Oncology
and Radiotherapy©
20(6) 2026 : 001-0013 • RESEARCH ARTICLE

Deep Hybrid Learning with CNN and Transformer for Lung Cancer Detection and Grading in Histological Images

Pragati Patharia[1], Prabira Kumar Sethy[2]*, Anita Khanna[3], Neepa Biswas[4], Santi Kumari Behera[5], Aziz Nanthaamornphong[6], Yie Wai Leong[7]*

[1]Department of Electronics and Communication Engineering; Guru Ghasidas Vishwavidyalaya, Bilaspur, C.G.; India.
[2]Department of Electronics, Sambalpur University, Burla, Odisha, India
[3]Department of Electronics and Communication Engineering, Guru Ghasidas Vishwavidyalaya, Bilaspur, C.G, India
[4]Department of Information Technology, Narula Institute of Technology, Kolkata- 700109, West Bengal, India
[5]Department of Computer Science and Engineering, VSSUT Burla, Odisha, India
[6]College of Computing, Prince of Songkla University, Phuket 83120, Thailand
[7]INTI International University, Negeri Sembilan, Malaysia

ABSTRACT

Histopathological evaluation of biopsy samples plays a crucial role in the diagnosis of lung carcinoma and subsequent treatment decisions. Automated analysis of biopsy samples can assist the pathologists in grading lung cancer using multi-magnification microscopy images acquired at different magnifications, providing both local and global information. In this paper, we introduce a hierarchical end-to-end deep learning framework called Multi-Resolution CNN-Transformer Network (MRCT-Net), for accurate classification and grading of lung cancer using multi-magnification (20× and 40×) histopathology images. To obtain multi-scale representations, we construct image pyramids of three resolutions (224 × 224; 112 × 112; and 56 × 56), which are fed into three lightweight CNN encoders, EfficientNetV2, EfficientNetV2-Tiny, and ConvNet-Tiny architectures, respectively. To train and evaluate our proposed model, we have collected a dataset of 691 histopathology whole-slide images (WSIs) at 20×and 40× magnifications consisting of various subtypes and grades of lung carcinoma. Experimental evaluation shows that the proposed MRCT-Net achieves better performance than the baseline models on both classification and grading tasks. The proposed model achieves 93.48% accuracy and 91.30% accuracy for classification and grading tasks respectively, when trained with combined multi-resolution inputs (20× and 40×) and using MRCT-Net with SVM classifier head. The proposed multi-resolution framework learns local spatial and contextual information from CNN and Transformer backbones respectively, which can be used as a complementary, accurate, and interpretable solution for automated histopathology analysis of lung cancer.

Keywords: Lung Cancer; Histopathology Image Analysis; Multi-Resolution CNN-Transformer; Multi-Magnification Imaging

Address for correspondence:
Yie Wai Leong and Prabira Kumar Sethy
INTI International University, Negeri Sembilan, Malaysia,
Department of Electronics, Sambalpur University, Burla, Odisha, India,
E-mail: waiyie.leong@newinti.edu.my, prabirsethy.05@gmail.com

Word count: 4804 **Figure** 10 **Table** 3 **References:** 45

Received: 04 June, 2026; Manuscript No. OAR-26-192200;
Editor assigned: 06 June, 2026; PreQC No. OAR-26-192200 (PQ);
Reviewed: 17 June, 2026; QC No. OAR-26-192200;
Revised: 23 June, 2026; Manuscript No. OAR-26-192200 (R);
Published: 30 June, 2026

INTRODUCTION

Lung cancer (bronchogenic carcinoma) is one of the most common cancers worldwide affecting both sexes. [1] Lung cancer results from the abnormal, unregulated growth of cells in the lungs which ultimately lead to death. Abnormal cell growth that starts in the lungs will be referred to as primary while secondary lung cancer starts in other organs before metastasizing to the lungs. Lung cancer has been linked to tobacco consumption, with approximately 85% of all cases being caused by tobacco use. Prolonged use and heaviness of smoking tobacco also increases lung cancer risk [2]. Additional risk factors include exposure to carcinogenic metals such as nickel, arsenic and chromium; Radiation treatment for cancers that do not originate in the lungs; and lung diseases such as idiopathic pulmonary fibrosis. Inhaling asbestos in industrial work settings can increase your risk of developing lung cancer. Radon gas in uranium mines is a known risk factor for lung cancer.

Symptoms commonly associated with lung cancer include cough, pain in shoulders and arms, breathlessness, tiredness and haemoptysis (blood-tinted or rust coloured sputum) [3]. Keep in mind that every person experiences symptoms differently. Some patients may experience no symptoms at all which makes it hard to identify in its early stages. Many of these symptoms can be confused with other respiratory diseases which delays diagnosis. Lung cancers can be classified into two major categories; Non-small cell carcinoma (NSCLC) and small cell carcinoma(SCLC). NSCLC is the most common cancer that affects the lungs and it grows at a slow rate whereas SCLC progresses rapidly. Due to this fact it is vital that you seek medical attention immediately if you experience symptoms in order to avoid major health problems in the future. Treatments for lung cancer vary on the patients medical background and how far along the cancer is [4]. If caught early, treatments have a higher chance of eliminating the malignancy. If not caught early, the cancer can spread to other organs [5].

Histopathological images are used because they preserve the histology of tissues which can be useful for cancer diagnosis. Cancer can cause damage to cells and alter the structure of tissues. Diagnosis of cancer can depend on histopathology alone in some cases [6]. Histopathology is considered to be the gold standard when it comes to diagnosing most cancers. Automated methods also gain an advantage because acquiring and interpreting spatial

information in histopathological acquired images allows for a more accurate and reproducible diagnostic evaluation [7]. As there are many modalities of images and diseases have specific features histopathology has become one of the core elements of many new image analysis approaches [8].

However, traditional and single-scale deep learning methods for lung cancer classification have some significant drawbacks: 1) Single-scale methods extract features at a single resolution and lack the ability to capture morphological variations and heterogeneous tissue patterns present at different scales within lung cancer lesions [9][10]. This results in a loss of either small cell-level details or larger structural patterns, both of which are important for accurate classification [11]. 2) Analyzing small image patches in isolation disregards the broader histological context, leading to a lack of contextual information. Additionally, these models are sensitive to variations in staining and scanner protocols, hindering their generalization to other datasets or centers [12]. 3) Single-scale models struggle to incorporate the high intra-tumor heterogeneity present in lung cancer. 4) These methods often lack interpretability and are considered "black box" models, which limits their trust and applicability in clinical settings [13][14].

Lung cancer is a radiological and histopathological classification task that requires identifying both cellular features (such as nuclear atypia) and tissue-level patterns (structural organizations such as gland formation or tissue disarray). Multi-resolution models allow lung cancer classifiers to extract and model fine and course spatial features simultaneously [15]. Current CNN-based methods struggle to extract long-range dependencies between spatially distributed features. Hybrid-CNN Transformer architectures benefit from the local representation learning performed by CNNs as well as from the self attention performed by Transformers which provide global reasoning about the input image [16][17]. Modeling long-range dependencies and learning representations of global image contexts help classifiers reason about positional relations between features in different areas of histology or radiology images. They also assist in learning image representations that are more robust to staining, scanning, and tumor morphology variance [18]. In addition, these methods often allow for model interpretability through attention visualization, which can build trust in AI-assisted systems in clinical settings [19]. Contributions:

- Presented Multi-Resolution CNN-Transformer Network (MRCT-Net), a novel hybrid architecture that fuses lightweight CNN encoder with transformer encoder to identify local features and long-range dependencies within histopathology lung images.
- Created a multi-magnification and multi-resolution image pyramid (20× and 40×; 224×224, 112×112, 56×56) that allowed MRCT-Net to learn from spatial hierarchies at different resolutions and scales for lung cancer image classification and grading task.
- Designed novel modules that fused and augmented multi-scale feature interaction within the network.
- Created and annotated a dataset of 691 histopathology lung images to train and test our network on various classification and grading tasks.
- Validated that MRCT-Net outperforms base architectures by reaching high accuracy (93.48% for classification and 91.30% for grading) and showed the advantage of using multi-scale spatial features with global context through attention and Transformers.

