Grading Invasive Ductal Carcinoma from Whole-Slide Histological Images using Deep Learning-Based Feature Encoding Techniques

Zhiguo Huang[1], Chatklaw Jareanpon[1], Phatthanaphong Chomphuwiset[2] and Rapeeporn Chamchong[1,*]

[1] Department of Computer Science, University of Mahasarakham, Maha Sarakham, Thailand
[2]MQ Square, Bangkok, Thailand

∗Corresponding Email: rapeeporn.c@msu.ac.th

Received February 4, 2025, Revised April 2, 2025, Accepted July 14, 2025, Published December 30, 2025

Abstract. *Invasive ductal carcinoma (IDC) grading is crucial for determining treatment and prognosis. However, the process of manual grading of whole-slide histological images (WSIs) is time-consuming and prone to variability. In this study, we propose a deep learning-based method aimed at automating the grading of breast cancer from WSIs. Unlike conventional approaches that directly process entire WSIs, our method divides them into smaller patches and employs an unsupervised autoencoder to extract pathological features from each patch. These features are then integrated into a comprehensive representation of the WSI. A classification model is subsequently utilized to assign one of three grades. The proposed approach effectively captures local pathological features while preserving spatial relationships between patches. This technique uniquely balances feature preservation with computational efficiency, addressing the challenges associated with the high resolution of WSIs. Experimental results on a breast cancer histological image dataset demonstrate that our method achieves an average accuracy of 71.43% while reducing training time by 50–67%. This performance outperforms the best results obtained using traditional feature extraction techniques. This highlights the robustness and reliability of our approach in reducing pathologists' workload and improving diagnostic consistency.*

Keywords: Slide grading, feature encoding, image-level classification, computer-aided diagnosis

1. Introduction

Digital histopathology has transformed medical diagnostics, especially in cancer research, by converting traditional glass slides into digitized Whole Slide Images (WSIs). This advancement enables pathologists to analyze tissue samples with greater precision and accessibility [1]. In breast cancer diagnosis, histologic grading remains a critical prognostic factor that directly influences treatment strategies and patient outcomes [2]. The Nottingham Histologic Grading (NHG) system represents the gold standard for classifying breast cancer severity, evaluating nuclear pleomorphism, tubular formation, and mitotic count to determine disease grade [3]. However, manual grading suffers from inter-observer and intra-observer variability, creating a pressing need for more standardized and reliable diagnostic approaches [4].

Computer-aided diagnosis (CAD) systems integrated with artificial intelligence (AI) and machine learning (ML) offer promising solutions to the subjectivity and inefficiency of manual grading. These systems can assist pathologists by automating tumor region identification, measuring histological characteristics, and assessing cancer severity [5]. Despite their potential, developing effective CAD systems for whole-slide histopathology presents significant challenges due to three primary factors: (1) the massive size of WSIs (often exceeding 10GB per slide), (2) the intricate morphological patterns that require precise feature identification, and (3) the scarcity of well-annotated training datasets [6].

The computational demands of processing entire high-resolution WSIs directly through deep learning models are prohibitive for most clinical applications [7]. Moreover, downsampling WSIs to accommodate model constraints inevitably results in the loss of critical histological details, compromising diagnostic accuracy [8]. While patch-based approaches have been proposed to address these challenges, they often struggle to maintain the contextual relationships between tissue regions that are essential for accurate grading [9, 10].

This study introduces a novel feature encoding approach that effectively bridges the gap between computational feasibility and diagnostic accuracy for IDC grading. Our approach offers three key advantages over existing methods:

- **Preservation of diagnostic information**: The autoencoder captures essential histopathological features while reducing dimensionality. By preserving spatial relationships and feature aggregation, it leverages deep learning to retain local cellular details and their global distribution on the slide, avoiding the information loss of direct downsampling.
- **Computational efficiency**: By encoding patches into a compact latent space representation, our method reduces the computational burden of processing gigapixel-scale WSIs by approximately 67% compared to baseline approaches.
- **Improved classification performance**: Our experiments demonstrate an average grading accuracy of 71.43%, surpassing conventional feature extraction techniques by 4.76% on the same dataset.

The benefits of our research extend beyond technical improvements to address critical clinical and research needs. Pathologists gain a reliable decision support tool that maintains diagnostic accuracy while significantly reducing analysis time. Healthcare systems benefit from more consistent diagnoses, reducing the need for secondary consultations and enabling more efficient resource allocation. Researchers gain a robust framework for extracting and analyzing histopathological features across large datasets, potentially enabling new discoveries in cancer biology.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature, critically analyzing existing approaches and identifying research gaps. Section 3 details our proposed methodology, including the patch-based feature encoding mechanism and the classification framework. Section 4 presents experimental results and comparative analysis, while Section 5 concludes with a discussion of implications and future research directions.

2. Related Work

Grading breast cancer (BC) from WSIs is a challenging and crucial undertaking in the field of digital pathology. Numerous computational methodologies have been devised to automate this task, with a specific emphasis on employing classification-based techniques that leverage machine and deep learning models. This section reviews relevant approaches for histopathological image analysis, particularly focusing on feature engineering and classification methods for invasive ductal carcinoma (IDC) grading.

2.1 Patch-based Analysis Approaches

Convolutional neural networks (CNNs) have emerged as a pivotal approach in addressing a wide array of medical applications, including the classification of breast cancer. A. Cruz-Roa et al. [11] utilized a CNN to directly identify invasive ductal carcinoma (IDC) from WSIs. Their method processes WSIs by dividing them into smaller patches, classifying each one, and reconstructing an IDC probability map based on patch coordinates and classification results. This map was used for the final grading prediction. The method achieved an accuracy of 84.23% in automatically detecting IDC regions within WSIs. However, this patch-based classification approach typically demands significant computational resources, especially when dealing with large WSIs [9].

