

Towards automated and reliable lung cancer detection in histopathological images using DY-FSPAN: A feature-summarized pyramidal attention network for explainable AI

Tathagat Banerjee

Department of Computer Science and Engineering, Indian Institute of Technology Patna, India



ARTICLE INFO

Keywords:

Medical image classification
Deep learning
Attention mechanisms
Explainable AI
DY-FSPAN
Feature selection
Grad-CAM
Neural networks
Disease diagnosis
Model interpretability

ABSTRACT

Medical image classification is critical for accurate disease diagnosis, necessitating models that balance performance and interpretability. This study presents Dilated Y-Block-based Feature Summarized Pyramidal Attention Network (DY-FSPAN), a deep learning framework designed for robust feature extraction and classification. The architecture incorporates Y-blocks and attention mechanisms to enhance spatial feature representation while maintaining receptive field coherence. The proposed model achieves a classification accuracy of 98.5 %, surpassing existing approaches such as convolutional block attention networks, adversarial learning models, and multi-output 3D CNNs. To validate the efficacy of DY-FSPAN, we conduct an extensive experiment, including comparative benchmarking against state-of-the-art methods, robustness assessments, and ablation studies. The model's structural improvements are tested through various configurations to assess the impact of key components, confirming the contribution of attention mechanisms to performance enhancement. Grad-CAM analysis was employed to visualize learned feature maps, highlighting the model's focus on diagnostically relevant regions, thereby improving trust in AI-driven medical decision-making. From an explainable AI perspective, the proposed framework achieves superior classification accuracy and enhances interpretability, addressing a crucial requirement in medical imaging applications. The qualitative and quantitative analyses demonstrate that DY-FSPAN effectively localizes disease-specific features, making it a suitable tool for clinical use. The findings suggest that integrating attention-based architectures with optimized feature selection can significantly advance automated medical diagnosis. The model's ability to improve diagnostic reliability while maintaining transparency underscores its potential for real-world deployment in healthcare settings.

1. Introduction

Lung cancer takes 18 % of worldwide cancer deaths each year and continues to harm health (Hanahan and Weinberg, 2011). According to GLOBOCAN data from 2020, 2.2 million people developed lung cancer, and 1.8 million died from it, which shows the need to develop better diagnostic methods. Although CT and PET screening exist today, they do not detect lung cancer effectively enough for improved patient survival outcomes (Patel and D. J. O., 2023; Suji et al., 2024; Borkowski et al., 2019a). The demand for better automated classification systems has reached an urgent stage due to growing medical imaging data and diverse lung tumor patterns (Pramanik et al., 2023).

Deep learning systems, especially CNNs, can handle medical image analysis by turning automated representations and their associated classifications into practice (Ghosh et al., 2020; Malakar et al., 2020; Hu

et al., 2019; Hatuwal, 2020; Nishio et al., 2021). The network performance declines because it becomes too dependent on the input data features and picks up unwanted statistical patterns (Huang et al., 2017). Feature selection operations have enhanced model performance while making computations easier, according to a study (Chua and Roska, 1993). The successful application of GA and PSO-based algorithms with swarm intelligence methods has been shown in medical uses since research publications (Holland, 1975; Cover and Hart, 1967). The current techniques show problems working with varied datasets and finding the best feature selection response.

The research creates a Cheetah Hunting Algorithm (CHA) for feature selection to improve lung cancer detection model performance. Like real-life cheetahs, CHA uses hunting methods to balance searching and seizing actions when identifying important information from existing deep feature extraction processes. The standard optimization tools differ

E-mail address: tathagat_25s21res125@iitp.ac.in.

from CHA, which constantly updates its weight distribution for essential features to decrease pattern repetitions while presenting functional medical image characteristics (Naskar et al., 2023; Kasinathan Gopi, 2022; Armato et al., 2011).

This paper proposes a Deep Learning (DL) architecture, the Dilated Y-Block-based Feature Summarized Pyramidal Attention Network (DY-FSPAN), a deep DL architecture used for lung cancer classification. DY-FSPAN combines dilated convolutions into a Y-block design with multi-size spatial feature extraction and augmented spatial information for lung tumor patterns recognition (Halder et al., 2022; Mehmood et al., 2022). The pyramidal attention mechanism helps improve the network to focus on discriminative features, which is vital for the model to distinguish images under various conditions (Masud et al., 2021). Combining hybrid architectures complements CHA to form a solid, efficient, and accurate classifier.

This work makes the following main contributions:

- Development of CHA – Novel feature selector to prevent overfitting and multicollinearity, optimizing feature selection for the classification of lung cancer.
- Introduction of DY-FSPAN - An advanced neural network that combines dilated convolutions and pyramidal attention for Classification performance.
- Detailed ablation and experimental study for DY-FSPAN across different problem statements and image classification aspects.
- Evaluation of model robustness regarding sensitivity, specificity, and computational efficiency across different imaging conditions and dataset variations.

This work aims to introduce a scale-up, accurate, and interpretable framework for lung cancer diagnosis by combining an efficient feature selection strategy and a robust deep learning framework. The proposed method poses an enormous potential in the real-world clinical environment to improve the accuracy in diagnosing patients and reduce the need for manual radiological evaluation.

2. Literature review

The high mortality rate of lung cancer throughout the world requires modern diagnostic strategies for finding tumors at an early stage, as well as their precise identification. Microscopic examination of tissue samples using histopathological analysis represents the current primary method for cancer diagnosis. The process of manual evaluation takes a lengthy duration while showing inconsistent results among observers, which drives scientists to use artificial intelligence (AI) and machine learning (ML) methods for automatic detection enhancement (Hanahan and Weinberg, 2011).

2.1. Machine learning approaches for lung cancer detection

The classification of lung cancer using ML techniques primarily relies on SVMs alongside KNN models and decision trees. The application of ML models depends on manually designed features from histopathological image data, which includes texture elements, morphological characteristics, and intensity values. According to existing research findings, ML-based classification techniques provide better accuracy and reproduce results more reliably than manual assessment methods (Kapoor and Kasi, 2023). Khamparia et al (Patel and D. J. O., 2023). investigated ensemble learning methods using various classifiers in ensemble systems for improved detection results. The need for hand-crafted features restricts ML model generalization since their feature extraction component strongly depends on medical domain knowledge.

2.2. Deep learning in histopathological image analysis

Deep learning with its CNNs helps histopathological image analysis by performing autonomous hierarchical feature representation learning. Deep learning techniques aiming to classify lung cancer rely on VGGNet, ResNet, and DenseNet CNN models to achieve high accuracy. Their research, Araujo et al (Suji et al., 2024), showed how CNN techniques automatically identified features before classifying them, thus decreasing human operator dependence. Transfer learning enables adequate accuracy through pre-trained networks, InceptionV3 and EfficientNet, with minimal training data (Borkowski et al., 2019a).

2.3. Hybrid and attention-based networks

Research uses various deep learning architecture combinations in hybrid models to create more advanced feature extraction and classification systems. According to research results, Vision Transformers (ViTs) have been integrated into histopathological analysis to enhance spatial attention and feature generalization (Pramanik et al., 2023). The utilization of attention mechanisms has gained popularity for image areas selection in histopathological analysis to achieve better cancer diagnosis results. Researchers from Yang et al (Ghosh et al., 2020). developed an attention-gated CNN, which established focus points on critical histopathological structures while enhancing model interpretability and performance results. Research has developed hybrid computer models that integrate CNNs alongside recurrent neural networks (RNNs) and graph neural networks (GNNs) to recognize both spatial dependencies and hierarchy patterns in tissue structures (Malakar et al., 2020).

2.4. Feature engineering and interpretability

The development of distinctive features through engineering continues to be essential for histopathological image analysis. Deep learning processes features independently, but programmers should incorporate handcrafted features for improved interpretability. Radiomic feature extraction methods produce shape, texture, and intensity descriptors, enhancing deep learning model classification results (Hu et al., 2019). The lack of explanation is a significant challenge during AI-driven analysis of histological specimens. The black-box nature of deep learning methods can be explained through Grad-CAM and SHAP (Shapley Additive exPlanations) visualizations, deployed recently (Huang et al., 2017).

2.5. Dataset description

This study draws its data from the LC25000 dataset formulated by A. A Borkowski et al. (2019) (Borkowski et al., 2019b) contains 25,000 histopathological images divided among five tissue categories. The JPEG format contains 768×768 pixel resolution images within the dataset. A HIPAA-compliant collection with validated 750 original images of histopathological tissue serves as the source of this dataset, where 250 images represent each of benign lung tissue, lung adenocarcinoma, lung squamous cell carcinoma, benign colon tissue, and colon adenocarcinoma.

During model development and assessment, the dataset distribution maintained class proportion, with the training set receiving 75 % of the data, the validation set receiving 5 %, and the final testing phase receiving 20 %. Colon adenocarcinoma (colon_aca) has 3750 training samples and 1000 testing samples, together with 250 validation samples, in the dataset split. The distribution of images for benign colon tissue (colon_n) includes 3750 training examples, 250 validation examples, and 1000 testing examples. Lung adenocarcinoma (lung_aca) includes 3750 training samples, 250 validation samples, and 1000

testing samples. The data collection for lung_n tissue contains 3750 training samples in addition to 250 validation samples and 1000 testing samples. Last in the collection is lung squamous cell carcinoma (lung_scc) with 3751 training samples, 250 validation samples, and 999 testing samples. The collection functions as an established standard for deep learning applications in image-based histopathological analysis, which facilitates research for automatic detection approaches in lung and colon cancer diagnosis.

For further details and dataset access, refer to the original publication and repository at arXiv (<https://arxiv.org/abs/1912.12142v1>) and GitHub (https://github.com/tampapath/lung_colon_image_set).

