JSCDM

JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 7 No. 2 (2026) 330-348
https://publisher.uthm.edu.my/ojs/index.php/iscdm

CTFN: A Hybrid CNN-Transformer Fusion Network with Explainability and Uncertainty Estimation for Enhanced Lung Cancer Classification

Akshay Bhuvaneswari Ramakrishnan[1], Meenalosini Vimal Cruz[2]*, Karthikeyan S[3]

[1] *School of Computing, Allen E. Paulson College of Engineering and Computing, Georgia Southern University, Statesboro, GA 30460, USA*
[2] *Department of Information Technology, Allen E. Paulson College of Engineering and Computing, Georgia Southern University, Statesboro, GA 30460, USA*
[3] *Head – Generative AI, GGS Information Services India Pvt. Ltd., Chennai 600100, Tamil Nadu, INDIA*

*Corresponding Author: mvimalcruz@georgiasouthern.edu
DOI: https://doi.org/10.30880/jscdm.2026.07.02.018

Article Info

Received: 10 October 2025
Accepted: 6 March 2026
Available online: 30 June 2026

Keywords

Lung cancer, histopathological images, convolutional neural network, one-API, transformer

Abstract

Timely identification of cancer significantly improves a patient's likelihood of survival. Medical professionals assert that early diagnosis of lung cancer can decrease mortality rates. Broadly used deep learning systems that depend on convolutional neural networks (CNNs) face multiple challenges, including domain discrepancies and limited dataset sizes combined with contextual information deficits and dependence on pre-trained weights, while dealing with data imbalance issues. We propose a new hybrid framework, Hybrid CNN-Transformer Fusion Network (CTFN), that combines local feature extraction through Convolutional Neural Networks into a system with Transformer modules specialized in capturing the global structural dependencies between features. In addition, the framework integrates explainability methods (Grad-CAM and occlusion sensitivity) and uncertainty estimation to improve interpretability and clinical reliability. The research utilizes Intel one-API to optimize model training on the LC25000 dataset that presents 15,000 histopathology images evenly distributed across three lung cancer classes. Our proposed Hybrid DenseNet-Transformer model achieves a peak test accuracy of 99.11% and a macro-averaged F1-score of 0.9911, demonstrating consistently high performance across all classes. The integration between CNN and Transformer layers offers superior classification results that effectively address existing research deficiencies.

1. Introduction

Among all cancer-related deaths worldwide, lung cancer remains a leading cause of cancer-related deaths, where adenocarcinoma and squamous cell carcinoma represent the leading types. Lung cancer detection, both precisely and in a timely manner, enables improved treatment results and prolonged patient survival [1]. The detection of adenocarcinoma and squamous cell carcinoma, along with benign tissues in histological images, proves

challenging due to the complex morphological characteristics of lung tissue [2]. Traditional deep learning techniques that include Convolutional Neural Networks (CNNs) effectively extract local information from medical imagery as described in reference [3]. Such approaches consistently fail to capture essential global dependencies between image regions, leading to diminished classification accuracy. Transformer models show exceptional performance in processing extended contextual interactions yet demonstrate limitations when differentiating fine details related to tissue localization. We propose a CNN-Transformer hybrid model to improve classification accuracy for histopathology images of lung cancer. In addition, we incorporate explainability tools such as Grad-CAM and occlusion sensitivity, alongside uncertainty estimation, to enhance clinical interpretability and reliability of predictions. Previous studies uncovered multiple performance limitations in existing systems that classify lung cancer pathology.

Standalone CNN models predominantly identify local texture patterns but are inadequate in modeling long-range contextual dependencies essential for histopathological interpretation [4], whereas transformer-based models effectively capture global context but compromise on precise tissue localization. Furthermore, the majority of current methodologies prioritize accuracy exclusively, resulting in restricted interpretability and a lack of mechanisms to assess prediction uncertainty, hence limiting clinical trust [5]. The proposed CTFN methodology addresses these deficiencies by combining CNN-based local feature learning with Transformer-driven global context modeling, while explicitly incorporating explainability and uncertainty estimates to enhance robustness and trustworthiness in lung cancer classification [6].

Additionally, training complex networks entails significant computational costs. Widespread use of pre-trained weights for shortening training time has been observed in studies [7], yet these models might not reach optimal performance for lung cancer detection. A significant problem affects model performance through biased outputs because it tends to ignore data distribution differences between samples. A novel hybrid system has been proposed that unifies CNNs for local feature detection with Transformer components for distant relationship understanding to resolve existing system deficiencies. Unlike prior works, our Hybrid CNN-Transformer Fusion Network (CTFN) explicitly integrates uncertainty-aware prediction and explainability, enabling both higher accuracy and improved trustworthiness. We also provide a comparative evaluation against established baselines (ResNet, DenseNet, and VGG16) to validate the robustness of our approach. The proposed research addresses domain shifting and transfer learning barriers specific to histopathological imaging and improves the efficiency of lung cancer subtype classification. Intel One-API training serves as the proposed method to reduce computational costs and improve operational efficiency. This paper is organized as follows: The second section outlines the methodology by explaining the architecture design and the training approaches. The paper presents experimental findings in Section 3, followed by Section 4, which specifies the effects of these discoveries. The complete analysis closes with Section 5, which suggests upcoming research directions.

2. Literature Review

Lung cancer is one of the most common and lethal malignancies worldwide and is still a challenging problem in early diagnosis and specific categorization. For defining appropriate treatment protocols and improving patient prognoses, the categorization of lung cancer is essential. In recent years, research has centered around the automation of such a classification process through the study of histopathological images and lung computed tomography (CT) scans. This paper attempts to integrate the current research on categorizing lung cancer based on the analysis of significant studies in both fields, focusing on what techniques they employ, what results they are able to achieve, and what limitations they encounter. This survey aims to identify research gaps and enhance diagnostic capabilities in lung cancer categorization. Histopathological image analysis presents unique computational challenges different from radiological modalities like CT. Whole-slide images (WSIs) are gigapixel in size, necessitating analysis at the patch level, which can complicate visualization of global tissue architecture. A fundamental preprocessing requirement is stain normalization [8, 9] to mitigate color variations from variation in tissue preparation, staining protocols, and scanner characteristics. Furthermore, robust evaluation requires WSI-level data splitting to prevent data leakage (where patches from the same patient slide appear in both training and test sets) and effective patch aggregation strategies to render a final, clinically relevant slide-level diagnosis. These domain-specific considerations form the critical context for deep learning in digital pathology. Several studies have focused specifically on lung cancer histopathology images. A deep learning approach in [10] utilized Efficient-Net variants (B0 to B7) to classify histopathology images at resolutions ranging from 224 × 224 pixels to 600 × 600 pixels. Parameters were optimized, and transfer learning was incorporated to enhance results and reduce overfitting. All model variants demonstrated strong accuracy, with EfficientNetB2 achieving 97% accuracy at a resolution of 260 × 260 pixels. Explainable AI frameworks have also been introduced. One study proposed an Efficient-Net-based pipeline integrated with Grad-CAM and SHAP, validated on TCGA-LUAD/LUSC and LC25000 datasets, demonstrating that combining explainability with high accuracy improves clinical trust [11]. Another study [12] introduced a mobile-embedded architecture integrating deep learning into digital pathology image analysis to detect squamous cell carcinomas and adenocarcinomas. The proposed design achieved high evaluation

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

performance while maintaining only 0.65 million trainable parameters and 6.4 million floating-point operations. In contrast, a parallel research thread utilizes CT scans for lung cancer diagnosis, employing distinct methodologies such as radiomics and 3D volume analysis. Chaunzwa et al. [13] proposed CT-based radiomics methods for tumor histology prediction in early-stage non-small cell lung cancer (NSCLC). Their study involved 311 patients and focused on distinguishing two common histological types: squamous cell carcinoma (SCC) and adenocarcinoma (ADC). CNN-based models achieved an AUC of 0.71 ($p = 0.018$), while k-NN and other machine learning classifiers obtained comparable AUC values of 0.71 ($p = 0.017$). SVM-based quantitative radiomics demonstrated similar discrimination performance. Shakeel et al. [14] described automated lung cancer detection from CT scans using an upgraded deep neural network combined with an ensemble classifier. The approach incorporated a multi-level brightness-preserving technique to minimize noise and enhance image resolution. A hybrid spiral optimization intelligence-generalized rough set system was used for affected area selection, followed by deep neural network segmentation, feature extraction, and ensemble-based classification. The method achieved 95.89% accuracy. Bridging these modalities, recent studies have begun to emphasize multi-modality fusion. For example, the ResoMergeNet (RMN) model combined CT, histopathology, and other imaging modalities, showing almost perfect performance on three-class lung cancer classification in controlled, dataset-specific evaluation settings. This shows how useful modality integration can be, but it also shows how protocol-dependent outcomes can be [15]. Other contemporary works highlight robust preprocessing, stratified sampling, and contrast enhancement techniques such as CLAHE combined with explainability evaluation by clinicians, further underlining the importance of trustworthiness in model outputs [16]. Table 1 shows the summary works related to lung cancer classification.