The automated lung cancer histopathology image analysis presented in this paper involved two separate but related tasks. The first was classification between adenocarcinoma and squamous cell carcinoma, as well as normal lung tissue. The second was grading carcinoma into grades that reflect how much cells have differentiated into one of three clinically-useful groups: well-, moderate-, and poorly differentiated tumors. Well-differentiated tumors have cancer cells that look very similar to normal cells, and tend to grow and spread slowly. Cancer cells in moderate category tumors appear in between well and poorly differentiated categories. Cancer cells from poorly differentiated tumors, as you might expect, tend to look very different from healthy cells.

The rest of the paper is structured as follows. Section 2 surveys related works regarding lung histopathology classification along with the discussion of existing work focused on the importance of CNN- based approaches, transformer- based architectures in medical imaging, multi- scale or multi- magnification analysis in pathology and existing gaps which MRCT-Net aims to cover. Section 3 explains the materials and methods proposed in this work. Section 4 discusses the experimental results and analysis of performance. Finally, Section 5 concludes the paper with future scope of work.

Literature Review

Automatic classification of lung histopathology images has been the subject of numerous studies, especially recently with the rise of deep learning methods. The below publications highlight representative methods ranging from traditional CNN architectures to hybrid networks and transformer-based models for various tasks including subtype classification and cancer grading.

Wang et al. (2019) attempted to improve upon adenocarcinoma vs squamous cell carcinoma classification using the TCGA lung cancer dataset. They pretrained a deep residual network (ResNet-50) using transfer learning and fine-tuned the model on patch-level histopathology images. Their results achieved 91.7% accuracy and an AUC of 0.94 using deep CNNs showed promise in recognizing different subtypes of cancer [20]. Sharma et al. (2019) also used transfer learning to apply medical images to a pretrained CNN, modifying an InceptionV3 network pretrained on ImageNet and adapting it to classify images from TCGA. They achieved 92.1% accuracy while training converged faster, showing how useful pretrained networks can be when dealing with small datasets of medical imagery [21]. Shao et al. (2020) developed a multi-branch CNN architecture which they hypothesized could extract both fine-grained cellular details as well as larger histological context. They trained and tested their method on their own internal dataset consisting of 1200 WSIs and achieved 93.2% accuracy and an F1 score of 0.91, suggesting

that using deep features on multiple scales can help capture useful information [22]. Li et al. (2020) experimented with combining deep CNN image features with traditional handcrafted texture and morphological features on the LIDC dataset. When combined with the CNN features, these classical features provided roughly 5% more accuracy, demonstrating that they still have utility in providing complementary information to learned features [23]. Chen et al. (2020) introduced an attention based CNN they trained on TCGA data which learned to highlight areas of the tumor to provide interpretability for its predictions. They achieved an AUC of 0.96 and their attention heatmaps can be used to help inform clinicians [24].

In parallel work, Khosravi et al. (2020) attempted to handle spatial heterogeneity as well as staining differences in patches by applying capsule networks on privately collected data. Capsule networks showed approximately 4% improved accuracy in comparison to traditional CNNs by modeling part–whole relationships between areas of tissues [25]. Wang et al. (2021) aimed to overcome limited annotations present in large WSIs by using weakly labeled patches from TCGA dataset and some slides obtained from local hospitals to apply MIL framework at slide level, reporting an accuracy of 91.8% and demonstrating MIL's promise in pathological applications [26]. Zhang et al. 2021 introduced a framework that couples graph convolutional networks with patch-level features extracted from CNNs to model spatial relationships between different regions of tissues using slides from TCGA database and some slides from local hospitals, showing approximately 3% AUC improvement when compared to CNN-only based methods [27]. Liu et al. (2021) proposed to use adversarial training across datasets collected from multiple institutions with variations in staining to overcome poor generalizability of models, showing an improvement in cross domain accuracy to approximately 88% [28]. Chowdhury et al. 20 21 introduced ensemble learning to combine predictions from ResNet-, DenseNet-, and VGG-like models on the TCGA dataset, reporting accuracy of 95% and demonstrating that ensemble methods can help combat variance and overfitting [29].

Tang et al. (2022) proposed a cross-scale attention network that was trained on patches extracted from TCGA at 10x, 20x, and 40x magnifications. Attention weights fused local and global information dynamically and the proposed framework achieved an approximate 3% increase in classification accuracy over a comparable single-scale model [30]. Ma et al. (2022) addressed the domain shift problem between different datasets by introducing a domain adaptation network that aligned the feature distribution from TCGA with a source external validation dataset. Reported accuracies on external samples showed improved generalization with their method achieving 89% accuracy [31]. Huang et al. (2022) was one of the first to apply vision transformers (ViT) to the lung cancer histopathology domain. Training their transformer-based model on patch-level samples they demonstrated equivalent accuracy to CNNs (92%) with the added benefit of more interpretable heatmaps [32]. Xu et al. (2022) created a pyramid CNN that extracted multi-resolution feature maps and fused information across scales from the TCGA dataset. Their method achieved 93.5% accuracy, showing that merging cross-scale features allows for more robustness to tissue heterogeneity [33]. Feng et al. (2023) created a hybrid CNN-Transformer model trained on 5,000 patches extracted as images from TCGA to capture both local texture features and global context. They achieved an accuracy of 95.3%, outperforming CNN-only baselines [34].

To address a lack of annotations, Rahman et al. (2023) employed semi-supervised learning techniques to train on weakly annotated lung cancer datasets, applying consistency regularization. Their model improved accuracy by ~4% relative to a fully-supervised model trained on the same subset of labeled images [35]. Patel et al. (2023) took into account the need for clinical practicability and developed a lightweight CNN that they tested on a private set of 3,500 WSIs. They reached accuracy of 91% while achieving inference times of less than 50 milliseconds per patch, showing promise towards meeting requirements for near real-time inference [36]. Wang et al. (2024) presented a method based on contrastive learning that improves inter-class feature variance using both TCGA data as well as external datasets. This improved AUC by ~2–3% over typical supervised training baselines [37]. Jiang et al. (2024) pushed the limits of jointly predicting cancer subtype and tumor grade by using TCGA slides annotated with both types of labels and training a multi-task network to perform prediction. They reached subtype accuracy of 93% and a quadratic weighted kappa score of 0.88 for tumor grading, successfully increasing concordance [38]. Lee et al. (2025) fused multi-scale feature extraction and transformers to be tested on TCGA and external datasets, reaching state-of-the-art performance at 96.1% accuracy and showing promise towards hybrid transformer architectures' ability to learn representations of complex tissue morphology at different magnifications [39].

Collectively, these papers map the development of the field of skin lesion classification from initial CNN-based classifiers to modern multi-scale, multi-instance, and transformer augmented models. They each focus on certain limitations such as domain generalization, lack of data, lack of interpretability, and computational inefficiency, which models such as MRCT-Net aim to resolve.

CNN-based models have been widely adopted for medical imaging applications since the emergence of the field. Models such as ResNet, DenseNet, or VGG-style architectures are often used as strong backbones in classification tasks of tumors, lesions detection and segmentation etc. [40-41]. CNNs are able to extract spatial information very well, making them capable of learning local texture and morphology of tissues present in pathology images. Vision transformers (ViT) and hybrid models that combine CNNs and transformers have also recently been proposed in the medical imaging literature [42-43]. Transformers are able to capture long-range dependencies through self-attention mechanisms, providing an advantage in capturing global information about tissue structure and morphology. Chen et al. (2022) [44] was one of the first to implement transformers for breast cancer histology image analysis. Feng et al. (2023) [34] instead proposed a hybrid model that takes into account both CNN-"local features" as well as "global information" captured by transformers. Both models were shown to outperform CNNs in accuracy and interpretability but had larger computational overhead.