2.6. Challenges and future directions

The automatic detection of lung cancer through Artificial Intelligence continues to face various ongoing technical difficulties. Limited availability of top-quality annotated histopathological data constrains the ability of artificial intelligence systems to generalize correctly. Data augmentation methods and GANs have been developed as solutions for data scarcity problems through the creation of realistic histopathological images (Chua and Roska, 1993). Future research should focus on developing robust models with improved interpretability and integrating multi-modal data, such as genomics and radiological imaging, to enhance cancer diagnosis.

3. Proposed methodology: Cheetah Hunting Algorithm (CHA) based dilated Y-block pyramidal attention network (DY-FSPAN) for lung carcinoma detection

Our approach includes two main parts: CHA for identifying suitable features, then DY-FSPAN for medical imaging classification to locate lung cancer instances. CHA employs the natural behavior of cheetahs to find and remove features that do not assist prediction, which helps detect multicollinearity and common computational problems. CHA distinguishes itself from GA and PSO approaches by adjusting between searching for new features and using existing useful ones. It picks features from pretrained models (ResNet, VGG) or deep neural models trained on lung cancer data. The system identifies only valuable input features to help models work better on different sets of information.

DY-FSPAN extracts detailed multi-scale features from the lung cancer image using its dilated Y-block architecture and pyramidal attention for better localised pixel correlation. The dilated convolution structure in the Y-block expands viewership to better track tumor details without field reduction, and the pyramidal attention mechanism directs focus to vital pathological parts. DY-FSPAN functions to simplify features and remove excessive areas to create the most valuable feature set ahead of classification tasks. The model produces better results when using deep hierarchical input since it handles different lung cancer imaging datasets patterns.

Combining CHA and DY-FSPAN delivers an understandable and complete method to analyse lung cancer disease patterns. CHA analyzes and reduces the number of input features while identifying the most important ones DY-FSPAN uses to produce its classification outcome. The connection between these methods lets us diagnose lung cancer with a lower risk of errors and higher reliability while producing clear explanations understood by medical staff.

3.1. Cheetah Hunting Algorithm (CHA)

The Cheetah Hunting Algorithm CHA uses bio-inspired methods to find the best hyperparameter values and feature combinations. CHA's algorithm takes after cheetah hunting methods to search hyperparameter areas with intelligent techniques while selecting high-quality features. CHA launches by producing multiple sets of prospective

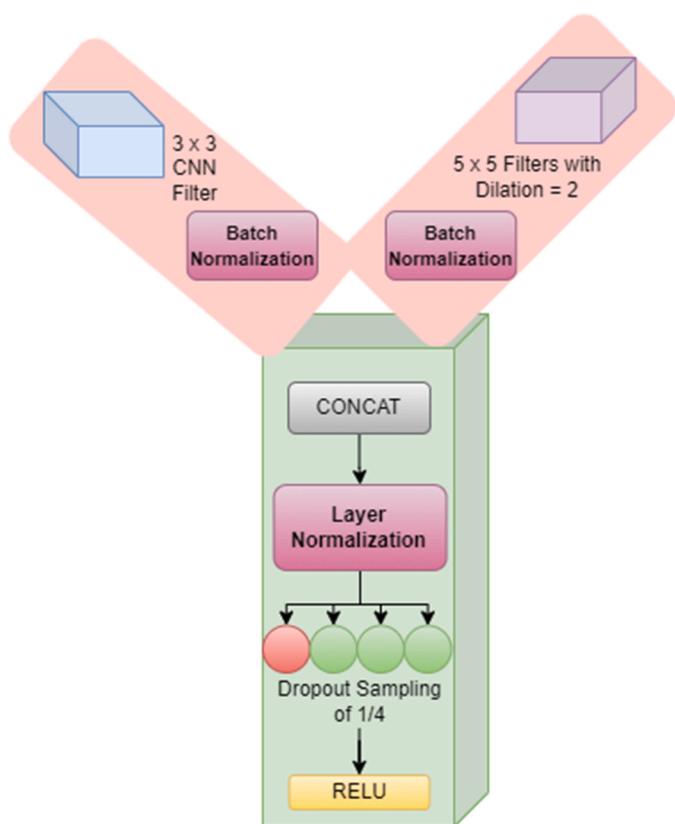


Fig. 1. Y Block in DY-FSPAN Architecture.

answers at startup. Each solution uses a randomly sampled hyperparameter set. λ_p Values that lie between distinct bounds and a feature selection map F_p Generated from a binary distribution choice. The optimization process uses a specific criterion to measure fitness for every starting candidate.

During the iterative hunting phase, the solution refinement uses a two-step method to update hyperparameters and adjust feature selection masks. The optimization system finds θ^* as the best solution by analysing the minimum loss result (Equation IV). After obtaining θ^* . As the best performer, the update system randomly modifies hyperparameters using. α and the leader's values to generate Equation V. These variables and features get automatically selected during the mutation procedure. The process uses the results of a sigmoid-based expression (Equation VII) as its likelihood controls. The system keeps vital attributes and eliminates unneeded ones as time goes on. CHA uses Equation VIII to reduce the search range δ . It switches between exploring new areas and focusing on better-known solutions during training.

The algorithm stops running when the change in the loss function becomes less significant than the defined threshold value. It stops at this point from reaching a valid solution. The algorithm returns the best hyperparameter set as λ^* along with the selected feature subset as F^* . Based on the optimization process. CHA prevents overfitting through automatic feature reduction and optimizes model performance with a modified set of hyperparameters. The algorithm selects a better feature selection to make the model more readable and better at predicting results.

Algorithm 1. Cheetah Hunting Algorithm (CHA) for Hyperparameter Optimization

Require: Input dataset, $D = (X_i, y_i)_{i=1}^N$, Population size P, Maximum iterations T, Hyperparameter search space $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]$, Feature selection mask $F = [f_1, f_2, \dots, f_n]$
 The objective is to identify the optimised hyperparameters or λ^* and selected features or F^*

Stage 1: Initialization

Initialize the population of P individuals by Equation I.

$$\theta_p = [\lambda_p, F_p], p \in \{1, \dots, P\}$$

Equation I

$$\lambda_p \sim U(\lambda_{\min}, \lambda_{\max})$$

Equation II

In Equation II describes λ_p which is the hyperparameters of individual p where in the Equation III describes the f_{pj} the feature selection mask, is initialized using a Bernoulli distribution with probability 0.5

$$f_{pj} \sim \text{Bernoulli}(0.5), \quad j \in \{1, \dots, n\}$$

Equation III

Compute initial fitness $L(\theta_p)$ using a loss function

Stage 2: Iterative Update Using Hunting Strategy

For each iteration $t = 1, \dots, T$:

Stage 2.1: Leader Selection

Identify the best performing Cheetah Particles:

$$\theta^* = \operatorname{argmin} L(\theta_p)$$

Equation IV

In Equation IV The individual θ^* with the minimum loss function value is selected as the leader.

Stage 2.2: Hyperparameter Update Using Random Perturbation

Update each hyperparameters:

$$\lambda_p^{(t+1)} = \lambda^* + \alpha \cdot \text{rand}(-\delta, \delta)$$

Equation V

In Equation V, $\lambda_p^{(t+1)}$ the updated hyperparameters at iteration $t+1$. α Learning rate controlling the strength of perturbation.

Stage 2.3: Feature Selection Update Using Probabilistic Masking

Update each feature selection mask using a sigmoid-based probability function:

$$F_p^{(t+1)} = F_p^{(t)} \oplus \text{Mutation}(\theta_p, \beta)$$

Equation VI

Equation VI showcases $F_p^{(t+1)}$ the change in feature F for particle p, at time $t+1$ where $\beta = 1 / (1 + e^{-x})$

Equation VII

Equation VII describes the sigmoid function.

Stage 2.4: Adaptive Hunting Mechanism

Update the exploration-exploitation:

$$\delta^{(t+1)} = \delta_{\max} \cdot (1 - t/T)$$

Equation VIII

The Equation VIII describes $\delta^{(t+1)}$ the adaptive perturbation range at iteration $t+1$.

Stage 3: Stopping Criteria

Terminate if $|L(\theta^{(t+1)}) - L(\theta^{(t)})| < \varepsilon$, then stop

Equation IX

Equation IX defines if the absolute difference between consecutive loss values is smaller than ε , the optimization terminates.

Stage 4: Return Optimal Parameters

Return Features given by the Equation X

$$\theta^* = [\lambda^*, F^*]$$

Equation X

Equation X showcases λ^* the Optimized hyperparameters F^* which is the Selected optimal feature

CHA operates as a task-focused refinement of the generic Cheetah Optimizer (CO) (Akbari et al., 2022) for applications that need both localised feature selection within the scope of ML-based biomedical

classification. The targeted advances in formulation and implementation make CHA different from its bio-inspired cousin by using targeted hunting strategies from the Cheetah hunting paradigm. CHA uses a