Table 1 *The summary works related to lung cancer classification*

Reference	Year	Model	Modality	Result (Accuracy)	Inference
[17]	2024	EfficientNetB7	Histopathology	99.77%	This particular work Utilized a pre-trained EfficientNetB7 model to categorize lung cancer histopathology images into primary malignancy categories: adenocarcinoma, squamous cell carcinoma, and giant cell carcinoma. The dataset comprises 15,000 entries.
[18]	2023	ResNet50	Histopathology	94.4%	The research adopts LungEffNet as its model for transfer learning applications. Makers designed this model with a modified version of EfficientNet while building its underlying architecture. The performance rates of LungEffNet match those of Convolutional neural networks while using fewer parameters for training procedures.
[19]	2023	XlmNet	CT scan	96.5%	The XlmNet model implemented ELM algorithms for classifying lung cancer histopathological images. Disturbing noise goes first, and standardization of histograms enables better image quality. Enhanced Profuse Clustering (EPC) uses superpixel clustering technology to break down regions of interest in computed tomography (CT) scans. PCA extracts statistical features from pictures through its analysis procedure. The ELM classifier uses system-generated data to identify lung nodules.
[20]	2024	Hybrid-LCSCDM	CT scan	98.54%	The study employs the Hybrid Lung Cancer Stage Classifier and Diagnosis Model (Hybrid-LCSCDM), utilizing a bifurcated methodology. The initial phase of the process involves feature

					extraction utilizing a pre-trained model known as VGG-16 to identify critical elements in lung CT scans that signify malignancy. The last phase entails the classification of these features with a machine learning method called XGBoost, which categorizes the scans into three categories: standard, benign, and malignant.
[21]	2023	Inception V3	CT scan	94.9%	Lung cancer detection uses InceptionV3 as its detection model in this research. The deep learning model achieves outstanding performance following training on computed tomography images from the dataset. Testing occurs through multiple performance metrics, such as accuracy alongside recall and precision.
[22]	2024	(SVM) with Histogram of Oriented Gradients (HOG) Features	CT scan	90%	The research concentrates on deriving both global and local characteristics from the CT scans. Each image is represented by a feature vector that encompasses eight separate categories of image attributes. Three distinct machine-learning techniques are utilized to develop detection models. The research emphasizes the application of Support Vector Machine (SVM) classifiers, which are trained with the retrieved characteristics.

Research papers show that lung cancer classification faces multiple persistent issues in current practice. ImageNet-pre-trained models encounter domain mismatch problems while analyzing the distinctive characteristics of histology and CT lung cancer images. Several methodologies experience dataset limitations, which produce overfitting effects and reduced generalization capacity. Global contextual information gets lost when classifiers at the patch level operate, which leads to classification errors. Both computational cost regulation linked with pre-trained weight usage and disproportionate class frequency distribution in lung cancer datasets cause bias to affect classification outcomes. While novel approaches such as RMN and explainable frameworks show promise, existing works still rarely integrate CNN–Transformer fusion with uncertainty estimation and systematic explainability. This gap motivates the proposed CTFN framework, which combines local–global feature learning with uncertainty-aware, explainable outputs for robust lung cancer subtype classification.

3. Methods and Materials

Our proposed hybrid architectural framework combines fine-tuned CNN backbones with domain-specific layers and Transformer modules tailored for pathology images. We applied stain normalization to address color variations and establish consistent image quality across histopathological slides across histopathological slides. We leveraged Intel oneAPI to accelerate computations, optimize memory usage, and improve training efficiency. Our design combines multi-scale strategies by using CNN classification with transformer-based attention to capture both local details and global contextual dependencies simultaneously. The approach successfully resolves earlier research limitations and contributes to building an efficient and durable system for lung cancer classification while also integrating explainability (Grad-CAM, occlusion sensitivity) and uncertainty estimation to enhance clinical trustworthiness.

3.1 Data Source

This study employs a collection of histopathological images of lung tissue samples, classified into three categories: Lung Squamous Cell Carcinoma (LUSC), Lung Adenocarcinoma (LUAD), and Lung Benign Tissue (LBT). A total of 15,000 histopathology image patches [23] were used, with 5,000 images per class to ensure balanced class distribution. As the LC25000 dataset is derived from a limited number of whole-slide images (WSIs) through extensive augmentation and does not provide explicit slide or patient identifiers, a leakage-aware data partitioning strategy was employed. Images originating from the same pre-augmentation source (i.e., an original image and its augmented variants) were treated as a single group and assigned exclusively to either the training, validation, or testing split to prevent patch-level leakage. The research obtained its images through public

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

institutions and clinical databases, providing an assorted collection of samples encompassing diverse demographic groups and disease conditions, thereby enhancing the generalizability and robustness of the proposed framework.

3.2 Data Splitting and Leakage Control

The LC25000 dataset originates from a limited number of whole-slide histopathology images that were expanded through aggressive data augmentation. Since explicit slide or patient identifiers are not available in public release, random patch-level splitting can lead to information leakage due to high similarity between augmented samples. To mitigate this risk, a grouped data splitting strategy was adopted, where each original image and its associated augmented variants were treated as a single group representing a proxy for slide origin. These groups were then stratified at the class level and assigned exclusively to training, validation, or test sets, ensuring that no augmented variants from the same source image appeared across different splits. This approach provides a conservative and leakage-aware evaluation protocol appropriate for augmented histopathology datasets.

3.3 Preprocessing

Our pre-processing pipeline adopted input image normalization alongside stratified dataset splitting and subset-specific transformations. Images were converted to the [0,1] range, followed by ImageNet-standard normalization to match pre-trained CNN specifications. We applied stain normalization methods to address color inconsistencies during image processing [24, 25]. The image transformation pipeline involved conversion to the LAB color space, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) applied to the L-channel, before recombination with the A and B channels to reconstruct the RGB image. CLAHE was applied using standard default parameters to enhance local contrast and reduce intensity variability. The experimental approach elevated image clarity while simultaneously reducing uncertainty caused by staining methods. The pre-processing workflow's assessment through Figure 1 reveals how stain normalization techniques deliver improved visual effects.

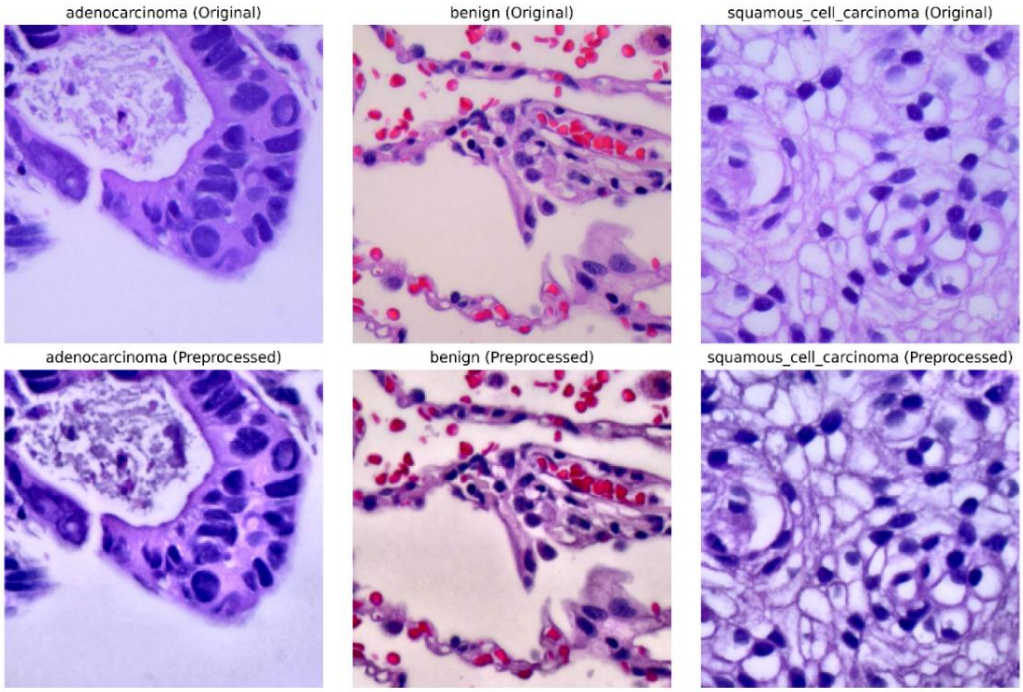

Fig. 1 *Before and after stain normalization*

Importantly, stain normalization and CLAHE were applied only to the training and validation subsets to prevent test-time distribution shift and data leakage; the test set underwent only resizing and normalization. For dataset partitioning, a leakage-aware stratified data partitioning strategy was employed, with a 70%/15%/15% split for training, validation, and testing. As the LC25000 dataset originates from a limited number of whole-slide images and lacks explicit slide or patient identifiers, image patches were grouped based on their pre-augmentation source, serving as a proxy for slide origin. These groups were stratified at the class level and assigned exclusively to a single subset to preserve class balance while preventing patch-level information leakage.

