up to 0.98, particularly for early-stage and small nodules. However, although it performed with high diagnostic accuracy and was externally validated, the approach required specialized molecular assays, increased clinical cost and complexity, and provided limited interpretability regarding the interaction between biological, imaging, and clinical risk factors

1.2.3. DeepFCM studies

Several studies have examined the incorporation of DL architectures with FCMs, leading to hybrid frameworks [32]–[35].

In this context, DeepFCM is an FCM-based multimodal methodology developed by our research team within the funded EMERALD project [43], integrating deep feature extraction with fuzzy cognitive reasoning to support interpretable medical diagnosis. The DeepFCM architecture has since been applied to multimodal settings that incorporate both clinical and imaging data for diseases such as CAD and SPN. By fusing clinical variables with CNN-derived imaging predictions, it achieved high diagnostic accuracy while maintaining strong interpretability through its underlying reasoning mechanisms. For all experiments, the initialization of the interconnections, experts' knowledge was applied in the form of linguistic values, which were later transformed

into a spectrum of numerical values. In these studies, the graph is built using both clinical and imaging characteristics. Then, meta-heuristic methods, including Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Extreme Learning Machine (ELM)-based strategies, are used to determine the optimal edge weights.

DeepFCM was first introduced in [31] to classify malignant SPNs using clinical data and PET scans. The clinical features included SUVmax uptake derived from the PET scan, SPN diameter, and the decision output of an RGB-CNN (originally introduced in [44]), which was responsible for analyzing the PET image independently. The DeepFCM model attained 94.71% accuracy with 2.99% standard deviation, assessed by 10-fold cross-validation, and displayed the relationships between input and output concepts through a concept graph.

DeepFCM achieved 77.95% accuracy in diagnosing CAD in [45], utilizing twenty-two clinical characteristics and RGB-CNN predictions as an additional concept attained from the training on MPI polar maps. The model's explainability was demonstrated by comparing the connections among input-output concepts with relationships initialized by experts and weights initialized randomly. In [30], DeepFCM's training functionalities were improved, achieving an accuracy of 83.07% for the same task. Using a Random Forest-based feature selection process, we identified a subset of eight informative clinical features. These eight features, along with the CNN predictions obtained from the RGB-CNN trained on Polar Maps, were used as the final input concepts. As an imaging XAI technique, Gradient-weighted Class Activation Mapping (Grad-CAM) highlighted the pixel areas that significantly influenced the RGB-CNNs' decisions. Additionally, Generative Pre-trained Transformer-Application Programming Interface (GPT-API) transformed DeepFCM outputs into clear, understandable language.

Deep-FCM-PSO was employed in [46] using an expanded dataset comprising nine clinical characteristics and PET images to effectively identify malignant SPNs, achieving 88.14% accuracy. Grad-CAM was used to highlight malignant areas in the PET images based on RGB-CNN feature maps. GPT-API provided natural language explanations for the DeepFCM-PSO outputs.

In study [47] Extreme Learning Machine (ELMs) was incorporated into the DeepFCM learning framework as the core training mechanism for estimating inter-concept relationships. Within this formulation, ELM was employed to analytically compute the output weight matrix of the FCM using the Moore–Penrose pseudoinverse, enabling fast learning of the interconnection weights that minimise the prediction error. The resulting model, termed DeepFCM-ELM, achieved an accuracy of $80.4\% \pm 4.97\%$ for CAD diagnosis and 90.57% for SPN identification. Model interpretability was supported through explicit visualisation of the weighted concept graph, allowing direct inspection of the influence of all clinical and imaging features. In parallel, Grad-CAM was applied to interpret the RGB-CNN component, highlighting image regions most influential for the CNN predictions. The DeepFCM-ELM outputs were further translated into structured natural-language explanations using the GPT-API. All components were integrated into a unified medical decision support system (MDSS) with a clinician-oriented interface and evaluated by medical experts.

1.2.4. Limitations of the literature studies

Existing multimodal approaches for CAD and SPN diagnosis have shown promising performance; however, important limitations remain. In many DeepFCM-based systems, multimodal fusion is performed by combining clinical variables with the final predictions of CNN models, rather than with structured or intermediate image representations. In such cases, imaging information is compressed into a single decision signal, limiting the richness of image-derived features and reducing the contribution of imaging evidence to the overall reasoning process. Conversely, in other multimodal frameworks, learning is primarily driven by imaging-derived features, with clinical variables incorporated only as secondary modifiers. This modality imbalance results in inconsistent integration strategies, where either clinical or imaging information dominates the decision process, thus limiting transparency and interpretability. Moreover, the FCM connections are usually initialized randomly or using domain knowledge, which may not accurately reflect data-driven relationships among features and can negatively affect convergence.

Interpretability is also commonly addressed in a limited or post hoc manner. Many approaches rely on a single explainability technique, such as saliency maps, which does not influence how clinical and imaging information are integrated during learning. Moreover, natural-language explanations are often generated using proprietary, token-based large language models, introducing cost, scalability, and reproducibility constraints. These

limitations motivate the development of a unified multimodal framework that integrates clinical variables with richer image representations, adopts data-driven initialization strategies, and supports transparent, interpretable, and reproducible clinical reasoning, which is the aim of the present study.

1.3. Aim of this study

This paper fills these gaps with several new contributions.

- Firstly, the proposed FCM-Encoder is constructed by employing modality-specific RGB-CNN encoders that transform imaging inputs into semantically meaningful concept representations, which are subsequently combined with structured clinical variables to form a unified multimodal concept space.
- Second, the proposed FCM-Encoder employs a PSO-based learning strategy with Lévy-flight perturbations to enhance global exploration during weight optimisation, improving convergence, while preserving sparsity and interpretability of the learned FCM structure.
- Furthermore, it replaces traditional expert-driven FCM weight initialization with data-driven Spearman correlation so that the initial adjacency really represents true dependency between features rather than subjective expert assumptions.
- Lastly, the framework replaces proprietary token-based NLG APIs with the open-source Qwen3-VL model, allowing the generation of transparent, clinical narratives derived from FCM reasoning and patient-specific activations.

2. Materials and Methods

2.1. CAD data

A total of 2036 patients underwent gated SPECT MPI between February 2018 and February 2022 at the Nuclear Medicine Department at the University Hospital of Patras. The ethical committee of the University General Hospital of Patras approved the data collection process. Attenuation correction was applied to all the gathered images. After ICA within 60 days and applying exclusion criteria, 594 patients were selected, with 43.82% CAD-positive. The twenty-two clinical characteristics considered are: 1. Sex, 2. Age, 3. body mass index (BMI), 4. known CAD, 5. previous acute myocardial infarction (AMI), 6. previous percutaneous coronary intervention (PCI), 7. previous coronary artery bypass grafting (CABG), 8. previous STROKE, 9. Diabetes, 10. Smoking, 11. Hypertension, 12. Dyslipidemia, 13. Angiopathy, 14. Chronic Kidney Disease, 15. Family History of CAD, 16. Asymptomatic, 17. Atypical Symptoms, 18. ANGINA LIKE, 19. Dyspnea on Exertion, 20. Incident of Precordial Pain, and 21. ECG, along with the 22. Expert Diagnosis with one output node, the CAD diagnosis, which is normal and pathological.

2.2. PET/CT data

The PET/CT image data have been recorded in the Clinical Sector of the Department of Nuclear Medicine at the University Hospital of Patras using a hybrid PET/CT scanner (Discovery iQ3 sl16, General Electric Healthcare). The data collection process was approved by the Ethics Committee of the University General Hospital of Patras. This system uses three detector rings with a 15cm field of view to reconstruct 35 axial images at 4.25mm intervals. 3D volumes are acquired to represent the whole body using various bed positions. At the same time, the patient was in a supine position. Two experienced human readers (N.P., 10 years of experience, D.J.A., 30 years of experience) characterized the SPN malignancy with patient follow-up. The study's nature waives the requirement for obtaining patients' informed consent. From 2020 to 2023, more than 800 PET/CT scans were reviewed to identify potential participants. Patients without detected SPNs or with SPNs with a diameter greater than 30mm were excluded. 456 patients with a single SPN were qualified. The total benign cases were 222, and the total malignant cases were 234. Experts annotated CT scan slices, noting the finding's type, location, margins, diameter, and SUVmax and SPN diameter, along with demographic information about each patient (gender, age, BMI). Also, the GLU variable is included. The SUVmax and diameter parameters were extracted from the PET scan. Each SPN finding is represented by a single 2D slice in which the full extent of the nodule is visible. Ultimately, the dataset's features were age, gender, BMI, SUVmax, diameter, and three categorical variables: location (left upper, right upper, left lower, right lower, middle, lingula), type (solid, ground-class, consolidated, speckled, calcified, cavitary), and margins (spiculated, lobulated, well-defined, ill-defined). Each categorical feature is one-hot encoded into binary variables.

2.3. Proposed methodology

In this study, we propose the FCM-Encoder framework for two diagnostic tasks: CAD diagnosis and SPN characterization. The proposed methodology represents clinical variables and image-derived features as the nodes (concepts) of an FCM. Clinical nodes correspond to explicit medical attributes, whereas image nodes are latent features extracted by the CNN encoders, whose clinical associations are characterised separately through correlation analysis.

For CAD diagnosis, structured clinical variables are initially normalised and transformed into clinical concept activations, forming the clinical input stream of the FCM. In parallel, myocardial perfusion polar-map images are trained by an RGB-CNN encoder to learn task-relevant imaging representations. The trained encoder compresses the polar maps into a set of latent imaging concepts suitable for integration within the FCM.

For SPN diagnosis, clinical and demographic variables are similarly normalised and represented as clinical concept activations, while CT and PET images are processed by modality-specific RGB-CNN encoders trained in the same manner.

For each task, the resulting imaging and clinical concepts are jointly integrated within the FCM-based architecture, enabling unified multimodal reasoning over heterogeneous data sources.

In both applications, the connection from each input concept to the diagnostic output is initialised from the Spearman rank correlation between that concept and the label, providing a data-driven prior, while all concept-to-concept interconnections are initialised at zero and refined during optimisation. During training, the FCM-Encoder adopts an enhanced particle swarm optimisation (PSO) strategy to iteratively refine the interconnection weight matrix, allowing clinical and imaging concepts to jointly influence the final prediction. Compared with conventional PSO, the proposed optimiser incorporates four key extensions: (i) a linearly decreasing inertia weight to regulate the exploration–exploitation balance; (ii) numerical gradient approximation for local solution refinement; (iii) Lévy flight–based perturbations to facilitate escape from local optima; and (iv) sparsity-aware regularisation to encourage clinically interpretable cognitive graphs.

The final system integrates multiple explainability-enhancing components. The learned interconnections among concepts are visualised, highlighting the most influential clinical characteristics guiding the prediction. A complementary Spearman correlation analysis is additionally performed between each learned imaging concept and the named clinical variables, assigning a derived clinical label to every informative image feature. Finally, the Qwen3-VL model, an open vision–language model (VLM), is employed to translate the numerical outputs of the FCM-Encoder into human-understandable clinical explanations. A schematic representation of the complete methodology is shown in Figure 1.

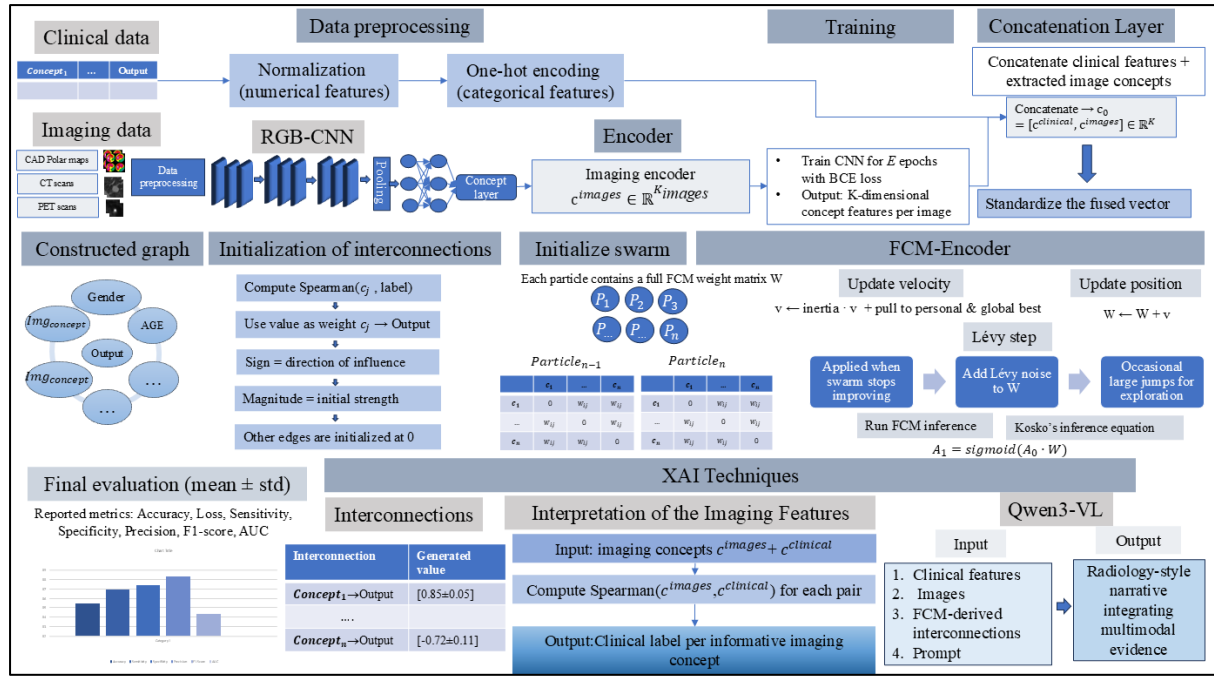


Figure 1: Representation of the pipeline of the FCM-Encoder multimodal approach

In the following subsections, we provide a detailed description of the proposed methodology.

2.3.1. Clinical data preprocessing

For CAD diagnosis, the continuous clinical variables were scaled to comparable ranges: age and BMI were normalised to $[0, 1]$ using min-max scaling, preventing variables with larger numeric ranges from dominating the learning process [48]. The categorical variable gender was binarised, and all remaining clinical variables were already encoded in binary form, indicating absence (0) or presence (1).

For SPN diagnosis, age was likewise min-max normalised, while body mass index (BMI), glucose (GLU), standardised uptake value (SUV), and nodule diameter were each scaled by dividing by a fixed maximum value (70, 192, 30, and 3, respectively) to bring all numerical columns onto a comparable scale. Categorical clinical features, such as nodule location, type, and margins, were one-hot encoded, converting each category into a separate binary column. This representation allows each clinical attribute to be treated as an independent concept within the FCM [48].

2.3.2. Imaging encoder

A lightweight RGB convolutional neural network (RGB-CNN), adapted from the architecture introduced in [44] for bone-metastasis diagnosis, where it is employed in our study as an imaging feature encoder. Its role is not to perform classification but to transform raw imaging inputs into numerical representations that summarise the most relevant visual information. The configuration used for each modality is described below.

For the CAD task, the RGB-CNN was trained on polar-map images. It comprises three convolutional blocks with 32, 64, and 128 filters (3×3 kernels), each followed by batch normalisation, ReLU activation, 2×2 max-pooling, and dropout (rate = 0.15); L2 weight regularisation (1×10^{-4}) is applied throughout. The feature maps are reduced by global average pooling and passed through a 128-unit ReLU dense layer, followed by a 8-unit concept layer whose activations form the image concepts, and finally a sigmoid output used for normal-pathological classification during RGB-CNN training. Images were resized to 300×300 and trained with a batch size of 16 using the Adam optimiser and binary cross-entropy loss (label smoothing 0.05). Light augmentation (random horizontal flips and random rotations) and early stopping on the validation loss (patience = 6, best weights restored) were used to limit overfitting.

For the SPN task, two separate RGB-CNN encoders were trained, one for CT and one for PET, sharing the same architecture but with modality-specific weights. Each comprises four convolutional blocks with 32, 64, 128, and 256 filters, each with batch normalisation, ReLU, 2×2 max-pooling, and dropout (rate = 0.25), followed by

global average pooling, a 128-unit ReLU dense layer, an 8-unit concept layer, and a sigmoid output for benign–malignant classification. Training used 100×100 inputs, a batch size of 16, the Adam optimiser, and binary cross-entropy loss (label smoothing 0.05), with random flip/rotation augmentation and early stopping (patience = 8). The CT and PET concept layers yield eight concepts each, giving sixteen imaging concepts in total.

During the training of the RGB-CNN encoder, each RGB-CNN is trained in a supervised manner on the diagnostic labels so that its concept layer learns task-discriminative features. After this phase, the encoder weights are frozen, and the sigmoid classification head is discarded; only the concept-layer activations are forwarded to the FCM. The standalone RGB-CNN accuracy reported in Table 2, therefore reflects the encoder used in isolation as a baseline classifier and does not represent the decision produced by the FCM-Encoder. Keeping the encoder fixed thereafter preserves the learned visual semantics and represents the imaging data in a reusable form for integration with the clinical concepts in the FCM reasoning layer.