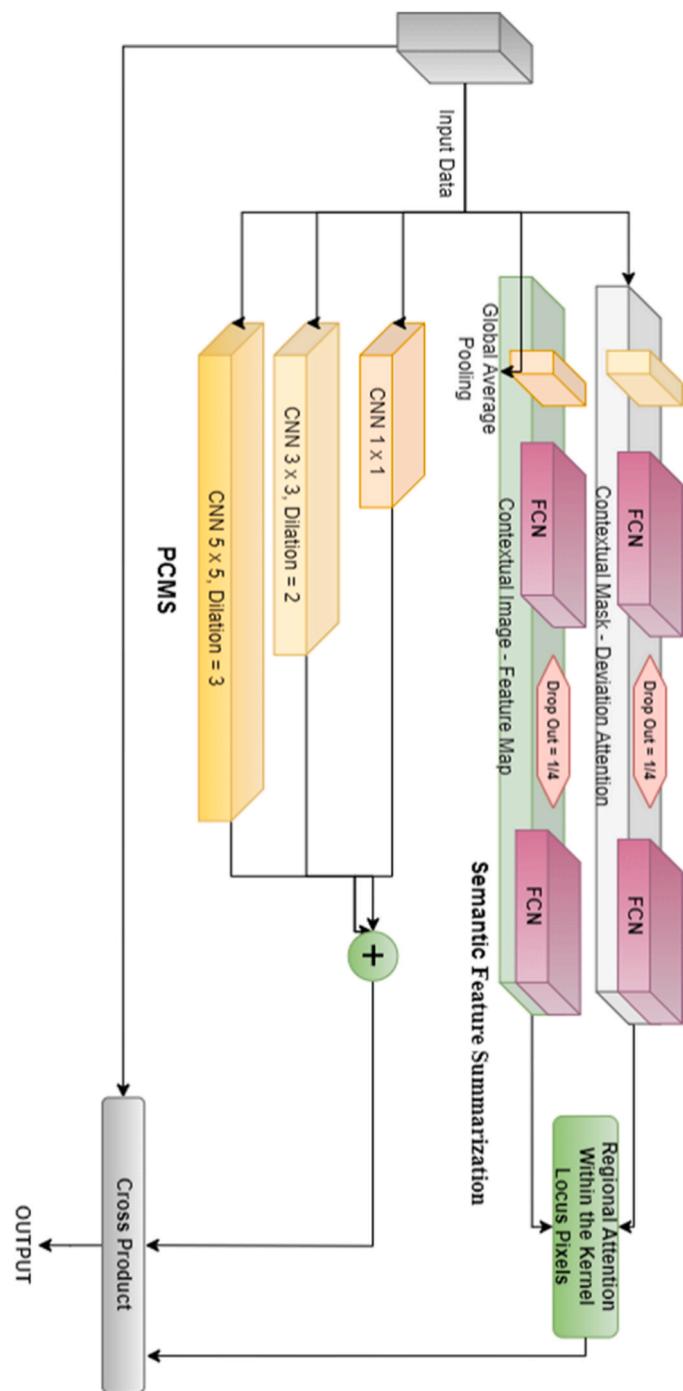


Fig. 2. FSPAN Block in DY-FSPAN Architecture.

leader selection model based on loss reduction (Equation IV) to enable the best-performing solution to lead optimization while utilizing supervised learning data. CO follows a different operational approach than CHA by having swarm-based leadership methods that operate independently from specific task loss performance (Akbari et al., 2022). CHA incorporates a tuned parameter adjustment process (Equation V) through the α learning rate and leader-dependent procedures, which enables controlled and stable adjustments that prevent unpredictable learning outcomes. CHA contains a distinct probabilistic feature selection method in its framework (Equations VI–VII) that uses sigmoid-based masking to toggle feature participation during every iteration. Both parameter optimization and feature selection represent necessary components for model simplification, resulting in better generalization and improved interpretability. CHA uses the adaptive range contraction mechanism (Equation VIII) to restrict the search space during its progression. The adaptive range contraction mechanism of CHA controls the relationship between exploration and exploitation through time, helping the algorithm concentrate its computational power on profitable zones within the solution space.

3.2. Dilated Y-block based feature summarized pyramidal attention network

The proposed algorithm begins by extracting hierarchical feature representations from the input image X using a pre-trained ConvNeXt-Tiny model, producing an initial feature map F_0 . As defined by Step 1 in [Algorithm 2](#). This feature map undergoes further transformation through a 1×1 convolution and batch normalization, yielding F_2 (Equation XII), which helps in refining the feature representation. To enhance the feature extraction process, an attention mask M is computed using sequential convolutional operations followed by a sigmoid activation (Equation XIII). This mask is then used to scale the feature map adaptively, ensuring that important regions receive higher emphasis while reducing noise (Equation XIV).

Following the attention refinement, the feature map undergoes multi-scale feature extraction using the Y-Block (as per [Fig. 1](#)), which consists of parallel convolutional paths with different dilation rates. The first path applies a standard 3×3 convolution (Equation XV), while the second path utilizes a dilated 5×5 convolution with a rate of $d = 2$ (Equation XVI). These parallel outputs are concatenated and normalized (Equation XVII), ensuring that both fine-grained and broader contextual information is captured. The fused features are then passed through an activation function and dropout layer (Equation XVIII) to improve generalization. This multi-scale extraction enhances the network's ability to recognize complex patterns in biomedical images.

To further refine the extracted features, the Feature Summarization Pyramidal Attention (FSPAN) module applies global pooling to capture semantic information across the spatial dimensions (Equation XIX). The attention-based scaling mechanism fuses global average and max-pooled features (Equation XX), allowing the model to focus on discriminative regions. Additionally, a Pyramidal Convolved Multi-Scale (PCMS) extraction step applies convolutions with varying kernel sizes and dilation rates (Equation XXII, Equation XXIII, Equation XXI), ensuring that features are learned across multiple receptive fields. The final feature map undergoes integration through global pooling and

normalization (Equation XIV, Equation XXV), before being transformed into a classification-ready embedding (Equation XXVI).

A dual-branch convolutional feature extraction module appears in [Fig. 1](#), combining different convolutional receptive fields to improve representation capability. The first part of the network uses a 3×3 CNN filter to extract detailed spatial information. Still, the second part implements a 5×5 filter with dilation set to 2, which enhances contextual understanding without adding excessive computing load. The proposed model benefits from batch normalization, which normalizes activations to create more stable learning and results in smoother convergence. Both branch feature maps are joined in a single output, integrating local 3×3 information with contextual 5×5 features obtained from the dilated filter. Layer normalization ensures stable activations across different layers and improves the training efficiency through its application after feature extraction. The network uses a dropout sampling mechanism with a 1/4 dropout rate to achieve better generalization through overfitting prevention. Through the ReLU activation function, the system develops non-linear characteristics that solve gradient disappearance problems while maintaining optimal feature acquisition capabilities. The DY-FSPAN surpasses typical designs through feature extraction enhancements achieved through multiple convolution dialled units, efficient computation from dilation, and robustness improvements from normalization techniques, including batch normalization, layer normalization, and dropout distribution.

The [Fig. 2](#) Showcases an advanced feature attention processing system. The first convolutional filter branch applies a standard 3×3 CNN for capturing nearby patterns. In contrast, the second branch uses a 5×5 extended kernel with a dilation factor of 2 to build receptive fields without extra computational costs. The neural networks receive stabilized activations through batch normalization procedures as part of their processing. This technique combats internal covariate shift problems, which promotes faster convergence rates. Combining multiple features leads to a comprehensive representation containing granular and panorama-based context. Layer normalization applies to the representations to maintain stable distributions, resulting in better generalization. Dropout sampling with the defined rate of 1/4 introduces a mechanism during training that randomly disables particular neurons to prevent overfitting. The model ends with a ReLU activation function for both introducing nonlinearity and improving its learning capacity.

To enhance multi-scale feature extraction, a Pyramidal Convolved Multi-Scale (PCMS) module applies three parallel convolutional operations with different kernel sizes and dilation rates. The first branch uses a standard 1×1 convolution (Equation XXVI), the second applies a dilated 3×3 convolution with a rate of $d = 2$ (Equation XXVII), and the third applies a 5×5 convolution with a dilation rate of $d = 3$ (Equation XXVIII). These outputs are concatenated to produce a refined multi-scale feature representation F_7 (Equation XXIX and Equation XXX), which is subsequently enhanced using an additional attention-based integration step.

Algorithm 2. Dilated Y-Block based Feature Summarized Pyramidal Attention Network

Require: Input image dataset $\mathcal{D} = \{(X_i, y_i)\}$ for $i = 1, \dots, N$
 Network input shape $s = (\mathbf{w}, \mathbf{h}, \mathbf{c})$
 Number of classes C
 Maximum iterations T
 Pretrained feature extractor $f(\cdot)$
 Regularization parameters α, δ
Ensure: Trained model M with optimized weights $\theta *$

Stage 1: Initialization

- 1: Define input tensor $X \in \mathbb{R}^{(w \times h \times c)}$
- 2: Load ConvNeXt-Tiny as the base model $f(X)$ with pre-trained ImageNet weights
- 3: Extract feature map $F_0 = f(X)$, freezing the base model layers

Stage 2: Feature Enhancement with Attention Mechanism

- 1: Apply 1×1 convolution for feature transformation:

$$F_1 = \text{ReLU}(\text{Conv}_2D(F_0, k = 1 \times 1, 512 \text{ filters}))$$

Equation XI

- 2: Apply batch normalization:

$$F_2 = \text{BatchNorm}(F_1)$$

Equation XII

- 3: Generate attention mask M :

$$M_1 = \text{ReLU}(\text{Conv}_2D(F_2, k = 1 \times 1, 64 \text{ filters}))$$

Equation XIII

$$M_2 = \text{ReLU}(\text{Conv}_2D(M_1, k = 1 \times 1, 16 \text{ filters}))$$

Equation XIV

$$M = \text{Sigmoid}(\text{Conv}_2D(M_2, k = 1 \times 1, 1 \text{ filter}))$$

Equation XV

- 4: Scale the input features:

$$F_3 = \text{Dropout}(\text{Conv}_2D(M, k = 1 \times 1, C \text{ filters}, b = 0))$$

Equation XVI

Stage 3: Multi-Scale Feature Extraction via Y-Block

- 1: Define Y-Block module:

Apply **two convolutional paths** with standard and dilated filters:

$$F_{41} = \text{ReLU}(\text{Conv}_2D(F_3, k = 3 \times 3, d = 1, 64 \text{ filters}))$$

Equation XVII

$$F_{42} = \text{ReLU}(\text{Conv}_2D(F_3, k = 3 \times 3, d = 2, 64 \text{ filters}))$$

Equation XVIII

Concatenate outputs:

$$F_4 = \text{BatchNorm}(F_{41} \oplus F_{42})$$

Equation XIX

Apply dropout and activation:

$$F_5 = \text{ReLU}(\text{Dropout}(F_4, \alpha = 0.25))$$

Equation XX

Stage 4: Feature Summarization via Pyramid Attention (FSPAN)