3.4 Architecture Design

This study develops a hybrid CFTN framework that combines Convolutional Neural Networks (CNNs) and Transformer models to boost image classification capacity through the combined utilization of local and global features. The CNN backbone extracts local edge and texture patterns through convolutions, producing feature maps that are subsequently processed by Transformer modules. The transformer analyzes image data as sequential patches through self-attention mechanisms, which detect and analyze long-range dependencies and spatial relationships between image regions. In the proposed architecture, CNN-extracted feature maps were partitioned into fixed-size patches of 16 × 16 pixels, which were flattened and linearly projected into a latent embedding space of dimension 768. These embeddings were processed by a Transformer encoder composed of 12 stacked layers with 12 attention heads per layer to model global contextual dependencies. Dropout with a rate of 0.1 was applied within the self-attention and feedforward layers to reduce overfitting during training. Both components produce merged outputs before they enter dense layers to run through classification while maintaining image comprehension quality. The hybrid architecture displayed in Figure 2 integrates Convolutional Neural Networks (CNNs) to extract regional features alongside Transformer blocks to understand global contexts for lung cancer histopathology image classification.

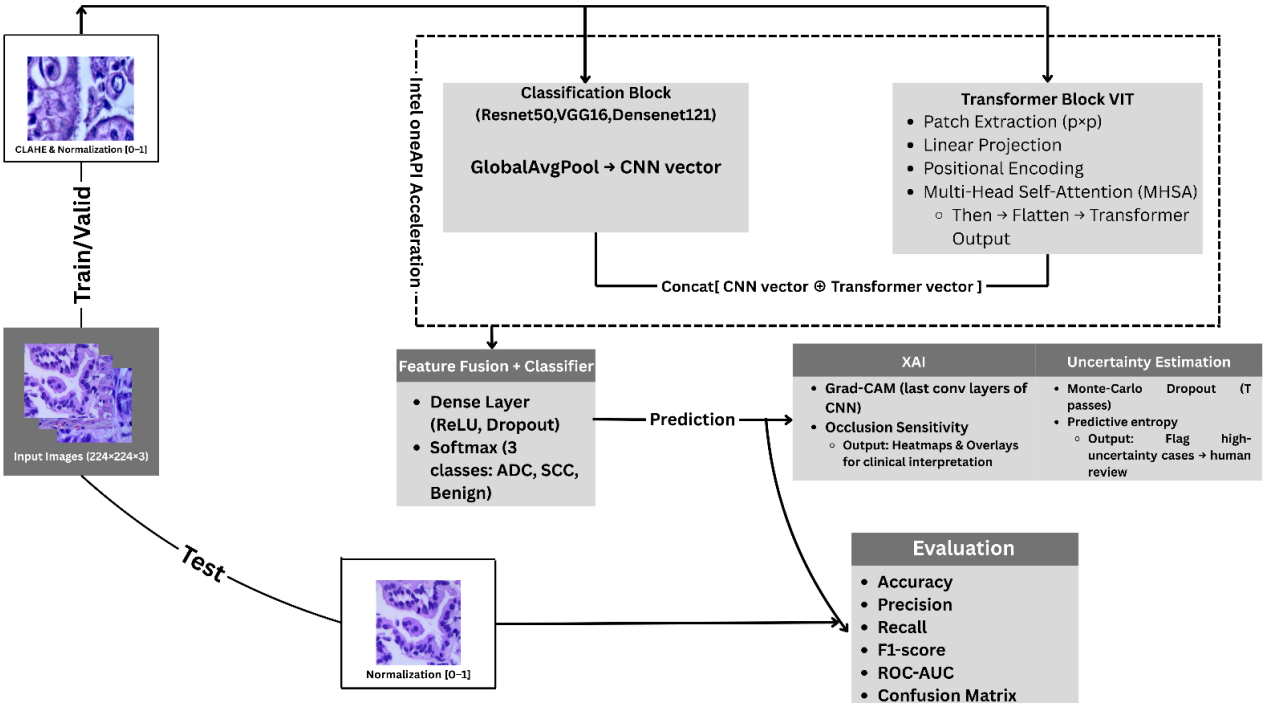

Fig. 2 *CNN-transformer fusion network (CTFN) architecture*

This combined approach boosts feature extraction through CNN network processing of local patterns and Transformer capabilities for complete and in-depth image composition understanding. With its capability to better understand both local and global image context, the integrated model provides excellent performance for tasks that require holistic interpretation, especially medical imaging and fine-grained visual categorization.

3.4.1 Formal Model Description

The input image is represented as $X \in R^{H*W*C}$, $C = 3$ where H represents image height, W denotes image width, and C denotes RGB channels. The images are subjected to stain normalization and contrast-limited adaptive histogram equalization (CLAHE) for the training and validation subsets, thereafter normalized to the range [0, 1]. Preprocessing guarantees uniform intensity distribution among samples, reducing color variations caused by staining.

3.4.2 CNN-Based Local Feature Extraction

The convolutional neural network backbone extracts the local (structural and textural) features. We utilized Densenet121, Resnet50 and VGG16 models due to their widespread adoption in medical imaging. Equation 1 represents the feature maps computed by the CNN backbone:

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

$$F_{CNN} = f_{CNN}(X), F_{CNN} \in R^{h*w*d} \quad (1)$$

Where f_{cnn} denotes the convolutional-pooling pipeline, h and w represent the spatial dimensions of the feature maps, and d represents the number of channels.

- DenseNet121 facilitates feature reuse by dense connectedness, promoting efficient gradient propagation.
- ResNet50 utilizes residual blocks to mitigate vanishing gradient problems in deep architectures.
- VGG16 offers a somewhat shallow baseline, with consistent 3×3 convolutions throughout its layers.

These options provide a range of model complexity and feature richness.

3.4.3 Transformer-Based Global Feature Encoding

The transformer model partitions the images into non-overlapping patches to capture contextual global information Equation 2.

$$P_i \in R^{p*p*C}, i = 1,2,3 \ldots . n \quad (2)$$

Where p is the patch size, $\boldsymbol{C}$ is the number of channels and $\boldsymbol{n} = \frac{\boldsymbol{HW}}{\boldsymbol{p^2}}$ is the number of patches. Each patch is then flattened and projected linearly into a latent embedding space Equation (3):

$$z_i = W_p . Flatten(P_i) + E_{pos}, z_i \in R^d \quad (3)$$

Where $\boldsymbol{W_p}$ denotes the linear projection matrix and $\boldsymbol{E_{pos}}$ represents the positional embeddings and $\boldsymbol{d}$ is the embedding dimension. This will result in a sequence of embeddings $\boldsymbol{z} = [\boldsymbol{z_1}, \boldsymbol{z_2} \ldots \ldots ., \boldsymbol{z_n}]$ which is then processed by the transformer encoder consisting of a multi-head-self-attention mechanism and a feed-forward neural network Equation (4):

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (4)$$

With queries $\boldsymbol{Q} = \boldsymbol{XW_Q}$, keys $\boldsymbol{K} = \boldsymbol{ZW_K}$ and Values $\boldsymbol{V} = \boldsymbol{ZW_V}$ are linear projections of the input embeddings $\boldsymbol{Z}$, and $\boldsymbol{d_k}$ denotes the dimensionality of the key vectors. This mechanism helps to model long range dependencies across image regions which is critical for detecting structural anomalies distributed over tissue areas.

3.4.4 CNN–Transformer Fusion

For integrating the individual representations extracted from CNN and Transformer we perform feature fusion Equation (5):

$$F = [GlobalAvgPool(F_{CNN}) \oplus F_{tr}] \quad (5)$$

Where $\oplus$ denotes the concatenation operation, GlobalAvgPool(.) reduces CNN features to compact descriptor and F_{tr} represents the Transformers global embedding. Micro and macro label tissue morphology will be considered by the fused network addressing the limitations of single branch models.