Patch-level analysis represents the foundational approach in computational histopathology for breast cancer, where high-resolution histopathological whole slide images (WSIs) are divided into smaller, manageable patches for feature extraction and classification. Transfer learning has proven particularly effective for histopathological image analysis due to the scarcity of annotated medical data. As demonstrated in the reference document, Bayramoglu et al. [12] achieved an accuracy of 88.03% in distinguishing between epithelial, inflammatory, and fibroblast nuclei using a fine-tuned VGG-16 architecture pre-trained on ImageNet.

Dash et al. [13] proposed a hybrid model called Mask-EffNet for lung cancer diagnosis through CT scan image analysis. Their model employs a masked autoencoder for feature extraction in the initial phase, followed by classification using a pre-trained EfficientNet model. This two-phase approach allows for effective feature acquisition before classification, which is conceptually similar to our patch-based feature encoding technique for WSI analysis. Their method achieved 98.98% test set accuracy with ROC scores of 0.9782-0.9872, demonstrating the effectiveness of leveraging deep transfer learning for medical image analysis.

2.2 Multi-scale and Attention-based Methods

M. Shanban et al. [10] developed a multi-scale convolutional neural network (MSCNN) for the classification of histopathology images. This innovative approach entails meticulously analyzing image patches at various scales to capture a wide range of multi-resolution features, thereby significantly enhancing the accuracy of grading. However, the model's heavy reliance on processing the entire slide presents a substantial computational burden and leaves it susceptible to the inadvertent loss of crucial details in the downsizing process.

The hierarchical nature of tissue morphology necessitates multi-scale feature extraction strategies. Khan et al. [14] proposed a multi-scale feature fusion model that effectively addresses the challenges of small object retention and domain adaptation in breast cancer classification. Their approach utilizes dilated layers to preserve fine-grained structures in deeper network layers and achieves 98.23% accuracy on the BreakHis dataset, demonstrating the effectiveness of multi-scale feature integration in preserving both microscopic and macroscopic tissue features.

Attention mechanisms have significantly enhanced feature relevance in histopathological image analysis. Wang et al. [15] introduced CTransPath, an innovative hybrid architecture that combines CNN's local feature extraction capabilities with a multi-scale Swin Transformer for capturing global contextual information. Their model employs a semantically relevant contrastive learning (SRCL) strategy that aligns multiple positive instances with similar visual concepts, enabling more robust feature representations without extensive manual annotations.

2.3 WSI-level Analysis Techniques

Sirinukunwattana et al. [16] proposed a Random Polygons Model (RPM) for glandular structures in histology images. Their approach models each gland as a polygon whose vertices represent epithelial nuclei locations, formulated as a Bayesian inference problem. Using Reversible-Jump Markov Chain Monte Carlo simulation, RPM can effectively segment glands across various differentiation grades of adenocarcinomas. The method doesn't assume architectural regularity of glandular structures, making it versatile for different cancer grades. However, its stochastic nature requires higher computational resources compared to deterministic approaches, which could potentially impact its clinical implementation.

Graph Neural Networks (GNNs) have proven effective in capturing spatial relationships between tissue regions [17]. Gao et al. [18] proposed an integrated CNN-GCN framework that captures spatial correlations in breast cancer histopathological images. Their architecture consists of an adaptive graph constructor and a novel graph learning module that eliminates the need for complex preprocessing like nuclei detection or tissue segmentation. This sophisticated design achieved superior performance on BioImaging 2015 (94.40% accuracy) datasets, significantly outperforming traditional CNN and GCN approaches.

Wang et al. [19] proposed a connectivity-aware graph transformer (CGT) for breast cancer classification from digital pathology images. Their approach constructs tissue graphs from histopathology images and employs a novel architecture that integrates connectivity embedding at every graph transformer layer through local connectivity aggregation. The model encodes spatial distance between nodes as connectivity bias in self-attention calculation, allowing it to distinctively capture topological relationships. Unlike methods that process entire whole slide images directly, CGT works with tissue graphs, which reduces computational requirements while maintaining the ability to represent complex cellular patterns. Experimental results on the

BRACS dataset demonstrated that CGT outperforms state-of-the-art methods and achieves classification performance comparable to domain expert pathologists for certain cancer subtypes, with lower standard deviations indicating more stable performance.

2.4 Recent Advances in IDC Grading

Voon et al. [20] evaluated the performance of seven CNNs for grading invasive ductal carcinoma in breast histopathological images. EfficientNetB0 [21], ResNet50 [22], and MobileNet [23] demonstrated outstanding performance. EfficientNetB0 employs a compound scaling method, achieving high accuracy and efficiency. ResNet50 alleviates the training difficulties of deep networks through residual connections, excelling in feature extraction and classification. MobileNet uses depthwise separable convolutions, making it lightweight and effective in resource-constrained environments. These models achieved over 90% accuracy on the Four Breast Cancer Grades (FBCG) dataset [24], establishing them as representative CNN methods for IDC grading tasks.

Similarly, Hattiya et al. [25] conducted a comparative study on seven CNN architectures for diabetic retinopathy detection, evaluating models such as ResNet50, DenseNet201, and InceptionV3. Their findings aligned with those of Voon et al., further reinforcing the strong performance of ResNet50, DenseNet201, and InceptionV3 in IDC grading tasks.

S. Sharma et al. [26] used an ensemble of CNN networks for the automatic grading of breast cancer, testing various combinations of base models—including DenseNet [27], ResNet, Inception V3 [28], MobileNet, and VGG [29]—which resulted in improved accuracy on the Databiox dataset [30], albeit with limited gains.

Furthermore, E. Kumaraswamy et al. [31] investigated the impact of feature extraction techniques on WSI-based cancer grade classification. Their work combined the VGG16 model with a classifier, demonstrating that applying this method to whole histopathological images yields superior results and offers valuable insights for high-resolution WSI classification.