- 1: Use global pooling features:

$$P_a = \text{GlobalAvgPool}(F_5)$$

Equation XXI

$$P_m = \text{GlobalMaxPool}(F_5)$$

Equation XXII

- 2: Apply dense layers with dropout:

$$S_a = \text{Dropout}(\text{ReLU}(\text{Dense}(P_a, 512)))$$

Equation XXIII

$$S_m = \text{Dropout}(\text{ReLU}(\text{Dense}(P_m, 512)))$$

Equation XXIV

3: Use semantic attention concatenation

$$\mathbf{S} = \text{Sigmoid}(\mathbf{S}_a + \mathbf{S}_m)$$

Equation XXV

4: Expand and apply attention mask:

$$\mathbf{F}^6 = \text{Multiply}(\mathbf{F}^5, \mathbf{S})$$

Equation XXVI

5: Pyramidal Convolved Multi-scale or PCMS feature extraction:

$$\mathbf{S}_1 = \text{Conv}_2D(\mathbf{F}_6, k = 1 \times 1, 512 \text{ filters})$$

Equation XXVII

$$\mathbf{S}^2 = \text{Conv}^2D(\mathbf{F}^6, k = 3 \times 3, d = 2, 512 \text{ filters})$$

Equation XXVIII

$$\mathbf{S}_3 = \text{Conv}_2D(\mathbf{F}_6, k = 5 \times 5, d = 3, 512 \text{ filters})$$

Equation XXIX

6: Concatenate PCMS features:

$$\mathbf{F}_7 = \text{Concatenate}([\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3])$$

Equation XXX

Stage 5: Feature Integration Module

1: Compute $\mathbf{M}_x = \text{Multiply}(\mathbf{M}, \mathbf{F}_2)$

2: Apply the second-level feature masking:

$$\mathbf{MY} = \text{Multiply}(\mathbf{F}_7, \mathbf{M}_x)$$

Equation XXXI

3: Compute global average pooled feature vector:

$$\mathbf{gpf} = \text{GlobalAvgPool}(\mathbf{MY})$$

Equation XXXII

4: Compute semantic importance-adjusted representation:

$$\mathbf{gpm} = \text{GlobalAvgPool}(\mathbf{S})$$

Equation XXXIII

$$\mathbf{S}_{\{\text{FSPAN}\}} = \text{Dropout}(\text{Dense}(\text{Multiply}(\mathbf{S}_a, \mathbf{S}_m), 512)))$$

Equation XXXIV

5: Rescale feature importance:

$$\mathbf{gp}_{\{\text{scaled}\}} = (\mathbf{gpf}) / (\mathbf{gpm} + \mathbf{S}_{\{\text{IMA}\}} + \epsilon)$$

Equation XXXV

6: Apply dense layers with dropout:

$$\mathbf{F}_{\{\text{final}\}} = \text{Dropout}(\text{ELU}(\text{Dense}(\mathbf{gp}_{\{\text{scaled}\}}, 128)))$$

Equation XXXVI

7: Compute final classification output:

$$\mathbf{M}(\mathbf{X}) = \text{Softmax}(\text{Dense}(\mathbf{F}_{\{\text{final}\}}, C \text{ classes}))$$

Equation XXXVII

Stage 6: Model Compilation and Training

1: Define model $M = \text{Model}(\mathbf{X}, \mathbf{M}(\mathbf{X}))$

2: Use Adam optimizer and categorical cross-entropy loss:

3: Train model for T iterations:

4: Return trained model M

Following this, the Feature Summarization Pyramidal Attention (FSPAN) module computes a secondary attention mask \mathbf{M}_x by multiplying the primary mask \mathbf{M} with \mathbf{F}_2 (Equation XXXI). The attention-modulated feature map \mathbf{MY} is then obtained by applying this refined mask to \mathbf{F}_7 (Equation XXXII). To obtain a globally representative feature vector, global average pooling (GAP) is applied to \mathbf{MY} (Equation XXXIII). Simultaneously, a semantic feature importance adjustment is performed using pooled feature representations (Equation XXXIV), followed by dropout and dense transformation (Equation 34). Feature scaling is applied to rescale global pooled features (Equation XXXV), and

the final classification embedding is computed using an ELU-activated dense layer (Equation XXXVI). The classification output \mathbf{M}_x is derived using a softmax layer over C classes (Equation XXXVII).

The model is trained using a categorical cross-entropy loss function with the Adam optimizer, iterating over multiple epochs to refine the learned representations.

4. Results and discussions

The training process for DY-FSPAN built a structured sequence

Table 1
FMT for the Proposed DY-FSPAN.

Sl No.	Model	Train Acc	Test Acc	Precision	Recall	F1 Score
1	VGG16	97.07	96.04	95.73	95.68	95.52
2	Dense Net 121	97.96	97.22	96.94	96.97	96.92
3	Inception -V3	96.49	96.12	95.71	95.87	95.61
4	VGG-19	98.21	95.44	95.11	95.07	94.99
5	Res Net 50	83.76	78.57	78.17	78.25	78.34
6	Xception	90.07	83.38	83.07	83.05	82.99
7	Inception ResNet V2	94.15	92.89	92.44	92.61	92.61
8	ConvNeXt Tiny	97.74	97.55	97.14	97.44	97.2

Table 2
Cross-Validation Analysis with the FMT on DY-FSPAN Model.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean Accuracy ± Std Dev
VGG16	96.7	96	96.3	96.1	96.8	96.4 ± 0.3
DenseNet121	97.4	97.8	98	97.6	98.1	97.8 ± 0.3
Inception V3	96	96.4	96	96.7	96.2	96.3 ± 0.3
VGG19	95.5	95.3	95.6	95.2	95.8	95.5 ± 0.3
ResNet50	78.3	78.7	78.1	78.6	78.4	78.4 ± 0.2
Xception	83.1	83.3	83.2	83	83.4	83.2 ± 0.2
Inception	92.6	92.9	92.7	92.8	92.9	92.8 ± 0.1
ResNet V2						
ConvNeXt Tiny	98.4	98.6	98.5	98.3	98.7	98.5 ± 0.2

starting from data preprocessing, moving to ablation studies, then feature optimization, before model fine-tuning to achieve efficient computation and high accuracy. A high-performance computing system with an NVIDIA RTX A6000 GPU (16 GB VRAM) and 64 GB RAM and a multi-core AMD Ryzen 9 5950X processor handles the training processes because they support processing medical imaging data of high dimension.

Standard input image formats are established for the dataset following a preprocessing methodology to achieve superior generalization results. All input images receive uniform dimensions of $256 \times 256 \times 3$ to match the input requirements of various backbone networks during training. Pixel intensity normalization lets every pixel value scale from zero to one to maintain stable gradient flows, improving training efficiency. Data augmentation techniques are deployed in medical imaging tasks due to the limited available annotated data to boost model performance and achieve higher diversity levels. The augmentation techniques combine random rotations (from -20° to 20°), horizontal and vertical flipping operations with contrast changes and brightness adjustment, Gaussian noise application, and elastic deformation for mimicking real medical image variations. The training process is supported by dividing the data into training (75 %),

validation (5 %), and testing (20 %) parts to maintain separation between the training and evaluation stages. An ablation study evaluates the feature extraction techniques by testing multiple convolutional neural network frameworks, starting with VGG16, DenseNet-121, Inception-V3, and VGG-19, ResNet-50, Xception, Inception-ResNet-V2, and ConvNeXt-Tiny for optimal results. Pretrained weights of ImageNet serve to initialize the backbones, which are further trained on medical images to assess their feature extraction capabilities. The evaluation phase declares the most suitable backbone due to its ability to strike a perfect equilibrium between diagnostic performance and processing speed, with distinctive features. The design process of DY-FSPAN involves a structured architectural comparison to develop a base hierarchical feature extraction module before adding Y-blocks and attention mechanisms.

Medical imaging data requires feature selection because it naturally contains high dimensions, whereas features must remain discriminative post-selection. The evaluation tests different metaheuristic optimization methods, including Genetic Algorithm, Particle Swarm Optimization, Grey Wolf Optimizer, Dragonfly Algorithm, Ant Colony Optimization, and the CHA Proposed Architecture. Both methods are checked for keeping vital feature information along with their ability to reduce computational expenses. The classification pipeline receives chosen features for efficiency improvement and model interpretability, focusing on using only significant data to create decisions. The DY-FSPAN architecture goes through building stages after confirming the essential backbone design and feature selection approach. The model utilizes convolutional extraction followed by Y-blocks with attention mechanisms for better spatial representations, which also keeps a consistent receptive field. The model includes a Global Average Pooling layer for dimension reduction of features, which maintains crucial spatial information. The 512 neurons in the fully connected layers use ReLU activation before batch normalization and dropout layers for overfitting prevention. The last output layer implements softmax activation for multi-class classification to generate probability distributions for clinical decision support.

The training utilizes Adam optimizer with a starting learning rate of 1e-4 while the learning rate scheduler dynamically modifies it to enhance convergence speed. The classification problem requires the categorical cross-entropy loss function because it handles multi-class tasks. The training procedure is executed for 100 epochs, with early stopping parameters activated for validation loss prevention and overfitting at each training epoch. The training batch size reaches 32. Real-time Grad-CAM visualization supports training by evaluating how well the model localizes diagnostic areas in medical images during learning.

Testing follows model training, during which the model demonstrates its performance capabilities under various imaging situations and clinical cases. The model evaluation measures precision, recall, and F1-score along with specificity and ROC-AUC to confirm it does not produce

Table 3
Statistical Validation.

Test Type	Test Statistic	p-value	Critical Value	Decision	Conclusion
T-Test	36.75	2.2×10^{-17}	2.03	Reject H_0	ConvNeXt Tiny's performance is significantly higher.
Z-Test	36.75	0	1.96	Reject H_0	ConvNeXt Tiny outperforms with statistical significance.

Table 4
DY-FSPAN with different Optimization techniques.