2.3.3. FCM-Encoder implementation: concept-level fusion of clinical and imaging streams

Following the preprocessing of the clinical variables and extraction of image-derived concepts by the frozen CNN encoder, the imaging stream is integrated at the concept level. For each subject, the normalised clinical feature vector is directly concatenated with the corresponding image-concept vector, and the combined vector is standardised to form the multimodal input representation that serves as the input layer of the FCM. This fused representation is augmented with a dedicated output concept initialized to 0.5, resulting in an initial concept state $A_0 \in \mathbb{R}^{N+1}$, which serves as the starting point for the FCM-Encoder. Through this formulation, both clinical and image-derived concepts are treated uniformly as input nodes within the same FCM structure, allowing the model to analyse associations between patient characteristics and imaging patterns in a shared concept space.

For the CAD classification task, eight image-derived concepts extracted from polar map images, and were concatenated with twenty-two clinical variables, forming a thirty-concept multimodal FCM, plus one diagnostic output node. For the SPN classification task, eight CT-derived concepts and eight PET-derived concepts were combined with nine clinical attributes, expanded to twenty-five concept nodes by one-hot encoding of the categorical variables, yielding a forty-one-concept multimodal input layer, again augmented with a single diagnostic output node.

At the fusion stage, standard feature normalization (StandardScaler) is applied to the combined concept vector to ensure that all clinical and imaging concepts lie on a comparable numerical scale. This step prevents dominance due to differences in feature magnitude and ensures that the relative influence of each concept is determined by the learned FCM interconnections rather than by scale effects.

Within the proposed framework, the fused multimodal representation is governed by a shared FCM concept layer, where each dimension corresponds either to a clinical variable or to an image-derived concept.

To clarify this distinction, two categories of input concepts are defined within the FCM. The first category contains named clinical concepts, that is, pre-defined and clinically interpretable variables such as age, diabetes, and smoking history, whose semantic meaning is fixed before training. The second category contains learned imaging features, that is, the continuous activations produced by the final concept layer of the RGB-CNN encoder (eight per branch). Unlike the clinical variables, the learned imaging features are not pre-specified medical attributes; they emerge from the RGB-CNN training (imaging encoder) and acquire their semantic content through their statistical association with the diagnostic label. The remainder of this paper retains the term "concepts" for both categories to preserve consistency with the FCM literature, but the distinction between named and learned concepts is maintained throughout. The interpretability of the learned imaging features is examined in Sections 3.1.6 and 3.1.12 for CAD and SPN diagnosis, respectively.

Finally, the full dataset was partitioned into an 80% training set and a 20% external hold-out test set. The training portion was further evaluated using 10-fold cross-validation (random seed = 42) to ensure reproducibility of the reported results.

2.3.4. Weight initialization

The model integrates a data-driven initialization strategy that aims to align each concept's initial influence on the diagnostic outcome. Rather than relying on expert-defined interconnections, the influence of each input concept on the outcome is computed. Spearman's rank correlation has been frequently applied as a global and non-parametric measure, ensuring robustness to non-linear and non-Gaussian feature distributions commonly encountered in clinical data [49, 50]. For every input concept j (clinical and imaging), the Spearman correlation coefficient p_j with respect to the outcome is estimated and used directly as the initial weight from that concept to the output node ($w_{\{j \rightarrow out\}} = p_j$). Its sign encodes the direction of influence, positive and negative values indicating an increasing or decreasing monotonic relationship, respectively, and its magnitude the initial strength. Undefined correlations (e.g., for constant features) are set to zero, so that no directional bias is imposed and the model determines their influence during learning [50, 51].

These correlation-based priors are applied to all input concepts, both clinical and imaging, affecting only the edges that link each input to the diagnostic output; all concept-to-concept edges are initialized at zero and learned during optimisation. In this way, the initialization provides an empirically grounded starting point for each concept's influence on the outcome without prescribing a fixed structure.

2.3.5. Particle Swarm Optimization

Following the proposed correlation-based initialization, the interconnection weights of the FCM must be further refined through a training procedure that can effectively search the high-dimensional and non-convex parameter space. In this work, Particle Swarm Optimisation (PSO) is adopted as a population-based learning strategy for adjusting the cognitive weight matrix in a fully data-driven manner. The main principles of PSO are briefly outlined next.

PSO [52] is a learning procedure that is inspired by the cooperative behavior of a flock of birds. The particles that represent candidate solutions are deployed across the search space. Each particle behaves as a potential solution has its position and velocity vector. These particles are distributed within the search space, and they change their trajectories over time after learning from their own experience and information shared with others. Each particle tracks its best-known position, denoted as p_i , which influences its future actions. Similarly, the swarm keeps a record of the global best position g found by any particle. This allows the particles to collaborate from their continuous interaction, toward a single best solution. A detailed description of the PSO approach is available in the literature [29, 30, 53].

The weight matrix that constitutes the total of interconnections in an FCM is highly influential, as it directly determines both the predictive performance of the model and the plausibility of the learned inter-concept relationships. Therefore, the training process must minimise classification error while ensuring that the resulting cognitive structure remains consistent with empirical clinical tendencies and historical evidence, when available [31]. In the present study, we propose the pioneering FCM-Encoder method for estimating the weight matrix in order to effectively capture the interconnections among concepts. In the following, we present the equations that constitute the proposed model.

2.3.6. FCM-Encoder Velocity Update

The proposed FCM-Encoder learning framework extends the classical PSO-based weight optimization paradigm by integrating multiple complementary feedback mechanisms within a unified velocity update rule. Beyond the standard cognitive and social components of PSO, the framework introduces a numerically estimated gradient-informed term and a sparsity-aware structural component, enabling both locally guided refinement and explicit control over the learned FCM topology.

The cognitive and social terms follow the standard PSO formulation introduced by Kennedy and Eberhart [52], which has been widely adopted in FCM learning for weight optimization [53]. The extended velocity update is defined as:

$$V_i(t+1) = \kappa \cdot \left[w \cdot V_i(t) + C^1 \cdot r^1 \cdot (P_i - X_i(t)) + C^2 \cdot r^2 \cdot (G - X_i(t)) + C^3 \cdot r^3 \cdot \nabla f_i - C^4 \cdot r^4 \cdot \text{sign}(X_i) \right] \quad (1)$$

where P_i and G denote the personal and global best positions, κ is the constriction factor, w the inertia weight, $C^1 - C^4$ the cognitive, social, gradient, and sparsity coefficients, $r^k \sim U(0,1)$ independent uniform random

scalars, ∇f_i the numerically estimated fitness gradient, and $\text{sign}(X_i)$ the element-wise sign of the particle's weight matrix. By explicitly incorporating gradient-informed and sparsity-constrained feedback in the swarm dynamics, the formulation optimises predictive accuracy and cognitive structure jointly within a single, unified update rule.

The inclusion of the gradient-informed term ∇f_i hybridizes the global exploration capability of PSO with the refinement behavior of gradient-assisted optimization strategies, which have been studied in generic swarm-based settings [54]. In the proposed framework, gradient information is estimated numerically through localized perturbations of the fitness function, allowing particles to exploit task-dependent slope cues while preserving the transparency and rule-based mechanism of FCM reasoning.

In addition, the sparsity-aware term $\text{sign}(X_i)$ introduces a topology-regularizing influence directly within the velocity update, encouraging structurally cognitive graphs during optimisation. The detailed definition and role of this sign-based term are provided in the following subsection.

Unlike conventional PSO-based FCM training approaches that rely solely on stochastic cognitive and social exploration, the proposed velocity update jointly incorporates cognitive, social, gradient-informed, and topology-aware feedback within a unified optimisation dynamic. To the best of our knowledge, this multifaceted formulation constitutes a key methodological contribution to the FCM-Encoder framework.

2.3.7. Sparsity Regularization via Sign-Based Pruning

To encourage interpretable cognitive graphs the velocity update includes a sparsity term based on the element-wise sign of each particle's weight matrix. This term is the subgradient of the L_1 norm $\|X_i\|$; when subtracted in the velocity update (Eq. 1), it drives weights of small magnitude toward zero, pruning weak or spurious interconnections while leaving strong, data-supported edges to be retained by the fitness term. It is defined element-wise as:

$$s(X_i) = \text{sign}(X_i) \quad (2)$$

where $\text{sign}(X_i)$ returns a matrix of the same dimensionality as X_i , replacing each entry with its sign (+1, -1, or 0). Because the term acts on the sign rather than the magnitude, it shrinks weak connections without flipping the polarity (direction of influence) of the retained edges, preserving the cause-effect interpretability required in medical FCM models. Unlike conventional approaches that impose sparsity as a static regularisation term within the objective function, the proposed formulation introduces it as an active force in the swarm dynamics, so that sparsity is optimised jointly with predictive accuracy at every iteration. To the best of our knowledge, integrating an L_1 - type pruning force directly into the velocity update of an FCM weight-learning procedure has not been previously reported.

2.3.8. Position Updates with Lévy-Driven Non-Local Exploration

When a particle stagnates, its position update incorporates a Lévy flight component $L'(\alpha, \gamma)$ that enables non-local exploration of the search space to escape local minima, a mechanism commonly used in metaheuristic optimization [55, 56]. While Lévy-based updates are well established in population-based optimisation, their integration within a clinical reasoning framework based on FCMs has, to the best of our knowledge, not been previously reported. This design choice enhances resistance to premature convergence and promotes diversity among candidate FCM weight matrices during optimisation:

$$X_i(t+1) = \tanh(X_i(t) + L') \quad (3)$$

where $X_i(t)$ is the current position of particle i , $X_i(t+1)$ its position at the next iteration, and $L'(\alpha, \gamma)$ a perturbation vector drawn from a Lévy distribution parameterised by the stability index α and the scale γ . By perturbing the particle around its own current position with a heavy-tailed displacement, the update applies small local moves with high probability while occasionally executing long-range jumps to distant regions of the weight space. The bounding $\tanh$ keeps the perturbed weights within $(-1, 1)$, and the heavy tail provides the non-local escape mechanism required to leave deep local minima of the FCM weight landscape without additional restart or diversification heuristics.

2.3.9. Lévy Flight Step Generation Using Mantegna's Algorithm

The Lévy flight step is constructed by Mantegna's algorithm, which efficiently simulates heavy-tailed step sizes from a Lévy distribution:

$$\text{Lévy}(\alpha, \gamma) = \frac{u}{|v|^{\frac{1}{\alpha}}} \quad (4)$$

where $v \sim N(0, 1)$ and the numerator u is a zero-mean normal variable whose standard deviation σ^u is given by Eq (5); the scale of the Lévy step is therefore controlled by σ^u through u . Though the implementation of the Mantegna algorithm itself is a straightforward procedure, we demonstrate that incorporating it in the FCM-Encoder optimization pipeline enhances the non-local exploration nature of our model as well as complements the aforementioned gradient-informed and sparsity-aware update mechanism.

2.3.10. Lévy Distribution Scaling for Stochastic Steps

The scale parameter σ^u of the Lévy distribution is computed using standard gamma-function-based expressions derived from Lévy distribution theory, ensuring that the generated stochastic steps conform to the required statistical properties [57]. Specifically:

$$\sigma^u = \left(\gamma \cdot \frac{\Gamma(1+\alpha) \cdot \sin \frac{\pi\alpha}{2}}{\Gamma\left(\frac{1+\alpha}{2}\right) \cdot a \cdot 2^{\frac{\alpha-1}{2}}} \right)^{\frac{1}{\alpha}} \quad (5)$$

where $\Gamma(\cdot)$ denotes the Gamma function and $\alpha \in (0, 2]$ is the stability index that controls the heaviness of the distribution tails. Smaller values of α produce more frequent long-range jumps, while $\alpha \rightarrow 2$ recovers Gaussian-like behaviour. This formulation guarantees that the sampled steps follow a true α -stable Lévy distribution with the correct variance scaling, providing the swarm with the heavy-tailed perturbations required to escape local minima while preserving fine-grained local refinement around the current best position.

2.3.11. Inertia Weight Schedule

The inertia weight w balances global exploration and local exploitation in the velocity update (Eq. 1). It follows a standard linearly decreasing schedule [58], reduced from an initial maximum w_{max} to a minimum w_{min} over the course of optimisation:

$$w(t) = w_{max} - (w_{max} - w_{min}) \cdot \frac{(t - 1)}{(T_{max} - 1)} \quad (6)$$

where t is the current iteration, T_{max} the maximum number of iterations, and w_{min} and w_{max} the lower and upper bounds of the inertia weight (set to 0.4 and 0.9, respectively). Early iterations use a high inertia that favours broad exploration of the FCM weight space, while later iterations use a low inertia that emphasises local exploitation and convergence refinement.

2.3.12. Fitness Evaluation Strategy

The quality of each candidate weight matrix is assessed by a fitness function combining a prediction term and a regularisation term. The prediction term is the binary cross-entropy (BCE) between the observed clinical outcome y_i and the predicted output $\hat{y}_i$ produced by the FCM-Encoder. To discourage overly large, dense weight matrices and favour cognitive graphs, a small L2 penalty on the weight matrix is added. The resulting objective is:

$$L = BCE(\hat{y}, y) + \lambda \|X\| \quad (7)$$

where $\|X\|$ denotes the Frobenius norm of the weight matrix and λ ($= 0.001$) controls the regularisation strength. This balances predictive accuracy against structural complexity, complementing the sign-based pruning applied within the velocity update.

2.3.13. Output Concept Prediction

The final prediction for a given input is produced by applying a sigmoid activation to the weighted sum of concept activations, yielding a bounded output in the interval $[0, 1]$. Following the standard fuzzy cognitive map activation rule [26], the predicted value of the output concept k is obtained in a single FCM update step::

$$\hat{y} = [\sigma(A_0 \cdot W)]_k \quad (8)$$

where $\sigma(\cdot)$ is the sigmoid function, W is the optimised (global best) weight matrix obtained from the PSO, k is the index of the output concept, and $A_0 = [x \mid 0.5]$ is the initial concept state formed by the input concept vector x augmented with the output concept initialised to 0.5.

2.3.14. Pseudocode of FCM-Encoder

Following, we present the pseudocode of the proposed model, FCM-Encoder.