The fused vector is then passed through fully connected layers followed by a SoftMax classifier Equation (6):

$$\hat{y} = Softmax(W_c F + b_c) \quad (6)$$

Where W_c *and* b_c represents trainable parameters. Probability distribution over three classes is represented by output $\hat{y}$.Categorical Cross-entropy loss is used to optimize the model Equation (7):

$$L = -\frac{1}{N}\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) \quad (7)$$

where $\boldsymbol{y_{i,c}}$ is the ground truth label (one-hot encoded), $\boldsymbol{\hat{y}_{i,c}}$ is the predicted probability for class c, $\boldsymbol{C}$ is the number of classes, and N is the number of training samples. Training employed Adam optimizer with learning rate $\boldsymbol{1\ X\ 10^{-4}}$, batch size of 32, and early stopping based on validation accuracy.

3.5 Explainability

Even though excellent metrics are required to validate the working of AI models to overcome the 'black box' issue the models we develop has to be interpretable. Especially in medical context it is important for the clinicians to interpret the working of our model to solve this issue we integrated several explainability strategies to our framework.

3.5.1 Gradient-based Class Activation Mapping (Grad-CAM)

Gradient-weighted Class Activation Mapping (Grad-CAM) can be used to understand the spatial regions that are highly influential for the model's decision [26]. We do this by applying Grad-CAM in the final convolutional layers of the CNN backbone [27]. For a given class score $\boldsymbol{y^c}$and feature maps $\boldsymbol{A^k}$ from the last convolutional layer, Grad-CAM weights are computed as Equation (8):

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (8)$$

The class descriptive heatmap is defined as Equation (9):

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (9)$$

The resulting heatmaps provide clinicians with interpretable cues that validate whether the model's decision-making aligns with pathological hallmarks.

3.5.2 Occlusion Sensitivity Analysis

In addition to Grad-CAM to measure the robustness of the predictions, occlusion sensitivity analysis was implemented. The classification confidence decrease was recorded by masking small image portions [28, 29]. Critical regions for decision-making were those whose removal significantly reduced prediction likelihood. This method provides an additional layer of validation by confirming that highlighted features align with genuinely significant structures rather than false correlations. Small areas of an image were systematically obscured, and the resultant decrease in classification confidence was documented. Areas whose elimination led to substantial decreases in prediction likelihood were recognized as essential for decision-making.

3.5.3 Uncertainty Estimation (for Clinical Safety)

Uncertainty estimates were used to determine the predictive reliability beyond just visual inspections [30]. Monte Carlo drop-out was employed at inference, whereby multiple stochastic forward passes through the network approximated the posterior distribution over predictions Equation (10). From these, predictive entropy Equation (11) and variance Equation (12) were computed as measures of epistemic and aleatoric uncertainty [31]. Let $\boldsymbol{p_t^c}$ be the softmax probability for class c and pass t:

$$\hat{P}^c = \frac{1}{T}\sum_{t=1}^{T} P_t^c \quad (10)$$

$$H(y|x) = -\sum_c \hat{P}^c \log \hat{P}^c \quad (11)$$

$$Var^c = \frac{1}{T}\sum_{t=1}^{T}\left(P_t^c - \hat{P}^c\right)^2 \quad (12)$$

3.5.4 Clinical Implications

The incorporation of Grad-CAM, occlusion sensitivity, and uncertainty estimates converts the proposed CNN–Transformer hybrid from a "black-box" classifier into an interpretable decision-support system. The paradigm addresses a major barrier to clinical implementation by highlighting relevant histological structures, validating their significance through disruption, and assessing the reliability of predictions. These explainability measures not only bolster pathologist trust but also function as a safeguard by referring doubtful or uncommon instances for human review. This enhances the translational significance of the approach in a standard lung cancer diagnosis.

3.6 One-API Integration

The proposed architecture uses Intel's oneAPI combined with the oneDNN library on an ASUS 14 NUC Pro AIPC device with an Intel Core Ultra 7 processor to improve hybrid architecture performance speeds. The one-API framework delivers effective parallelization and resource management for CPU cores, which results in improved performance across training and inference procedures [32]. With optimized deep learning primitives from oneDNN, we achieved speed and efficiency enhancements, reaching 60% during training and 47% during inference. The incorporation of one API optimizes the development process and guarantees effective scalability across various hardware configurations, rendering it appropriate for real-time medical diagnostics applications. One-API Deep Neural Network Library (oneDNN) offers highly optimized implementations of fundamental components for deep learning [33]. The library provides optimized implementations of essential components such as convolution, matrix multiplication, pooling, batch normalization, and activation functions. Enhance the efficacy of existing frameworks, including the OpenVINO™ toolkit, Intel's AI Tools, PyTorch*, and TensorFlow*. Figure 3 illustrates the Intel architecture and the oneDNN library utilization.

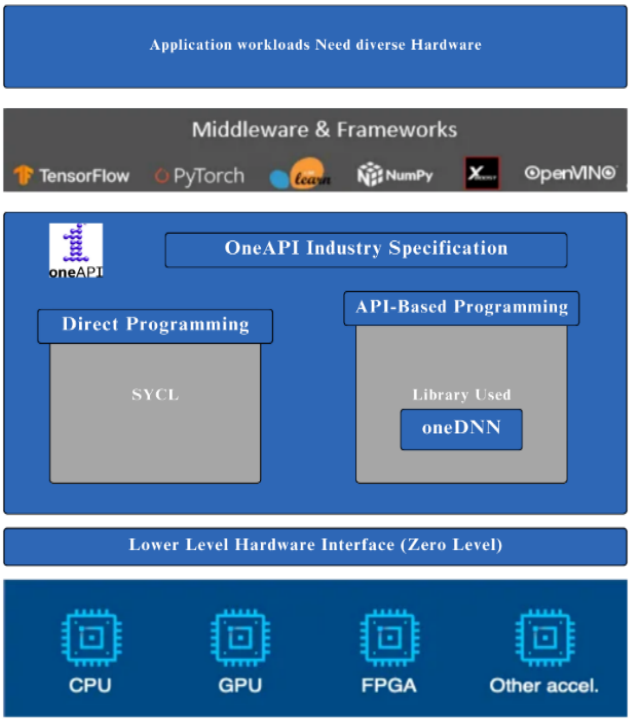

Fig. 3 *One-API architecture*

4. Results and Discussion

The performance of our proposed hybrid CFTN architecture was compared against baseline convolutional architectures (ResNet50, DenseNet121, and VGG16). Models were assessed using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices across the three histological categories: adenocarcinoma (ADC), squamous cell carcinoma (SCC), and benign.

4.1 Overall Performance Comparison

The aggregated overall performance metrics of baseline and hybrid models are presented in Table 2. Overall the baseline models achieved high accuracy (0.89–0.98). The metrics reported are micro-averaged across classes. As the evaluation dataset is perfectly balanced, micro-averaged precision, recall, and F1-score coincide numerically with overall accuracy. To avoid masking class-specific behavior that may not be reflected in micro-averaged metrics, detailed class-wise precision, recall, and F1-scores are presented in Section 4.2 (Tables 3–5). We can clearly see that the classification results increased after the addition of the transformer module to the baseline architecture.

Table 2 *Overall performance of baseline and hybrid models*

Model	Accuracy	Precision	Recall	F1-score	Macro PR-AU
ResNet50	0.8185	0.8425	0.8185	0.8165	-
Hybrid ResNet	0.8553	0.8776	0.8553	0.8441	0.945
DenseNet121	0.9876	0.9876	0.9876	0.9876	-
Hybrid Dense	0.9911	0.9911	0.9911	0.9911	0.994
VGG16	0.9400	0.9500	0.9400	0.9400	-
Hybrid VGG16	0.9700	0.9700	0.9700	0.9700	0.987

Across all the models, the DenseNet family demonstrated superior performance. The baseline DenseNet121 achieved an impressive accuracy of 98.76%, which was further improved by the hybrid CNN–Transformer variant to 99.11%, highlighting the role of Transformers in refining global contextual understanding and reducing residual misclassifications. The VGG16 model emerged as the second-best performer. The baseline VGG16 achieved 94% accuracy, whereas its hybrid extension enhanced classification to 97% accuracy, demonstrating significant advantages of fusion. In comparison, ResNet50 exhibited inferior performance overall, with the baseline model achieving merely 81.85% accuracy and the hybrid variant increasing little to 85.53%. Macro-averaged PR-AUC is reported for the hybrid models to assess confidence-based robustness of the proposed CNN–Transformer architectures, while baseline models are retained primarily for comparative accuracy benchmarking.