2.5 Challenges in Automated IDC Grading

Despite significant advancements in computational pathology, automated IDC grading remains a challenging task. According to the NHG standard, IDC grading requires a more nuanced feature analysis than subtype classification. While subtype classification primarily depends on distinct cellular and architectural patterns, grading necessitates the evaluation of subtle variations in nuclear pleomorphism, mitotic activity, and tubule formation. These morphological characteristics often exhibit continuous rather than discrete variations, making traditional classification methods less effective.

Furthermore, WSI-level grading introduces substantial computational and methodological challenges due to the gigapixel scale of whole-slide images and the need to integrate contextual information across tissue regions. The scarcity of well-annotated IDC grading datasets further complicates the task, as creating grade-annotated collections requires extensive expert review and consensus among pathologists.

This study proposes a novel approach to mitigate the challenges of direct whole-slide grading. By dividing WSIs into manageable patches and using an autoencoder to extract condensed feature representations, this approach preserves vital diagnostic details. The encoded patch features are systematically aggregated to maintain spatial relationships, creating an efficient yet comprehensive representation of the entire slide for subsequent classification. This feature-focused method provides an effective and computationally practical solution for grading breast cancer, closing the divide between thorough analysis and achievable scalability, as explained in the following sections.

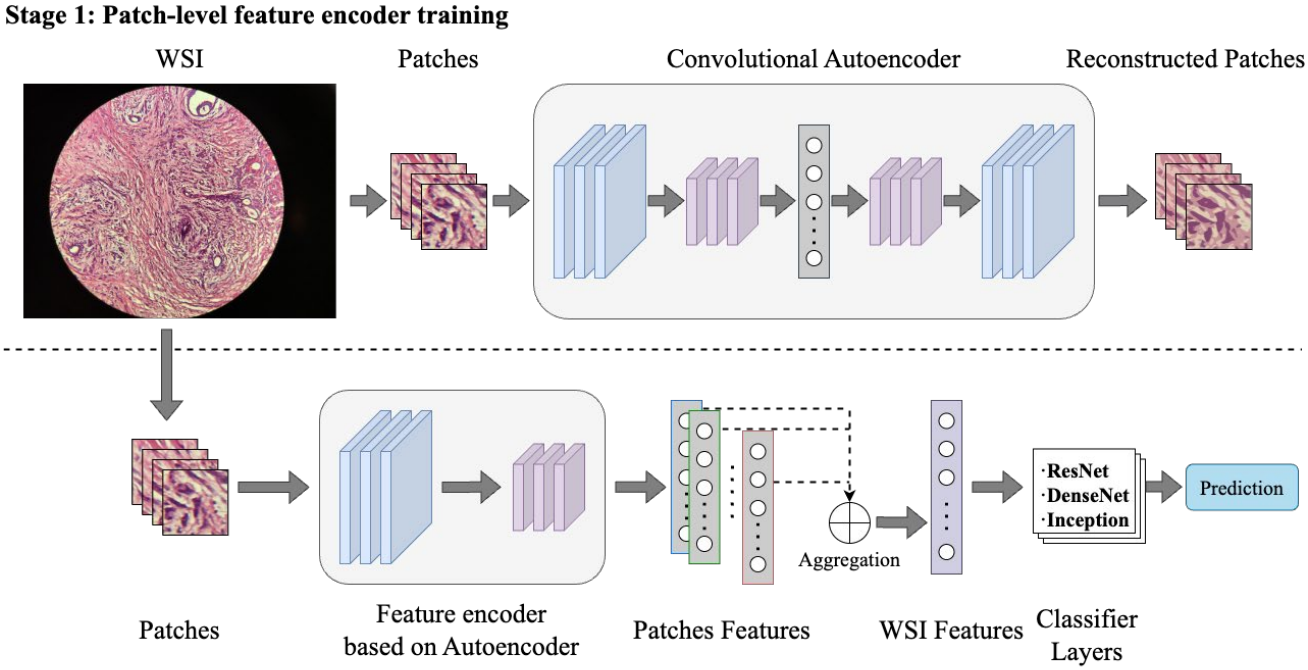


Figure 1. Overall process of proposed feature encoding for grade classification.

3. Methodology

The accurate grading of Invasive Ductal Carcinoma (IDC) is crucial for evaluating disease progression and tailoring effective treatment strategies. This innovative study introduces a novel feature encoding technique for analyzing high-resolution Whole Slide Images (WSIs). Our technique is specifically engineered to address the complexities inherent in IDC grading, thereby enhancing the precision and reliability of the grading process. The diagram in Fig. 1 provides an overview of the proposed method for encoding features and its corresponding application steps. The main objective of this approach is twofold: to transforming high-resolution WSIs into compact, information-rich representations that preserve diagnostic features and to achieve precise classification of IDC severity.

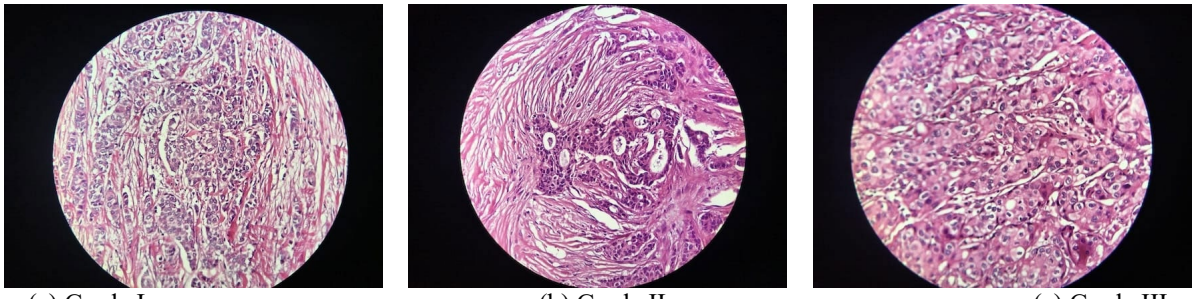
The approach to addressing this problem begins by dividing the WSI into a series of smaller patches. Following this, a training process is carried out using a dataset of these patches to train an Autoencoder network. The autoencoder employs convolutional layers to compress input images into a lower-dimensional latent space representation. Subsequently, a reconstruction process involving deconvolution is used to generate an image that closely resembles the original input.