FCM-Encoder-Pseudocode

Algorithm: FCM-Encoder

Input: Clinical dataset D , paired image set I , labels y

Preprocessing: normalise numerical features, one-hot encode categoricals

Train RGB-CNN encoders on images; extract image concepts; fuse with clinical features

1: Function FCM-Encoder(D, I, y)

2: // ----- Step 1 : feature fusion -----

3: Train RGB-CNN encoders to obtain image-concept extractors $\varphi_{modality}$

4: For each patient k :

5: $x^{(k)} \leftarrow [\text{clinical}^{(k)}, \varphi_1(I_1^k), \dots, \varphi_M(I_M^k)]$ // concept vector for patient k

6: $x \leftarrow [x^{(1)}; \dots; x^{(P)}]$ // data matrix: P patients $\times n_{concepts}$

7: $N \leftarrow n_{concepts} + 1$ // $n_{concepts}$ concept nodes + 1 output node

8: $output_{idx} \leftarrow N$ // output is the last node

9:

10: // ----- Step 2 : initialise weight matrix $W \in R^{(N \times N)}$ -----

11: Initialise $W \leftarrow \text{zeros}(N, N)$

12: // Spearman-based prior for all concepts

```

13:   For j = 1 ... N-1:
14:      $p_j \leftarrow \text{SpearmanCorr}(x[:, j], y)$ ;   if  $p_j$  is NaN:  $p_j \leftarrow 0$ 
15:      $W[j, \text{output}_{idx}] \leftarrow \text{clip}(p_j, -1, 1)$ 
16:      $\text{diag}(W) \leftarrow 0$ ;    $W[\text{output}_{idx}, :] \leftarrow 0$ 
17:   // ----- Step 3 : initialise PSO swarm of K particles -----
18:   For i = 1 ... K:
19:      $X_i \leftarrow W + \text{Normal}(0, \sigma_{init}^2)$    // particle position
20:      $V_i \sim \text{Uniform}(-1, 1)^{(N \times N)}$ ;    $\text{diag}(V_i) \leftarrow 0$ ;    $V_i[\text{output}_{idx}, :] \leftarrow 0$ 
21:      $P_i \leftarrow X_i$ ;    $\text{best\_score}_i \leftarrow +\infty$ 
22:    $G \leftarrow X_1$ ;    $\text{global\_best\_score} \leftarrow +\infty$ 
23:
24:   // ----- Step 4 : PSO optimisation (per CV fold) -----
25:   For fold f = 1 ... 10:
26:     Split into  $x_{train}$ ,  $x_{test}$ ,  $y_{train}$ ,  $y_{test}$ 
27:     For epoch t = 1 ...  $T_{max}$ :
28:        $w \leftarrow w_{max} - (w_{max} - w_{min})(t-1)/(T_{max}-1)$    // Eq. (6)
29:       For each particle i:
30:         // fitness with L2 magnitude penalty
31:          $\hat{y}_i \leftarrow [\sigma(A_0 \cdot X_i)]_{\text{output}_{idx}}$ ;    $A_0 = [x_{train} \mid 0.5]$ 
32:          $\text{score}_i \leftarrow \text{BCE}(y_{train}, \hat{y}_i) + \lambda \cdot ||X_i||$ ;    $\lambda = 0.001$    // Eq. (7)
33:         If  $\text{score}_i < \text{best\_score}_i$ :    $P_i \leftarrow X_i$ ;    $\text{best\_score}_i \leftarrow \text{score}_i$ ;    $\text{no\_improve}_i \leftarrow 0$ 
34:         Else:    $\text{no\_improve}_i \leftarrow \text{no\_improve}_i + 1$ 
35:         If  $\text{score}_i < \text{global\_best\_score}$ :    $G \leftarrow X_i$ ;    $\text{global\_best\_score} \leftarrow \text{score}_i$ 
36:
37:       // hybrid velocity update (always applied)
38:        $\text{grad}_i \leftarrow \text{FiniteDiffGradient}(\text{Fitness}, X_i, \text{EPS\_GRAD})$ 
39:        $r_1, r_2, r_3, r_4 \sim \text{Uniform}(0, 1)$ 
40:        $V_i \leftarrow \kappa \cdot [w \cdot V_i + C^1 r_1 (P_i - X_i) + C^2 r_2 (G - X_i) + C^3 r_3 \cdot \text{grad}_i - C^4 r_4 \cdot \text{sign}(X_i)]$ 
41:       // Eq. (1); sign per Eq. (2)
42:        $V_i \leftarrow \text{clip}(V_i, -V_{clip}, +V_{clip})$ 
43:        $X_i \leftarrow \tanh(X_i + V_i)$    // Eq. (1b)
44:        $\text{diag}(X_i) \leftarrow 0$ ;    $X_i[\text{output}_{idx}, :] \leftarrow 0$ 
45:
46:       // Lévy-driven escape - only when stagnating, added on top
47:       If  $\text{no\_improve}_i \geq \text{STALL}$ :

```

```

47:       $L'_i \leftarrow \text{Mantegna}(\alpha, \gamma)$  // Eqs. (4)-(5)
48:       $X_i \leftarrow \tanh(X_i + L'_i)$  // Eq. (3)
49:       $\text{diag}(X_i) \leftarrow 0$ ;  $X_i[\text{output}_{idx}, :] \leftarrow 0$ 
50:       $\text{no\_improve}_i \leftarrow 0$ 
51:
52:  // ----- Step 5 : inference -----
53:   $A_0 = [x_{test} \mid 0.5]$ 
54:   $\hat{y} \leftarrow [\sigma(A_0 \cdot G)]_{\text{output}_{idx}}$  // Eq. (8), for all test patients
55:  Evaluate: Accuracy, Loss (cross-entropy), Sensitivity, Specificity, Precision, F1,
  AUC
56:  Return G, evaluation metrics

```

2.4. Methods for Interpreting Model Outputs

In this section, we describe the explainable methods implemented in the proposed diagnostic framework. These methods are designed to improve the transparency and interpretability of model by associating multimodal inputs with concept interactions.

2.4.1. Generated Concept Interconnections (Weight Matrix)

Within the proposed framework, interpretability of the clinical reasoning pathway is grounded in the weighted interconnections between clinical concepts and the model's diagnostic output. The output edge of every input concept is initialised from its Spearman rank correlation with the label, computed on the training split. For the named clinical concepts, these correlation-based priors are directly interpretable: the sign indicates whether the variable has a positive or negative monotonic association with malignancy, and the magnitude its initial strength. They initialise the corresponding outgoing edges to the diagnostic output, providing a clinically structured starting point aligned with the associations observed in the training data.

Each weighted edge connecting a clinical concept to the output node is characterized by both a magnitude and a sign. The absolute value indicates relative importance with respect to prediction, while the sign indicates whether increasing the variable is expected to increase or decrease the estimated malignancy probability. During FCM-Encoder optimization, these correlation-informed initial weights may be altered when supported by the data, allowing the final structure to reconcile prior statistical evidence with learned multimodal interactions.

After training, the learned clinical weight matrix provides an explicit representation of the direction and relative strength of influence between individual clinical concepts and the diagnostic output. Each weight represents both the direction and strength of influence, allowing the FCM structure to be inspected directly without post-hoc explanation methods.

2.4.2. Clinical Interpretation of the Learned Imaging Features

Unlike the named clinical variables, the CNN-derived imaging features enter the FCM as data-driven activations without a priori medical attributes. A complementary analysis is performed at the end of training to associate each informative imaging feature with a named clinical concept. The imaging features are first ranked by the absolute magnitude of their learned FCM weight to the diagnostic output, identifying those on which the model most strongly relies. Each feature is then correlated with every named clinical variable using Spearman's rank coefficient, and the clinical variable with the strongest absolute Spearman correlation is adopted as the derived label, with its associated p-value reported. Visual confirmation is finally obtained by inspecting, for each informative feature, the input images (polar maps for CAD, CT/PET images for SPN) of the patients with the highest activation.

2.4.3. Natural Language Generation

FCMs provide an explicit structure through concept nodes and weighted interconnections, however, expressing these numerical influences in clinically meaningful language requires an additional interpretive layer. For this purpose, an open multimodal large language model (Qwen3-VL), as introduced by Bai et al. in [59], was employed as a narrative generation component and deployed via the Ollama inference framework. The model operates strictly in inference mode and is not involved in prediction, classification, or diagnostic decision-making. In the proposed study, it is used exclusively to convert model outputs into structured, clinician-oriented explanations.

At inference time, the prompt to Qwen3-VL includes four elements: (i) the FCM-Encoder output label and its full set of concept weights, (ii) the patient-specific clinical concept activations, (iii) the polar map for CAD or CT/PET scans for SPN diagnosis, (iv) the RGB-CNN prediction for the image instances according to the diagnostic task and (v) guidance about the structure of the generated response. These inputs are provided as contextual evidence for explanation, not for learning. The prompt includes the following rules. For the clinical section, only activated concepts with a positive influence are eligible for explanation. Demographic variables, numerical encodings, and shortcut diagnostic inputs (e.g., sex, age, BMI, expert diagnosis) are explicitly excluded. Each selected feature must be described using disease-specific medical terminology without reference to binary or numerical values. For the imaging section, Qwen3-VL analyses the provided image together with domain-specific instructions. In CAD cases, explanations focus on perfusion defects, ischemic patterns, stress–rest reversibility, and correspondence to coronary territories (LAD, RCA, LCx). In SPN cases, explanations focus on lesion morphology, margins, uptake behaviour, and spatial context. Visual layout artefacts and embedded text are ignored. This design separates prediction and reasoning (performed by CNN and FCM-Encoder) from language generation, ensuring that explanations are derived from model outputs and images, while remaining clinically structured, reproducible, and constrained by fixed prompting rules. The prompt that was inserted into Qwen3-VL for inference of the CAD instance is available in the Appendix.

3. Results

3.1. Classification results

This section presents the experimental evaluation of the proposed FCM-Encoder framework for CAD diagnosis and SPN identification. The classification performance of the full configuration is first compared with representative literature studies on the same source datasets, grouped by input category (clinical, imaging, and multimodal). Systematic ablation studies then quantify the contribution of each architectural component, including the multimodal fusion, the FCM reasoning module, the Spearman-correlation-based initialisation, and the Lévy-flight perturbation inside the PSO refinement. Two further controlled comparisons evaluate alternative image-encoder backbones against the RGB-CNN used by the proposed framework, and three representative graph neural network baselines against the FCM reasoning module. The interpretation of the trained model is finally examined through two complementary analyses: the learned FCM weights between the named clinical concepts and the diagnostic output, and a Spearman correlation analysis that assigns a derived clinical label to each informative learned imaging feature. All results are obtained under 10-fold cross-validation, with every metric reported as the mean and standard deviation across folds.

3.1.1. FCM-Encoder Classification Results – CAD

Table 1 reports the classification performance of the proposed FCM-Encoder on the CAD task under its full configuration, namely multimodal inputs (polar maps and 22 clinical features) with Spearman-correlation-based initialisation of the FCM interconnections, together with representative literature studies on the same source dataset grouped by input category: clinical-only, imaging-only (polar maps), and multimodal. Alternative configurations of the proposed framework, such as random initialisation and clinical-only inputs, are reported separately in the ablation study in Section 3.1.2 (Table 2). Standard evaluation metrics are used for the comparison, including accuracy, loss, sensitivity, specificity, precision, F1-score, and AUC.

Table 1: Comparison of the proposed FCM-Encoder with literature studies on the CAD dataset, grouped by input category (clinical, imaging, multimodal).

Models	Input	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
--------	-------	----------	------	-------------	-------------	-----------	----------	-----

Literature studies (Clinical)								
Samaras et al. (2023) [60]	Clinical	78.82%	-	74.63%	82.05%	-	75.32%	-
Apostolopoulos et al. (2024) [61]	Clinical	80.17%	-	82.86%	78.05%	-	78.61%	89.17%
Samaras et al. (2023) [62]	Clinical	78.29%	-	76.21%	80.8%	-	78.82%	-
Literature studies (Imaging)								
Papandrianos et al. (2022) [6]	Polar maps	95.83%	0.21	97%	94%	-	-	96%
Apostolopoulos et al. (2020) [63]	Polar maps	74.53%	-	75%	73.43%	-	-	-
Apostolopoulos et al. (2023) [64]	Polar maps	78.9%	-	87.32%	72.16%	-	88.99%	-
Literature studies (Multimodal)								
Apostolopoulos et al. (2021) [65]	Polar maps/clinical (22 features)	78.44%±3.71%	-	77.36%±3.87%	79.25%±5.12%	-	75.5%±2.89%	79.26%±3.66%
Feleki et. al (2023) [45]	Polar maps/clinical (22 features)	77.95%±5.58%	0.22±0.05	76.98%±8.27%	77.39%±7.13%	73.97%±0.09%	-	-
Feleki et. al (2023) [30]	Polar maps/clinical (8 features)	83.07%±4.72%	0.17	86.61%	79.99%	81.78%	-	-
Papageorgiou et. al (2025)[47]	Polar maps/clinical (22 features)	80.4%±4.97%	0.19±0.05	75.47%±11.07%	83.75%±14.33%	73.98%±8.49%	-	-
Proposed model (FCM-Encoder with Spearman correlation)								
FCM-Encoder	Polar maps/clinical (22 features)	81.82% ± 4.51%	0.19±0.04	86.23%±9.68%	76.17%±11.16%	74.79%±6.99%	79.52%±4.67%	86.01%±4.53%

In the clinical-only setting, Samaras et al. [60] (CatBoost) and Apostolopoulos et al. [61] (Random Forest) rely on black-box classifiers with post-hoc SHAP or feature-importance explanations, while Samaras et al. [62] use a reduced subset of clinical features. In all three cases, interpretability is introduced retrospectively and is detached from the learning mechanism.

For imaging-only configurations, Papandrianos et al. [6] report a high accuracy of 95.83% on polar maps, but on a substantially smaller dataset and without clinical fusion. The remaining polar-map-only studies [63] [64] fall below 79%, indicating that a purely image-based input does not consistently match multimodal fusion.

In the multimodal category, Apostolopoulos et al. [65] and Feleki et al. [45] reached approximately 78% accuracy using conventional feature-level fusion. Feleki et al. [30] obtained a higher accuracy $83.07\% \pm 4.72\%$, but their architecture relies on only ten input concepts (nine clinical parameters derived from feature selection with Random Forest and the scalar RGB-CNN prediction), so the two modalities are fused at a single representation level rather than at the concept-interaction level. Papageorgiou et al. [47], using the same 22 clinical features and polar maps as the present study, reported $80.4\% \pm 4.97\%$ proposing DeepFCM-ELM.

The proposed FCM-Encoder reaches the highest accuracy ($81.82\% \pm 4.51\%$), sensitivity ($86.23\% \pm 9.68\%$), and AUC ($86.01\% \pm 4.53\%$) among the studies that use the full clinical feature set, together with the lowest binary cross-entropy loss (0.19 ± 0.04) of the multimodal block. This is achieved on the same multimodal input as [47], [65] and [45] (polar maps and twenty-two clinical features) but with a different fusion strategy: clinical and imaging concepts interact directly inside the FCM rather than being concatenated at the prediction or feature level, the interconnections are initialised from a Spearman correlation prior on the training set without recourse

to expert rules, and the weight matrix is refined by a Lévy-flight-enhanced PSO scheme that explicitly minimises a regularised cross-entropy objective.

3.1.2. Ablation Analysis of the FCM-Encoder Modules for CAD diagnosis

To isolate the contribution of each architectural element of the proposed FCM-Encoder, a systematic component ablation was conducted under the identical 10-fold cross-validation protocol, training/validation partitions, preprocessing pipeline, and RGB-CNN backbone, so that the only factor varying between rows of Table 2 is the component under test. Six configurations were compared against the full proposed model: Clinical only (image branch disabled, FCM inferencing on clinical features alone); Images only (CNN to Dense classifier) (clinical branch and FCM reasoning removed, predictions taken directly from the CNN sigmoid head); Clinical + Images + FCM, no PSO (random FCM initialisation, PSO refinement skipped); Clinical + Images + FCM, PSO (Lévy-flight escape step disabled); and No Spearman initialisation (PSO initialised from a zero-weight FCM rather than the Spearman-correlation prior).

Table 2: Component ablation of the proposed FCM-Encoder on the CAD dataset, reported as mean and standard deviation over 10-fold cross-validation. Each row disables or replaces a single module while keeping all other components identical to the full proposed configuration.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
Clinical only	79.2% $\pm$ 4.53%	0.23 $\pm$ 0.05	78.4% $\pm$ 4.9%	82.1% $\pm$ 8.7%	72.3% $\pm$ 10.4%	71.1% $\pm$ 6.5%	77.10% $\pm$ 4.8%
Images only (CNN→Dense classifier)	74.58% $\pm$ 5.25%	0.45 $\pm$ 0.04	81.72% $\pm$ 5.59%	65.48% $\pm$ 9.52%	72% $\pm$ 7.85%	68.31% $\pm$ 7.76%	73.6% $\pm$ 5.63%
Clinical + Images + FCM, no PSO (random init)	79.59% $\pm$ 3.49%	0.65% $\pm$ 0.16	77.82% $\pm$ 16.22%	79.69% $\pm$ 8.4%	76.09% $\pm$ 4.84%	75.8% $\pm$ 8.06%	84.11% $\pm$ 4.7%
Clinical + Images + FCM, PSO (no Lévy mutation)	80.86% $\pm$ 4.15%	0.49 $\pm$ 0.06	80.42% $\pm$ 15.31%	81.38% $\pm$ 10.43%	79.21% $\pm$ 6.56%	78.57% $\pm$ 6.8%	85.15% $\pm$ 4.01%
No Spearman initialization	76.89% $\pm$ 5.29%	0.15 $\pm$ 0.02	80.92% $\pm$ 6.07%	74.07% $\pm$ 7.46%	70.87% $\pm$ 9.13%	75.15% $\pm$ 6.32%	85.25% $\pm$ 4.02%
FCM-Encoder	81.82% $\pm$ 4.51%	0.19 $\pm$ 0.04	86.23% $\pm$ 9.68%	76.17% $\pm$ 11.16%	74.79% $\pm$ 6.99%	79.52% $\pm$ 4.67%	86.01% $\pm$ 4.53%

The results in Table 2 confirm that every component contributes positively, with the full proposed model attaining the highest accuracy (81.82%) and the lowest loss (0.19). Using the clinical variables or the images alone lowers performance, with the images-only settings showing the largest decline, which indicates that the two modalities are most effective when combined. Likewise, removing any of the optimisation components, the Spearman initialisation, the PSO refinement, or the Lévy-flight step, weakens the model by lowering accuracy or increasing the loss. Overall, each component contributes to a distinct improvement, and only their combination yields the highest accuracy and the lowest loss of the proposed framework.

3.1.3. Comparison with Alternative Image Encoder Architectures for CAD diagnosis

To verify whether more recent pretrained networks could

replace the RGB-CNN used in the proposed framework, five alternative encoders were evaluated under the identical 10-fold cross-validation protocol, clinical preprocessing pipeline, and FCM-Encoder reasoning module, so that only the encoder backbone varies between rows of Table 3.