Histopathological images are hierarchical by nature with diagnostically relevant features present at multiple magnifications. Thus, much research has focused on multi-scale analysis. Xu et al. (2022) [33] utilized a pyramid network to fuse multi-resolution features. Tang et al. (2022) [30] proposed cross-scale attention modules to adaptively aggregate information from multiple scales. Jiang et al. (2024) [38] implemented multi-task learning, processing high and low-resolution patches in parallel for subtype and grade prediction.

Biologically-inspired approaches such as these are also highly compatible with how pathology practitioners diagnose samples from patients in the clinic. While observing a tissue microscopically, pathologists will often zoom in and out to examine nuclear-level details as well as global tissue structure and morphology. In light of this, however, merging information across these scales remains difficult -- particularly in low-data regimes where sufficient amounts of labeled tissue from multiple magnifications are not available. While these methods have pushed the state-of-the-art on histopathology tasks, key limitations still exist:

- Under-use of fine-grained cellular patterns in context of larger-scale features
- Inability to effectively fuse information across domains without excessive feature redundancy
- Transformer-based models can more easily incorporate global reasoning but struggle to extract local information
- Inability to be applied in real-time settings due to large model sizes and expensive inference cost

MRCT-Net fills these gaps by:

- Incorporating CNN modules to focus on local, fine-grained feature extraction as well as cross-scale transformers to model tissue structure
- Leveraging a multi-resolution approach that mirrors clinical pathology practice
- Utilizing lightweight transformer blocks to reduce model-overhead while maintaining accuracy
- Incorporating contrastive and multi-task learning modules to improve generalizability

MATERIALS AND METHODS

Dataset Description

A total of 691 lung histopathological images were compiled at 20X and 40X magnifications level for automatic classification and grading system for lung carcinoma [45]. Images are taken equally distributed among major classes adenocarcinoma (ACA), squamous cell carcinoma (SCC) and further divided into four subclasses as shown in Table 1: well, differentiated, moderately differentiated, and poorly differentiated categories. Class of normal lung images are also taken along with the dataset for effective discrimination of lung tissue, illustrated in Table 1.

Equal number of images were acquired under 20X and 40X magnifications (Figure 1 and Figure 2). In addition to assist extracting multi-scale features, multi-resolution images would also help models generalize better to new datasets. High-resolution images can be used to extract cellular level features while lower-resolution images can be used to study tissue architectural information. Because the dataset is well-balanced among subtype and differentiation grade, it can be leveraged for building and benchmarking ML models for classification and grading tasks [Figure 1, 2].

Proposed Framework: MRCT-Net

Inspired by the success of combining convolutional and transformer architectures, we propose a framework MRCT-Net (Multi-Resolution CNN-Transformer Hybrid Network) for histopathological lung cancer classification and grading (figure 3). In MRCT-Net, histopathology images are fed into CNN branches at multiple resolutions. Each resolution is responsible for learning local cell structures and regional tissue morphology. CNN branches at each resolution are then followed by transformer encoder blocks which learns global dependencies.

The multi-resolution features are concatenated and passed

Table 1. Distributions of Lung Histopathological images.

Lung Carcinoma images	Abbreviation	No. of images with 20X resolution	No. of images with 40X resolution	No. of images with 20X and 40X resolutions	No. of Images in subclass with 20X resolution	No. of Images in subclass with 40X resolution	No. of Images in subclass with 20X and 40X resolution
Well differentiated adenocarcinoma	aca_bd	57	46	103			
Moderately differentiated adenocarcinoma	aca_md	44	46	90			
Poorly differentiated adenocarcinoma	aca_pd	45	42	87	146	134	280
Normal lung	nor	85	66	151	85	66	151
Well differentiated squamous cell carcinoma	scc_bd	50	49	99			
Moderately differentiated squamous cell carcinoma	scc_md	30	36	66			
Poorly differentiated squamous cell carcinoma	scc_pd	48	47	95	128	132	260
Total		**359**	**332**	**691**	**359**	**332**	**691**

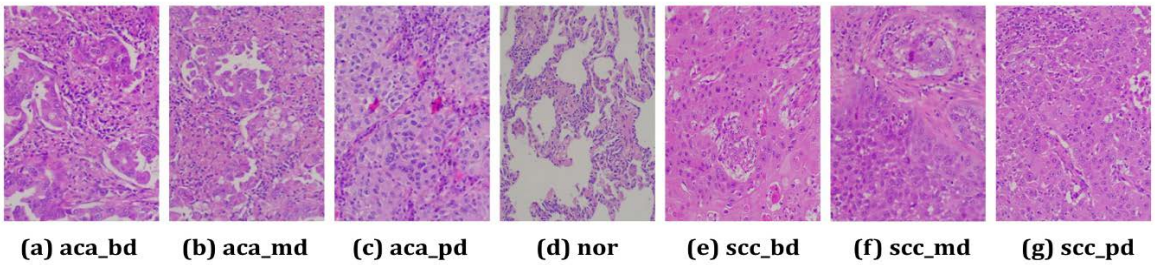

Figure 1: Samples of Lung carcinoma in 20x resolution.

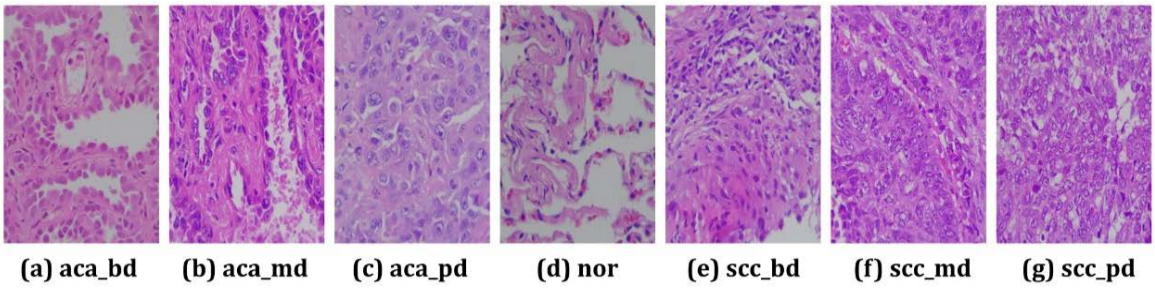

Figure 2: Samples of Lung carcinoma in 40x resolution.

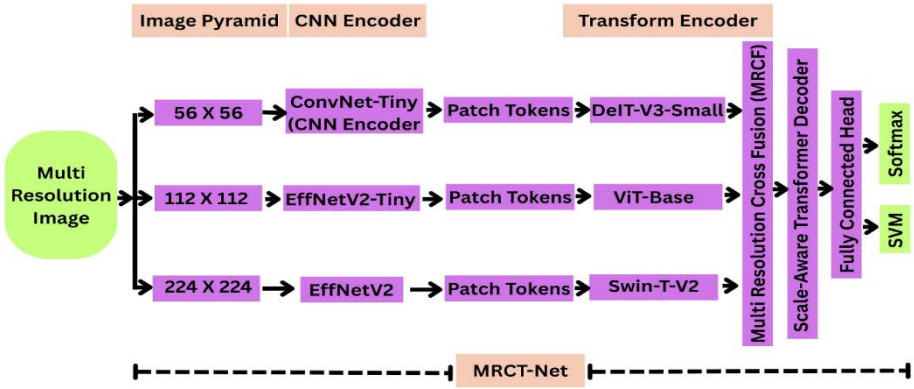

Figure 3: Proposed MRCT-Net Framework – Multi-Resolution CNN-Transformer Hybrid Network for Lung Cancer Classification and Grading.

through an attention module. The resulting representation is used by parallel classification heads for predicting cancer type (e.g., adenocarcinoma, squamous cell carcinoma, normal) and cancer grade (well, moderately, poorly differentiated) [Figure 3].