4.2 Class-Wise Performance

To understand the model's reliability and its understanding across the histopathological subtypes, we did a detailed analysis on class-wise performance. The Hybrid VGG16 (Table 3) demonstrated strong performance with F1-scores of 0.96 for adenocarcinoma and squamous cell carcinoma, achieving perfect classification of benign tissue. The most notable misclassifications in our model were inside the squamous cell carcinoma category, with 54 samples erroneously categorized as adenocarcinoma, resulting in a recall of 0.93.

To further evaluate robustness, each hybrid model was trained and evaluated across 10 independent runs with different random seeds. Hybrid VGG16 achieved a macro-F1 of 0.972 ± 0.0025 (95% CI: 0.970–0.974). The class-wise recall across runs was 0.986 ± 0.006 for adenocarcinoma, 0.998 ± 0.002 for benign tissue, and 0.932 ± 0.010 for squamous cell carcinoma, confirming stable performance despite mild SCC variability. The average total misclassifications per run were 68 ± 11, predominantly SCC to adenocarcinoma errors.

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

Table 3 *Class-wise performance of Hybrid VGG16*

Class	Precision	Recall	F1-score	Support
Adenocarcinoma	0.93	0.99	0.96	751
Benign	0.99	1.00	1.00	751
Squamous Cell Carcinoma	0.99	0.93	0.96	751

Models based on ResNet Table 4 encountered significant challenges in the detection of adenocarcinoma. The baseline ResNet erroneously classified 258 adenocarcinoma cases as benign, resulting in a recall of only 0.65. In the hybrid ResNet model, the recall for adenocarcinoma decreased to 0.58, while the precision increased to 0.98. This indicates that whereas the hybrid architecture enhanced the detection of benign and squamous carcinomas, it compromised sensitivity for adenocarcinoma, resulting in a notable imbalance.

Across 10 runs, Hybrid ResNet achieved a macro-F1 of 0.845 ± 0.021 (95% CI: 0.830–0.860), indicating substantially higher variability compared to DenseNet and VGG16 variants. Adenocarcinoma recall showed marked instability (0.590 ± 0.040), explaining the performance gap. The model averaged 314 ± 32 misclassifications per run, primarily driven by adenocarcinoma to benign errors.

Table 4 *Class-wise performance of hybrid ResNet*

Class	Precision	Recall	F1-score	Support
Adenocarcinoma	0.98	0.58	0.73	751
Benign	0.88	1.00	0.93	751
Squamous Cell Carcinoma	0.78	0.99	0.87	751

Near-perfect recall and precision for all categories were achieved by the hybrid DenseNet model with benign tissue, achieving 100% recall and precision, and both adenocarcinoma and squamous cell carcinoma maintaining F1 scores above 0.98 (Table 5).

Robustness analysis demonstrated exceptional stability for Hybrid DenseNet, achieving a macro-F1 of 0.993 ± 0.0011 (95% CI: 0.992–0.994). Class-wise recall across runs was 0.987 ± 0.004 for adenocarcinoma, 0.999 ± 0.001 for benign tissue, and 0.990 ± 0.003 for squamous cell carcinoma. The model averaged only 15 ± 5 misclassifications per run, primarily between SCC and adenocarcinoma, confirming both high accuracy and low variance across independent trials.

Table 5 *Class-wise performance of hybrid DenseNet.*

Class	Precision	Recall	F1-score	Support
Adenocarcinoma	0.99	0.98	0.99	751
Benign	1.00	1.00	1.00	751
Squamous Cell Carcinoma	0.98	0.99	0.99	751

Based on the class-wise F1-scores, the Hybrid DenseNet achieves a macro-averaged F1-score of 0.993, followed by Hybrid VGG16 (0.973) and Hybrid ResNet (0.843), confirming consistent performance across classes.

4.3 ROC-AUC Analysis

The ROC curve analysis presented in Figure 4 validated these findings. DenseNet121 and VGG16, in both baseline and hybrid configurations, attained nearly flawless class separability with AUC scores nearing 1.00 across all categories. The hybrid DenseNet and hybrid VGG16 achieved flawless AUC values, guaranteeing exceptional discriminative performance. ResNet50 exhibited diminished separability for adenocarcinoma, achieving AUC values near 0.90, although its hybrid variant enhanced AUC but did not attain the robustness of DenseNet and

VGG16. While ROC-AUC evaluates class separability across thresholds, macro-averaged PR-AUC (reported in Table 2) further assesses confidence ranking under class-balanced conditions, providing complementary validation of model robustness.

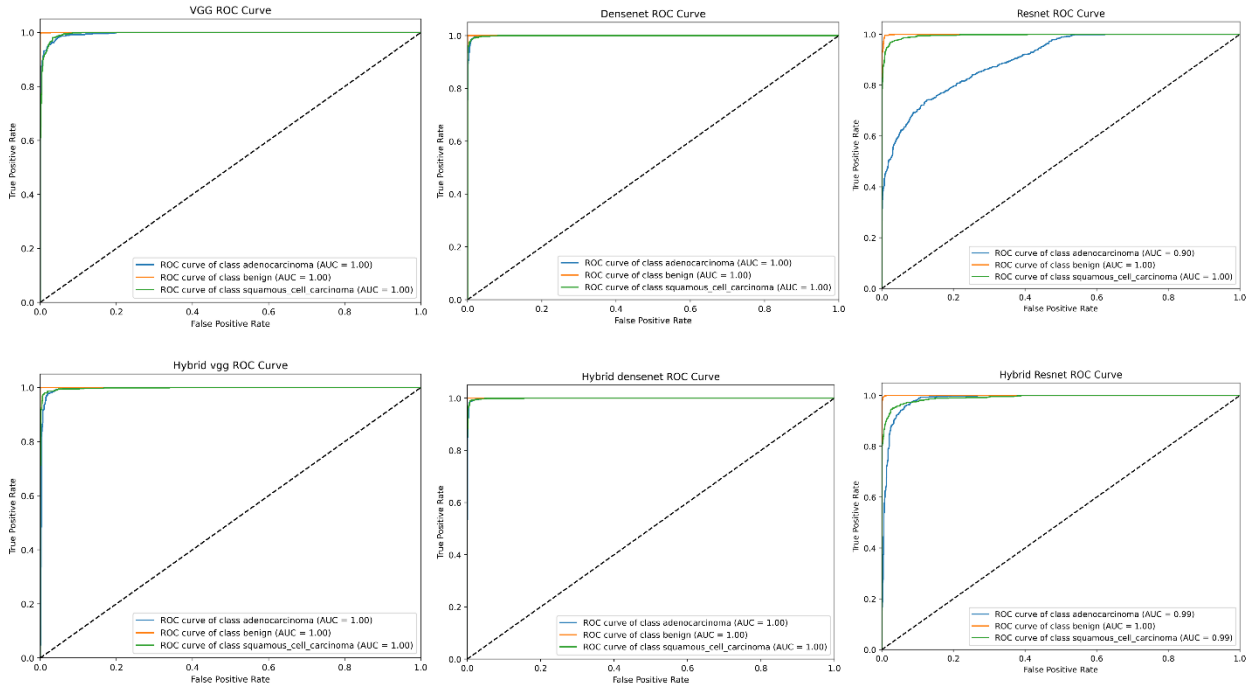

Fig. 4 *ROC–AUC curves for baseline and hybrid models. Hybrid DenseNet121 and Hybrid VGG16 achieved near-perfect separability with AUC ≈ 1.00 across all classes, demonstrating the effectiveness of transformer integration in capturing global dependencies*

4.4 Confusion Matrix Analysis

The examination of the confusion matrix elucidates mistake patterns. As shown in Figure 5 ResNet50 had a pronounced inclination to misclassify cancer as benign, whereas hybrid ResNet partially rectified this issue; however, significant deficiencies in recall persisted. VGG16 exhibited robust generalization yet displayed ambiguity between squamous cell carcinoma and adenocarcinoma, accounting for its marginally diminished SCC recall. DenseNet, conversely, consistently achieved the most distinct distinction, with the hybrid variation minimizing errors to merely 8 misclassified squamous carcinoma samples, demonstrating its efficacy in differentiating nuanced histological characteristics.