In order to facilitate WSI classification and IDC grading, the proposed feature encoding technique harnesses the power of a trained encoder component to extract features from each patch. These extracted patch features are then amalgamated to construct a comprehensive feature representation of the WSI. The aggregation method takes into consideration the original spatial relationships between the patches, effectively preserving the density and distribution characteristics of the cancerous regions. Subsequently, the aggregated WSI feature representation is channeled into a deep learning-based classifier, which is adept at performing the grading task. The input layer of the classifier is adeptly adjusted to accommodate the dimensionality of the WSI encoding. Through rigorous training, the deep learning classifier learns both patch-level features and the spatial relationships between patches. This allows it to leverage multi-scale features from high-resolution WSIs, thereby enhancing classification performance.

The proposed feature encoding technique offers several advantages that could greatly benefit our work. First, a patching and feature reconstruction process is implemented. This process effectively addresses the challenge of handling high-resolution WSIs, which typically demands significant computational resources. For feature extraction, a pre-trained encoder is employed. This encoder processes small patch images to extract and compress features. Such an approach prevents the loss of microscopic details - a crucial aspect of pathological analysis. Furthermore, the deep learning-based classifier has two key strengths. It examines micro-level features while simultaneously analyzing their large-scale relationships. Through this dual approach, a more comprehensive integration of multi-scale features into the grading decision is achieved. It is expected that the proposed approach can significantly improve the accuracy. In the following subsections, a detailed theoretical explanation is provided. Each step of the method is thoroughly described to ensure a clear understanding for all readers.

3.1 Data and Preprocessing

In this research study, a comprehensive analysis of histopathological images of invasive ductal carcinoma (IDC) was conducted with the aim of classifying cancer grades. The dataset comprises images of Grade I, II, and III breast cancer obtained from 124 patients at the Anahid Clinic and Pour Sina Hakim Digestive Diseases Research Center, Isfahan University of Medical Science [30]. The distribution of patients across the grades includes 37 for Grade I, 43 for Grade II, and 44 for Grade III. The whole-slide images were captured at 4x, 10x, 20x, and 40x magnifications, all in RGB format, and saved as JPEG files. For our research, we concentrated on a subset of 40x magnification images with a resolution of 4032×3024 pixels. This subset includes 131 Grade I, 180 Grade II, and 143 Grade III images. An example of histology is illustrated in Fig. 2. Based on our investigation, this represents the only open access breast cancer WSI dataset with grade labeling.



(a) Grade I (b) Grade II (c) Grade III

Figure 2. Sample histology slides of IDC (graded by the expert pathologists) from the dataset at a magnification of 40x [30].

To standardize image dimensions and minimize the impact of microscope imaging, the original dataset was cropped using an inscribed square method, as shown in Fig. 3. The resulting cropped images all have a consistent resolution of 2048×2048 pixels, which represents the maximum feasible size while ensuring uniformity across all samples. These cropped images form the dataset for the baseline method.

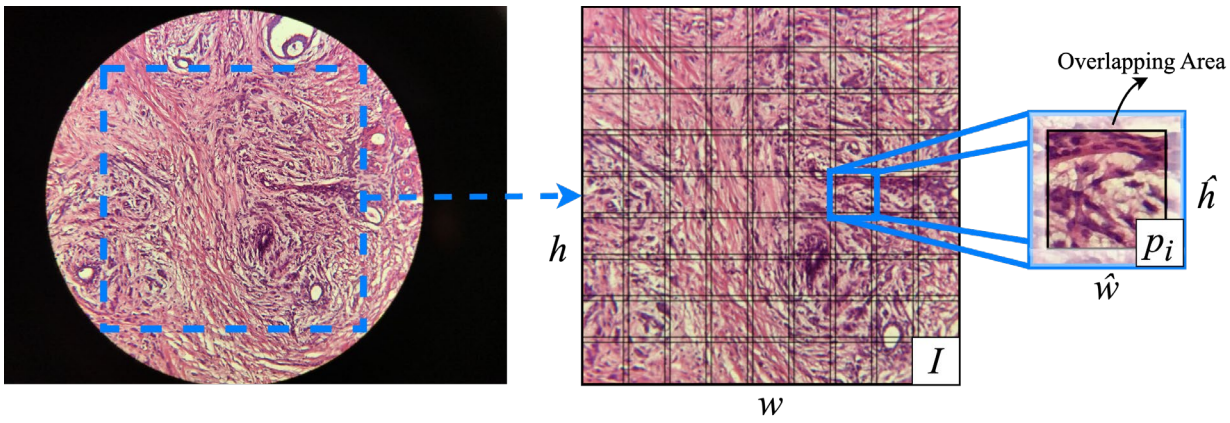


Figure 3. WSI patching with overlap.

3.2 Patch Encoding

Let $I \in \mathbb{R}^{w \times h}$ represents an image of dimension $w \times h$. The image I is divided into n small size, overlapping patches, each of dimension $\hat{w} \times \hat{h}$. The set of patches can be denoted as $I = p_1, p_2, ..., p_n$, where each patch $p_i \in \mathbb{R}^{\hat{w} \times \hat{h}}$. The process is illustrated in Fig. 3.

To train a model for encoding the histology patches, the patches are firstly collected from a separate set of data. A patch encoding function $f: p \rightarrow t$ is then constructed, where t represents the encoded feature vector for each patch p. The goal of this encoding function is to map each patch p to its corresponding representation $t \in \mathbb{R}^d$, where d is the dimensionality of the encoded feature space.