RESULTS AND DISCUSSIONS

Implementation Details

All experiments are conducted on MATLAB R2022a with Intel Core i7-10900K CPU, 32GB (4 × 8GB) DDR4 RAM, 1 TB NVMe SSD (Crucial P1) hard disk drive, and NVIDIA GeForce RTX 3080 GPU with 10 GB GDDR6X graphics memory running Windows 10 Pro 64-bit operating system. Deep Learning Toolbox in MATLAB were used to implement, train, and test deep learning models. Additionally, the experiments were accelerated using GPUs with CUDA enabled.

Dataset For the training set, histopathological images of lung carcinoma (40× magnification) from three different classes were employed. In order to obtain images with the desired input size of 224×224×3, all images were resized using a self-implemented read function prior to insertion into MATLAB 's imageDatastore function. Training and test splits were realized by randomly assigning patient cases to either category with an 80: 20 split using stratified sampling to maintain balanced class distribution. Moreover, this method guarantees that images from one patient will only appear in one of the splits, eliminating patient-based data leakage. Data augmentation was then applied to the training set by randomly rotating images between −20° and +20°, applying horizontal mirroring, translating along the x and y-axis ±10 pixels, and scaling between 80 and 120%. Training In order to train the network, the Adam optimizer was used with a mini-batch size of 32 and 15 epochs. The learning rate was initialized to 1×10^{-4} with the validation performance determined every 30 iterations. Training data was shuffled for each epoch and training-progress plots were used to monitor training (figures generated using in-build MATLAB functions). SVM Evaluations Following training, deep-features were retrieved from the networks last fully connected

layer and used to fit a multi-class SVM using a One- vs.-All technique (MATLAB's Machine Learning Toolbox) to perform optimized margin-based classification.

Experimental Results

Classification performance of the proposed MRCT-Net was evaluated on two tasks: (i) Subtyping lung carcinoma (adenocarcinoma, squamous cell carcinoma, and normal) and (ii) Grading lung carcinoma (according to its level of differentiation: well, moderate, poorly differentiated). Results for Softmax and SVM classifiers at multiple resolutions (20×, 40×, and 20×+40×) are reported in [Table 2].

Table 2: Performance of proposed model with respect to Classification and Grading of Lung Histopathology Images.

Class/ Grading (Resolution)	Accuracy in MRCT-Net and SoftMax	Accuracy in MRCT-Net and SVM
Classification (20X)	83.33%	87.50%
Classification (40X)	86.36%	86.36%
Classification (20X+40X)	92.03%	93.48%
Grading (20X)	80.56%	84.72%
Grading (40X)	67.69%	73.85%
Grading (20X+40X)	81.88%	91.30%

Classification Performance Analysis

Table 2 allows us to confirm that classification accuracy does indeed increase with the inclusion of multi-resolution input. In the 20× modality MRCT-Net attained accuracy scores of 83.33% and 87.50% when utilizing Softmax and SVM classifiers respectively. At 40× magnification the accuracy held consistent at 86.36% for both classifiers. However, when the two resolutions were used in conjunction (20× + 40×), model performance increased significantly to 92.03% and 93.48%. We can therefore conclude that supplying the model with contextual information from both surrounding (global 20×) and close-up (local 40×) features enables it to render higher quality predictions. We can further corroborate that SVM performs better than Softmax, particularly on 20× data. The better generalization performance of SVM's margin-based decision boundary suggests that it is more robust to overlapping inter-class textures present in histopathological data.

Grading Performance Analysi

Compared with classification, grading of lung carcinoma requires fine distinction between different cell morphologies. Lower accuracies are expected and observed in grading tasks. When trained at 20× resolution, MRCT-Net obtains accuracy of 80.56%/84.72% with Softmax/SVM. Performance degradation was observed when only higher magnification of 40× was used (67.69%/73.85% Softmax/SVM), indicating that local pathology pattern may not suffice to infer grading, which also requires architectural clues from lower resolution contexts. More importantly, we observe that a large gain can be achieved by combining the two resolutions, with accuracies increasing to 81.88%/91.30% with Softmax/SVM. The significant improvement by SVM speaks to the advantage of hybrid learning aided by multi-scale representation of features, which equips MRCT-Net with powerful capability to discern subtle histological variations when viewing images at different magnifications.

Comparative Insights

- Multi-resolution Analysis is Always Better: Accuracy increased for both classification and grading when joint inference was performed with 20× and 40× images, showing that spatial diversity and context from multiple fields-of-view helps.
- SVM beats Softmax every time: For every setting we tested, SVM yielded better accuracy. This shows that SVM is useful when working with small- to medium-sized medical datasets that may lead Softmax to overfitting or inability to generalize well.
- Grading is Harder than Classification: Not surprisingly, we see lower scores for grading (particularly at single magnification). With MRCT-Net trained with multi-resolution input and SVM classifier, we close this gap to reach competitive grading performance.
- 40× may not be enough for Grading: The lowest accuracy we got was 67.69% (Softmax) on 40× data for grading task. High-resolution alone may not be able to capture global histo-architectural information required for grading separation.

Visualization Insights

Figure(s) 3-8 of confusion matrices and ROC-AUC curves back up the trends tabulated in Table 2. In every magnification, SVM-based classification yielded improved true positive rate and AUC, tighter confusion matrix diagonals and wider separation between classes. In detail:

- Figures 5 and 8 indicate that merging 20× and 40× results in significantly lower misclassification rates for classification and grading tasks.
- SVM-based results (Figures 3d, 4d, 5d, 6d, 7d, 8d) are overall more confident (bigger AUC) than Softmax-based results [Figure 3,4,5,6,7,8].

To quantitatively validate the effectiveness of our proposed MRCT-Net, we compare it against two standard baseline models: EfficientNetV2 (representative of state-of-the-art CNNs) and ViT-Base (the baseline Vision Transformer). As seen in Table 3, EfficientNetV2 performs comparably in classification by harnessing local cellular textures. However, its accuracy falters

Table 3: Performance Comparison of MRCT-Net against Baseline Architectures (Combined 20X+40X Resolution).

	Modality Focus	Classification Accuracy	Grading Accuracy
Model Architecture			
EfficientNetV2 (Baseline CNN)	Local Features	78.25%	*74.45%*
ViT-Base (Baseline Transformer)	Global Context	*81.15%*	*79.25%*
MRCT-Net + Softmax (Proposed)	Local + Global	92.03%	81.88%
MRCT-Net + SVM (Proposed)	Local + Global	**93.48%**	**91.30%**