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

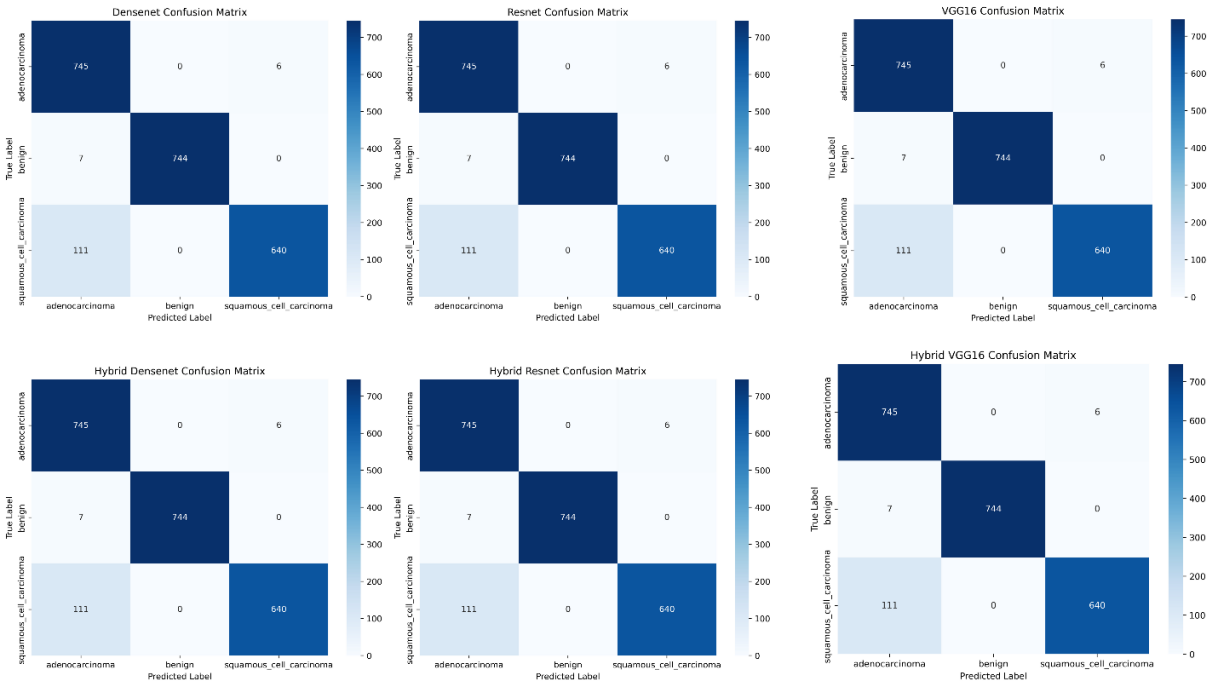

Fig. 5 *Confusion matrices of baseline and hybrid models (ResNet50, DenseNet121, and VGG16) on the lung cancer histopathology dataset. The diagonal elements indicate correct predictions, while off-diagonal cells highlight common misclassifications, particularly between adenocarcinoma and squamous cell carcinoma*

In conclusion, the DenseNet-based hybrid model demonstrated superior efficacy, with near-perfect classification for all three lung cancer subtypes, closely succeeded by the hybrid VGG16, which provided consistent and equitable enhancements over its baseline performance. Hybrid ResNet, despite a marginal enhancement in accuracy, continued to be constrained by its inadequate sensitivity to adenocarcinoma. The results collectively affirm that hybrid CNN–Transformer architectures significantly improve histopathology image categorization, with DenseNet–Transformer fusion emerging as the most dependable configuration for clinical use.

5. Explainability and Uncertainty Analysis

Explainability evaluations were performed using Grad-CAM and occlusion sensitivity to analyze model decisions. Figure 6. Grad-CAM heatmaps frequently emphasized nuclear and cytoplasmic regions exhibiting aberrant morphology, corroborating the histopathological characteristics utilized by pathologists. Occlusion sensitivity maps corroborated the reliability of predictions by revealing localized decreases in confidence when diagnostically pertinent areas were obscured. In cancer, Grad-CAM underscored dense cellular aggregates, whereas benign samples showcased structured tissue margins. Visualizations of squamous cell carcinoma indicated a significant focus on keratinizing areas, thereby substantiating the biological validity of the acquired features.

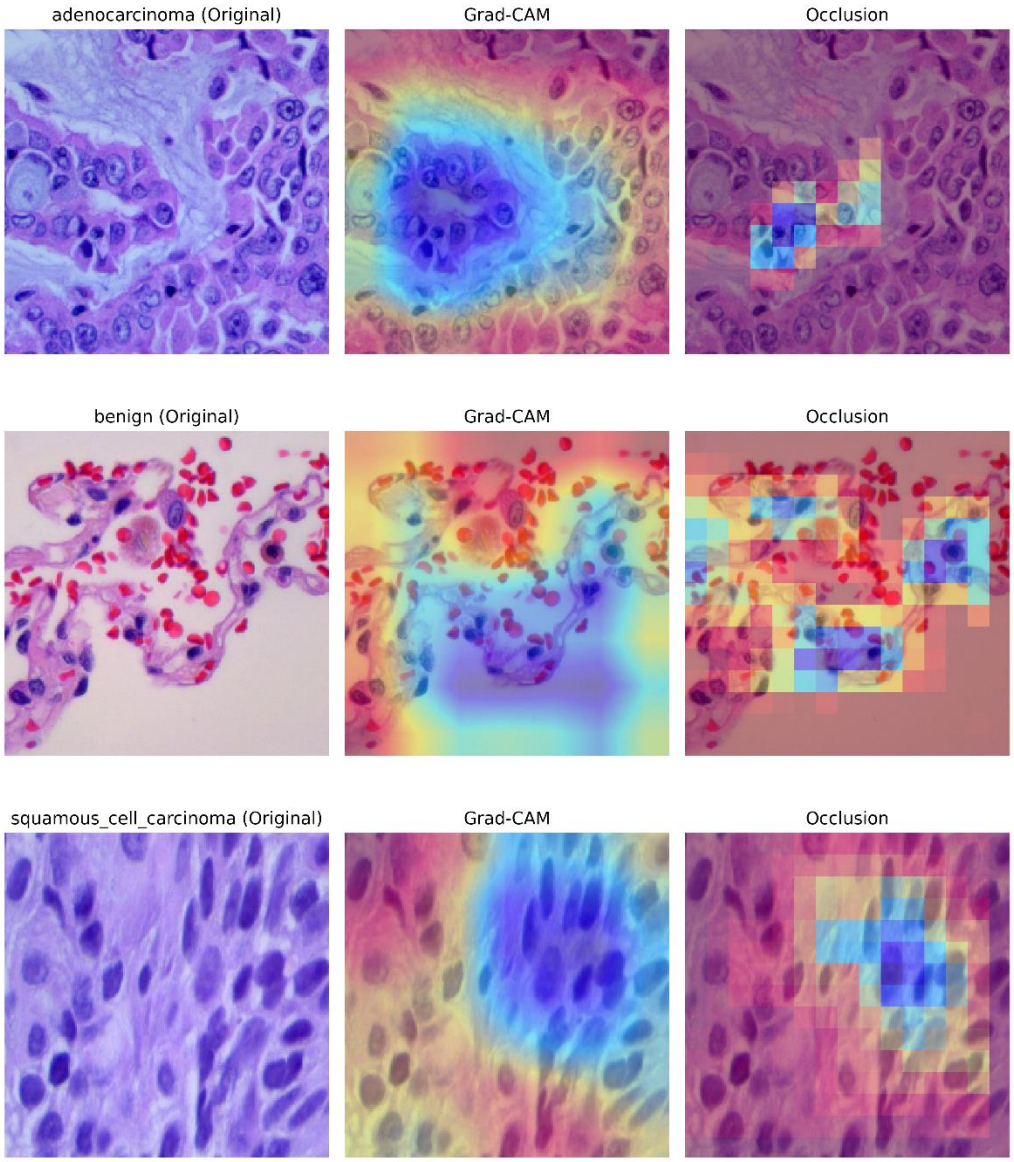

Fig. 6 *Results of explainability demonstrating original pictures, Grad-CAM visualizations, and occlusion sensitivity overlays for adenocarcinoma, benign, and squamous cell cancer specimens*

In addition to interpretability, uncertainty estimation was utilized to evaluate dependability in clinical applications. Predictive entropy distributions Figure 7a demonstrated that the majority of test samples displayed minimal entropy (<0.05), signifying high-confidence predictions. Box plots comparing accurate and incorrect predictions Figure 7b indicated that misclassified instances exhibited significantly greater uncertainty (mean entropy = 0.4355) than correctly classified instances (0.0228). The calibration analysis using a reliability diagram Figure 7c indicated that the hybrid DenseNet model is predominantly well-calibrated, exhibiting only minor overconfidence in mid-probability bins.