Two categories were considered: (i) lightweight CNNs trained end-to-end from ImageNet weights (MobileNetV3-Small, EfficientNetB0, DenseNet121), and (ii) frozen feature extractors used without fine-tuning, comprising EfficientNetB0 (frozen) as a standard ImageNet baseline and BiomedCLIP (ViT-B/16) as a state-of-the-art biomedical foundation model.

All backbones receive 224×224 RGB inputs and are trained for up to 40 epochs with a batch size of 16, the Adam optimiser at a learning rate of 1×10^{-4} , binary cross-entropy loss with label smoothing of 0.05, and early stopping on the inner validation split (patience 6). For BiomedCLIP, the 512-dimensional embeddings are projected to the 16-dimensional concept space used by all other encoders via PCA fitted on each training fold, ensuring identical feature dimensionality across configurations.

The FCM-Encoder reasoning module is shared across all configurations and uses 60 particles, 40 PSO epochs, inertia annealed from 0.9 to 0.4, cognitive and social coefficients $C_1 = C_2 = 1.6$, a Lévy-flight perturbation coefficient $C_3 = 0.5$, a sparsity coefficient $C_4 = 0.1$, and a stall counter of 10 iterations before triggering Lévy escape.

Table 3: Image encoder comparison on the CAD dataset, reported as mean $\pm$ standard deviation over 10-fold cross-validation. Only the encoder backbone varies between rows; the clinical pipeline, splits, and FCM-Encoder module are identical across all configurations.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC	Params
Lightweight architecture								
MobileNetV3-Small	81.44% $\pm$ 4.76%	0.49 $\pm$ 0.08	87.92% $\pm$ 7.38%	75.53% $\pm$ 10.54%	75.01% $\pm$ 7.56%	80.46% $\pm$ 4.29	86.37% $\pm$ 5.13	2.5 M
EfficientNetB0	81.67% $\pm$ 6.7%	0.52% $\pm$ 0.12	76.02% $\pm$ 18.36%	84.39% $\pm$ 9.8%	81.43% $\pm$ 6.12%	77.11% $\pm$ 10.3%	84.81% $\pm$ 6.72	5.3 M
DenseNet121	80.2% $\pm$ 5.07%	0.54% $\pm$ 0.1	85.42% $\pm$ 7.44%	75.81% $\pm$ 8.9%	73.99% $\pm$ 7.28%	78.92% $\pm$ 5.27%	84.07% $\pm$ 5.17%	8.1 M
Frozen ImageNet feature extractor								
EfficientNetB0	81.44% $\pm$ 5.41%	0.49% $\pm$ 0.08	85.96% $\pm$ 9.09%	76.58% $\pm$ 11.44%	75.78% $\pm$ 5.91%	80.17% $\pm$ 5.22%	85.65% $\pm$ 5.39	5.3 M (frozen)
BiomedCLIP (ViT-B/16)	81.05% $\pm$ 3.59%	0.48% $\pm$ 0.05	88.54% $\pm$ 9.04%	76.43% $\pm$ 9.79%	75.24% $\pm$ 8.36%	80.72% $\pm$ 4.9%	86.56% $\pm$ 3.61%	86 M (frozen)
Proposed								
FCM-Encoder	81.82% $\pm$ 4.51%	0.19 $\pm$ 0.04	86.23% $\pm$ 9.68%	76.17% $\pm$ 11.16%	74.79% $\pm$ 6.99%	79.52% $\pm$ 4.67%	86.01% $\pm$ 4.53%	0.11 M

The proposed FCM-Encoder achieves the highest accuracy (81.82%) and by far the lowest loss (0.19) of all configurations. The alternative encoders fall between 80.20% and 81.67% accuracy and between 0.48 and 0.54 loss, which corresponds to a 2.5- to 2.8-fold worse loss despite using 22 to 780 times more parameters per branch. BiomedCLIP (86 M parameters) reaches accuracy and AUC values that are statistically indistinguishable from those of the much smaller frozen EfficientNetB0 and from the proposed 0.11 M-parameter encoder, suggesting that its biomedical pretraining corpus, dominated by figures and radiographs, transfers poorly to cardiac polar perfusion maps.

Overall, the comparison supports the choice of our encoder, since it outperformed the rest encoders.

3.1.4. Comparison with Graph Neural Network Baselines for CAD diagnosis

Because the FCM-Encoder and graph neural networks (GNNs) both learn weighted relations between feature nodes, three representative GNN baselines were compared against the proposed model while the rest of the multimodal pipeline was preserved. The three span the main families of graph learning. The GCN of Kipf and Welling [66] represents the spectral family: each layer updates a node through a symmetrically normalised, self-loop-augmented averaging of its neighbours' features followed by a learnable linear transformation, and in its original form operates transductively over a fixed graph. The Graph Attention Network (GAT) of Veličković et al. [67] represents the attention-based family, replacing these fixed aggregation weights with learned per-edge attention coefficients (four heads in the first layer) so that informative neighbours contribute more strongly, while additionally yielding edge-level importance scores. GraphSAGE, of Hamilton et al. [68], represents the

inductive family, learning a mean aggregator over a sampled neighbourhood that generalises naturally to nodes unseen during training.

All baselines operate on the same feature stream as the proposed model: the RGB-CNN concept vector concatenated with the clinical variables. A population graph is built inductively per fold, with each patient connected to its five nearest training neighbours by cosine similarity and no edges between validation nodes, so that label information cannot propagate among validation patients.

The three GNNs share a two-layer architecture with 24 hidden units and a dropout rate of 0.6, and are trained with the Adam optimiser (learning rate 1×10^{-2} , weight decay 2×10^{-3}) under cross-entropy loss for up to 150 epochs with early stopping (patience 20) on an inner 80:20 split of each training fold; the best-weight model is restored before predictions are produced for the held-out validation patients. The 10-fold cross-validation protocol and the reported metrics are identical to those used for the proposed FCM-Encoder, so that only the reasoning module varies between the rows of Table 4.

Table 4: Comparison with graph neural network baselines on the CAD dataset (mean $\pm$ standard deviation over 10-fold cross-validation). The three graph baselines (GCN, GAT, GraphSAGE) use the same fused feature stream (clinical variables and RGB-CNN concepts) and the same inductive k -nearest-neighbours population graph, differing only in the graph neural network architecture; the FCM-Encoder is the proposed method.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
GCN (Kipf & Welling 2017) [66]	79.77% $\pm$ 6.17%	0.51 $\pm$ 0.09	82.07% $\pm$ 15.99%	77.29% $\pm$ 9.34%	74.52% $\pm$ 8.31%	77.1% $\pm$ 9.87%	82.82% $\pm$ 6.99%
GAT (Veličković 2018) [67]	81.44% $\pm$ 4.23%	0.5 $\pm$ 0.05	86.32% $\pm$ 7.3%	77.84% $\pm$ 7.99%	75.69% $\pm$ 6.79%	80.22% $\pm$ 4.24%	84.37% $\pm$ 3.11%
GraphSAGE (Hamilton 2017) [68]	80.85% $\pm$ 3.79%	0.51 $\pm$ 0.08	85.54% $\pm$ 12.5	78.03% $\pm$ 10.45%	76.87% $\pm$ 7.34%	80.01% $\pm$ 5.35%	84.94% $\pm$ 4.34%
FCM-Encoder	81.82% $\pm$ 4.51%	0.19 $\pm$ 0.04	86.23% $\pm$ 9.68%	76.17% $\pm$ 11.16%	74.79% $\pm$ 6.99%	79.52% $\pm$ 4.67%	86.01% $\pm$ 4.53%

The proposed FCM-Encoder achieves the highest accuracy (81.82%) and AUC (86.01%) among the four configurations, exceeding GCN, GAT, and GraphSAGE by 2.05, 0.38, and 0.97 accuracy points, respectively, and exhibits the lowest cross-fold variance. The gap over GAT lies within one standard deviation, but no GNN baseline outperforms the proposed model on either metric. The largest difference is in loss: the FCM-Encoder reaches a loss of 0.19 against 0.50–0.51 for the GNN baselines, an approximately 2.6-fold reduction consistent with the PSO refinement explicitly minimising a regularised cross-entropy objective over the FCM-induced probability outputs.

3.1.5. Analysis of FCM-Encoder clinical weights- CAD

This section analyses the interconnections learned by FCM-Encoder between clinical input concepts and the output concept representing CAD diagnosis. The model is trained without expert-informed priors, using the proposed Spearman-based initialization. Expert linguistic values are reported solely for comparative and interpretative purposes and are not used during the training process, as they were defined for previous literature studies for CAD diagnosis, for the initialization of interconnections among concepts [30, 45, 47]. The available linguistic categories include very weak (VW), weak (W), medium (M), strong (S), and very strong (VS), which represent increasing degrees of influence between concepts. Table 5 reports the resulting weights for each connection to the output node, expressed as mean $\pm$ standard deviation across cross-validation folds. These values describe both the direction of influence (positive or negative) of each learned association.

Table 5: Demonstration of FCM-Encoder-generated values applied to CAD dataset

Interconnection	Linguistic values provided by nuclear experts	Transformed to a spectrum of values	FCM-Encoder with Spearman initialization	FCM-Encoder random initialization
Gender $\rightarrow$ Output	M	[0.35, 0.65]	[-0.47 $\pm$ 0.09]	[-0.26 $\pm$ 0.07]

Age→Output	W	[0.1, 0.4]	$[-0.08 \pm 0.09]$	$[0.11 \pm 0.09]$
BMI→Output	VW	[0, 0.25]	$[-0.2 \pm 0.09]$	$[-0.08 \pm 0.12]$
known CAD→Output	S	[0.5, 0.75]	$[0.47 \pm 0.04]$	$[0.29 \pm 0.11]$
previous AMI→Output	VW	[0, 0.25]	$[-0.12 \pm 0.13]$	$[0.03 \pm 0.19]$
previous PCI→Output	W	[0.1, 0.4]	$[-0.09 \pm 0.07]$	$[0.05 \pm 0.15]$
previous CABG→Output	W	[0.1, 0.4]	$[0.29 \pm 0.08]$	$[0.23 \pm 0.1]$
previous STROKE→Output	M	[0.35, 0.65]	$[-0.09 \pm 0.08]$	$[0.1 \pm 0.18]$
Diabetes→Output	S	[0.5, 0.75]	$[0.4 \pm 0.05]$	$[0.29 \pm 0.07]$
Smoking→Output	M	[0.35, 0.65]	$[0.09 \pm 0.1]$	$[0.06 \pm 0.12]$
Hypertension→Output	M	[0.35, 0.65]	$[0.11 \pm 0.08]$	$[-0.07 \pm 0.13]$
Dyslipidemia→Output	M	[0.35, 0.65]	$[0.08 \pm 0.08]$	$[0.15 \pm 0.06]$
Angiopathy→Output	M	[0.35, 0.65]	$[0.25 \pm 0.07]$	$[0.26 \pm 0.08]$
Chronic Kidney Disease→Output	W	[0.1, 0.4]	$[-0.03 \pm 0.1]$	$[-0.02 \pm 0.15]$
Family history of CAD→Output	W	[0.1, 0.4]	$[0.05 \pm 0.08]$	$[0.11 \pm 0.08]$
Asymptomatic→Output	-S	$[-0.75, -0.5]$	$[-0.08 \pm 0.1]$	$[-0.04 \pm 0.22]$
Atypical Symptoms→Output	VS	[0.75, 1]	$[-0.18 \pm 0.1]$	$[-0.06 \pm 0.2]$
Angina like→Output	S	[0.5, 0.75]	$[0.18 \pm 0.06]$	$[0.27 \pm 0.06]$
Dyspnoea on exertion→Output	M	[0.35, 0.65]	$[-0.16 \pm 0.08]$	$[-0.05 \pm 0.15]$
Incident of precordial pain→Output	S	[0.5, 0.75]	$[-0.09 \pm 0.11]$	$[0.12 \pm 0.17]$
Ecg→Output	M	[0.35, 0.65]	$[0.08 \pm 0.1]$	$[0.09 \pm 0.11]$
Expert Diagnosis Binary→Output	VS	[0.75, 1]	$[0.54 \pm 0.06]$	$[0.4 \pm 0.04]$

As shown in Table 5, the FCM-Encoder model with the proposed Spearman-based initialization learns a clinically coherent and highly interpretable structure, despite not using any expert knowledge during training. When compared with the expert linguistic values (included only for reference), the generated weights are closely aligned in both direction and relative magnitude. The strongest positive influence on the CAD output is given by Expert Diagnosis (0.54 ± 0.06), followed by known CAD (0.47 ± 0.04) and diabetes (0.40 ± 0.05), matching expert expectations that these are dominant risk drivers. Prior CABG also shows a positive effect (0.29 ± 0.08), while angiography (0.25 ± 0.07) contributes moderately, again consistent with clinical reasoning.

In contrast, the random initialization in the interval $[-1, 1]$ produces weights that are less compatible with expert knowledge, with more frequent sign inconsistencies and larger deviations across folds. Overall, the small standard deviations for the main risk determinants (e.g., known CAD, diabetes, expert diagnosis) indicate consistent learning behaviour across folds. Even without expert priors, the proposed Spearman-initialized FCM-Encoder generates a weight structure that closely matches expert reasoning, whereas random initialization does not perform adequately in terms of clinical coherence.

3.1.6. Clinical Interpretation of the Learned Imaging Features — CAD

While Section 3.1.5 analysed the FCM weights of the named clinical concepts, whose meaning is fixed before training, the eight imaging features extracted by the RGB-CNN are learned numerical activations rather than pre-defined medical attributes. To make their clinical content explicit, we carried out a complementary analysis at the end of training, correlating each learned imaging feature (Spearman, across all patients) with the named clinical variables. Expert diagnosis and prior known CAD status were excluded from this correlation, since both

are disease-status proxies, allowing the features to map to specific physiological attributes rather than to the diagnosis itself. This analysis is distinct from the initialization step, where the Spearman correlation between each concept and the diagnostic label sets the prior weights to the output node. Among the informative features, we report only those whose FCM weight sign agrees with the sign of their strongest clinical correlation, because when the learned weight and the correlation point in opposite directions, the label derived is ambiguous. Three features met this criterion and are reported in Table 6.

Table 6: Clinical interpretation of the informative learned image features (CAD dataset). For each image concept, the modality branch, mean FCM weight to the diagnostic output, the strongest clinical correlate (Spearman ρ , p -value), and the derived clinical label are reported. Only features with statistically significant correlations ($p < 0.01$) are shown.

Image feature	FCM weight to output	Top clinical correlate	Spearman ρ	p-value	Derived clinical label	
Img-1	+0.2	ECG	+0.4	<0.001	Perfusion-defect (abnormal-ECG, associated)	signature disease-
Img-6	-0.1	ECG	-0.4	<0.001	Normal-perfusion (normal-ECG)	signature
Img-3	-0.37	ECG	-0.4	<0.001	Normal-perfusion (normal-ECG)	signature

Across the informative features, the strongest clinical correlate is the ECG, and for the dominant detectors, the sign of this correlation matches the sign of the FCM weight. Img-1 carries the largest positive weight and correlates positively with an abnormal ECG ($\rho = +0.40$), so it acts as the main CAD-pattern detector and raises the predicted probability of disease when an ischemic perfusion pattern is present. Img-3 carries the most negative weight and correlates negatively with the ECG ($\rho = -0.40$), so it acts as the main normal-perfusion detector and lowers the predicted probability when perfusion is uniform. Img-6 follows the same direction as Img-3, with a negative ECG correlation and a smaller negative weight, serving as a secondary normal-perfusion detector that reinforces this signal. Img-1 and Img-3, which the FCM weighs most strongly and in opposite directions, capture the central contrast between ischemic and normal myocardial perfusion. Figure 2 illustrates this contrast using the eight highest-activation polar maps of each feature.

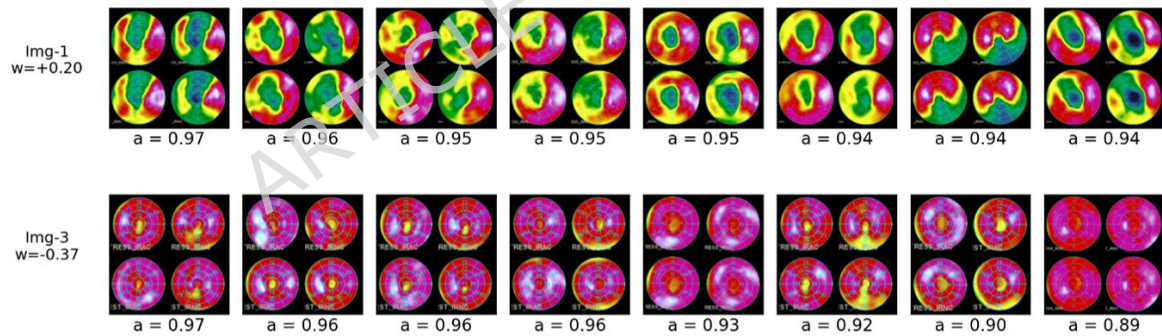


Figure 2: Top 8 polar maps for two contrasting imaging features learned. Img-1 (positive FCM weight) is activated by localized perfusion defects; Img-3 (negative FCM weight) is activated by uniform high perfusion.