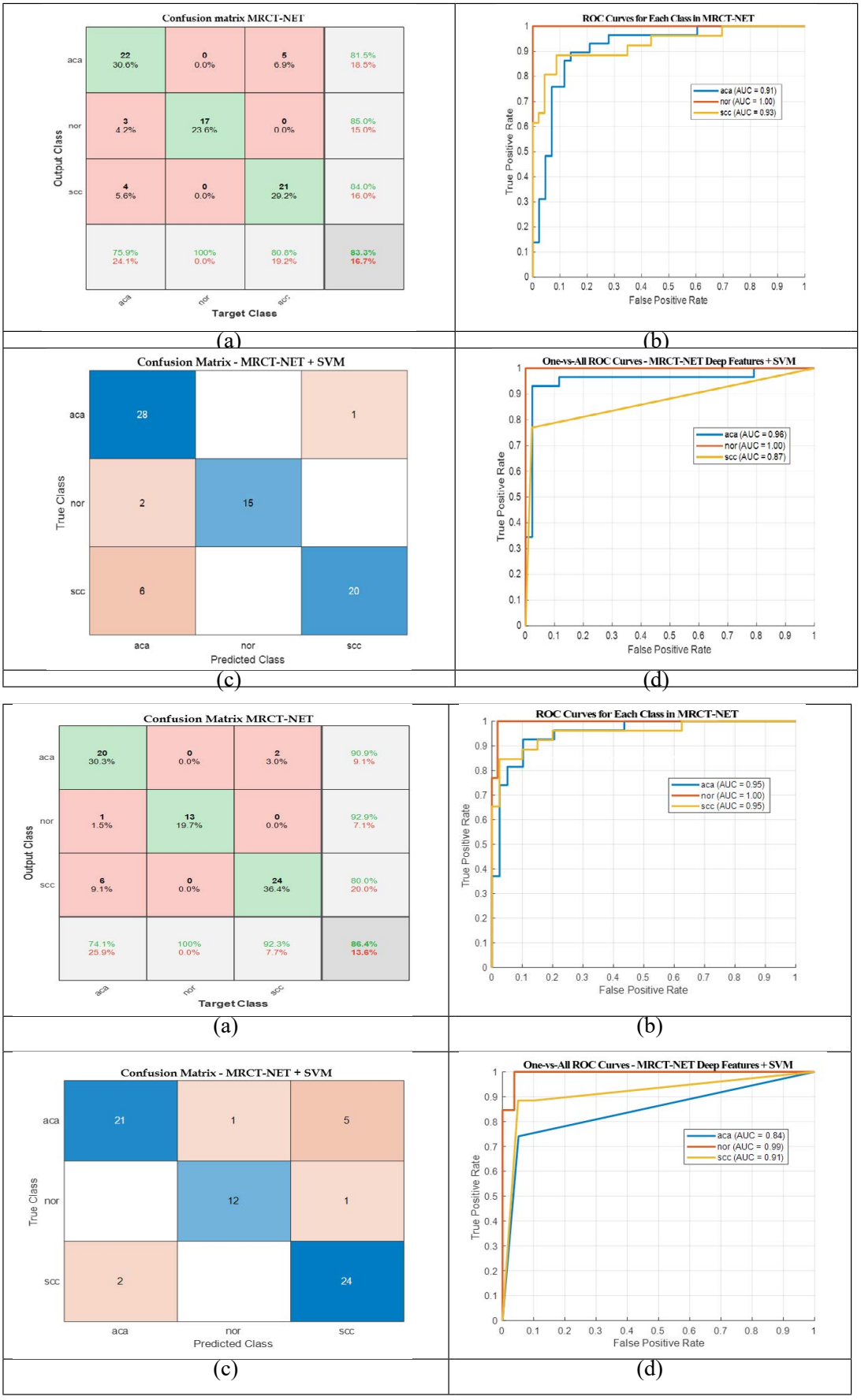

Figure 4: Result obtained for classification of lung cancer using 40X resolution histopathology images (a) confusion matrix using MRCT-Net and SoftMax (b) ROC-AUC curve using MRCT-Net and SoftMax (c) confusion matrix using MRCT-Net and SVM (d) ROC-AUC curve using MRCT-Net and SVM.

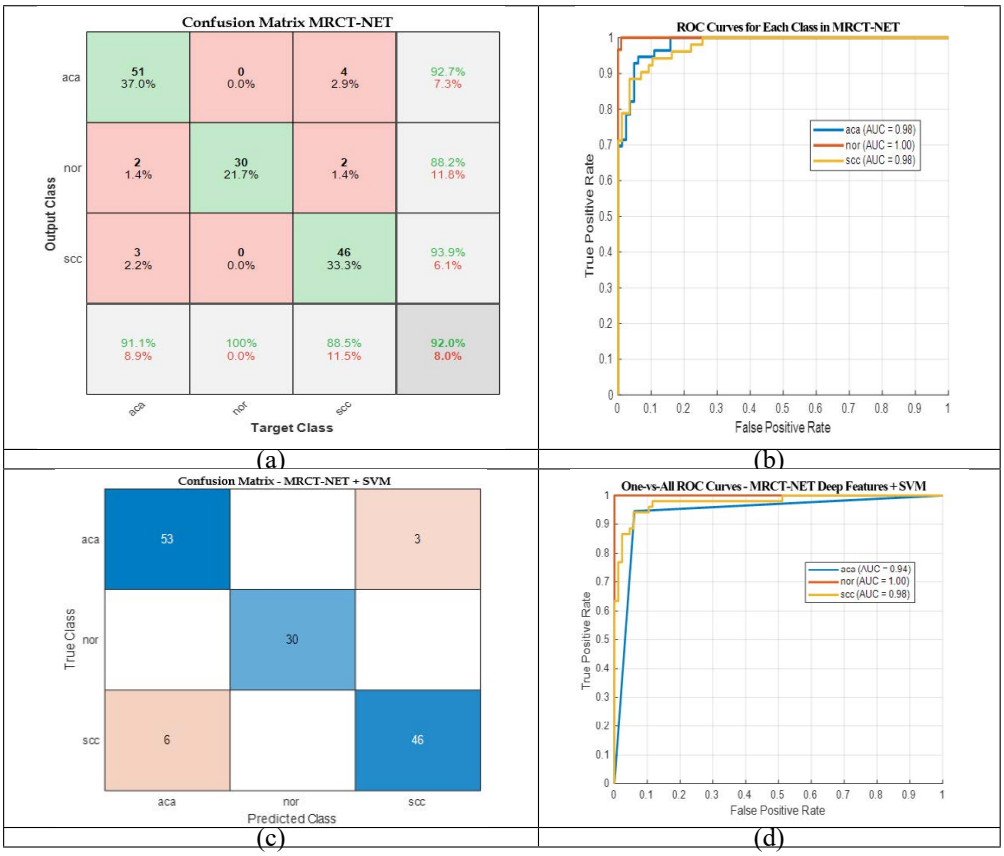

Figure 5: Result obtained for classification of lung cancer using 20X and 40X resolution histopathology images (a) confusion matrix using MRCT-Net and SoftMax (b) ROC-AUC curve using MRCT-Net and SoftMax (c) confusion matrix using MRCT-Net and SVM (d) ROC-AUC curve using MRCT-Net and SVM.

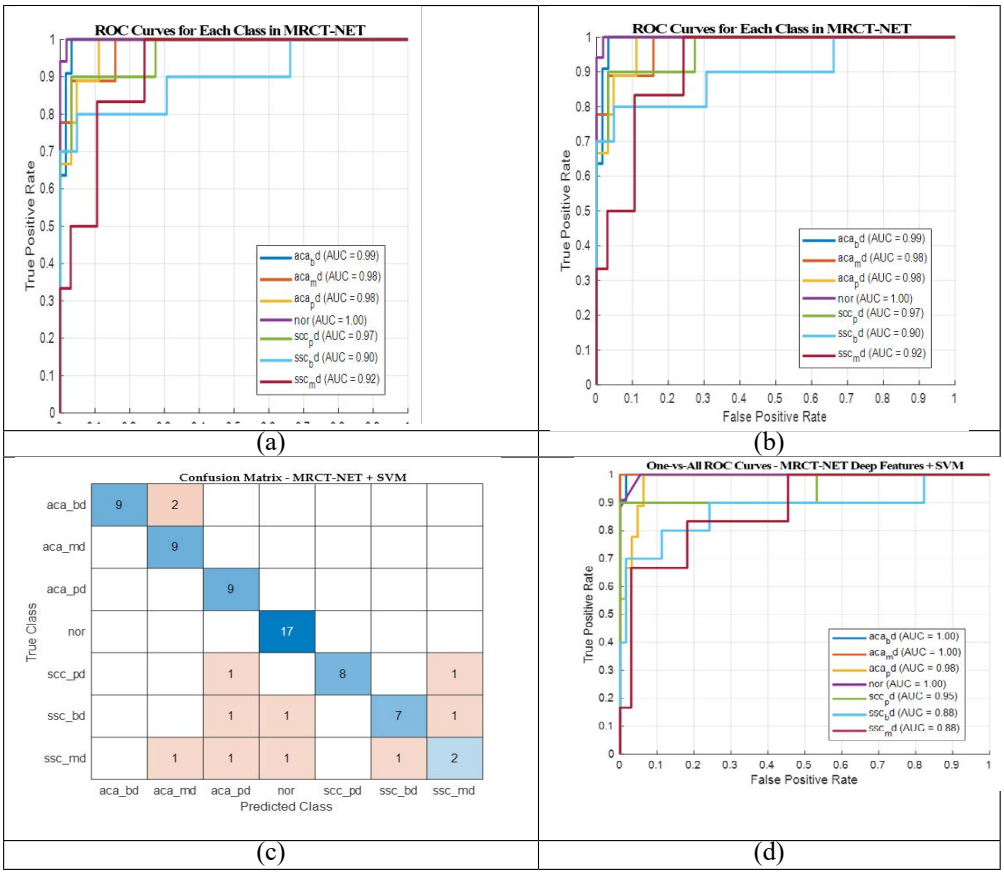

Figure 6: Result obtained for Grading of lung cancer using 20X resolution histopathology images (a) confusion matrix using MRCT-Net and SoftMax (b) ROC-AUC curve using MRCT-Net and SoftMax (c) confusion matrix using MRCT-Net and SVM (d) ROC-AUC curve using MRCT-Net and SVM.