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 330-348

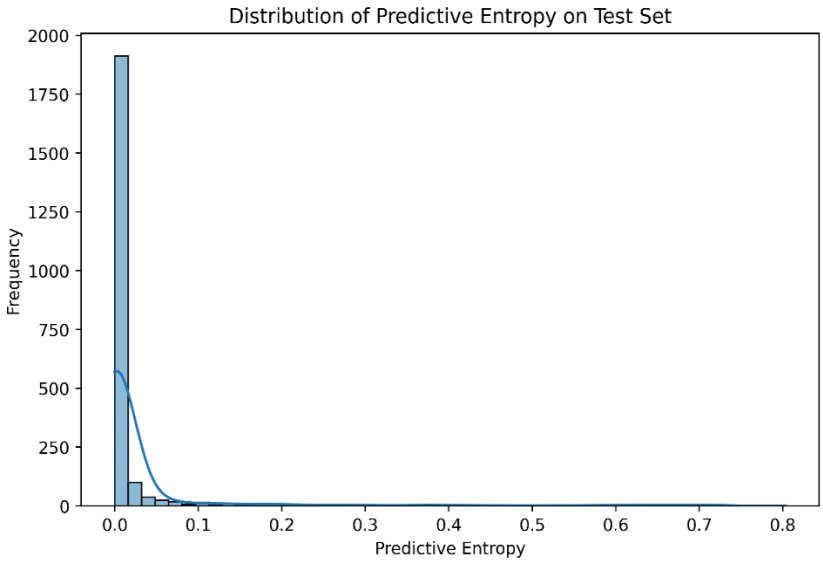

Fig. 7 *a) Distribution of predictive entropy values across the test set*

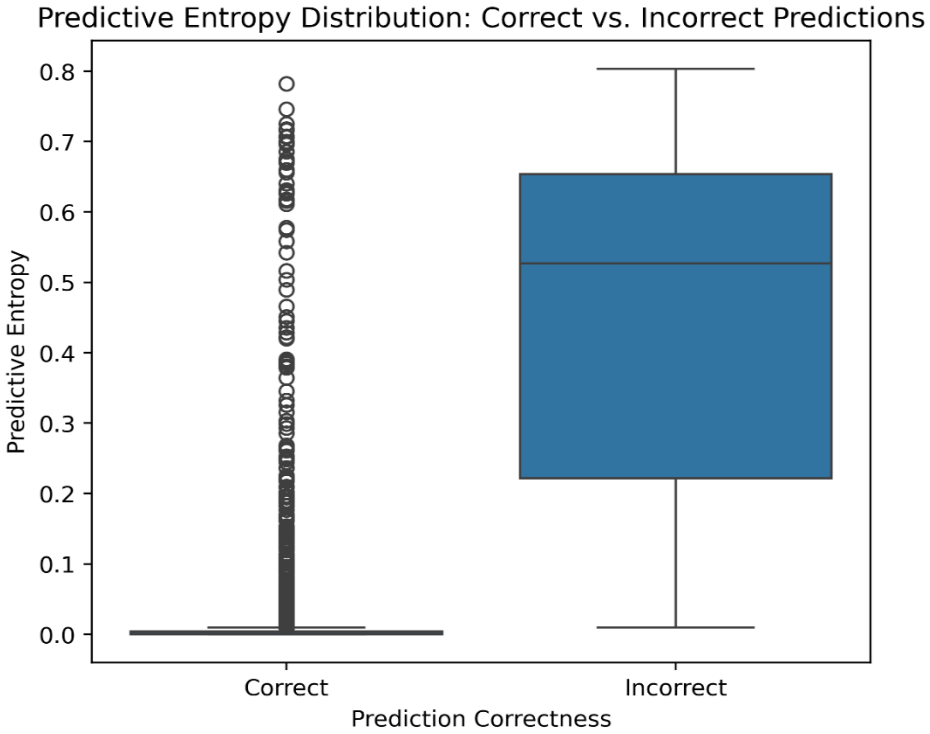

Fig. 7 *b) Predictive entropy comparison between correct and incorrect predictions*

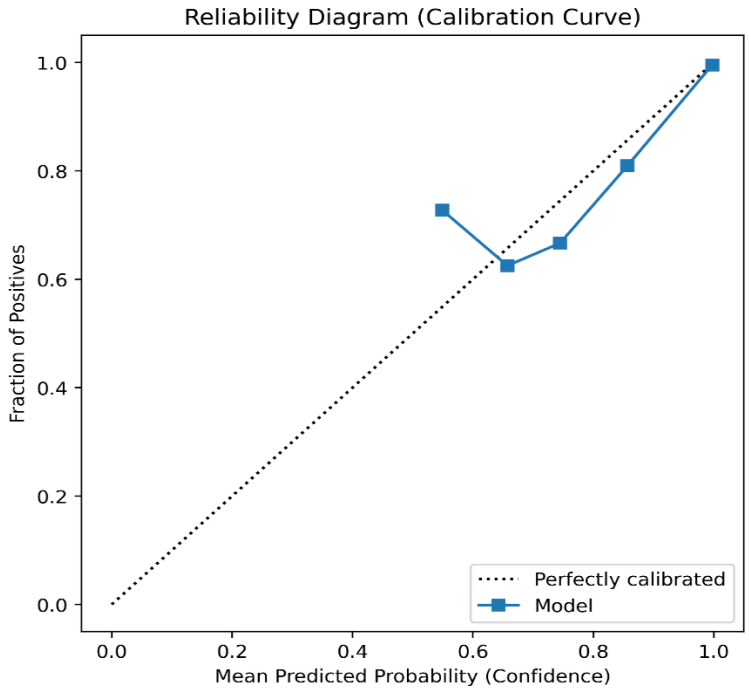

Fig. 7 *c) Reliability diagram showing model calibration across probability bins*

The aggregated results indicated that 91.1% of test samples were categorized in the low-uncertainty category (entropy < 0.05), with an accuracy of 99.9%, whilst the remaining 8.9% of high-uncertainty samples were classified with diminished accuracy (87.0%). This robustly endorses the efficacy of uncertainty estimate as a triaging tool to identify confusing cases for additional human evaluation.

6. Future Works

Subsequent study will concentrate on enhancing the clinical interpretability of the hybrid architecture. Although Grad-CAM offered valuable spatial insights, supplementary explainable AI (XAI) methodologies, like attention visualization from Transformer heads, integrated gradients, and saliency maps, may facilitate the connection between model predictions and pathologist thinking.

A significant focus is multimodal integration. Integrating supplementary information, like genetic profiles, radiological scans, and clinical metadata, could produce a more comprehensive diagnostic model. This corresponds directly with our next initiative, where multimodal survival analysis and prognostic modeling are being formulated. Expanding the hybrid CNN–Transformer framework to accommodate multi-source data might enhance diagnostic precision and prognostic potential in lung and other malignancies.

7. Conclusion

This research introduced a hybrid CNN–Transformer architecture for the categorization of lung cancer histopathology, combining CNNs for local feature extraction with Transformers for global contextual modeling. The method exhibited enhanced performance relative to baseline CNNs, with the hybrid DenseNet–Transformer attaining 99.11% accuracy, precision of 1.000 for benign class, and balanced recall for adenocarcinoma and squamous cell carcinoma classes.

Utilization of Intel oneAPI and oneDNN facilitated effective parallelization, decreasing training duration by 60% and inference expenses by 47%, highlighting the framework's scalability for high-performance computing settings. Grad-CAM and occlusion sensitivity demonstrated that model attention corresponded with histological features, while predictive uncertainty estimates introduced an additional safety measure by identifying ambiguous cases for human evaluation.

All experiments were conducted using Python 3.10, PyTorch 2.1.0, and Intel oneAPI 2024.0 with oneDNN optimizations under CPU-only execution. Training utilized 16 OpenMP threads with MKL acceleration. Hybrid DenseNet required approximately 42 seconds per epoch (batch size 32), compared to 68 seconds for the baseline model. Inference averaged 3.1 ms per sample versus 5.9 ms for the baseline. All training, evaluation, and benchmarking scripts will be publicly released upon acceptance to ensure reproducibility. Despite issues in interpretability and computational demands, the results underscore the promise of hybrid

architectures in enhancing digital pathology. The suggested system, by integrating strong classification efficacy, uncertainty recognition, and hardware optimization, signifies a substantial advancement towards clinically applicable AI technologies. This paradigm, through enhanced explainability, multimodal data integration, and streamlined deployment, can facilitate expedited, precise, and dependable lung cancer diagnostics, hence enhancing patient outcomes.