The eight polar maps with the highest activation for Img-1 exhibit clear perfusion defects, with marked stress–rest mismatch patterns consistent with ischemia ($a \approx 0.94\text{--}0.97$), so Img-1 acts as a CAD-pattern detector that amplifies the diagnostic output when such defects are present. Conversely, the eight polar maps with the highest activation for Img-3 show homogeneous uptake without focal defects ($a \approx 0.89\text{--}0.97$), so Img-3 acts as a normal-pattern detector that suppresses the diagnostic output when uniform perfusion is detected. The opposite signs of their FCM weights to the output are therefore consistent with the visual content of the respective top-activation images, providing a clinically interpretable account of how the FCM integrates imaging information. The cross-fold stability analysis confirms that Img-1 and Img-3 are among the most stable informative features across the ten folds, supporting the robustness of the derived clinical labels.

3.1.7. FCM-Encoder Classification Results - SPN

Table 7 reports the classification performance of the proposed FCM-Encoder under the full multimodal configuration (clinical, CT, and PET features with Spearman-correlation-based initialisation), together with representative literature studies on the same source dataset covering clinical-only, imaging-only, and multimodal approaches. Results across alternative input settings of the proposed framework (clinical only, clinical + CT, clinical + PET) and the contribution of each architectural component are reported separately in the ablation study in Section 3.1.8 (Table 8).

Table 7: Comparison of the proposed FCM-Encoder with literature studies on the SPN dataset, grouped by input category (clinical, imaging, multimodal).

Models	Input	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
Literature studies (Clinical)								
Samaras et al. (2023) [69]	Clinical	94.33%	-	95.71%	92.96%	-	-	-
Apostolopoulos et al. (2024) [70]	Clinical	95.39%	-	96.88%	93.83%	-	95.57%	-
Literature studies (Imaging)								
Samaras et al. (2024) [71]	CT scans	92.3%	-	93.62%	90.91%	-	-	-
Samaras et al. (2024) [71]	PET scans	89%	-	80.85%	97.73%	-	-	-
Literature studies (Multimodal)								
Feleki et al. (2023) [31]	PET/clinical (3 features)	94.71%±2.99%	0.05	92.06%	96.98%	91.91%	-	-
Papandrianos et al. (2025) [46]	PET/clinical (9 features)	88.14%±3.8%	0.12	88.36%±5.23%	87.29%±7.48%	91.27%±5.28%	-	-
Papageorgiou et al. (2025) [47]	CT/clinical (9 features)	90.57%±3.07%	0.08±0.03	81.9%±10.56%	96%±5.47%	89.21%±5.06%	-	-
Proposed model (FCM-Encoder with Spearman initialization)								
FCM-Encoder	CT/PET/clinical (9 features)	93.39%±3.09%	0.06±0.003	91.84%±6.48%	94.78%±3.09%	95.12%±2.92%	93.29%±3.57%	96.71%±2.66%

In the clinical-only setting, Samaras et al. [69] reached 94.33% accuracy with a black-box AdaBoost classifier explained post-hoc via SHAP, while Apostolopoulos et al. [70] reported 95.39% by including the expert diagnosis as an input feature, which is not directly comparable to a purely data-driven setting.

For imaging-only configurations, Samaras et al. [71] obtained 92.30% on CT and 89.00% on PET with a YOLOv8-based pipeline, but without clinical fusion.

In the multimodal category, Feleki et al. [31] reached 94.71% ± 2.99% using only three input concepts that fuse modalities at the prediction level rather than at the concept-interaction level. Papandrianos et al. [46] and

Papageorgiou et al. [47] reported $88.14\% \pm 3.80\%$ and $90.57\% \pm 3.07\%$, respectively under comparable multimodal inputs.

The proposed FCM-Encoder reaches $93.39\% \pm 3.09\%$ accuracy and outperforms every fully data-driven literature study in Table 7. The bi-modal configurations of [46] (PET/clinical) and [47] (CT/clinical) are addressed within the proposed framework as ablations (Clinical + PET and Clinical + CT) in Table 8. The key methodological distinction with respect to the multimodal literature baselines is that the proposed model integrates the imaging-derived concepts and clinical variables at the concept-interaction level inside the FCM, with interconnections initialised by Spearman rank correlations on the training set rather than by expert-defined priors or specialised tuning stages.

3.1.8. Ablation Analysis of the FCM-Encoder Modules for SPN diagnosis

To isolate the contribution of each architectural element of the proposed FCM-Encoder, a systematic component ablation was conducted under the identical 10-fold cross-validation protocol, training/validation partitions, preprocessing pipeline, and RGB-CNN backbones, so that the only factor varying between rows of Table 8 is the component under test. Eight configurations were compared against the full proposed model: Clinical only (both image branches disabled); Clinical + CT and Clinical + PET (one imaging modality disabled in turn); Images only (CNN to Dense classifier), CT and Images only (CNN to Dense classifier), PET (clinical branch and FCM reasoning removed, predictions taken directly from the CNN sigmoid head); Clinical + Images + FCM, no PSO (random FCM initialisation, PSO refinement skipped); Clinical + Images + FCM, PSO (no Lévy mutation) (Lévy-flight escape step disabled); and No Spearman initialisation (PSO initialised from a zero-weight FCM rather than the Spearman-correlation prior).

Table 8: Component ablation of the proposed FCM-Encoder on the SPN dataset, reported as mean and standard deviation over 10-fold cross-validation. Each row disables or replaces a single module while keeping all other components identical to the Full proposed configuration.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
Clinical	88.39%±5.05%	0.12±0.05	89.07%±7.27%	87.86%±10.17%	89.52%±8.23%	88.78%±4.29%	91.99%±3.86%
Clinical+CT	92.84%±3.77%	0.07±0.04	91.26%±8.05%	94.11%±4.01%	94.64%±3.06%	92.69%±4.43%	96.85%±2.82%
Clinical+PET	92.84%±3.57%	0.07±0.03	90.86%±6.61%	94.86%±4.1%	92.77±3.91%	96.95%±2.1%	92.85%±2.85%
Images only (CNN→Dense classifier) CT	85%±8.7%	0.29±0.12	88.59%±7.33%	83.52%±11.89%	88.41%±7.08%	85.28%±7.63%	86.05%±6.78%
Images only (CNN→Dense classifier) PET	83.12%±6.43%	0.34±0.06	92.21%±6.18%	72.±9.21%	91.31%±5.75%	80.04%±5.63%	82.13%±4.26%
Clinical + Images + FCM, no PSO (random init)	91.47%±5.19%	0.33±0.1	89.63%±6.02%	93.14%±6.94%	93.55%±6.74%	91.41%±5.29%	93.65%±5.54%
Clinical + Images + FCM, PSO (no Lévy mutation)	92.85%±2.52%	0.28±0.05	91.37%±6.3%	94.3%±3.41%	94.65%±3.55%	92.27%±3.01%	95.63%±2.46%
No Spearman initialization	90.36%±5.16%	0.09±0.02	87.7%±8.98%	93.13%±5.87%	93.36%±5.22%	90.17%±5.86%	95.22%±2.75%
FCM-Encoder	93.39%±3.09%	0.06±0.003	91.84%±6.48%	94.78%±3.09%	95.12%±2.92%	93.29%±3.57%	96.71%±2.66%

The results confirm that every component contributes positively, with the full proposed model attaining the highest accuracy (93.39%) and the lowest loss (0.06). Using a single imaging modality together with the clinical data, or the images alone, lowers performance, with the images-only configuration showing the largest decline. Likewise, removing any of the optimisation components, the Spearman initialisation, the PSO refinement, or the Lévy-flight step, weakens the model by lowering accuracy or increasing the loss.

Overall, each component contributes a distinct improvement, and only their combination yields the highest accuracy and the lowest loss of the proposed framework.

3.1.9. Comparison with Alternative Image Encoder Architectures for SPN diagnosis

The same panel of alternative encoders introduced in Section 3.1.3 was evaluated for the SPN task while preserving the rest of the multimodal pipeline. Each encoder is instantiated twice per fold, once for the CT branch and once for the PET branch, so that only the backbone varies between rows of Table 9. The CNN training settings and the FCM-Encoder reasoning module hyperparameters are identical to those reported in Section 3.1.3, with three adjustments specific to the SPN task: the concept dimensionality is set to 8 rather than 16, the training is extended to 60 epochs, and the early-stopping patience is raised to 8. For BiomedCLIP, the 512-dimensional embeddings of each modality branch are projected to the same 8-dimensional concept space via PCA fitted on each training fold.

Table 9: Image encoder comparison on the SPN dataset, reported as mean $\pm$ standard deviation over 10-fold cross-validation. Only the encoder backbone varies between rows; the clinical pipeline, splits, and FCM-Encoder module are identical across all configurations.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC	Params
Lightweight architecture								
MobileNetV3-Small	92.28% $\pm$ 3.91%	0.28 $\pm$ 0.11	96.13% $\pm$ 4.54%	87.95% $\pm$ 10.17%	90.39% $\pm$ 7.72%	92.82% $\pm$ 3.24%	95.78% $\pm$ 3.11	2.5 M
EfficientNetB0	90.35% $\pm$ 5.88%	0.29 $\pm$ 0.1	95.72% $\pm$ 4.69%	84.74% $\pm$ 10.8%	87.48% $\pm$ 8.23%	91.16% $\pm$ 5.15%	95.41% $\pm$ 3.74%	5.3 M
DenseNet121	91.73% $\pm$ 4.53%	0.3 $\pm$ 0.18	90.77% $\pm$ 10.9%	92.63% $\pm$ 5.96%	93.23% $\pm$ 5.08%	91.46% $\pm$ 6.48%	94.98% $\pm$ 5.15%	8.1 M
Frozen ImageNet feature extractor								
EfficientNetB0	92.27% $\pm$ 3.91%	0.31 $\pm$ 0.16	93.41% $\pm$ 5.53%	90.32% $\pm$ 9.08%	92.28% $\pm$ 5.64%	92.63% $\pm$ 3.54%	95.35% $\pm$ 3.61%	5.3 M (frozen)
BiomedCLIP (ViT-B/16)	92.4% $\pm$ 3.07%	0.25 $\pm$ 0.06	95.88% $\pm$ 4.4%	91.13% $\pm$ 6.73%	92.07% $\pm$ 5.53%	93.74% $\pm$ 2.81%	96.74% $\pm$ 1.83%	86 M (frozen)
Proposed								
FCM-Encoder	93.39% $\pm$ 3.09%	0.06 $\pm$ 0.003	91.84% $\pm$ 6.48%	94.78% $\pm$ 3.09%	95.12% $\pm$ 2.92%	93.29% $\pm$ 3.57%	96.71% $\pm$ 2.66%	0.11 M

The proposed FCM-Encoder achieves the highest accuracy (93.39%) and the lowest loss (0.06) of all configurations. The alternative encoders fall between 90.35% and 92.40% accuracy and between 0.25 and 0.31 loss. The proposed model also delivers the highest specificity (94.78%) and precision (95.12%) together with the lowest cross-fold variance in accuracy ($\pm$ 3.09%), indicating fewer false positives and greater stability.

The slightly higher sensitivity of MobileNetV3-Small and BiomedCLIP reflects a different operating-point trade-off, favouring recall over specificity. Notably, BiomedCLIP (86 M parameters) reaches accuracy and AUC values that are statistically indistinguishable from those of the much smaller frozen EfficientNetB0 and from the proposed 0.11 M-parameter encoder, suggesting that its biomedical pretraining corpus, dominated by figures and radiographs, transfers poorly to CT and PET inputs.

Overall, the comparison demonstrated that FCM-Encoder matches or exceeds the accuracy of larger alternatives while producing substantially lower loss.

3.1.10. Comparison with Graph Neural Network Baselines for SPN diagnosis

The same three GNN baselines introduced in Section 3.1.4 (GCN [66], GAT [67], and GraphSAGE [68]) were evaluated for the SPN task while preserving the rest of the multimodal pipeline. All baselines operate on the same feature stream, the CT-CNN and PET-CNN concept vectors concatenated with the clinical variables, and use the same architectural and training settings described in Section 3.1.4 (two-layer architecture, 24 hidden units, dropout 0.6, Adam at 1×10^{-2} , weight decay 2×10^{-3} , 150 epochs with early stopping). The inductive five-nearest-neighbour population graph and the 10-fold cross-validation protocol are also identical, so that only the classifier varies between rows of Table 10.

Table 10: Comparison with graph neural network baselines on the SPN dataset (mean $\pm$ standard deviation over 10-fold cross-validation). The three graph baselines (GCN, GAT, GraphSAGE) use the same fused feature stream (clinical variables and RGB-CNN concepts) and the same inductive k -nearest-neighbours population graph, differing only in the graph neural network architecture; the FCM-Encoder is the proposed method.

Models	Accuracy	Loss	Sensitivity	Specificity	Precision	F1-score	AUC
GCN (Kipf & Welling 2017) [66]	90.92% $\pm$ 4.5%	0.26 $\pm$ 0.08	92.43% $\pm$ 6.35%	89.49% $\pm$ 7.42%	90.36% $\pm$ 6.56%	91.12% $\pm$ 4.45%	96.03% $\pm$ 2.08%
GAT (Veličković 2018) [67]	90.93% $\pm$ 4.97%	0.29 $\pm$ 0.11	88.53% $\pm$ 8.25%	93.21% $\pm$ 5.99%	93.44% $\pm$ 6%	90.66% $\pm$ 5.54%	94.21% $\pm$ 3.98%
GraphSAGE (Hamilton 2017) [68]	92.32% $\pm$ 2.95%	0.24 $\pm$ 0.07	94.45% $\pm$ 7.76%	89.77% $\pm$ 4.59%	91.03% $\pm$ 4.44%	92.39% $\pm$ 3.56%	96.7% $\pm$ 2.35%
FCM-Encoder	93.39% $\pm$ 3.09%	0.06 $\pm$ 0.003	91.84% $\pm$ 6.48%	94.78% $\pm$ 3.09%	95.12% $\pm$ 2.92%	93.29% $\pm$ 3.57%	96.71% $\pm$ 2.66%

The proposed FCM-Encoder achieves the highest accuracy (93.39%), specificity (94.78%), precision (95.12%), and F1-score (93.29%), exceeding GCN, GAT, and GraphSAGE by 2.47, 2.46, and 1.07 accuracy points, respectively. AUC is statistically tied with GraphSAGE (96.71% vs. 96.70%), and GCN and GraphSAGE show marginally higher sensitivity, in both cases within one standard deviation.

The most pronounced difference between the proposed model and the GNN baselines is observed in the binary cross-entropy loss: the FCM-Encoder reaches 0.06 against 0.24–0.29 for the three baselines, which indicates that the proposed model produces better results than any of the graph-based alternatives.

3.1.11. Analysis of FCM-Encoder clinical weights- SPN

Table 11 presents the FCM-Encoder weight values learned from the SPN data using two initialization strategies: (i) the proposed Spearman-based initialization and (ii) fully random initialization in the range $[-1,1]$. For each interconnection, the mean weight and its standard deviation across all cross-validation folds are reported. Expert linguistic weights and their transformation to numerical intervals are included only for comparison and were not used as priors or constraints during training.