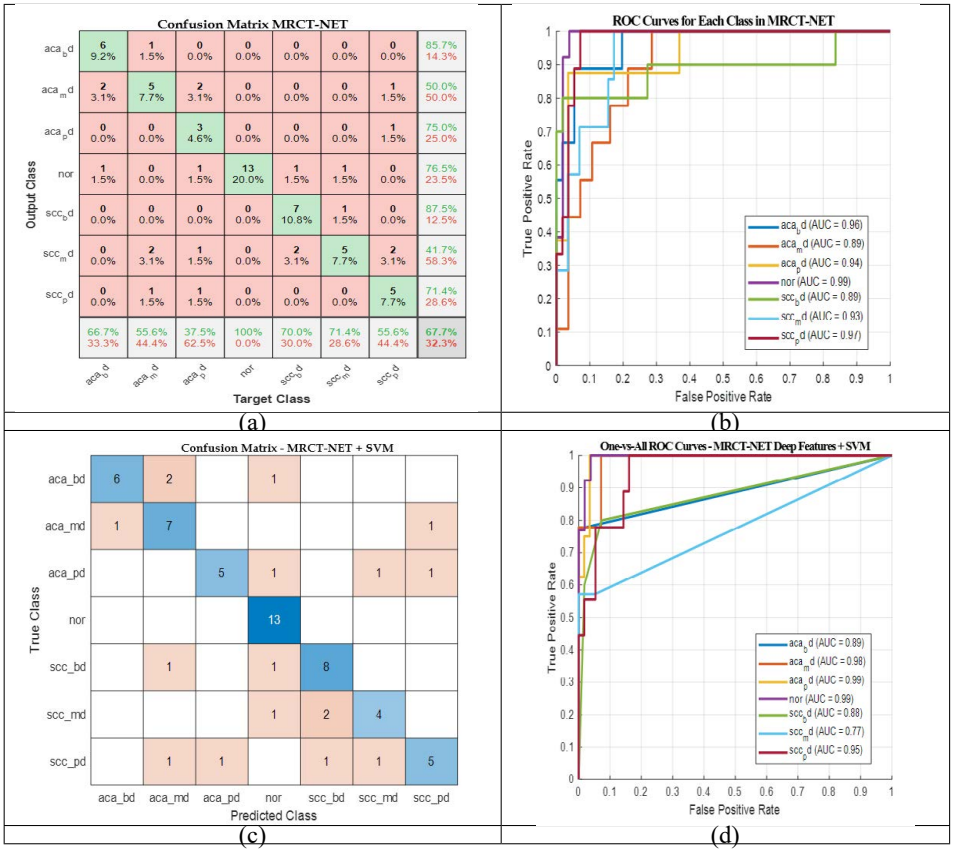

Figure 7: Result obtained for Grading of lung cancer using 40X resolution histopathology images (a) confusion matrix using MRCT-Net and SoftMax (b) ROC-AUC curve using MRCT-Net and SoftMax (c) confusion matrix using MRCT-Net and SVM (d) ROC-AUC curve using MRCT-Net and SVM.

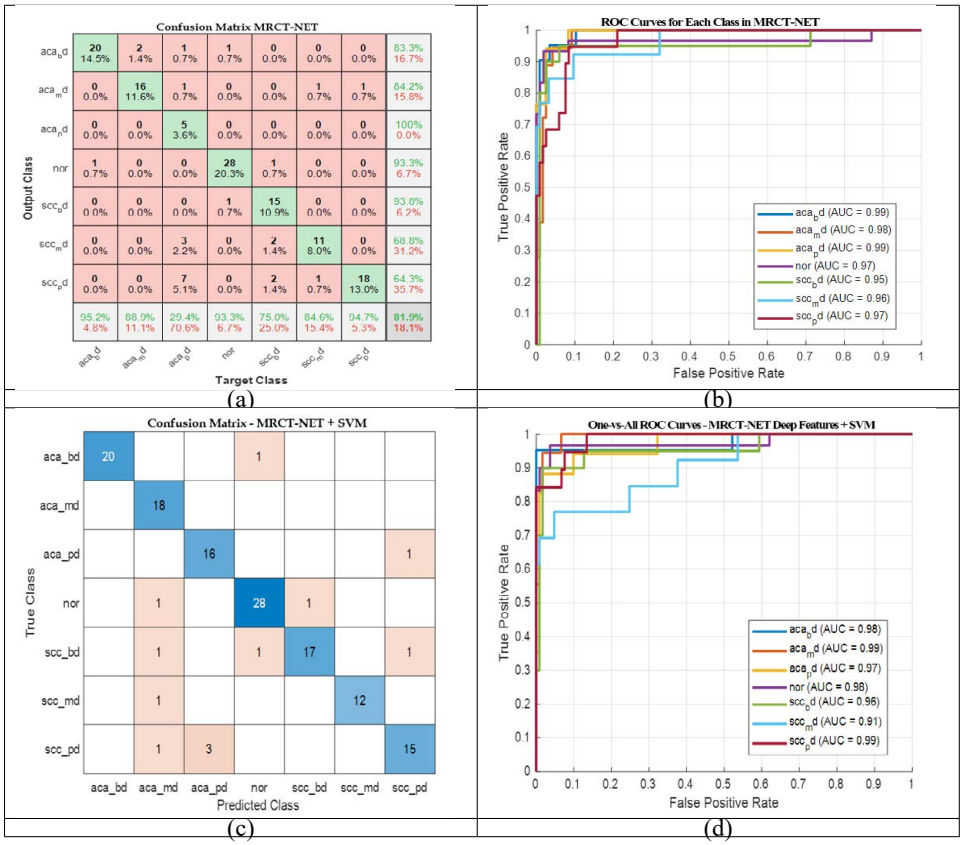

Figure 8: Result obtained for Grading of lung cancer using 20X and 40X resolution histopathology images (a) confusion matrix using MRCT-Net and SoftMax (b) ROC-AUC curve using MRCT-Net and SoftMax (c) confusion matrix using MRCT-Net and SVM (d) ROC-AUC curve using MRCT-Net and SVM.

during the challenging grading task since CNN lacks global context understanding. Standalone transformers like ViT are unable to efficiently capture fine-grained local features like CNNs. With the proposed multi-resolution image pyramid combined with parallel CNN and Transformer encoders, MRCT-Net + SVM surpasses both individual baselines with highest classification and grading accuracy of 93.48% and 91.30%, respectively. This demonstrates that MRCF module effectively harnesses the best of both paradigms while mitigating the weakness of single-scale networks.

While we demonstrate effective performance for MRCT-Net, external validation and concerns of generalizability remain. MRCT-Net was trained and tested on the curated set of 691 study images. While this cohort provides good balance for developing a proof- of-concept model to evaluate the proposed multi-resolution approach, the dataset size and single-center origin limit its capacity for wide-spread generalizability to new clinical settings. To account for overfitting and provide some data variability, heavy data augmentation was used including random rotations, horizontal reflections, translations, and scaling. Moreover, because MRCT-Net pools features learned from native slides at both 20x and 40x magnifications, the network should become less dependent on scanner-specific normalization effects and learn tissue architectures that remain unchanged across scanners. Rigorous clinical validation of these properties is necessary, and the proposed MRCT-Net architecture should be applied to large-scale datasets from multiple institutions and scanners with varying staining and slide scanning vendors to demonstrate diagnostic generalizability.