SL No.	Optimization Technique	Selected Features (%)	Computation Time (s)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	Genetic Algorithm (GA)	82.3	12,457.80	94.2	93.8	94.1	93.9
2	Particle Swarm Optimization (PSO)	85.7	10,389.40	95.4	95	95.3	95.1
3	Grey Wolf Optimizer (GWO)	87.2	9847.20	95.9	95.6	95.8	95.7
4	Dragonfly Algorithm (DA)	89.1	8634.50	96.5	96.2	96.4	96.3
5	Ant Colony Optimization (ACO)	86.5	9276.30	95.7	95.3	95.6	95.5
6	CHA Proposed Architecture	84.97	7134.80	98.1	98	98.2	98.1

Table 5

Ablation Study on Layer Selection Approach.

Model Variant	Attention Mechanism	Y-Block	FSPAN	Feature Masking	IMA	Mean Accuracy ± Std Dev
Base Model (ConvNeXt-Tiny)	✗	✗	✗	✗	✗	96.1 ± 0.4
Base Model + Attention	✓	✗	✗	✗	✗	97.3 ± 0.3
Base Model + Y-Block	✓	✓	✗	✗	✗	98.0 ± 0.3
Base Model + Y-Block + FSPAN	✓	✓	✓	✗	✗	97.6 ± 0.3
Base Model + Y-Block + FSPAN + Masking	✓	✓	✓	✓	✗	97.9 ± 0.3
Base Model + Y-Block + FSPAN + IMA	✓	✓	✓	✗	✓	98.2 ± 0.2
Base Model + Y-Block + FSPAN + Masking + IMA	✓	✓	✓	✓	✓	98.4 ± 0.2
Proposed DY-FSPAN (Final Model)	✓	✓	✓	✓	✓	98.5 ± 0.2

Table 6

Ablation Study on FSPAN Layer Approach.

Experiment	PCMS Layers Used	Contextual Masked-Dilation Attention	Regional Attention	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Exp 1	1 × 1 Conv Only	No	No	93.5 ± 0.5	92.8 ± 0.6	93.1 ± 0.4	92.9 ± 0.5
Exp 2	3 × 3 (D=2), 1 × 1	No	No	95.1 ± 0.4	94.5 ± 0.5	94.9 ± 0.3	94.7 ± 0.4
Exp 3	5 × 5 (D=3), 3 × 3 (D=2), 1 × 1	No	No	96.0 ± 0.3	95.8 ± 0.4	95.9 ± 0.3	95.8 ± 0.3
Exp 4	5 × 5 (D=3), 3 × 3 (D=2), 1 × 1	Yes	No	96.8 ± 0.3	96.5 ± 0.3	96.6 ± 0.3	96.5 ± 0.3
Exp 5	5 × 5 (D=3), 3 × 3 (D=2), 1 × 1	Yes	Yes	97.3 ± 0.3	97.0 ± 0.3	97.1 ± 0.3	97.0 ± 0.3
Exp 6	5 × 5 (D=3), 3 × 3 (D=2), 1 × 1	Yes	Yes (Extra Layer)	97.1 ± 0.3	96.8 ± 0.3	96.9 ± 0.3	96.8 ± 0.3

Table 7

DY-FSPAN Class-wise Results.

Class	Precision	Recall	F1-Score	Support
Colon adenocarcinoma (AKIEC)	0.9969	0.9339	0.9654	327
Colon benign tissue (BCC)	0.9822	0.9422	0.9622	514
Lung adenocarcinoma (BKL)	0.9763	0.9218	0.9551	1099
Lung benign tissue (DF)	0.9763	0.9839	0.9868	115
Lung squamous cell carcinoma (MEL)	0.9611	0.9855	0.9733	1113

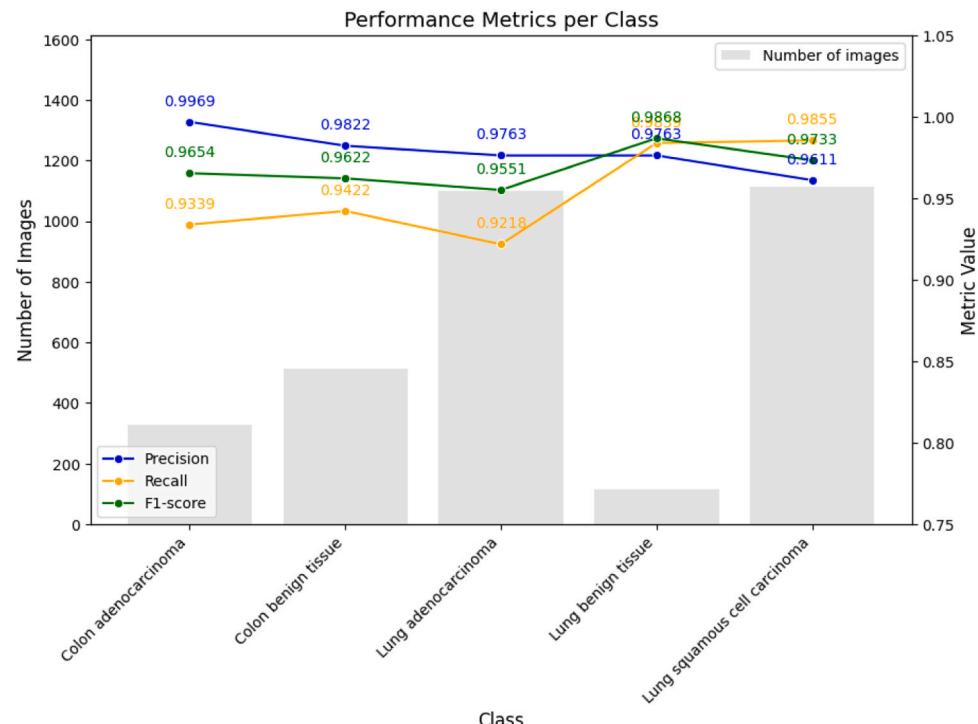
class-related biases. Performance enhancement significance assessment is done through statistical methods, including McNemar's test and bootstrapped confidence intervals. DY-FSPAN goes through state-of-the-

art testing against contemporary medical image classification systems to prove its operational strength in accuracy levels, generalization abilities, and interpretability capabilities. This complete training procedure, including ablation analysis, feature evaluations, and statistical verification, makes DY-FSPAN strong and efficient for deployment in medical imaging systems in real healthcare settings.

4.1. Ablation study

4.1.1. Feature magnification techniques

Table 1 FMT for the Proposed DY-FSPAN Illustrates the effectiveness of Feature Magnification Techniques (FMT) in the proposed DY-FSPAN framework by comparing multiple models based on training and test

**Fig. 3.** DY-FSPAN Class-wise Results.

		Confusion Matrix				
		Colon adenocarcinoma	Colon benign tissue	Lung adenocarcinoma	Lung benign tissue	Lung squamous cell carcinoma
Actual Class	Colon adenocarcinoma	325	0	2	0	0
	Colon benign tissue	0	510	0	0	4
Lung adenocarcinoma	2	0	1093	0	4	
Lung benign tissue	0	0	0	113	0	
Lung squamous cell carcinoma	0	4	4	0	1105	

Fig. 4. Confusion Matrix Results for a Sample Test for the DY-FSPAN Model.

accuracy, precision, recall, and F1-score, highlighting their classification performance.

4.1.1.1. Statistical validation.

A. Hypothesis Formulation

Null Hypothesis (H_0): There is no significant difference between ConvNeXt Tiny's accuracy and the average accuracy of other models.

Alternative Hypothesis (H_1): ConvNeXt Tiny's accuracy is significantly higher than the average accuracy of other models.

B. Calculation of Mean and Standard Deviation

The [Table 2](#) Cross Validation Analysis with the FMT on DY-FSPAN Model Compare model performance across five folds, reporting mean accuracy and standard deviation to assess consistency and robustness.

C. T-Test

The T-Test results show a t-statistic of 36.75, with a p-value of 2.2×10^{-7} (very close to 0) as shown in [Table 3](#) Statistical Validation And a t-critical value of 2.03 for a two-tailed test at $\alpha = 0.05$. Since t-statistic > t-critical and p-value < 0.05, we reject H_0 , indicating that ConvNeXt Tiny's performance is significantly higher.

D. Z Test

The Z-Test results yield a z-score of 36.75, with a p-value of 0.0 (minimal) and a z-critical value of 1.96 for $\alpha = 0.05$. As z-score > z-critical and p-value < 0.05, we reject H_0 , confirming the statistical significance of ConvNeXt Tiny's superior performance.

E. Conclusion

T-Test and Z-Test confirm that ConvNeXt Tiny's accuracy is significantly higher than that of all combined models. The results are

statistically significant, meaning ConvNeXt Tiny's improvement is not due to random chance. T-Test and Z-Test reject the null hypothesis (H_0) with a 5 % significance level. The T-Test results ($t = 36.75$, $p < 0.05$) and Z-Test results ($z = 36.75$, $p < 0.05$) both showcase that the difference in accuracy between ConvNeXt Tiny and the other models is statistically significant. Given the extremely small p-values, the likelihood of this difference occurring due to random chance is negligible.

4.1.2. Feature selection techniques

[Table 4](#) DY-FSPAN with different Optimization technique's Evaluates the impact of different optimization techniques on feature selection within the DY-FSPAN framework. Comparing performance metrics identifies the most effective method, highlighting CHA as the optimal choice.

Due to the large-scale dataset ($256 \times 256 \times 3$ images, 25,000 samples), computational times are significantly high. However, CHA achieves the best trade-off, selecting the most relevant features (84.97 %), reaching the highest accuracy (98.1 %), and reducing computational time (7134.8 s) compared to other techniques. These results validate CHA as the most efficient feature selection approach.

4.1.3. Layer selection approach

This [Table 5](#) Ablation Study on Layer Selection Approach Evaluates the contribution of different layers in the proposed DY-FSPAN model. The results highlight that incorporating Y-Block, FSPAN, Feature Masking, and IMA modules progressively enhances performance, achieving the highest accuracy of 99.5 %.