Table 11: Demonstration of FCM-Encoder-generated weighted values applied in SPN data

Interconnection	Linguistic values provided by experts	Transformed to a spectrum of values	FCM-Encoder with Spearman initialization	FCM-Encoder random initialization
Gender $\rightarrow$ Output	W	[0.1, 0.4]	[-0.02 $\pm$ 0.09]	[-0.07 $\pm$ 0.18]
Age $\rightarrow$ Output	M	[0.35, 0.65]	[0.04 $\pm$ 0.08]	[-0.2 $\pm$ 0.2]
BMI $\rightarrow$ Output	W	[0.1, 0.4]	[-0.09 $\pm$ 0.07]	[-0.32 $\pm$ 0.11]
GLU $\rightarrow$ Output	VW	[0, 0.25]	[0.21 $\pm$ 0.12]	[0.2 $\pm$ 0.12]
SUVmax $\rightarrow$ Output	VS	[0.75, 1]	[0.67 $\pm$ 0.12]	[0.55 $\pm$ 0.04]
Diameter $\rightarrow$ Output	S	[0.55, 0.85]	[0.59 $\pm$ 0.05]	[0.55 $\pm$ 0.06]

Location_Left_Upper_Lobe→Output	M	[0.35, 0.65]	[0.3±0.1]	[0.2±0.26]
Location_Right_Upper_Lobe→Output	S	[0.55, 0.85]	[-0.21±0.08]	[-0.11±0.28]
Location_Left_Lower_Lobe→Output	W	[0.1, 0.4]	[-0.009±0.14]	[0.01±0.2]
Location_Right_Lower_Lobe→Output	W	[0.1, 0.4]	[0.11±0.09]	[0.29±0.27]
Location_Middle→Output	M	[0.35, 0.65]	[-0.14±0.13]	[-0.19±0.31]
Location_Lingula→Output	W	[0.1, 0.4]	[-0.02±0.13]	[0.13±0.28]
Type_Solid→Output	-M	[-0.65, -0.35]	[0.34±0.11]	[0.36±0.17]
Type_Ground_Class→Output	VS	[0.75, 1]	[-0.25±0.09]	[-0.2±0.2]
Type_Consolidated→Output	S	[0.55, 0.85]	[0.22±0.09]	[0.28±0.22]
Type_Speckled→Output	-M	[-0.65, -0.35]	[-0.38±0.07]	[-0.39±0.14]
Type_Semi_Solid→Output	M	[0.35, 0.65]	[0.01±0.11]	[0.12±0.23]
Type_Calcified→Output	-M	[-0.65, -0.35]	[-0.16±0.06]	[-0.29±0.24]
Type_Cavitory→Output	M	[0.35, 0.65]	[0.21±0.09]	[0.25±0.18]
Margins_Spiculated→Output	S	[0.55, 0.85]	[0.47±0.04]	[-0.05±0.07]
Margins_Lobulated→Output	M	[0.35, 0.65]	[0.44±0.1]	[0.52±0.1]
Margins_Well_Defined→Output	-M	[-0.65, -0.35]	[-0.64±0.05]	[-0.53±0.04]
Margins_Ill_Defined→Output	M	[0.35, 0.65]	[-0.03±0.008]	[-0.04±0.25]

As shown in Table 11, the proposed FCM-Encoder with Spearman-based initialization produces weight patterns that are largely consistent with expert linguistic assessments, even though expert knowledge was not used in training. For example, SUVmax is assigned a strong positive weight, compatible with a “very strong” expert evaluation, thus validating the model’s capability to strongly correlate malignancy to high radiotracer uptake. Similarly, diameter and margin-related features, such as spiculation and lobulation, are assigned positive weights, in agreement with expert expectations and established clinical practice. These connections also exhibit relatively small standard deviations, indicating consistent learning behaviour across folds.

In contrast, the random initialization in $[-1,1]$ yields less consistent and clinically inaccurate weights. Several interconnections show weaker magnitudes, sign changes, or much larger deviations across folds, indicating reduced interpretability. For instance, some location and margin-related features display inconsistent directions compared with both expert knowledge and the Spearman-initialized model. Overall, the proposed Spearman-based initialization enables FCM-Encoder to converge toward expert-like, interpretable interconnections, whereas random initialization performs more poorly in terms of clinical coherence.

3.1.12. Clinical Interpretation of the Learned Imaging Features — SPN

As in Section 3.1.6 for CAD, we assessed the clinical meaning of the learned SPN image features by correlating each feature’s activation with the named clinical variables across all patients (Spearman). Each imaging branch contributes eight learned features, sixteen in total, and all were examined. A feature is labelled only when its activation is significantly correlated with a named clinical variable ($p < 0.01$) and the sign of that correlation matches the sign of its FCM weight, so that its clinical association and its diagnostic role agree; the remaining features were left unlabelled rather than assigned a potentially misleading description. Seven features satisfy both criteria, five from the CT branch and two from the PET branch and are reported in Table 12 with their strongest clinical correlate and derived label.

Table 12: Clinical interpretation of the informative learned image features (SPN dataset). For each image concept, the modality branch, mean FCM weight to the diagnostic output, the strongest clinical correlate (Spearman ρ , p -value), and the derived clinical label are reported. Only features with statistically significant correlations ($p < 0.01$) are shown.

Image feature	Modality	FCM weight	Top clinical correlate	Spearman ρ	P-value	Derived clinical label
CT-8	CT	+0.12	Nodule diameter	+0.32	<0.001	Large-nodule (malignancy-associated) signature
CT-6	CT	+0.07	Nodule diameter	+0.34	<0.001	Large nodule signature
CT-4	CT	+0.04	Nodule diameter	+0.37	<0.001	Large nodule signature
CT-3	CT	-0.08	Nodule diameter	-0.29	<0.001	Small-nodule (benign-associated) signature
CT-2	CT	-0.07	Nodule diameter	-0.3	<0.001	Small-nodule signature
PET-1	PET	+0.24	SUV uptake (FDG uptake)	+0.41	<0.001	High-uptake (malignancy-associated) signature
PET-2	PET	-0.12	SUV uptake (FDG uptake)	-0.46	<0.001	Low-uptake (benign-associated) signature

Table 12 shows a clear division between the two modalities: the CT concepts key on nodule diameter (lesion size), whereas the PET concepts key on FDG uptake (metabolic activity). The sign of each FCM weight is consistent with this reading: positive-weight concepts correspond to larger or more avid, malignancy-associated lesions, and negative-weight concepts to smaller or less avid, benign lesions. The learned imaging concepts therefore recover two complementary and clinically established malignancy markers, nodule size and metabolic activity, rather than redundant or opaque features.

Figure 3 provides visual confirmation for the most influential positive- and negative-weight concept of each branch.

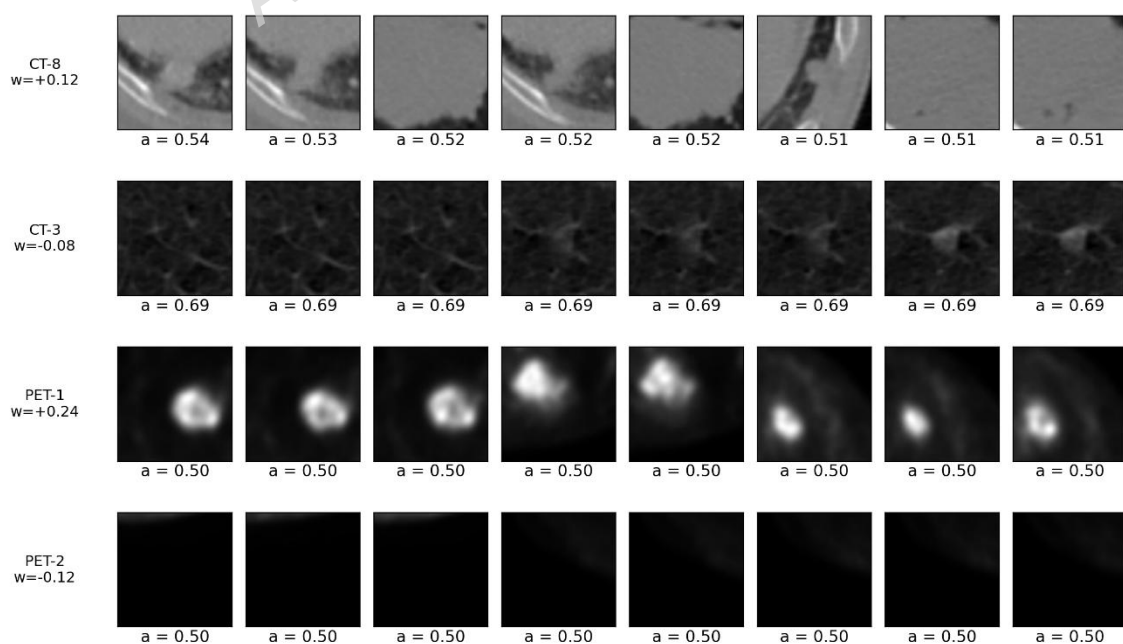


Figure 3: Eight highest-activation patches for the four most influential SPN image features, identified from the learned FCM-Encoder weights. For each modality, the top row is the strongest positive-weight feature (CT-8, PET-1) and the bottom row the strongest negative-weight feature (CT-3, PET-2), with the learned weight w indicated. CT-8 activates on large soft-tissue lesions and CT-3 on small low-density nodules; PET-1 activates on lesions with intense focal FDG uptake and PET-2 on lesions with minimal or absent uptake.

For each modality, Figure 3 shows the eight highest-activation patches of its most influential positive-weight and most influential negative-weight image feature, identified directly from the learned FCM-Encoder weights. The patches confirm that these features capture clinically coherent morphology rather than arbitrary texture.

In the CT branch, the positive-weight feature CT-8 ($w = +0.12$) responds to large, soft-tissue–dense lesions, consistent with its positive correlation with nodule diameter and with the higher malignancy risk associated with larger nodules, whereas the negative-weight feature CT-3 ($w = -0.08$) responds to small, low-density nodules, a pattern typical of benign findings.

In the PET branch, the positive-weight feature PET-1 ($w = +0.24$) activates on lesions with intense focal FDG uptake (SUV), a recognised hallmark of malignancy, while the negative-weight feature PET-2 ($w = -0.12$) activates on lesions with minimal or absent uptake, characteristic of benign nodules.

3.2. Natural Language Generation

In this section, we report the results of the LLM-assisted clinical narrative generation framework to interpret the proposed FCM-Encoder model and RGB-CNN decision-making process. The FCM-Encoder provides diagnostic predictions for both CAD (using clinical variables and polar maps) and SPN diagnosis (using clinical variables together with CT and PET scans), while the RGB-CNN is used exclusively for imaging-based inference (polar maps for CAD and CT/PET scans for SPN). In addition to classification outputs, the FCM-Encoder produces interconnections among concepts for each diagnostic setting.

Qwen-3VL is employed as a vision–language model to transform these quantitative outputs into structured, clinically readable narratives. Specifically, the generated FCM interconnections and activated concepts are used to identify the most influential clinical contributors for each patient, based on their relative activation strength and relational importance within the cognitive graph. These contributors are then described textually, without altering the underlying model predictions. In parallel, Qwen-3VL interprets imaging inputs, polar maps in the CAD task, and CT/PET scans in the SPN task by translating salient visual patterns identified by the RGB-CNN into qualitative radiological or metabolic descriptors.

3.2.1. NLG application for CAD diagnosis

First, we demonstrate the application of Qwen3-VL to CAD diagnosis by interpreting the decision outputs of the proposed FCM-Encoder and the RGB-CNN. The prompt provided to Qwen3-VL is reported in the Appendix. In the representative case, the patient presented with known CAD, angina-like symptoms, diabetes, chronic hypertension, active smoking status, angiopathy, and dyspnea on exertion. Based on the integrated outputs of the FCM-Encoder (using clinical variables and polar maps) and the RGB-CNN (using polar-map representations), the case was consistently classified as pathological.

Overall Summary

Both the fuzzy cognitive mapping (FCM) and deep learning-based RGB-CNN model classified the patient's cardiovascular status as **pathological**, with concordant findings across clinical and imaging modalities. The agreement between these methodologies, leveraging complementary diagnostic frameworks, substantially reinforces confidence in the presence of **severe coronary artery disease (CAD)**. This convergence reduces diagnostic uncertainty and supports targeted therapeutic intervention.

Most Important Clinical Contributors

- **Established coronary artery disease (CAD):** Reflects prior atherosclerotic damage, significantly elevating CAD risk through chronic coronary stenoses and fixed obstructions, contributing to irreversible myocardial ischemia.
- **Diabetes mellitus:** Accelerates atherosclerosis via endothelial dysfunction and hyperglycemia-induced plaque vulnerability, increasing the likelihood of progressive coronary obstruction and fixed ischemia.
- **Smoking:** Promotes chronic inflammation and oxidative stress, accelerating atherosclerotic progression and heightening the risk of coronary stenosis through enhanced plaque development.
- **Chronic hypertension:** Induces sustained mechanical stress on coronary arteries, facilitating plaque formation and progression, thereby worsening the severity of coronary obstructions.
- **Microvascular complications (angiopathy):** Impair myocardial perfusion due to chronic vascular damage, increasing ischemic susceptibility and supporting fixed perfusion defects on imaging.
- **Exertional dyspnea (dyspnoea on exertion):** Reflects impaired myocardial blood flow during increased cardiac demand, directly attributable to fixed coronary obstructions and chronic ischemia.

RGB-CNN Interpretation

The polar maps reveal **significant fixed perfusion defects** in the **right coronary artery (RCA) territory**, corresponding to the inferior wall, and the **left anterior descending artery (LAD) territory**, affecting the anterior septum. These defects exhibit **no reversible component**, indicating chronic ischemic injury resulting from irreversible coronary stenoses.

Specifically, the right inferior and anterior regions demonstrate persistently reduced perfusion, consistent with severe RCA and LAD obstruction. This pattern of non-reversible defects is highly suggestive of **advanced chronic coronary disease**, directly supporting the pathological classification and reinforcing the need for **urgent revascularization assessment**.

Figure 4: Application of Qwen to CAD data to justify the decision-making process of FCM-Encoder and RGB-CNN

As shown in Figure 4, Qwen3-VL translates the diagnostic outputs of the FCM-Encoder and the RGB-CNN into a clear clinical explanation. It identifies the main factors specific to each patient that contribute to CAD. The generated narrative highlights known risk factors captured by the activated FCM concepts, while also interpreting the myocardial perfusion polar map produced by the RGB-CNN. The model accurately describes fixed perfusion defects in the right coronary artery (RCA) and left anterior descending (LAD) territories, which are consistent with chronic, non-reversible ischemia, and the explanation is in accordance with clinical experts.

3.2.2. NLG application for SPN identification

We further evaluated the LLM-assisted Clinical Narrative Generation module (Qwen3-VL) on the SPN dataset to determine its ability to produce radiologically interpretable narratives from multimodal clinical and imaging data. In this case, the system received clinical variables, CT and PET scans, and the FCM-Encoder-weighted concept interconnections, along with the activated concepts. In the representative case, the patient had a SPN measuring 2.7 cm, located in the left upper lobe, with a consolidated component, spiculated margins, and a SUV of 25.5.

Overall Summary

Both the FCM-Encoder and the RGB-CNN classified this case as malignant. The agreement between clinical reasoning integrating metabolic activity, nodular morphology, and anatomical location and image-based reasoning leveraging radiographic and metabolic signatures provides strong diagnostic confidence for a malignant solitary pulmonary nodule.

Most Important Clinical Contributors

The following clinical and morphological factors were identified as key contributors to malignancy probability, based on positive concept weights and patient-specific findings:

- **High standardized uptake value (SUV)** reflects increased metabolic activity characteristic of aggressive tumor proliferation.
- **Large nodule diameter** corresponds to increased tumor burden and invasive potential.
- **Upper-lobe location (left lung)** is associated with higher malignancy risk due to favorable vascular and lymphatic pathways.
- **Lobulated margins** indicate heterogeneous growth patterns commonly observed in malignant nodules.
- **Consolidated nodule morphology** suggests infiltrative growth into surrounding lung parenchyma.

CT-Based Findings

The CT scan demonstrates a solitary pulmonary nodule (SPN) exhibiting characteristics suggestive of malignant potential. Key imaging features include:

- **Increased tissue density**, reflecting abnormal cellular proliferation and deviation from normal lung parenchyma.
- **Textural heterogeneity**, observed as variation in grayscale intensity, associated with malignant lesions.
- **Irregular margins**, suggesting invasive growth behavior and potential extension beyond the primary lesion.
- **Associated vascular changes**, potentially reflecting increased contrast enhancement and angiogenesis driven by tumor metabolic demands.

PET-Based Findings

The PET scan reveals a substantial area of increased radiotracer uptake concentrated within the pulmonary nodule. This uptake pattern is highly suggestive of elevated metabolic activity, a hallmark of malignant tissue, and reflects increased cellular demand for glucose associated with rapid tumor proliferation.

Key metabolic findings include:

- **Intense focal tracer uptake** within the nodule, exceeding that typically observed in normal lung parenchyma or benign pulmonary nodules.
- **High metabolic burden**, indicating aggressive cellular activity and elevated energy consumption.
- **Elevated signal-to-noise ratio** in the region of interest, reinforcing the presence of localized, intense metabolic demand.

Overall, the observed radiotracer distribution and intensity are consistent with the aggressive growth potential commonly associated with Non-Small Cell Lung Cancer (NSCLC), supporting the need for further diagnostic investigation.

Figure 5: Application of Qwen3-VL to SPN data to justify the decision-making process of FCM-Encoder and RGB-CNN

Figure 5 illustrates the application of Qwen3-VL to the outputs of the FCM-Encoder and the RGB-CNN for SPN diagnosis. The output response provides an example of how patient-specific concept activation of FCM and learned cross-concept connections are transformed into a structured clinical summary, emphasizing the most influential clinical contributions to the probability of malignancy. Concurrently, imaging data from the provided CT/ PET scans for this case is translated into CT-based and PET-based findings that represent salient radiographic and metabolic attributes of malignant pulmonary nodules.