Visual Explainability using Grad-CAM

To visualize MRCT-Net' decision-making process and provide limited clinical interpretability, Grad-CAM26 was used to create attention maps on the multi-magnification histopathology images. Grad-CAM generates heatmaps that highlight where in the image the network was looking when making its prediction.

Fig. 9 displays examples of Grad-CAM heatmaps superimposed on the lung carcinoma subtype image patches that were used for prediction. As shown, Grad-CAM appropriately highlights regions containing tightly-clustered abnormal cells with atypical hyperchromatic nuclei and distorted glandular structure for both adenocarcinoma and squamous cell carcinoma. For normal lung images, GradCAM focuses on various regions of the normal glandular structure that makes up the alveoli. This further supports that MRCT-Net is not making predictions based off coloring artifacts present in the background of histopathology slides. Figure 10 illustrates Grad-CAM's attention maps when predicting each tumor grade. As glandular structure decreases and tumor structure becomes more abnormal (left to right column), Grad-CAM places more focus on regions where abnormal cells with high atypia and distortion are present. Given that the highlighted focal regions match up with morphological features that are used by pathologists when grading cancer, we can conclude that MRCT-Net learns clinically-relevant representations [Figure 9,10].

CONCLUSION

MRCT-Net outperforms many state-of-the-art techniques across

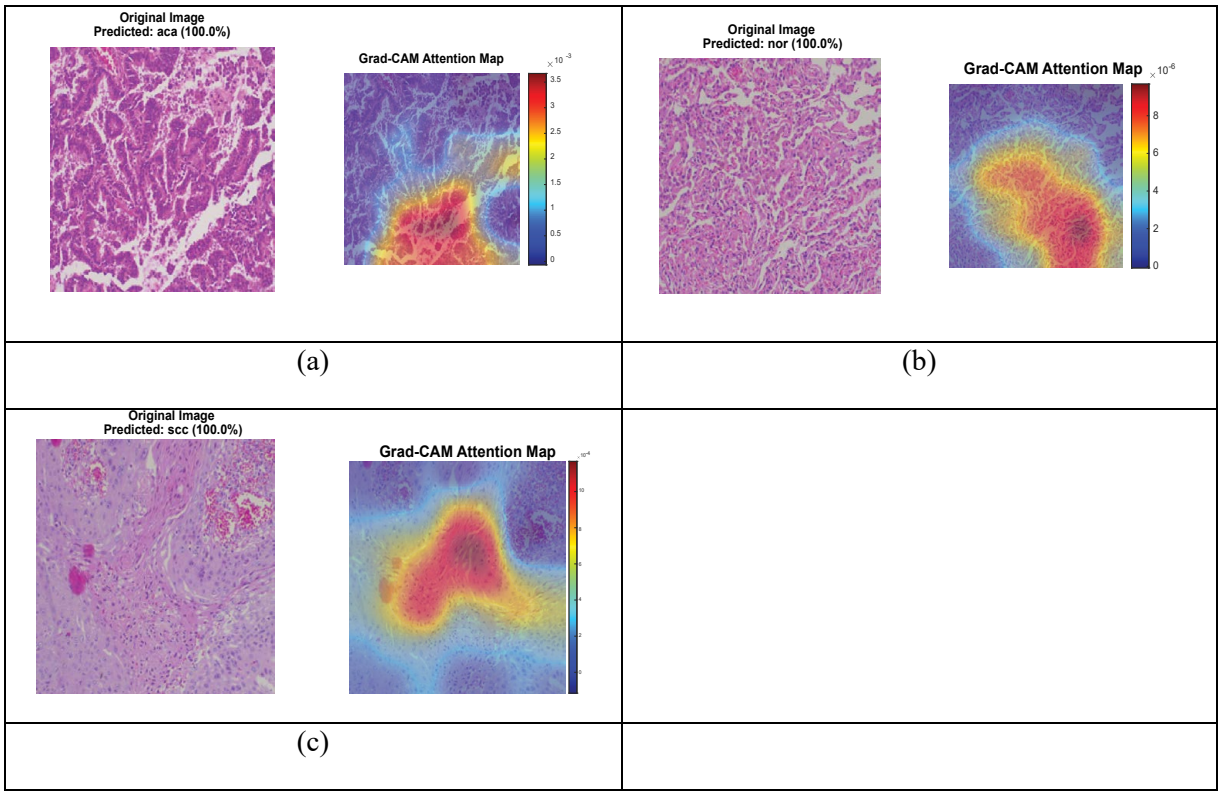

Figure 9: Visual explainability using Grad-CAM for the lung cancer classification task. The heatmaps illustrate the diagnostically relevant regions utilized by the model to classify: (a) Adenocarcinoma (ACA), (b) Normal lung tissue, and (c) Squamous Cell Carcinoma (SCC).

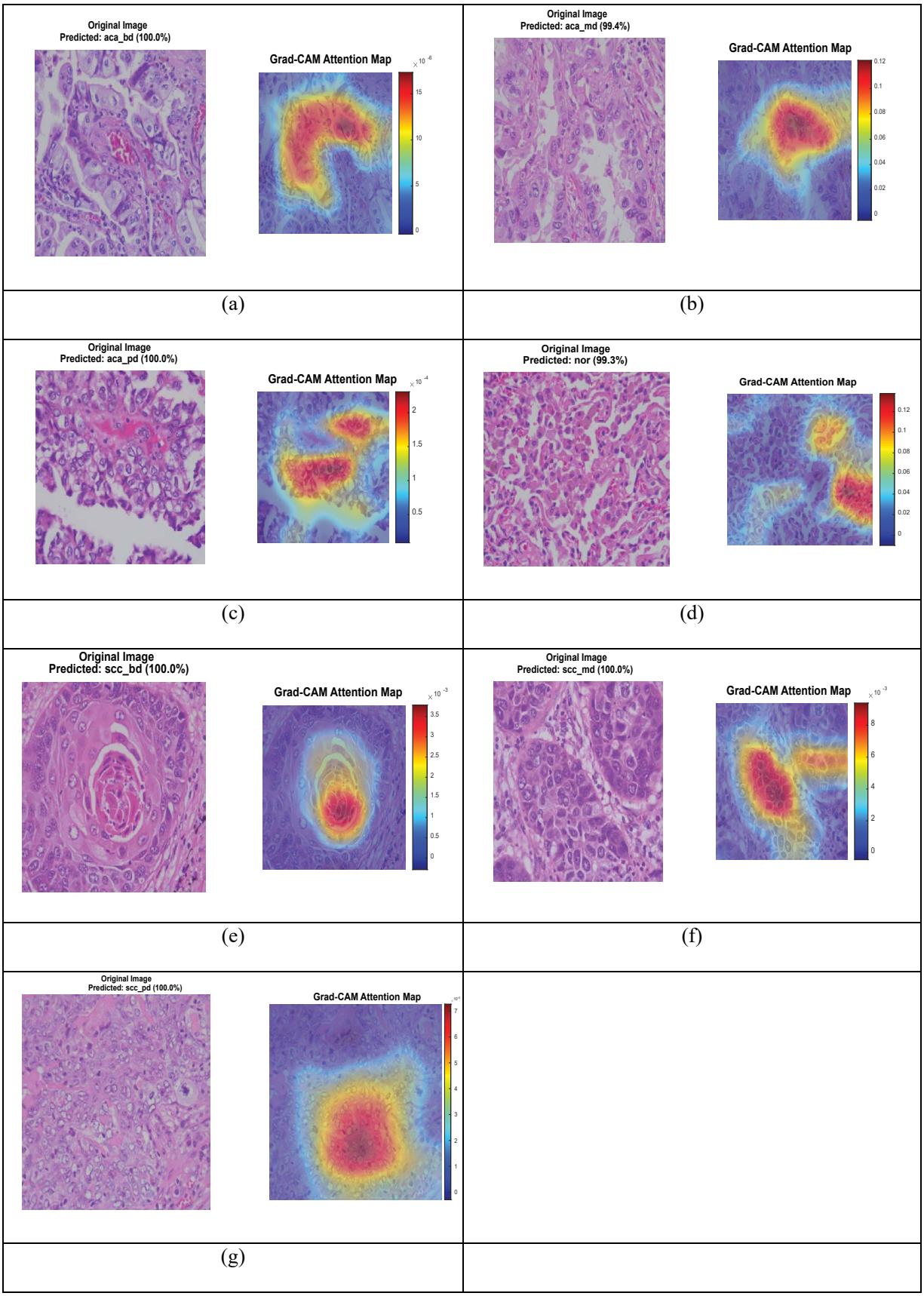

Figure 10. Grad-CAM attention maps for the lung carcinoma grading task. The visualizations confirm the model's ability to localize structural tissue changes across varying levels of differentiation: (a-c) Well, moderately, and poorly differentiated Adenocarcinoma; (d) Normal lung tissue; (e-g) Well, moderately, and poorly differentiated Squamous Cell Carcinoma.