An autoencoder model is utilized as the encoding function f [32]. The autoencoder consists of two main components: an encoder function f_{enc} and a decoder function f_{dec}. An overview of the autoencoder structure is shown in Fig. 4. The encoder f_{enc} compresses each patch p_i into a lower-dimensional latent representation $t_i \in \mathbb{R}^d$, where d is the dimensionality of the latent space. The decoder f_{dec} reconstructs the original patch from the latent representation. For a given patch p_i, the encoding and decoding processes can be represented as:

$$t_i = f_{enc}(p_i) \tag{1}$$

$$\hat{p_i} = f_{dec}(t_i) \tag{2}$$

The autoencoder undergoes training through the minimization of the reconstruction loss, which quantifies the difference between the original patch p_i and its reconstruction $\hat{p_i}$. A commonly employed approach for quantifying reconstruction loss is using the mean squared error (MSE) as the designated metric stipulated by equation (3). This loss function is frequently employed in training autoencoders to minimize the differences between the input and output data.

$$L = \frac{1}{n} \sum_{i=1}^{n} \left|| p_i - \hat{p_i} \right||^2 \tag{3}$$

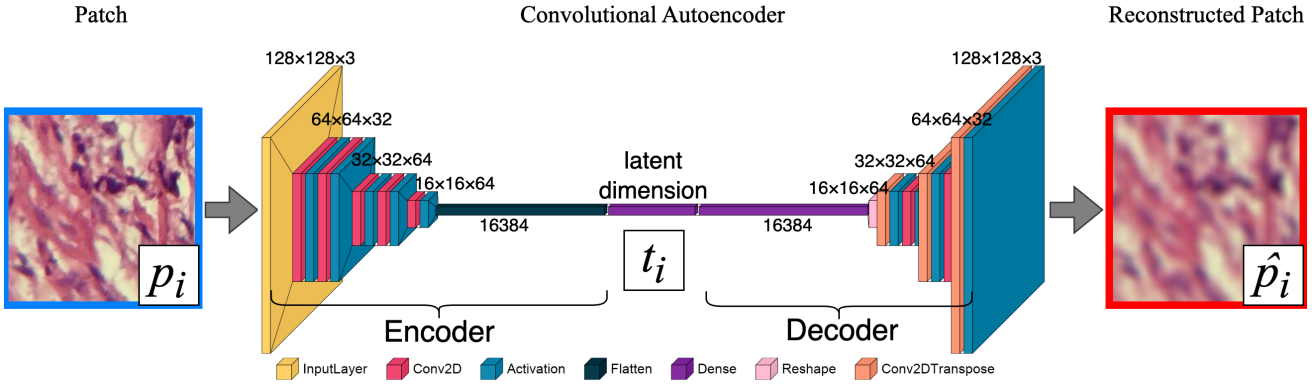


Figure 4. Overall process of Autoencoder training.

3.3 Slide Feature Generation

With a trained autoencoder model in hand, our objective is to generate features for the full slide image $I \in \mathbb{R}^{w \times h}$. The procedure has three distinct steps.

The initial step involves image patching and feature extraction, which is clearly depicted in Fig. 5. The image I is divided into n overlapping patches p_i. This patch division is identical to the one used during the training process. For each patch p_i, the trained encoder f_{enc} is used to generate the latent representation (feature vector) $t_i \in \mathbb{R}^d$, refer to equation (1). The set of features for all patches of the image I can be written as: $T_I = t_1, t_2, ..., t_n$, $T \in \mathbb{R}^{n \times d}$, and T_I represents the concatenation of the feature vectors for all patches of I.

To effectively analyze WSIs, we need to consolidate the features extracted from individual patches into a unified representation. This WSI-level feature aggregation utilizes feature concatenation, which is a straightforward yet effective technique. It effectively preserves spatial relationships while ensuring computational efficiency. This method creates a unified feature matrix by sequentially combining latent representations (t_i) of individual patches. The resulting matrix, while substantially smaller than the original WSI resolution, retains the essential spatial structure of the image. The concatenation process can be formally described as:

$$T_I = [t_1, t_2, \dots, t_n] = \begin{bmatrix} t_1 & t_2 & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & t_n \end{bmatrix} \tag{4}$$

The WSI feature representation T_I is transformed from a set of t_i in $\mathbb{R}^{1\times d}$ into a comprehensive feature vector in $\mathbb{R}^{nrow\times ncol\times d}$. This ensures that the final feature vector contains rich spatial and contextual information from the original WSI.

3.4 Slide Grading

This research endeavor is focused on transforming the slide grading task into an image classification problem. The approach of generating features from the autoencoder model for classification offers a distinct advantage over directly classifying whole slide images. It enables the implementation of a more advanced and refined deep learning-based technique. This method enables a meticulous linking of the extracted feature representations with the corresponding slide grades, ultimately leading to a more nuanced and accurate classification process. Let $I\in\mathbb{R}^{w\times h}$ denote the whole slide image, and $t_I\in\mathbb{R}^{nrow\times ncol\times d}$ be the corresponding feature vector extracted from the patches of I using the encoder function f_{enc} (see the previous section). The objective of the grading task is to classify image I into one of C possible grades (e.g., severity levels).

This process can be formulated as learning a mapping function $g: \mathbb{R}^{nrow\times ncol\times d}\rightarrow 1, 2, \dots, C,$ where $g(t_I)$ represents the predicted grade for the image I. Mathematically, this can be expressed as:

$$\hat{y} = g(t_I) \tag{5}$$

where $\hat{y}\in 1, 2, \dots, C$ is the predicted class label (grade) for the image I, and t_I is the feature vector representing the image. This work utilizes a neural network g_θ, parameterized by θ, mapping the input feature vector t_I to the predicted class label $\hat{y}$. This neural network can be organized as a sequence of fully connected layers, followed by a softmax activation function to produce the probabilities for each class.