4. Discussion

This study introduced the FCM-Encoder framework as a multimodal, interpretable model for CAD and SPN diagnosis. Imaging data are processed through modality-specific RGB-CNN encoders and combined with normalised clinical concept activations, representing both data streams in a unified conceptual space. Spearman rank correlations between each input concept and the diagnostic label were used to derive data-driven priors on the concept-to-output connections, while the concept-to-concept connections were initialised at zero so that their influence could emerge through optimisation. The FCM weights were then updated by a hybrid PSO scheme enriched with Lévy-flight exploration to escape local minima, gradient-based local adjustments, and sparsity-aware regularisation.

Relative to existing multimodal CAD and SPN studies, the proposed FCM-Encoder overcomes key limitations of conventional classical machine-learning approaches. Prior frameworks typically rely on decision- or feature-level concatenation, which yields restricted interpretability. DeepFCM-based methods introduced reasoning and improved transparency but often depended on expert-assigned weights, and the fusion relied on CNN predictions rather than imaging representations. In contrast, the FCM-Encoder integrates deep imaging representations and clinical variables within a unified concept-level space, using data-driven, sign-coherent clinical initialisation and neutral imaging concepts whose influence emerges through optimisation. This design delivers competitive diagnostic performance and clinically interpretable multimodal reasoning, addressing the transparency and robustness gaps observed in prior literature. Across the two diagnostic tasks, the FCM-Encoder reached $81.82\% \pm 4.51\%$ accuracy for CAD with clinical data and polar maps, and $93.39\% \pm 3.09\%$ accuracy for SPN characterisation with clinical data and CT/PET scans, assessed by 10-fold cross-validation.

The ablation studies confirmed that the performance of the FCM-Encoder is the product of the combination of its components rather than the dominance of any single one. Restricting the model to the clinical branch or to the imaging branch in isolation produced the largest reductions in accuracy and loss on both datasets, which indicates that the two modalities encode complementary rather than redundant diagnostic information and that the FCM reasoning layer is essential to integrate them. The removal of the optimisation components produced the following pattern: disabling the PSO stage and relying on a randomly initialised FCM yielded the highest cross-entropy loss of any configuration despite a comparable accuracy, which demonstrates that PSO contributes less to classification correctness than to the calibration of the resulting probability estimates. Disabling only the Lévy-flight escape step within PSO produced a smaller loss, indicating that the global-search component of the

optimiser contributes meaningfully when the loss surface contains local minima that gradient-style refinement alone would not escape. Replacing the Spearman correlation prior with a zero-weight initialisation lowered accuracy and increased loss, which suggests that the prior shapes the region of the weight space. Taken together, the configurations show that the full model attains its performance because these contributions are complementary.

The image-encoder comparison further supported the choice of the RGB-CNN. The FCM-Encoder paired with the RGB-CNN outperformed both end-to-end lightweight ImageNet-pretrained CNNs and frozen feature extractors, while using substantially fewer parameters per modality branch, which indicates that the advantage does not derive from encoder capacity but from the alignment between the visual representation and the reasoning module that consumes it. Lightweight architectures such as MobileNetV3-Small and EfficientNetB0 are optimised for general-purpose natural-image classification and produce embeddings whose internal structure reflects the statistics of ImageNet rather than those of nuclear medicine imagery, with the result that any subsequent classifier must compensate for this mismatch through additional fine-tuning. Frozen feature extractors are subject to the same limitation in a more pronounced form, since their representations are fixed and cannot adapt to the target modality at all. The RGB-CNN, by contrast, is trained directly on the target imagery and produces concept-level features, attaining the lowest loss of all encoders examined.

The FCM-Encoder was benchmarked against representative graph neural network baselines (GCN, GAT, and GraphSAGE) operating on the same concept-level feature stream. Across both datasets, the FCM-Encoder matched or exceeded the best graph baseline in accuracy and AUC while remaining competitive across sensitivity, specificity, precision, and F1-score, which indicates that the advantage arises from the manner in which inter-concept influence is modelled. Graph neural networks propagate information along edges that are fixed prior to training, and their reasoning is therefore constrained by the quality of the pre-specified similarity graph and by the assumption that patients proximate in concept space should also influence one another's predictions. The FCM-Encoder does not rely on this assumption: it learns directed, signed influences between the concepts themselves, which constitutes a more faithful representation of clinical reasoning. The most pronounced difference was in loss, where the FCM-Encoder achieved a substantially lower cross-entropy than every graph baseline, which is consistent with the PSO stage optimization. A further practical benefit is that the FCM learns its own connectivity during optimisation, removing the need to pre-specify a population graph and the associated hyperparameter tuning required by graph-based models.

The interpretability of the FCM-Encoder is supported by several complementary mechanisms. The Spearman-based initialisation generated clinically coherent positive and negative pathways consistent with established risk patterns across both disease domains, whereas random initialisation within $[-1, 1]$ produced trainable but clinically incoherent interconnections. Moreover, a complementary interpretability mechanism is provided by the analysis of the learned imaging features, which in both tasks could be linked to recognised clinical findings. For CAD, the Spearman labelling procedure assigned each informative polar-map concept a named clinical attribute, and visual inspection of the most strongly activated images confirmed that the signed FCM weights are consistent with the underlying clinical content: the positive-weight concept responds to perfusion-defect patterns indicative of disease, whereas the negative-weight concept responds to normal-perfusion patterns. For SPN, the two imaging branches captured separate diagnostic features established in clinical practice for malignancy assessment: the CT concepts encoded lesion size (nodule diameter), with positive-weight concepts corresponding to larger, malignancy-prone nodules and negative-weight concepts to smaller, benign nodules, while the PET concepts encoded metabolic activity (FDG uptake), with positive-weight concepts corresponding to high-uptake, malignancy-associated lesions and negative-weight concepts to low-uptake, benign lesions. The learned image concepts could therefore be tied to clinically meaningful markers in both applications, reinforcing the transparency of the framework.

Finally, the Qwen3-VL model translated the FCM interconnections, clinical attributes, and imaging evidence into structured, clinician-readable narratives without altering or inferring diagnostic decisions, contextualising rather than replacing the FCM-Encoder and RGB-CNN predictions. The combined use of interpreted interconnections among concepts and vision-language explanation establishes a multi-perspective interpretability framework that enhances transparency and supports clinical validation.

While the framework demonstrates strong interpretability and competitive diagnostic performance, several methodological considerations should be noted. First, the proposed multimodal optimisation procedure requires training of two modality-specific CNN encoders and performing population-based PSO weight search within

each cross-validation fold. Although this design enables fully data-driven multimodal concept integration, it increases runtime compared with single-stage gradient-trained classifiers and may restrict applicability in settings where computational resources are limited. Second, the experimental evaluation is confined to a defined set of datasets and diagnostic tasks, selected to enable a focused and systematic analysis of the proposed methodology. Extending the evaluation to additional datasets and clinical scenarios represents a natural progression of this work and would better support the generalisation claims of the framework. As future work, all developed components, including the FCM-Encoder, optimisation strategies, explainability and narrative-generation modules, will be consolidated into a unified software package, enabling systematic reuse, extensibility, and consistent evaluation across a broader range of multimodal diagnostic tasks.

5. Conclusions

The proposed FCM-Encoder framework introduces a novel and interpretable approach to multimodal medical decision support for CAD diagnosis and SPN characterisation. By unifying modality-specific imaging encoders with structured clinical variables inside an FCM-based cognitive reasoning model, the method achieves competitive diagnostic performance while maintaining explicit structural transparency throughout the decision process. The FCM weight matrix is initialised through a Spearman correlation-guided strategy, aligning each concept's initial influence on the diagnostic output with the associations found in the training data, and is then refined through a hybrid PSO scheme enriched with Lévy-flight exploration. Systematic ablation, image-encoder, and graph neural network comparisons confirmed three findings. First, each architectural element of the framework contributes positively to performance. Second, the RGB-CNN encoder matches or exceeds substantially larger pretrained alternatives. Third, the FCM-based reasoning module outperforms representative graph-based architectures while preserving signed, concept-level interpretability. A complementary correlation analysis further assigned a derived clinical label to each informative learned image concept, ensuring that the imaging-derived nodes of the FCM carry a meaningful interpretation on the same footing as the named clinical variables. The integration of concept-level interaction reasoning, weight-based saliency interpretation, and vision-language narrative generation via Qwen3-VL provides clinicians with clinically meaningful insight into both imaging evidence and multimodal concept dependencies. Given the critical need for reliability, transparency, and interpretability in medical AI, the FCM-Encoder constitutes a reliable and practically relevant infrastructure for multimodal diagnostic reasoning in healthcare.

Declarations

Abbreviations

The following table (Table 13) lists the abbreviations that are used in this manuscript:

Table 13: List of abbreviations and acronyms

Acronym	Description
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
PSO	Particle Swarm Optimization
FCM	Fuzzy Cognitive Map
BMI	Body Mass Index
GLU	Glucose Level
SUVmax	Standardized Uptake Value
NSCLC	Non-small cell Lung Cancer
SCC	Squamous cell carcinoma

ADC	Adenocarcinoma
CT	Computed Tomography
PET	Positron Emission Tomography
CNN	Convolutional Neural Network
MDSS	Medical Decision Support System
ELM	Extreme Learning Machine
Grad-CAM	Gradient-weighted Class Activation Mapping
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
NLG	Natural Language Generation
GPT-API	Generative Pre-trained Transformer - Application Programming Interface
AUC	Area Under Curve
ROC	Receiver operating characteristic
GNN	Graph Neural Networks
GCN	Graph Convolutional Networks
RELU	Rectified Linear Unit
LIDC-IDRI	Lung Image Database Consortium and Image Database Resource Initiative
BN	Batch Normalization
TCIA	Cancer Imaging Archive
MAE	Mean Absolute Error
MSE	Mean Squared Error
LLM	Large language model
RNN	Recurrent Neural Network
VW	Very Weak
W	Weak
M	Medium
S	Strong
VS	Very Strong
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
GPU	Graphics Processing Unit
GB	Gigabyte
RAM	Random Access unit
H	Hertz
AMD	Advanced Micro Devices

Ethics approval and consent to participate

All procedures in this study were in accordance with the Declaration of Helsinki. This study was approved on the 3rd of March 2022 by the ethical committee of the University General Hospital of Patras (Ethical & Research Committee of University Hospital of Patras—protocol number 108/10–3–2022). The requirement to obtain informed consent was waived by the director of the diagnostic center due to its retrospective nature.

Consent for publication

Not applicable. The study used anonymized retrospective data with no identifiable personal information.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to ethical and legal restrictions associated with the use of anonymized clinical data.

Competing interests

The authors declare no competing interests.

Funding: No funding.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

Design and conduct of the study (A.F., N.I.P.); model development (A.F.); data collection (I.A., N.P.); data processing (A.F., I.A.); Validation (I.A., N.I.P.); manuscript preparation (A.F., N.I.P.); manuscript review (E.P., N.I.P., I.A.); medical expertise (D.A., N.P., N.I.P.), Supervision (N.I.P.).

Acknowledgements

Not applicable.

References

- Shahjehan, R.D., Sharma, S., Bhutta, B.S.: Coronary Artery Disease. In: StatPearls. StatPearls Publishing, Treasure Island (FL) (2025)
- Channon, K.M., Newby, D.E., Nicol, E.D., Deanfield, J.: Cardiovascular computed tomography imaging for coronary artery disease risk: plaque, flow and fat. *Heart*. 108, 1510–1515 (2022). <https://doi.org/10.1136/heartjnl-2021-320265>
- Papandrianos, N.I., Feleki, A., Papageorgiou, E.I., Martini, C.: Deep Learning-Based Automated Diagnosis for Coronary Artery Disease Using SPECT-MPI Images. *Journal of Clinical Medicine*. 11, 3918 (2022). <https://doi.org/10.3390/jcm11133918>
- Lai, Y.-H., Chen, W.-N., Hsu, T.-C., Lin, C., Tsao, Y., Wu, S.: Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Sci Rep*. 10, 4679 (2020). <https://doi.org/10.1038/s41598-020-61588-w>
- Han, Y., Ma, Y., Wu, Z., Zhang, F., Zheng, D., Liu, X., Tao, L., Liang, Z., Yang, Z., Li, X., Huang, J., Guo, X.: Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur J Nucl Med Mol Imaging*. 48, 350–360 (2021). <https://doi.org/10.1007/s00259-020-04771-5>
- Papandrianos, N.I., Apostolopoulos, I.D., Feleki, A., Apostolopoulos, D.J., Papageorgiou, E.I.: Deep learning exploration for SPECT MPI polar map images classification in coronary artery disease. *Ann Nucl Med*. 36, 823–833 (2022). <https://doi.org/10.1007/s12149-022-01762-4>
- Betancur, J., Commandeur, F., Motlagh, M., Sharir, T., Einstein, A.J., Bokhari, S., Fish, M.B., Ruddy, T.D., Kaufmann, P., Sinusas, A.J., Miller, E.J., Bateman, T.M., Dorbala, S., Di Carli, M., Germano, G., Otaki, Y., Tamarappoo, B.K., Dey, D., Berman, D.S., Slomka, P.J.: Deep Learning for Prediction of Obstructive Disease from Fast Myocardial Perfusion SPECT: A Multicenter Study. *JACC Cardiovasc Imaging*. 11, 1654–1663 (2018). <https://doi.org/10.1016/j.jcmg.2018.01.020>
- Betancur, J., Otaki, Y., Motwani, M., Fish, M.B., Lemley, M., Dey, D., Gransar, H., Tamarappoo, B., Germano, G., Sharir, T., Berman, D.S., Slomka, P.J.: Prognostic Value of Combined Clinical and Myocardial Perfusion Imaging Data Using Machine Learning. *JACC: Cardiovascular Imaging*. 11, 1000–1009 (2018). <https://doi.org/10.1016/j.jcmg.2017.07.024>
- Apostolopoulos, I.D., Pintelas, E.G., Livieris, I.E., Apostolopoulos, D.J., Papathanasiou, N.D., Pintelas, P.E., Panayiotakis, G.S.: Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. *Med Biol Eng Comput*. 59, 1299–1310 (2021). <https://doi.org/10.1007/s11517-021-02378-y>
- Wen, Q., Yue, Y., Shang, J., Lu, X., Gao, L., Hou, Y.: The application of dual-layer spectral detector computed tomography in solitary pulmonary nodule identification. *Quant Imaging Med Surg*. 11, 521–532 (2021). <https://doi.org/10.21037/qims-20-2>
- Alves, V.M., dos Santos Cardoso, J., Gama, J.: Classification of Pulmonary Nodules in 2-[18F]FDG PET/CT Images with a 3D Convolutional Neural Network. *Nucl Med Mol Imaging*. 58, 9–24 (2024). <https://doi.org/10.1007/s13139-023-00821-6>
- Zhang, L., Jin, X., Zheng, Y., Ye, X.: Analysis of the imaging and clinical features of subsolid pulmonary nodules in stage IA non-small cell lung cancer. *Inflammation and Cell Signaling*. 10, (2023)
- Wang, H., Li, Y., Han, J., Lin, Q., Zhao, L., Li, Q., Zhao, J., Li, H., Wang, Y., Hu, C.: A machine learning-based PET/CT model for automatic diagnosis of early-stage lung cancer. *Front Oncol*. 13, 1192908 (2023). <https://doi.org/10.3389/fonc.2023.1192908>
- Dey, D., Slomka, P.J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P.P., Marwick, T.H.: Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 73, 1317–1335 (2019). <https://doi.org/10.1016/j.jacc.2018.12.054>
- Zreik, M., Lessmann, N., van Hamersvelt, R.W., Wolterink, J.M., Voskuil, M., Viergever, M.A., Leiner, T., Išgum, I.: Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Medical Image Analysis*. 44, 72–85 (2018). <https://doi.org/10.1016/j.media.2017.11.008>
- Lee, J.S., Lisker, S., Vittinghoff, E., Cherian, R., McCoy, D.B., Rybkin, A., Su, G., Sarkar, U.: Follow-up of incidental pulmonary nodules and association with mortality in a safety-net cohort. *Diagnosis (Berl)*. 6, 351–359 (2019). <https://doi.org/10.1515/dx-2019-0008>
- McWilliams, A., Tammemagi, M.C., Mayo, J.R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., Atkar-Khattra, S., Berg, C.D., Evans, K., Finley, R., Yee, J., English, J., Nasute, P., Goffin, J., Puksa, S., Stewart, L., Tsai, S., Johnston, M.R., Manos, D., Nicholas, G., Goss, G.D., Seely, J.M., Amjadi, K., Tremblay, A., Burrowes, P., MacEachern, P., Bhatia,