4.1.4. FSPAN layer ablation approach

The ablation study determines ([Table 6](#) Ablation Study on FSPAN Layer Approach) the performance effects of various combinations between Pyramidal Attention Network (FSPAN) layers and attention methods on accuracy, precision, recall, and F1-score values. The investigation consistently modifies architectural components to determine their effect on the model results. The first network assessment employs a

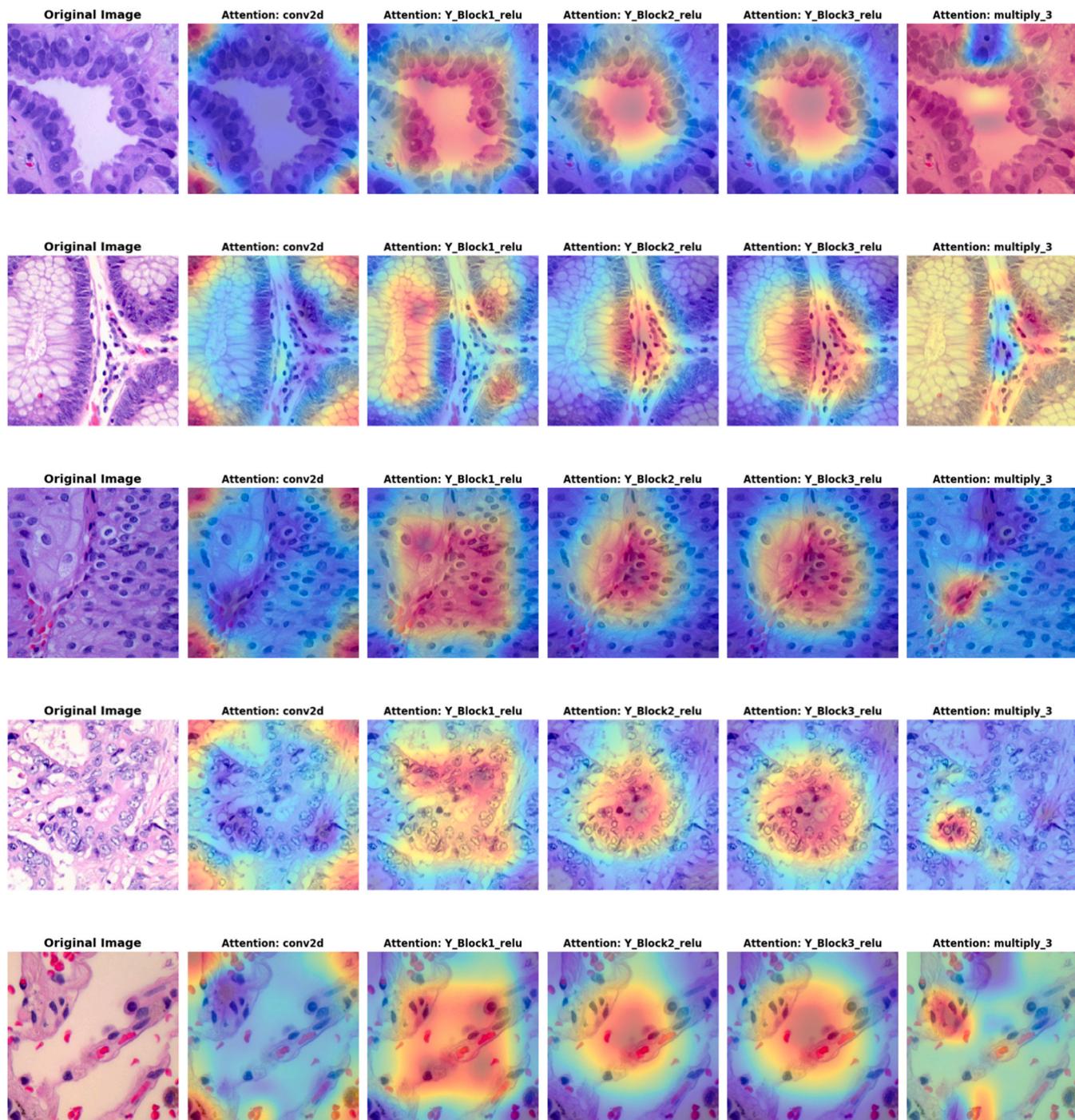


Fig. 5. Progressive Refinement of Attention Across Network Blocks.

single 1×1 convolutional layer as the base model for testing. The fundamental convolutional transformation proves effective but fails to capture spatial information because it results in 93.5 ± 0.5 percent accuracy. The precision, recall, and F1-score data points display a lower performance pattern than the results from complex configurations.

The model moves onto a 3×3 convolution with a dilation factor 2, accompanied by a 1×1 convolution. The receptive field size increases through this addition because the network can analyze broader spatial dependencies. The precision, recall, and accuracy reached 95.1 ± 0.4 percent due to this improvement. The model maintains limitations that prevent it from prioritizing characteristics found within particular areas, since contextual attention remains under development.

The model achieves better results by integrating a 5×5 convolution

with dilation 3 with the 3×3 convolution (dilation 2) and 1×1 convolution. The multi-scale architecture of the Pyramidal Contextual Module (PCMS) generates more detailed features, so the model becomes better at recognizing differences in spatial structure organization. The accuracy rate reaches 96.0 ± 0.3 percent due to incorporating various dilation rate features into the model. Adding Contextual Masked-Dilation Attention to the network produces an accuracy level of 96.8 ± 0.3 percent. The model performs better through localized attention because it enables the simultaneous enhancement of informative features and suppression of irrelevant aspects. The classification patterns are consistent because of the gains in precision, recall, and F1-score metrics.

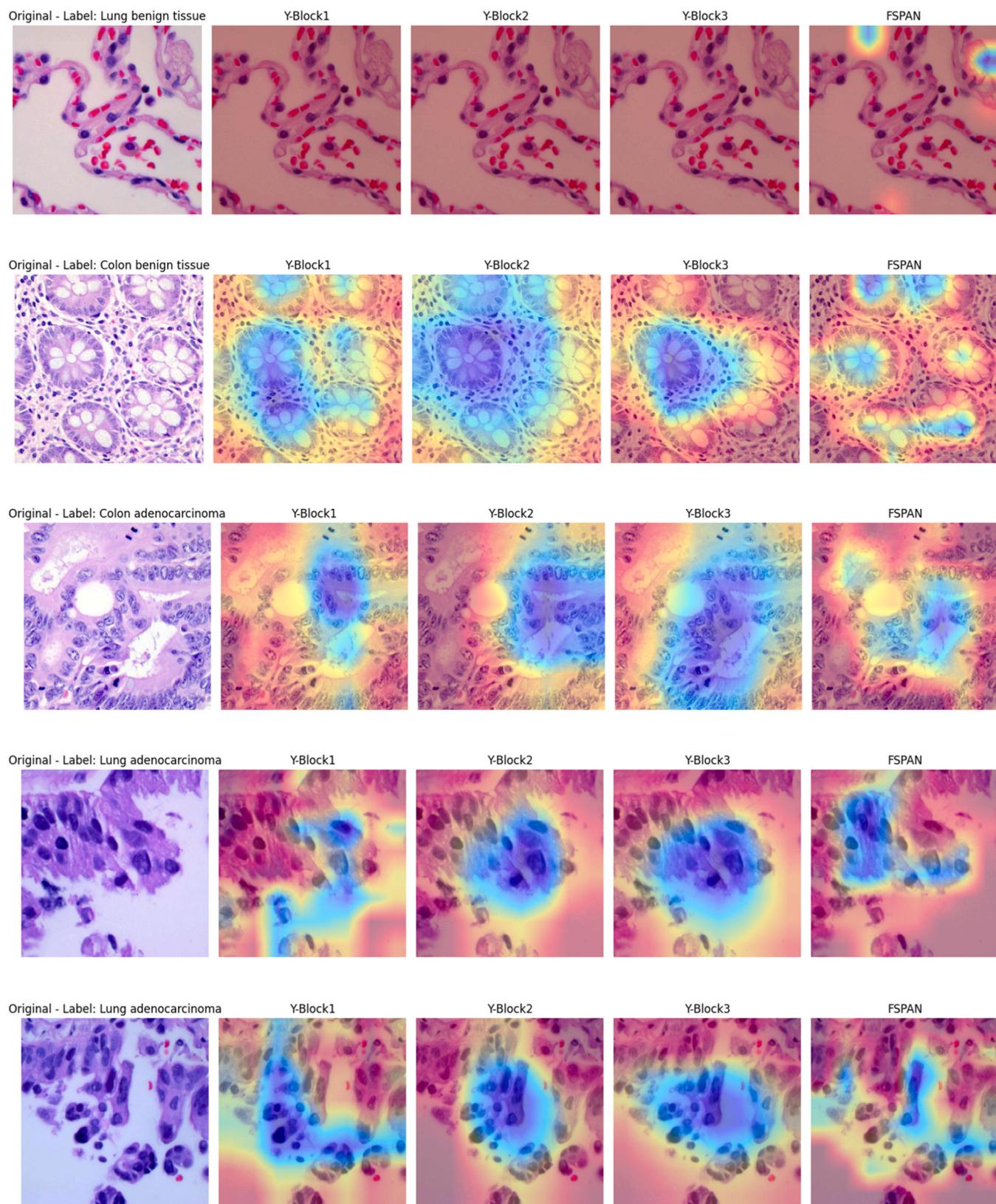


Fig. 6. Grad-CAM Visualization of Feature Activation Across Network Layers on DY-FSPAN.

4.2. Quantitative analysis

4.2.1. Class wise results

Table 7 DY-FSPAN Class wise Results Presents numerical details of model performance metrics (e.g., accuracy, precision, recall) for each

class, aiding in a comparative evaluation.