Acknowledgement

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, for providing the research and publication support. The authors also acknowledge the computational resources provided through Intel oneAPI and ASUS AI PC (Intel Core Ultra 7 processor) for model optimization and experimentation.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design****: Akshay Bhuvaneswari Ramakrishnan, Meenalosini Vimal Cruz, Karthikeyan S;* ***data collection****: Akshay Bhuvaneswari Ramakrishnan;* ***analysis and interpretation of results****: Akshay Bhuvaneswari Ramakrishnan, Meenalosini Vimal Cruz, Karthikeyan S;* ***draft manuscript preparation****: Akshay Bhuvaneswari Ramakrishnan, Meenalosini Vimal Cruz. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Charpidou, A., Hardavella, G., Boutsikou, E., Panagiotou, E., Simsek, G. Ö., Verbeke, K., ... & Domagała-Kulawik, J. (2024). Unravelling the diagnostic pathology and molecular biomarkers in lung cancer. Breathe, 20(2).

[2] Cocuz, I. G., Popelea, M. C., Niculescu, R., Manea, A., Sabău, A. H., Tinca, A. C., ... & Cotoi, O. S. (2024). Pathophysiology, Histopathology, and Differential Diagnostics of Basal Cell Carcinoma and Cutaneous Squamous Cell Carcinoma—An Update from the Pathologist's Point of View. International Journal of Molecular Sciences, 25(4), 2220.

[3] Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. IEEE Access, 12, 41180-41218.

[4] Faria, N., Campelos, S., & Carvalho, V. (2023). A novel convolutional neural network algorithm for histopathological lung cancer detection. Applied Sciences, 13(11), 6571.

[5] Guo, Y., Song, Q., Jiang, M., Guo, Y., Xu, P., Zhang, Y., ... & Yao, X. (2021). Histological subtypes classification of lung cancers on CT images using 3D deep learning and radiomics. Academic radiology, 28(9), e258-e266.

[6] Hatuwal, B. K., & Thapa, H. C. (2020). Lung cancer detection using convolutional neural network on histopathological images. Int. J. Comput. Trends Technol, 68(10), 21-24.

[7] Šarić, M., Russo, M., Stella, M., & Sikora, M. (2019, June). CNN-based method for lung cancer detection in whole slide histopathology images. In 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech) (pp. 1-4). IEEE.

[8] Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J. M., Ciompi, F., & Van Der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical image analysis, 58, 101544.

[9] Salvi, M., Michielli, N., & Molinari, F. (2020). Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. Computer methods and programs in biomedicine, 193, 105506.

[10] Sunila, Anjum., Imran, Ahmed., Muhammad, Asif., Hanan, Aljuaid., Fahad, Alturise., Yazeed, Yasin, Ghadi., Rashad, Elhabob. (2023). Lung Cancer Classification in Histopathology Images Using Multiresolution Efficient Nets. Computational Intelligence and Neuroscience, doi: 10.1155/2023/7282944

[11] Faria, N., Campelos, S., & Carvalho, V. (2025). Enhancing Histopathological Lung Cancer Diagnosis Through Explainable Deep Learning Models: A Methodological Framework Proposal. Information, 16(9), 740.

[12] Zhao, H., Su, Y., Wang, M., Lyu, Z., Xu, P., Jiao, Y., ... & Fu, P. (2022). The machine learning model for distinguishing pathological subtypes of non-small cell lung cancer. *Frontiers in oncology*, *12*, 875761.

[13] Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., Christiani, D. C., Mak, R. H., & Aerts, H. J. W. L. (2021). Deep learning classification of lung cancer histology using CT images. Scientific Reports, 11(1), 5471. https://doi.org/10.1038/s41598-021-84630-x

[14] Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2022). Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Computing and Applications, 34(12), 9579–9592. https://doi.org/10.1007/s00521-020-04842-6

[15] Ejiyi, C. J., Cai, D., Fiasam, D. L., Adjei-Arthur, B., Obiora, S., Ayekai, B. J., ... & Qin, Z. (2025). Multi-modality medical image classification with ResoMergeNet for cataract, lung cancer, and breast cancer diagnosis. Computers in Biology and Medicine, 187, 109791.

[16] Adil, N., Singh, P., & Nagwani, N. K. (2025). LungSE-Net: Enhanced Lung Cancer Diagnosis via Lightweight CNN Model using Histopathological Images. *Procedia Computer Science*, *260*, 118-125.

[17] Anandhavalli, Muniasamy., Salma, Abdulaziz, Alquhtani., Syeda, Meraj, Bilfaqih., Prasanalakshmi, Balaji., G., Karunakaran. (2023). Lung cancer histopathology image classification using transfer learning with convolution neural network model. Technology and Health Care, doi: 10.3233/thc-231029

[18] Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence, 126, 106902. https://doi.org/10.1016/j.engappai.2023.106902

[19] Sujatha, Baskar. (2023). Extreme learning machine algorithm-based model for lung cancer classification from histopathological real-time images. Computational intelligence, doi: 10.1111/coin.12576

[20] Abdalbasit, Mohammed, Qadir., Peshraw, Ahmed, Abdalla., Dana, Faiq. (2024). A Hybrid Lung Cancer Model for Diagnosis and Stage Classification from Computed Tomography Images. Al-mağallaẗ al-ʻirāqiyyaẗ al-handasaẗ al-kahrabāʾiyyaẗ wa-al-ilikttrūniyyaẗ, 20(2):266-274. Doi: 10.37917/ijeee.20.2.23

[21] Meena, S., Kumar, A., Sood, M., & Meena, R. K. (2023, June). Lung cancer detection and classification model using inception V3 algorithm. In International Conference on Data Analytics and Management (pp. 423–433). Springer Nature Singapore.

[22] Murugaiyan, Suresh, Kumar., Panneerselvam, Deepak., Parthasarathy, Vasanthan., Kandasamy, Vijayakumar. (2024). Integrated global and local feature extraction and classification from computerized tomography (CT) images for lung cancer classification. Bratislavské lekárske listy, 125(4):223-232. doi: 10.4149/bll_2024_34

[23] Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset (LC25000). arXiv:1912.12142v1 [eess.IV], 2019

[24] Hoque, M. Z., Keskinarkaus, A., Nyberg, P., & Seppänen, T. (2024). Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion*, *102*, 101997.

[25] Dzulkifli, F. A., Mashor, M. Y., Raof, R. A. A., & Jaafar, H. (2024). An introduction to double stain normalization technique for brain tumour histopathological images. *International Journal of Electrical and Computer Engineering (IJECE)*, *14*(1), 375-388.

[26] Ennab, M., & Mcheick, H. (2025). Advancing AI interpretability in medical imaging: a comparative analysis of pixel-level interpretability and Grad-CAM models. Machine Learning and Knowledge Extraction, 7(1), 12.

[27] Raveenthini, M., Lavanya, R., & Benitez, R. (2025). Grad-CAM based explanations for multiocular disease detection using Xception net. Image and Vision Computing, 154, 105419.

[28] Ali, S. A., Arain, K. R., Mushtaq, N. A., & ul Rehman, O. (2025). Interpretable Deep Learning for Brain Tumor Diagnosis: Occlusion Sensitivity-Driven Explainability in MRI Classification. VFAST Transactions on Software Engineering, 13(2), 135-146.

[29] Cheng, Z., Wu, Y., Li, Y., Cai, L., & Ihnaini, B. (2025). A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision. Sensors, 25(13), 4166.

[30] Lindenmeyer, A., Blattmann, M., Franke, S., Neumuth, T., & Schneider, D. (2025). Towards Trustworthy AI in Healthcare: Epistemic Uncertainty Estimation for Clinical Decision Support. Journal of Personalized Medicine, 15(2), 58.

[31] Ojha, J., Presacan, O., G. Lind, P., Monteiro, E., & Yazidi, A. (2025). Navigating uncertainty: A user-perspective survey of trustworthiness of ai in healthcare. ACM Transactions on Computing for Healthcare, 6(3), 1-32.

[32] Liang, W., Fujita, N., Kobayashi, R., & Boku, T. (2024, January). Using Intel oneAPI for Multi-hybrid Acceleration Programming with GPU and FPGA Coupling. In *Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region Workshops* (pp. 69-76).

[33] Goli, M., Narasimhan, K., Reyes, R., Tracy, B., Soutar, D., Georgiev, S., ... & Chereshnev, E. (2020, November). Towards cross-platform performance portability of dnn models using sycl. In *2020 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC)* (pp. 25-35). IEEE.