- R., Tsao, M.-S., Lam, S.: Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *New England Journal of Medicine*. 369, 910–919 (2013). <https://doi.org/10.1056/NEJMoa1214726>
18. Cho, S.M.J., Koyama, S., Honigberg, M.C., Surakka, I., Haidermota, S., Ganesh, S., Patel, A.P., Bhattacharya, R., Lee, H., Kim, H.C., Natarajan, P.: Genetic, sociodemographic, lifestyle, and clinical risk factors of recurrent coronary artery disease events: a population-based cohort study. *Eur Heart J*. 44, 3456–3465 (2023). <https://doi.org/10.1093/eurheartj/ehad380>
19. McWilliams, A., Tammemagi, M.C., Mayo, J.R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., Atkar-Khattra, S., Berg, C.D., Evans, K., Finley, R., Yee, J., English, J., Nasute, P., Goffin, J., Puksa, S., Stewart, L., Tsai, S., Johnston, M.R., Manos, D., Nicholas, G., Goss, G.D., Seely, J.M., Amjadi, K., Tremblay, A., Burrowes, P., MacEachern, P., Bhatia, R., Tsao, M.-S., Lam, S.: Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 369, 910–919 (2013). <https://doi.org/10.1056/NEJMoa1214726>
20. Detterbeck, F.C., Boffa, D.J., Kim, A.W., Tanoue, L.T.: The Eighth Edition Lung Cancer Stage Classification. *CHEST*. 151, 193–203 (2017). <https://doi.org/10.1016/j.chest.2016.10.010>
21. Zreik, M., van Hamersvelt, R.W., Khalili, N., Wolterink, J.M., Voskuil, M., Viergever, M.A., Leiner, T., Isgum, I.: Deep Learning Analysis of Coronary Arteries in Cardiac CT Angiography for Detection of Patients Requiring Invasive Coronary Angiography. *IEEE Trans Med Imaging*. 39, 1545–1557 (2020). <https://doi.org/10.1109/TMI.2019.2953054>
22. Liu, X., Mo, X., Zhang, H., Yang, G., Shi, C., Hau, W.K.: A 2-year investigation of the impact of the computed tomography-derived fractional flow reserve calculated using a deep learning algorithm on routine decision-making for coronary artery disease management. *Eur Radiol*. 31, 7039–7046 (2021). <https://doi.org/10.1007/s00330-021-07771-7>
23. Callister, M.E.J., Baldwin, D.R., Akram, A.R., Barnard, S., Cane, P., Draffan, J., Franks, K., Gleeson, F., Graham, R., Malhotra, P., Prokop, M., Rodger, K., Subesinghe, M., Waller, D., Woolhouse, I., Group, B.T.S.P.N.G.D.: British Thoracic Society guidelines for the investigation and management of pulmonary nodules: accredited by NICE. *Thorax*. 70, ii1–ii54 (2015). <https://doi.org/10.1136/thoraxjnl-2015-207168>
24. Goldin, J.G., Brown, M.S., Petkovska, I.: Computer-aided diagnosis in lung nodule assessment. *J Thorac Imaging*. 23, 97–104 (2008). <https://doi.org/10.1097/RTI.0b013e318173dd1f>
25. Apostolopoulos, I.D., Papandrianos, N.I., Papathanasiou, N.D., Papageorgiou, E.I.: Fuzzy Cognitive Map Applications in Medicine over the Last Two Decades: A Review Study. *Bioengineering*. 11, 139 (2024). <https://doi.org/10.3390/bioengineering11020139>
26. Kosko, B.: Fuzzy cognitive maps. *International Journal of Man-Machine Studies*. 24, 65–75 (1986). [https://doi.org/10.1016/S0020-7373\(86\)80040-2](https://doi.org/10.1016/S0020-7373(86)80040-2)
27. Papageorgiou, E.I., Salmeron, J.L.: A Review of Fuzzy Cognitive Maps Research During the Last Decade. *IEEE Transactions on Fuzzy Systems*. 21, 66–79 (2013). <https://doi.org/10.1109/TFUZZ.2012.2201727>
28. Papageorgiou, E.I., Froelich, W.: Application of Evolutionary Fuzzy Cognitive Maps for Prediction of Pulmonary Infections. *IEEE Transactions on Information Technology in Biomedicine*. 16, 143–149 (2012). <https://doi.org/10.1109/TITB.2011.2175937>
29. Feleki, A., Apostolopoulos, I., Papageorgiou, K., Papageorgiou, E., Apostolopoulos, D., Papandrianos, N.: A Fuzzy Cognitive Map Learning Approach for Coronary Artery Disease Diagnosis in Nuclear Medicine. Presented at the August 21 (2023)
30. Feleki, A., Apostolopoulos, I., Moustakidis, S., Papageorgiou, E., Papathanasiou, N., Apostolopoulos, D., Papandrianos, N.: Explainable Deep Fuzzy Cognitive Map Diagnosis of Coronary Artery Disease: Integrating Myocardial Perfusion Imaging, Clinical Data, and Natural Language Insights. *Applied Sciences*. 13, 11953 (2023). <https://doi.org/10.3390/app132111953>
31. Feleki, A., Apostolopoulos, I.D., Papageorgiou, E., Moustakidis, S., Papathanasiou, N.D., Apostolopoulos, D.J., Kokkinos, K., Papandrianos, N.: Deep Fuzzy Cognitive Map methodology for Non-Small Cell Lung Cancer diagnosis based on Positron Emission Tomography imaging. In: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–6 (2023)
32. Apostolopoulos, I.D., Groumpos, P.P.: Non-invasive modelling methodology for the diagnosis of coronary artery disease using fuzzy cognitive maps. *Computer Methods in Biomechanics and Biomedical Engineering*. 23, 879–887 (2020). <https://doi.org/10.1080/10255842.2020.1768534>
33. Apostolopoulos, I.D., Apostolopoulos, D.I., Spyridonidis, T.I., Papathanasiou, N.D., Panayiotakis, G.S.: Multi-input deep learning approach for Cardiovascular Disease diagnosis using Myocardial Perfusion Imaging and clinical data. *Phys Med*. 84, 168–177 (2021). <https://doi.org/10.1016/j.ejmp.2021.04.011>
34. Arsanjani, R., Xu, Y., Dey, D., Fish, M., Dorbala, S., Hayes, S., Berman, D., Germano, G., Slomka, P.: Improved accuracy of myocardial perfusion SPECT for the detection of coronary artery disease using a support vector machine algorithm. *J Nucl Med*. 54, 549–555 (2013). <https://doi.org/10.2967/jnumed.112.111542>
35. Otaki, Y., Singh, A., Kavanagh, P., Miller, R.J.H., Parekh, T., Tamarappoo, B.K., Sharir, T., Einstein, A.J., Fish, M.B., Ruddy, T.D., Kaufmann, P.A., Sinusas, A.J., Miller, E.J., Bateman, T.M., Dorbala, S., Di Carli, M., Cadet, S., Liang, J.X., Dey, D., Berman, D.S., Slomka, P.J.: Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease. *JACC Cardiovasc Imaging*. 15, 1091–1102 (2022). <https://doi.org/10.1016/j.jcmg.2021.04.030>
36. Liu, A., Wang, Z., Yang, Y., Wang, J., Dai, X., Wang, L., Lu, Y., Xue, F.: Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram. *Cancer Commun (Lond)*. 40, 16–24 (2020). <https://doi.org/10.1002/cac2.12002>
37. Mhaske, D., Rajeswari, K., Tekade, R.: Deep Learning Algorithm for Classification and Prediction of Lung Cancer using CT Scan Images. In: 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). pp. 1–5 (2019)
38. Lathakumari, K.R., Ramachandra, A.C., Avanthi, U.C., Basil Ronald, C., Bhavatharani, T.: Classification of Non-Small Cell Lung Cancer Using Deep Learning. In: 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC). pp. 1–5 (2023)
39. Yang, K., Liu, J., Tang, W., Zhang, H., Zhang, R., Gu, J., Zhu, R., Xiong, J., Ru, X., Wu, J.: Identification of benign and malignant pulmonary nodules on chest CT using improved 3D U-Net deep learning framework. *Eur J Radiol*. 129, 109013 (2020). <https://doi.org/10.1016/j.ejrad.2020.109013>
40. Wu, P., Sun, X., Zhao, Z., Wang, H., Pan, S., Schuller, B.: Classification of Lung Nodules Based on Deep Residual Networks and Migration Learning. *Comput Intell Neurosci*. 2020, 8975078 (2020). <https://doi.org/10.1155/2020/8975078>
41. Wang, W., Chakraborty, G.: Deep Learning for Automatic Identification of Nodule Morphology Features and Prediction of Lung Cancer. In: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST). pp. 1–6 (2019)
42. He, J., Wang, B., Tao, J., Liu, Q., Peng, M., Xiong, S., Li, J., Cheng, B., Li, C., Jiang, S., Qiu, X., Yang, Y., Ye, Z., Zeng, F., Zhang, J., Liu, D., Li, W., Chen, Z., Zeng, Q., Fan, J.-B., Liang, W.: Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study. *Lancet Digit Health*. 5, e647–e656 (2023). [https://doi.org/10.1016/S2589-7500\(23\)00125-5](https://doi.org/10.1016/S2589-7500(23)00125-5)
43. HOME, <https://emerald.uth.gr/>

44. Papandrianos, N., Papageorgiou, E., Anagnostis, A., Feleki, A.: A Deep-Learning Approach for Diagnosis of Metastatic Breast Cancer in Bones from Whole-Body Scans. *Applied Sciences*. 10, 997 (2020). <https://doi.org/10.3390/app10030997>
45. Feleki, A., Apostolopoulos, I.D., Papageorgiou, K., Papageorgiou, E.I., Apostolopoulos, D.J., Papandrianos, N.I.: A Fuzzy Cognitive Map Learning Approach for Coronary Artery Disease Diagnosis in Nuclear Medicine. In: *Fuzzy Logic and Technology, and Aggregation Operators: 13th Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT 2023, and 12th International Summer School on Aggregation Operators, AGOP 2023, Palma de Mallorca, Spain, September 4–8, 2023, Proceedings*. pp. 14–25. Springer-Verlag, Berlin, Heidelberg (2023)
46. Feleki, A., Papandrianos, N., Apostolopoulos, I., Papageorgiou, E., Papathanasiou, N., Apostolopoulos, D., Alonso Moral, J.M., Andreu-Perez, J.A.-P.: A Medical Decision Support System for Explainable Multimodal Detection of Non-Small Cell Lung Cancer Using Clinical and PET Data. Presented at the EXPLAINABILITY 2024, The First International Conference on Systems Explainability November 17 (2024)
47. Papageorgiou, E., Feleki, A., Papandrianos, N., Apostolopoulos, I., Papageorgiou, K., Papathanasiou, N., Apostolopoulos, D.: Multimodal Diagnosis using Deep Fuzzy Cognitive Map with Extreme Learning Machine Integrated into a Medical Decision Support System for Coronary Artery Disease and Non-Small Cell Lung Cancer Detection. In: *Proceedings of the 28th Pan-Hellenic Conference on Progress in Computing and Informatics*. pp. 400–406. Association for Computing Machinery, New York, NY, USA (2025)
48. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer New York, New York, NY (2009)
49. Spearman, C.: The proof and measurement of association between two things. *International Journal of Epidemiology*. 39, 1137–1150 (2010). <https://doi.org/10.1093/ije/dyq191>
50. Conover, W.J.: *Practical nonparametric statistics*. New York : Wiley (1980)
51. Gupta, M., Cotter, A., Pfeifer, J., Voevodski, K., Canini, K., Manglyov, A., Moczydlowski, W., Esbroeck, A. van: Monotonic Calibrated Interpolated Look-Up Tables, <http://arxiv.org/abs/1505.06378>, (2016)
52. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. pp. 1942–1948 vol.4 (1995)
53. Papageorgiou, E.I., Parsopoulos, K.E., Stylios, C.S., Groumpos, P.P., Vrahatis, M.N.: Fuzzy Cognitive Maps Learning Using Particle Swarm Optimization. *J Intell Inf Syst*. 25, 95–121 (2005). <https://doi.org/10.1007/s10844-005-0864-9>
54. Zahara, E., Kao, Y.-T., Su, J.-R.: Enhancing Particle Swarm Optimization with Gradient Information. In: *2009 Fifth International Conference on Natural Computation*. pp. 251–254 (2009)
55. Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Murphy, E.J., Prince, P.A., Stanley, H.E.: Lévy flight search patterns of wandering albatrosses. *Nature*. 381, 413–415 (1996). <https://doi.org/10.1038/381413a0>
56. Yang, X.-S., Deb, S.: Cuckoo Search via Levy Flights, <http://arxiv.org/abs/1003.1594>, (2010)
57. Mantegna, R.N.: Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. *Phys. Rev. E*. 49, 4677–4683 (1994). <https://doi.org/10.1103/PhysRevE.49.4677>
58. Shi, Y., RC, E.: A Modified Particle Swarm Optimizer. Presented at the *Proceedings of the IEEE Conference on Evolutionary Computation, ICEC June 4* (1998)
59. Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., Hui, B., Ji, L., Li, M., Lin, J., Lin, R., Liu, D., Liu, G., Lu, C., Lu, K., Ma, J., Men, R., Ren, X., Ren, X., Tan, C., Tan, S., Tu, J., Wang, P., Wang, S., Wang, W., Wu, S., Xu, B., Xu, J., Yang, H., Yang, J., Yang, S., Yao, Y., Yu, B., Yuan, H., Yuan, H., Zhang, J., Zhang, X., Zhang, Y., Zhang, Z., Zhou, C., Zhou, J., Zhou, X., Zhu, T.: Qwen Technical Report, <http://arxiv.org/abs/2309.16609>, (2023)
60. Samaras, A.-D., Moustakidis, S., Apostolopoulos, I.D., Papageorgiou, E., Papandrianos, N.: Uncovering the Black Box of Coronary Artery Disease Diagnosis: The Significance of Explainability in Predictive Models. *Applied Sciences*. 13, 8120 (2023). <https://doi.org/10.3390/app13148120>
61. Apostolopoulos, I.D., Papandrianos, N.I., Apostolopoulos, D.J., Papageorgiou, E.: Between Two Worlds: Investigating the Intersection of Human Expertise and Machine Learning in the Case of Coronary Artery Disease Diagnosis. *Bioengineering*. 11, 957 (2024). <https://doi.org/10.3390/bioengineering11100957>
62. Samaras, A.-D., Moustakidis, S., Apostolopoulos, I.D., Papandrianos, N., Papageorgiou, E.: Classification models for assessing coronary artery disease instances using clinical and biometric data: an explainable man-in-the-loop approach. *Sci Rep*. 13, 6668 (2023). <https://doi.org/10.1038/s41598-023-33500-9>
63. Apostolopoulos, I.D., Papathanasiou, N.D., Spyridonidis, T., Apostolopoulos, D.J.: Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. *Hell J Nucl Med*. 23, 125–132 (2020). <https://doi.org/10.1967/s002449912101>
64. Apostolopoulos, I.D., Papathanasiou, N.D., Papandrianos, N., Papageorgiou, E., Apostolopoulos, D.J.: Innovative Attention-Based Explainable Feature-Fusion VGG19 Network for Characterising Myocardial Perfusion Imaging SPECT Polar Maps in Patients with Suspected Coronary Artery Disease. *Applied Sciences*. 13, 8839 (2023). <https://doi.org/10.3390/app13158839>
65. Apostolopoulos, I.D., Apostolopoulos, D.I., Spyridonidis, T.I., Papathanasiou, N.D., Panayiotakis, G.S.: Multi-input deep learning approach for Cardiovascular Disease diagnosis using Myocardial Perfusion Imaging and clinical data. *Phys Med*. 84, 168–177 (2021). <https://doi.org/10.1016/j.ejmp.2021.04.011>
66. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks, <https://arxiv.org/abs/1609.02907v4>
67. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks, <http://arxiv.org/abs/1710.10903>, (2018)
68. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive Representation Learning on Large Graphs, <http://arxiv.org/abs/1706.02216>, (2018)
69. Samaras, A.-D., Apostolopoulos, I.D., Moustakidis, S., Papageorgiou, E., Papathanasiou, N.D., Apostolopoulos, D.J., Papandrianos, N.: Explainable Classification for Non-Small Cell Lung Cancer Based on Positron Emission Tomography Features and Clinical Data. Presented at the 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA) July 1 (2023)
70. Apostolopoulos, I.D., Papathanasiou, N.D., Apostolopoulos, D.J., Papandrianos, N., Papageorgiou, E.I.: Integrating Machine Learning in Clinical Practice for Characterizing the Malignancy of Solitary Pulmonary Nodules in PET/CT Screening. *Diseases*. 12, 115 (2024). <https://doi.org/10.3390/diseases12060115>
71. Samaras, A.-D., Apostolopoulos, I., Papageorgiou, E., Moustakidis, S.P., Papathanasiou, N., Apostolopoulos, D., Papandrianos, N.: Explainable Yolov8 Model for Solitary Pulmonary Nodules Classification Using Positron Emission Tomography and Computed Tomography Scans, <https://papers.ssrn.com/abstract=4824618>, (2024)