The final output layer of the network is denoted by z, while the weight matrix and bias vector are represented by W and b, respectively. The output layer applies to a linear transformation followed by a softmax function to produce the class probabilities:

$$p = \text{softmax}(Wz + b) \tag{6}$$

where $p = [p_1, p_2, \dots, p_C]$ is the vector of class probabilities, and the probability of the image is classified into grade j. The softmax function is defined as p_j:

$$p_j = \frac{\text{EXP}((Wz+b)_j)}{\sum_{k=1}^{C}\text{EXP}((Wz+b)_k)} \tag{7}$$

The predicted class label $\hat{y}$ is then given by the class with the highest probability as:

$$\hat{y} = \arg\max_j(p_j) \tag{8}$$

The subsequent section encompasses an in-depth examination of the training procedure, evaluation metrics, and the acquired results. A comparative analysis evaluates the proposed approach compared to baseline methods, aiming to provide a comprehensive understanding of its inherent advantages and limitations.

4. Experiment and Results

This section outlines the performance metrics, experimental setup, and results for assessing the proposed WSI grading method. The experiments encompass two crucial stages: firstly, training an autoencoder for feature extraction, and secondly, training a classification model utilizing the extracted features.

4.1 Performance Metrics

To comprehensively evaluate the proposed model, two sets of evaluation metrics are employed: feature encoder metrics to assess the quality of extracted features and classification metrics to evaluate the diagnostic performance. The performance assessment of our approach is grounded in information theory principles and statistical learning theory. Information theory guides our feature encoding evaluation, measuring how effectively the autoencoder preserves critical diagnostic information while reducing dimensionality. Statistical learning theory provides the foundation for evaluating how well our model generalizes from training to testing data through cross-validation protocols.

A. Feature Encoder Metrics

Two standard metrics are utilized to evaluate the quality of the feature encoder. Peak Signal-to-Noise Ratio (PSNR) measures the pixel-level fidelity of reconstructed images compared to the original ones, defined as:

$$\text{PSNR} = 10 * \log_{10} \frac{max(I)^2}{\text{MSE}} \tag{9}$$

where *max(I)* is the maximum possible pixel value (255 for 8-bit images), and MSE is the Mean Squared Error between the original and reconstructed images, reference equation (3). Higher PSNR values indicate better reconstruction quality.
Structural Similarity Index (SSIM) assesses the perceived quality of digital images by measuring the structural similarity between two images:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{10}$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y.

B. Classification Metrics

The diagnostic performance of this model is rigorously evaluated using a comprehensive suite of classification metrics. These metrics encompass accuracy, precision, recall, F1 score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), each providing unique insights into different aspects of model performance.

The fundamental metrics are derived from the confusion matrix elements: True Positives (TP, correctly identified positive cases), True Negatives (TN, correctly identified negative cases), False Positives (FP, incorrectly identified positive cases), and False Negatives (FN, incorrectly identified negative cases). The basic performance metrics are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

$$\text{F1} = \frac{2*TP}{2*TP+FP+FN} \tag{14}$$

To comprehensively assess the model's discriminative capability across different decision thresholds, Receiver Operating Characteristic (ROC) analysis is employed. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), defined as:

$$TPR = \frac{TP}{TP+FN} \tag{15}$$

$$FPR = \frac{FP}{FP+TN} \tag{16}$$

The integral of this curve yields the AUC-ROC score, providing a threshold-independent performance metric. This score ranges from 0.5, indicating performance equivalent to random chance, to 1.0, representing perfect classification. For the evaluation of our multiclass classification model, we employed the one-vs-rest strategy to calculate AUC-ROC scores. This approach iteratively treats each class as positive while considering all other classes as negative. The final results of all metrics were aggregated across different labels using the macro-averaging method.

4.2 Experiment Settings

The experiment consisted of three main stages: autoencoder training, feature extraction, and classification. The experimental design followed established machine learning evaluation principles to ensure robustness and reliability. We employed stratified 5-fold cross-validation to maintain class distribution consistency across training and validation splits, addressing potential bias from the dataset's moderate imbalance (131 Grade I, 180 Grade II, and 143 Grade III images).

To evaluate the generalizability of our feature encoding approach, two experimental settings were compared: (1) a baseline approach using downsampled WSIs as direct inputs for classification and (2) our proposed method, where encoded feature vectors served as classifier inputs. This comparison helped quantify the advantages of feature encoding. We selected diverse deep learning architectures: ResNet-50, DenseNet-201/DenseNet-121, and Inception V3. These models were chosen for their distinct feature extraction strategies—ResNet's residual connections enhance gradient flow in deep networks, DenseNet's dense connectivity promotes feature reuse, and Inception's multi-scale processing captures spatial features at different resolutions. And they have shown strong performance in related IDC grading tasks [17]. This architectural diversity allowed us to assess whether the benefits of feature encoding are consistent across different classification paradigms.

For implementation, WSIs were resized to match model input requirements (224×224 for most networks, 299×299 for Inception V3). To ensure compatibility with lower-dimensional latent features, we replaced DenseNet-201 with DenseNet-121 and applied zero-padding to match Inception V3's input size. Data augmentation was also used to mitigate overfitting while

preserving critical pathological characteristics.

4.3 Hyperparameters Configuration

Extensive experiments were conducted to determine the optimal hyperparameter settings for this model. Table 1 presents the comprehensive hyperparameter configuration used in our experiments.

Table 1 Hyperparameter settings

Hyperparameter	Value
Model input size	128×128(Autoencoder) 224×224(Classifier)
Number of epochs	100 (Autoencoder) 100 (Classifier)
Batch size	64
Optimizer	Adam
Learning rate	0.001
Data augmentation	Random resized crop Rotation 90°, 180°, 270° Random vertical flip Random horizontal flip
Number of fold	5

4.4 Results

A. Autoencoder Training

To access the performance of feature encoding techniques for WSI of IDC, we first preprocessed the data into 128×128 image patches and trained an autoencoder. The encoder compresses features through several convolutional layers to generate latent space representations of patch. The decoder employs transposed convolutional layers to reconstruct patch image from the latent representations. The experiments were conducted with various latent dimensions to determine the optimal latent space dimension for this specific dataset.