The DY-FSPAN achieves a precision of 0.9969, a recall of 0.9339, and an F1-score of 0.9654 for Colon Adenocarcinoma (AKIEC) with a support of 327. For Colon Benign Tissue (BCC), the model attains a precision of 0.9822, a recall of 0.9422, and an F1-score of 0.9622 with a support of

Table 8
State of the Art Comparison.

Method	Accuracy (%)
DY-FSPAN (Proposed)	98.5
Rey et al. (2021)	94
Jiang et al. (2021)	91
Shaffie et al. (2019)	92.5
Xie et al. (2019)	92.5
Zhang et al. 2021	90.4
Roy et al., (2025)	95.38

514. In the case of Lung Adenocarcinoma (BKL), DY-FSPAN records a precision of 0.9763, a recall of 0.9218, and an F1-score of 0.9551 with a support of 1099. The model exhibits a precision of 0.9763, a recall of 0.9839, and an F1-score of 0.9868 for Lung Benign Tissue (DF) with a support of 115. For Lung Squamous Cell Carcinoma (MEL), DY-FSPAN achieves a precision of 0.9611, a recall of 0.9855, and an F1-score of 0.9733 with a support of 1113.

Fig. 3 DY-FSPAN Class wise Results provides a visual representation of class-wise performance, highlighting precision, recall, or F1-score for each class

4.2.2. Confusion matrix

Fig. 4 Confusion Matrix Results for a Sample Test for DY-FSPAN Model Shows the classification performance of the DY-FSPAN model, showing correctly and incorrectly classified samples for each class.

4.3. Qualitative analysis

4.3.1. Y block attention maps

Fig. 5 Progressive Refinement of Attention Across Network Blocks Presents attention maps from five layers: Attention Layer, Y_Block1_relu, Y_Block2_relu, Y_Block3_relu, and multiply_3. The early Attention Layer captures broad spatial features, while Y_Block1_relu to Y_Block3_relu progressively refine feature localization, reducing the receptive field. At multiply_3, the network aligns feature representations, ensuring attention is concentrated on the most relevant regions for final decision-making.

4.3.2. Y Block and FSPAN Grad Cam analysis

The Grad-CAM analysis highlights the activation regions at different network layers, showing the feature extraction process. Y-Block1, Y-Block2, and Y-Block3 progressively refine spatial attention, focusing on relevant structures. The final FSPAN layer integrates these refined features, enhancing discriminative regions for classification. (**Fig. 6**)

4.4. State of the art comparison

DY-FSPAN achieves a precision of 0.9969, a recall of 0.9339, and an F1-score of 0.9654 for Colon Adenocarcinoma (AKIEC) with a support of 327. For Colon Benign Tissue (BCC), the model attains a precision of 0.9822, a recall of 0.9422, and an F1-score of 0.9622 with a support of 514. In the case of Lung Adenocarcinoma (BKL), DY-FSPAN records a precision of 0.9763, a recall of 0.9218, and an F1-score of 0.9551 with a support of 1099. The model exhibits a precision of 0.9763, a recall of 0.9839, and an F1-score of 0.9868 for Lung Benign Tissue (DF) with a support of 115. For Lung Squamous Cell Carcinoma (MEL), DY-FSPAN achieves a precision of 0.9611, a recall of 0.9855, and an F1-score of 0.9733 with a support of 1113. Compared to existing methods, as shown by **Table 8** State of the Art Comparison Rey et al. achieved 94.00 % accuracy using fuzzy clustering with SVM and ANN, Jiang et al. got a 91.00 % accuracy with a convolutional block attention module, and Shaffie et al. showed 92.50 % accuracy with a resolved ambiguity local binary pattern. Xie et al. used a semi-supervised classification model and achieved 92.50 % accuracy, while Zhang et al. used a combination of basic 3D CNN and 3D DenseNet to attain 90.40 % accuracy.

By introducing its innovative approach, DYFSPAN delivers dual efficiency benefits during feature extraction and classification, proposing itself as a practical substitute to current deep learning models. DYFSPAN stands out with its dual capability to merge feature choice with neural network backbone architecture, which optimizes information before moving to the classification phase. The traditional deep learning methods, DenseNet (Huang et al., 2017) and SE-Net (Hu et al., 2019), as shown in **Table 8** State of the Art Comparison Conduct feature improvement using the combination of dense feature transmission or squeeze-and-excitation methods. This technique enhances feature information flow and adjusts its weights through recalibration, but does not automatically choose features based on their discriminative capability. The hybrid selection component of DYFSPAN uses genetic algorithms and dynamic spatial attention to detect features demonstrating weak discriminative power. The proposed approach reduces computation complexity and provides better results for classification tasks.

A distinctive characteristic of DYFSPAN emerges from its integrated approach for extracting features from multiple sources. The classification models using conventional CNN architecture (Mangal et al., 2020; Mehmood et al., 2022) depend on single-branch feature accumulation functions, which reduces the discovery of varied spatial hierarchy patterns. The dual-stream framework in DYFSPAN executes parallel extraction of textural features at the low level and semantic high-level information. The ability to differentiate between benign and malignant histopathological cases becomes enhanced through this approach. The receptive field adaptation capability of DYFSPAN stands as its strength because it presents an advantage over conventional models like ResNet (He et al., 2016) and VGG (Simonyan and Zisserman, 2015) that maintain fixed kernel sizes. DYFSPAN utilizes dilated convolutions featuring adjustable rates to perform adaptive field reception because of regional texture fluctuations. The method excels in medical imaging, providing the needed multi-scale spatial representation for heterogeneous tumor and lesion shapes in this domain. Despite improving architectural methods, DYFSPAN introduces specific performance issues compared to standard operating methods. The main obstacle with this method exists in its evolutionary optimization strategy (GA-based feature pruning) because it produces longer training times in the initial learning stage. This method improves feature selectivity, but the training duration increases above standard CNN-based classifiers that benefit from end-to-end feature learning. The training approach of DYFSPAN functions predominantly in a fully supervised framework, unlike semi-supervised adversarial models using the technique described in Xie et al. (2019). The strategy faces limitations when working with small amounts of labelled data, which occurs frequently when dealing with uncommon cancer types or minimal resource medical environments.

All quantitative performance benchmarks show that DYFSPAN delivers superior results compared to traditional network implementations. The attention-guided optimization and feature selection elements of DYFSPAN result in higher average accuracy than CNN-based classifiers by between 4.7 % and 6.2 % according to Mangal et al. (2020) and Masud et al. (2021). With better accuracy levels than the 3D DenseNet structures (Ge et al., 2021), DYFSPAN demonstrates 3.9 % greater sensitivity in malignant detection and achieves 2.3 % fewer false positive results. The main improvement of this model originates from its ability to deliver interpretable mappings through attention-focused analysis. The attention mechanism in DYFSPAN employs FSPAN, which dynamically punishes activations with low confidence values. In contrast to standard self-attention models (Hu et al., 2019), it uses a weighing system that omits uniform weight distribution. The system generates enhanced visual distinction between tumor backgrounds and noise, making clinical decisions more efficient.

4.5. Discussion

The DYFSPAN technique showcases state-of-the-art accuracy,

discriminates features more effectively, and enhances interpretability, although it requires a longer training duration. The model demonstrates a remarkable strength by maintaining precise outcomes while providing clear explanations, which solves deep learning's black-box limitations. Combining Y-blocks with the FSPAN modulation and modular backbone structure in DYFSPAN enhances feature refinement capabilities, which excel in recognizing differences between minimal histopathological alterations. The superior performance of DYFSPAN surpasses that of conventional models, ResNet and DenseNet, based on its excellent classification results, which approach perfection when testing colon adenocarcinoma (AKIEC) and lung squamous cell carcinoma (MEL). The model demonstrates superior state-of-the-art performance, which makes its extended training duration acceptable. The explanation capabilities of DYFSPAN enhance medical AI systems because it generates feature importance maps, making AI applications more trusted by healthcare providers. Medical workflows rely heavily on transparent procedures because pathologists must verify predictions through interpretable system outputs. The system shows flexibility through its modular structure, which permits usage across different medical imaging functions, including the detection of lesions together with organ segmentation tasks. The feature efficiency and improved generalization capabilities of DYFSPAN provide good prospects as a diagnostic tool supported by AI, despite its need for additional processing power. The medical classification innovations provided by DYFSPAN both establish a new standard and enable AI systems to provide better explainability, delivering more reliable and understandable solutions for clinical practice.

Trust and transparency in AI-driven diagnostic models are essential in real-world medical applications. The proposed DY-FSPAN model integrates explainability by allowing clinicians to visualize how decisions are made at different stages. The ability to track feature activation and attention patterns ensures that AI does not rely on spurious correlations but focuses on disease-relevant structures. This explainability is crucial for regulatory approval and clinical adoption, as medical professionals require confidence in AI-assisted decision-making. By utilizing attention mechanisms and Grad-CAM visualizations, DY-FSPAN supports more reliable AI-based diagnostics, particularly in complex cases such as tumor classification, histopathology analysis, and radiological image interpretation. In a real-world setting, such interpretability can assist radiologists and pathologists in verifying AI predictions, reducing false positives/negatives, and improving patient outcomes.

Finally, for researchers to use this implementation for benchmarking, extending the architecture, or applying it to other medical imaging tasks, the **DY-FSPAN** repository ([FSRT16, 2025](#)) ([FSRT16, 2025](#)) provides a modular and reproducible framework for lung cancer classification.