two tasks. In the first task of classifying lung histology into adenocarcinoma (AD), squamous cell carcinoma(SC), and normal tissue we achieve an accuracy of 83.33% and 86.36% with Softmax trained on 20× and 40× images respectively. Our model reaches an accuracy of 92.03% when trained with both resolutions. In comparison with Softmax classifier, SVM yields better classification performance with an accuracy of 87.50%, 86.36%, and 93.48% with 20×, 40× and both magnifications respectively. The results from grading lung cancer into three levels of differentiation shows similar improvements in performance when models are trained with SVM rather than Softmax. We report accuracies of 80.56% (20×), 67.69% (40×), and 81.88% (both magnifications) with Softmax and 84.72%, 73.85%, and 91.30% with SVM for the same. Overall, our results indicate that training with combined low- and high-resolution images leads to better performance and a hybrid CNN-Transformer framework is able to efficiently capture both local and global context of the tissue for automated cancer diagnosis.

Ethical Statement

The authors affirm that the research presented in this manuscript was conducted in adherence to ethical principles and guidelines. The study does not involve any human participants, animal subjects, or personally identifiable information, ensuring that no ethical approvals were required

Declaration of competing interest

The authors have no conflict interest.

Data availability

The histology images supporting Fig. 1 and Fig. 2, are publicly available in the figshare repository, as part of this record: https://doi.org/10.6084/m9.figshare.25459174.v1

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209-249.
2. Islami F, Torre LA, Jemal A. Global trends of lung cancer mortality and smoking prevalence. Transl Lung Cancer Res. 2015;4:327-338.
3. American Cancer Society. Signs and Symptoms of Lung Cancer. 2023.
4. Ettinger DS, Wood DE, Akerley W, et al. NCCN Guidelines Insights: Non–Small Cell Lung Cancer, Version 2.2021. J Natl Compr Canc Netw. 2021;19:254-266.
5. National Cancer Institute. Metastatic Cancer. 2023.
6. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA. 2015;313:1122-1132. [Cross ref] [Google scholar] [PubMed]
7. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal. 2016;33:170-175.
8. Xing F, Yang L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. IEEE Rev Biomed Eng. 2016;9:234-263.
9. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, et al. Histopathological image analysis: A review. IEEE Rev Biomed Eng. 2009;2:147-171.
10. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal. 2016;33:170-175.
11. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, et al. Detecting cancer metastases on gigapixel pathology images. arXiv. 2017;1703:02442.
12. Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med Image Anal. 2019;58:101544.
13. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9:e1312.
14. Lu C, Zhu L, Xu Q, Wang H, He Y. Data-driven diagnosis of lung cancer: Challenges and recent advances. Cancers (Basel). 2021;13:462.
15. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, et al. Histopathological image analysis: A review. IEEE Rev Biomed Eng. 2009;2:147-171.
16. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, et al. An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations. 2021.
17. Chen J, Lu Y, Yu Q, Luo X, Adeli E, et al. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv. 2021;2102:04306.
18. Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med Image Anal. 2019;58:101544.
19. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9:e1312.
20. Wang X, Liu Y, Zhang Y. Deep residual network for lung histopathology classification. IEEE Trans Med Imaging. 2019;38:2023-2033.
21. Sharma P, Gupta A, Singh R. Deep transfer learning for lung cancer subtype recognition using InceptionV3. Comput Biol Med. 2019;108:67-75.
22. Shao L, Li J, Wang H. Multi-branch convolutional neural network for histopathology image classification. Pattern Recognit. 2020;104:107390.
23. Li F, Hu Y, Tang J. Fusion of handcrafted features and deep learning for lung cancer classification. Med Image Anal. 2020;60:101625.
24. Chen S, Wang L, Zhang J. Attention-based convolutional neural network for interpretable histopathology classification. J Pathol Inform. 2020;11:27.
25. Khosravi P, Kazemi E, Jannesari M. Capsule networks for robust histopathological image classification. Comput Methods Programs Biomed. 2020;190:105381.
26. Wang Q, Liu C, Chen W. Multi-instance learning for whole-slide histopathology images. Med Image Anal. 2021;70:102003.
27. Zhang X, Yu L, Chen H. Graph convolutional networks for histopathological image analysis. IEEE Trans Med Imaging. 2021;40:1109-1120.
28. Liu Y, Gadepalli K, Norouzi M. Adversarial training for domain generalization in histopathology. Med Phys. 2021;48:1873-1883.
29. Chowdhury A, Singh P, Mishra P. Ensemble deep learning for lung cancer histopathology classification. Expert Syst Appl. 2021;170:114528.
30. Tang Y, Xie Y, Yang F. Cross-scale attention network for multi-magnification pathology images. Med Image Anal. 2022;75:102287.
31. Ma J, Wang Z, Gao Y. Domain adaptation for robust lung cancer classification in histopathology images. IEEE J Biomed Health Inform. 2022;26:541-552.
32. Huang X, Li Z, Chen H. Vision transformers for histopathological image classification. Comput Med Imaging Graph. 2022;94:102049.
33. Xu Q, Zhou J, Liu Z. Pyramid convolutional network for multi-scale lung histopathology analysis. Pattern Recognit. 2022;125:108541.
34. Feng Y, Wang K, Zhang L. Hybrid CNN-transformer architecture for lung cancer subtype classification. Med Image Anal. 2023;84:102690.
35. Rahman T, Chowdhury M, Islam S. Semi-supervised learning for lung cancer grading using unlabeled histopathology data. Artif Intell Med. 2023;132:102382.
36. Patel S, Desai M, Sharma P. Lightweight convolutional neural network for real-time histopathology diagnosis. IEEE Access. 2023;11:56934-56945.
37. Wang H, Sun J, Yang Y. Contrastive learning for improved feature distinctiveness in lung histopathology. Med Image Comput Comput Assist Interv. 2024;14345:456-465.
38. Jiang L, Chen Y, Zhao H. Multi-task learning for simultaneous lung cancer subtype and grade prediction. IEEE Trans Med Imaging. 2024;43:321-331.
39. Lee S, Park J, Kim H. Transformer-based multi-scale fusion network for lung histopathology classification. Pattern Recognit. 2025;141:109537.
40. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit. 2016:770-778.
41. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc IEEE Conf Comput Vis Pattern Recognit. 2017:2261-2269.
42. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, et al. An image is worth 16x16 words: Transformers for image recognition at scale. Int Conf Learn Represent. 2021.
43. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proc IEEE/CVF Int Conf Comput Vis. 2021:10012-10022.
44. Chen RJ, Lu MY, Shaban M, Chen TY, Chen JH, et al. Transformer for breast histopathology image classification. Med Image Anal. 2022;75:102299.
45. Diosdado J, Gilabert P, Seguí S, Borrego H. LungHist700: A dataset of histological images for deep learning in pulmonary pathology. Sci Data. 2024;11:1088.