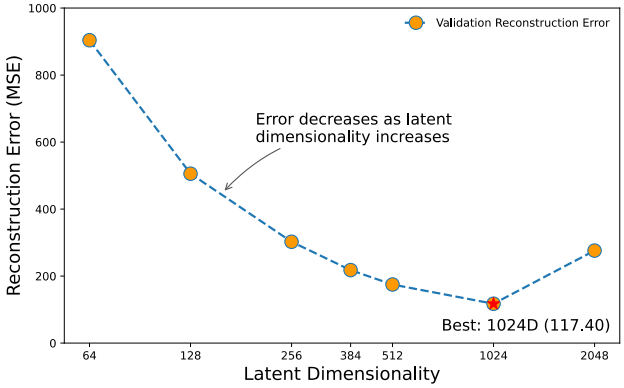


Fig. 6 Impact of latent dimensionality on reconstruction quality

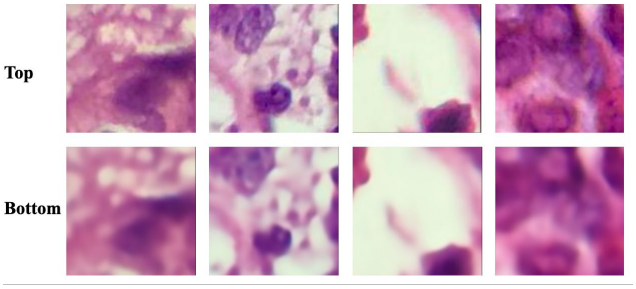


Fig. 7 Visual comparison between original patch and reconstructed patch for 1024 latent dim (**Top**: the original patch images and **Bottom**: the corresponding decoded images)

The reconstruction error for various dimensions can be observed in Fig. 6, while Fig. 7 displays sample original and reconstructed images. Upon visual inspection, it is evident that the reconstructed images maintain the crucial cellular structural features required for pathological diagnosis.

Table 2 Performance metrics comparison over latent dimensionality of Autoencoder

Dim	64	128	256	384	512	1024	2048
PSNR(dB)	30.37	31.51	32.83	33.68	34.26	**34.86**	31.90
SSIM	0.70	0.76	0.81	0.84	0.86	**0.90**	0.85

Reconstruction quality was assessed using 100 original images and their corresponding reconstructed images, evenly distributed across all classes in the test set. The comparison of results can be found in Table 2.
In our study, PSNR increased with higher latent space dimensions, peaking at 34.8617dB for 1024 dimensions. This suggests that the reconstructed images had minimal distortion, making the differences almost imperceptible to the human eye. SSIM also improved with higher dimensions, reaching a maximum of 0.8981 at 1024 dimensions, indicating high similarity to the original images (refer to Fig. 7). A slight drop in PSNR and SSIM was observed at 2048 dimensions, which could be due to insufficient training dataset.

B. Performance on Slide Grading

The comprehensive evaluation of our proposed feature encoding approach against baseline methods revealed significant performance improvements across all key metrics. As shown in Table 3, our feature encoding method consistently outperformed baseline models using the resizing approach across all evaluation metrics. Among the tested architectures, the ResNet-50 classifier with our feature encoding method achieved the highest accuracy of 71.43% and AUC-ROC score of 0.8940, representing a meaningful improvement over the baseline ResNet-50 model (66.67% accuracy, 0.8894 AUC-ROC).

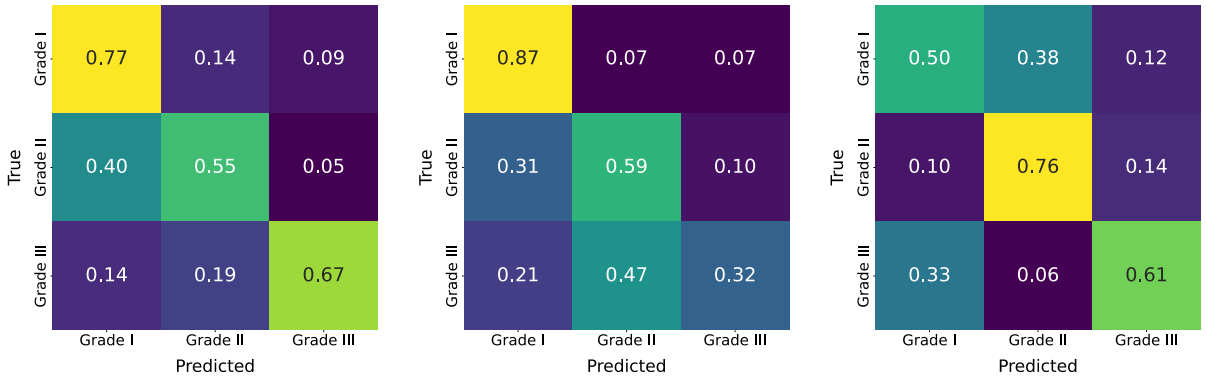
Of particular clinical significance is the improvement in recall (sensitivity), which increased from 0.6631 in the baseline ResNet-50 to 0.7246 in our feature encoding-enhanced model. This 6.15% improvement in recall is especially critical in IDC grading, as it directly relates to the model's ability to correctly identify higher-grade malignancies. In clinical practice, missing a high-grade tumor (false negative) carries substantially greater risk than over-grading a lower-grade tumor, as under-treatment of aggressive cancers can lead to poorer patient outcomes. The enhanced recall of our model means fewer high-grade tumors would be misclassified as lower grade, potentially reducing the risk of insufficient treatment planning.

Our model also demonstrated improved F1 score (0.7184 versus 0.6653 in the baseline), which represents a balanced measure of precision and recall. This balanced performance is crucial for optimizing resource allocation in clinical settings while maintaining patient safety. While maximizing recall remains the primary concern in cancer grading, the improved F1 score indicates that our model achieves this without excessive false positives that could lead to unnecessary aggressive treatments, patient anxiety, and medical resource wastage.