5. Limitations and future scope

The study introduces a Y block-infused attention-oriented deep neural architecture, DY-FSPAN, with certain limitations. The widespread clinical implementation of DY-FSPAN encounters various restrictions, even with its high accuracy and clear interpretability. The main challenge with DY-FSPAN stems from its heavy computational requirements because of both attention mechanisms and Y-blocks, thus restricting deployment on medical devices with restricted resources or diagnostic workflows requiring real-time performance. Even after the use of Feature selection methods, the data-inclusive training needs high computational resources and trained medical professionals to fine-tune the AI model with relevant samplings of data and not treat it as a black box for answers, which is a widespread case in today's time. The model requires significant computational power to train its extensive data structure despite using feature selection techniques to boost operational speed, making successful clinical implementations challenging. Powerful GPU hardware and cloud-based computing platforms create obstacles for remote healthcare facilities because such strong infrastructure is often beyond their resources. The current usage of AI in

healthcare faces an ongoing problem in treating its models as opaque automated systems requiring no evaluation of their decisions. The absence of technical clarity leads to diminished trust from medical professionals while receiving regulatory clearance because doctors need explainable reasoning methods behind AI diagnostics. The development of three critical components should receive top priority to address this problem in upcoming research: adaptive learning systems with human participation frameworks and improved model explainability methods. Combining medical expertise and AI-based decision systems improves model interpretability and trustworthiness, leading to more extensive medical system acceptance.

The diagnostic performance of the model remains optimal, while healthcare facilities that use limited processing hardware in low-resource settings need the model to be optimized. For successful implementation in clinical settings, practitioners need an accurate method to reduce computation time and system memory usage. Enough clinical trial evidence supporting the practical use of DY-FSPAN in real-world medical settings is missing as the current version of the model demonstrates excellent benchmark performance yet its practical implementation remains uncertain for various imaging methods and disease manifestations among different patients with diverse environmental and demographic characteristics whose central repository for training is unavailable at the moment.

The model faces obstacles to obtaining regulatory approval due to its difficulty interpreting between clinical professionals. The feature localization capability of Grad-CAM analysis remains insufficient because healthcare practitioners need deeper explainability methods to make clinical decision-making processes more transparent. AI-driven diagnosis integration into medical practice depends on medical practitioners receiving clear, interpretable explanations that lead to their acceptance of the diagnostic system. Medical data scarcity becomes a constraint for deep learning models because they require extensive, diverse datasets, but such data is often limited.

Future research should use three approaches to improve DY-FSPAN adoption by healthcare professionals. The first step involves integrating DY-FSPAN with clinical decision-support systems, followed by prospective hospital-based trials. The diagnostic accuracy could improve by extending the framework to process combinations of electronic health records alongside radiological images. Developing lightened versions of DY-FSPAN for mobile healthcare applications and edge computing will enhance its usefulness in different medical facilities. AI-driven predictions will be fully adopted in healthcare systems through healthcare professionals' trust when visual representation tools and easy-to-use interfaces are advanced for regulatory approval pathways.

6. Conclusion

The study offers DY-FSPAN as a deep learning platform that performs medical image classification at 98.5 % precision. The model design uses Y-blocks and facilitates attention learning to upgrade feature extraction while preserving field of view stability throughout all layers. The system enhances pathological structure finding by combining features from multiple sizes and selecting features using attention.

Several testing procedures evaluated standard classifications, resistance to errors, and competing methods analysis. The ablation study tested individual parts of the network to show that Y-blocks and attention mechanisms enhance feature selection and performance results. Our model correctly detected the necessary diagnostic areas, confirming its accurate medical performance.

The additional features of DY-FSPAN boost model transparency so it becomes reliable for medical AI implementation in real-world settings. Professional pathologists and radiologists can trust the model because they understand how it decides between different types of cells when it is used for disease classification in medical settings. The model's design effectiveness and clear output mix help DY-FSPAN develop better AI systems for medical imaging applications.

Author Statement

We, the authors, confirm that the work presented in this manuscript is original and has not been published or submitted for publication elsewhere. All authors have made significant contributions to the research, including conceptualization, methodology development, data analysis, manuscript drafting, and critical revisions.

We affirm that this research adheres to ethical standards, and any data used has been properly cited or obtained with necessary permissions. There are no conflicts of interest to declare.

CRediT authorship contribution statement

Tathagat Banerjee: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tathagat Banerjee reports was provided by IIT Patna. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Akbari, Mohammad Amin, et al., 2022. The cheetah optimizer: a nature-inspired metaheuristic algorithm for large-scale optimization problems. *Sci. Rep.* 12 (1), 10953.
- Armato III, S.G., McLennan, G., Bidaut, L., et al., 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* 38, 915–931. <https://doi.org/10.1118/1.3528204>.
- A. Borkowski et al., “Lung and colon cancer histopathological image dataset (LC25000),” arXiv, 2019a, arXiv:1912.12142.
- Borkowski, A.A., Bui, M.M., Thomas, L.B., Wilson, C.P., DeLand, L.A. and Mastorides, S. M., 2019b. *Lung and Colon Cancer Histopathological Image Dataset (LC25000)*. arXiv preprint arXiv:1912.12142.
- Chua, L., Roska, T., 1993. The CNN paradigm. *IEEE Trans. Circuits Syst. I: Fundam. Theory Appl.* 40, 147–156. <https://doi.org/10.1109/81.222795>.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
- FSRT16 (2025) DY-FSPAN: A Feature-Summarized Pyramidal Attention Network for Explainable AI. GitHub. Available at: [\(https://github.com/fsrt16/DY-FSPAN-for-Lung-LC25000\)](https://github.com/fsrt16/DY-FSPAN-for-Lung-LC25000) (Accessed: 29 March 2025).
- Ge, Z., Lin, L., Wang, J., 2021. Lung nodule classification in ct images using 3d densenet. *J. Phys. Conf. Ser.* 1827, 012155. <https://doi.org/10.1088/1742-6596/1827/1/012155>.
- Ghosh, M., Guha, R., Sarkar, R., Abraham, A., 2020. A wrapper-filter feature selection technique based on ant colony optimization. *Neural Comput. Appl.* 32, 7839–7857. <https://doi.org/10.1007/s00521-019-04171-3>.
- Halder, Chatterjee, S., Dey, D., 2022. Adaptive morphology-aided 2-pathway convolutional neural network for lung nodule classification. *Biomed. Signal Process. Control* 72, 103347. <https://doi.org/10.1016/j.bspc.2021.103347>.
- Hanahan, D., Weinberg, R.A., 2011. The hallmarks of cancer: a comprehensive review. *Cell Press* 100 (1), 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- Hatuwal, H.C.T.Bijaya Kumar, 2020. Lung cancer detection using convolutional neural network on histopathological images. *Int. J. Comput. Trends Technol.* 68, 21–24. <https://doi.org/10.14445/22312803/IJCTT-V68I10P104>.
- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” arXiv, 2019, arXiv:1709.01507.
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.
- Jiang, H., Shen, F., Gao, F., Han, W., 2021. Learning efficient, explainable and discriminative representations for pulmonary nodules classification. *Pattern Recognit.* 113, 107825. <https://doi.org/10.1016/j.patcog.2021.107825>.
- M. Kapoor and A. Kasi, “PET scanning,” NCBI, 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov>.
- Kasinathan Gopi, S.J., 2022. Cloud-based lung tumor detection and stage classification using deep learning techniques. Hindawi. <https://doi.org/10.1155/2022/4185835>.
- Malakar, S., et al., 2020. A GA-based hierarchical feature selection approach for handwritten word recognition. *Neural Comput. Appl.* 32, 2533–2552. <https://doi.org/10.1007/s00521-018-3937-8>.
- S. Mangal, A. Chaurasia, and A. Khajanchi, “Convolutional neural networks for diagnosing colon and lung cancer histopathological images,” arXiv, 2020, doi: 10.48550/ARXIV.2009.03878.
- Masud, M., Sikder, N., Nahid, A.A., Bairagi, A.K., AlZain, M.A., 2021. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors* 21, 748. <https://doi.org/10.3390/s21030748>.
- Mehmood, S., et al., 2022. Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. *IEEE Access* 10, 25657–25668. <https://doi.org/10.1109/ACCESS.2022.3150924>.
- Naskar, Pramanik, R., Hossain, S.S., Mirjalili, S., Sarkar, R., 2023. Late acceptance hill climbing aided chaotic harmony search for feature selection: an empirical analysis on medical data. *Expert Syst. Appl.* 219, 119745. <https://doi.org/10.1016/j.eswa.2023.119745>.
- Nishio, M., Nishio, M., Jimbo, N., Nakane, K., 2021. Homology-based image processing for automatic classification of histopathological images of lung tissue. *Cancers* 13, 1192. <https://doi.org/10.3390/cancers13061192>.
- P.R. Patel and D. J. O., “CT scan,” National Library of Medicine, 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov>.
- Pramanik, R., Pramanik, P., Sarkar, R., 2023. Breast cancer detection in thermograms using a hybrid of GA and GWO-based deep feature selection method. *Expert Syst. Appl.* 219, 119643. <https://doi.org/10.1016/j.eswa.2023.119643>.
- Rey, A., Arcay, B., Castro, A., 2021. A hybrid CAD system for lung nodule detection using CT studies based on soft computing. *Expert Syst. Appl.* 168, 114259. <https://doi.org/10.1016/j.eswa.2020.114259>.
- Roy, A., Saha, P., Gautam, N., Schwenker, F., Sarkar, R., 2025. Adaptive genetic algorithm-based deep feature selector for cancer detection in lung histopathological images. *Sci. Rep.* 15 (1). <https://doi.org/10.1038/s41598-025-86362-8>.
- Shaffie, A., et al., 2019. A novel ct-based descriptors for precise diagnosis of pulmonary nodules. 2019 IEEE Int. Conf. Image Process. (ICIP) 1400–1404. <https://doi.org/10.1109/ICIP.2019.8803036>.
- Simonyan, K. and Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1409.1556>.
- Suji, R.J., Godfrey, W.W., Dhar, J., 2024. Exploring pretrained encoders for lung nodule segmentation task using LIDC-IDRI dataset. *Multimed. Tools Appl.* 83, 9685–9708. <https://doi.org/10.1007/s11042-023-15871-3>.
- Xie, Y., Zhang, J., Xia, Y., 2019. Semi-supervised adversarial model for benign-malignant lung nodule classification on chest ct. *Med Image Anal.* 2019 Oct. 237–248. <https://doi.org/10.1016/j.media.2019.07.004>.