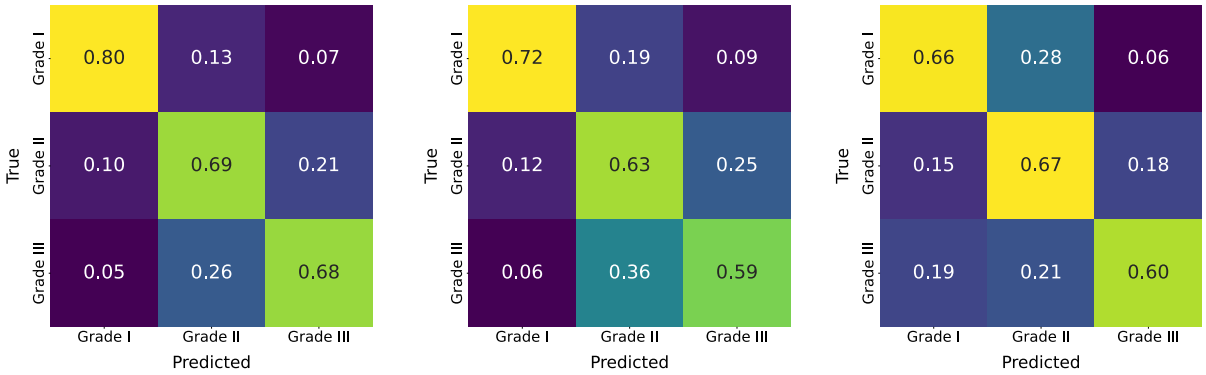
The precision metric also showed improvement (0.7136 versus 0.6806 in the baseline), indicating better specificity in identifying true high-grade cases. This enhanced precision helps avoid over-treatment scenarios, reducing unnecessary treatment toxicity and improving resource utilization. In clinical implementation, higher precision would increase pathologists' confidence in the system's recommendations, potentially improving adoption rates of AI-assisted diagnostic tools.

Table 3 Performance metrics comparison: baseline model vs feature encoding models

Model	Method	Accuracy	Precision	Recall	F1 Score	AUC-ROC
ResNet-50	Resize	0.6667	0.6806	0.6631	0.6653	0.8894
DenseNet-201	Resize	0.5714	0.5765	0.5896	0.5517	0.8308
Inception V3	Resize	0.6190	0.6208	0.6243	0.6183	0.7656
ResNet-50	Feature Encoding	**0.7143**	**0.7136**	**0.7246**	**0.7184**	**0.8940**
DenseNet-121	Feature Encoding	0.6467	0.6592	0.6482	0.6518	0.8037
Inception V3	Feature Encoding	0.6435	0.6501	0.6437	0.6444	0.8004



(a) ResNet-50 (b) DenseNet-201 (c) Inception V3



(d) ResNet-50 (FE) (e) DenseNet-121 (FE) (f) Inception V3 (FE)

Fig. 8 Confusion matrix comparison: baseline model vs feature encoding models.

The confusion matrix comparison in Fig. 8 provides further evidence of our approach's superior performance. Feature encoding-based models exhibited fewer misclassifications, particularly between Grade I and Grade III—the most clinically significant distinction where misclassification would have the most severe treatment implications. Among feature encoding models, ResNet-50 and DenseNet-121 demonstrated superior discriminative ability for Grade I, while Inception-V3 achieved more balanced accuracy across all three grades. The persistent classification challenges primarily stem from the complex similarities between intermediate and advanced IDC stages (Grades II and III), which frequently cause divergent opinions even among experienced pathologists.

The improved recall for higher grades in our model aligns with clinical practice, where pathologists tend to upstage borderline cases to ensure adequate treatment. This conservative approach to cancer grading is appropriately reflected in our model's performance characteristics, making it more suitable for real-world clinical application where patient safety is paramount. To provide pathology experts with more intuitive insights into our model's performance, Fig. 9 presents representative examples of both correctly and incorrectly classified samples across all three grades. For visualization clarity, we display the square-cropped WSIs. While our pathological expertise is limited, we offer observations from a computational classification perspective.

The correctly classified Grade I samples (Fig. 9, top row, left) exhibit well-formed tubular structures with relatively uniform nuclei and minimal pleomorphism. These images display characteristic low-grade features including regular cellular organization and abundant stromal elements. The model successfully captures these defining Grade I attributes, particularly the tubular formation patterns that pathologists consider during grading.

For Grade II (Fig. 9, top row, middle), correctly classified images show intermediate features with moderate nuclear pleomorphism and partial tubule formation. These samples present a balance of organized and disorganized tissue architecture - a challenging intermediate pattern that our feature encoding method effectively distinguishes from both low and high grades.

Correctly classified Grade III samples (Fig. 9, top row, right) display the hallmark features of high-grade malignancy: marked nuclear pleomorphism, absence of tubule formation, and sheets of disorganized tumor cells. The model successfully identifies these aggressive patterns, which represent cases where proper classification is most clinically critical.

The misclassified samples (Fig. 9, bottom row) reveal several interesting patterns. Grade I samples misclassified as Grade II (bottom row, left) typically contain areas with higher cellularity or slight nuclear atypia that mimic intermediate-grade features. For Grade II samples (bottom row, middle), misclassifications often occur in cases with heterogeneous features - some regions appearing more like Grade I while others resemble Grade III. This heterogeneity reflects the inherent challenge pathologists face when encountering borderline cases.

Most notably, Grade III samples misclassified as Grade II (bottom row, right) often display areas of residual tubule formation alongside highly pleomorphic regions. These challenging cases highlight the clinical reality that cancer grades exist on a continuum rather than as discrete categories, explaining the persistent challenge of distinguishing borderline cases even with advanced computational approaches.

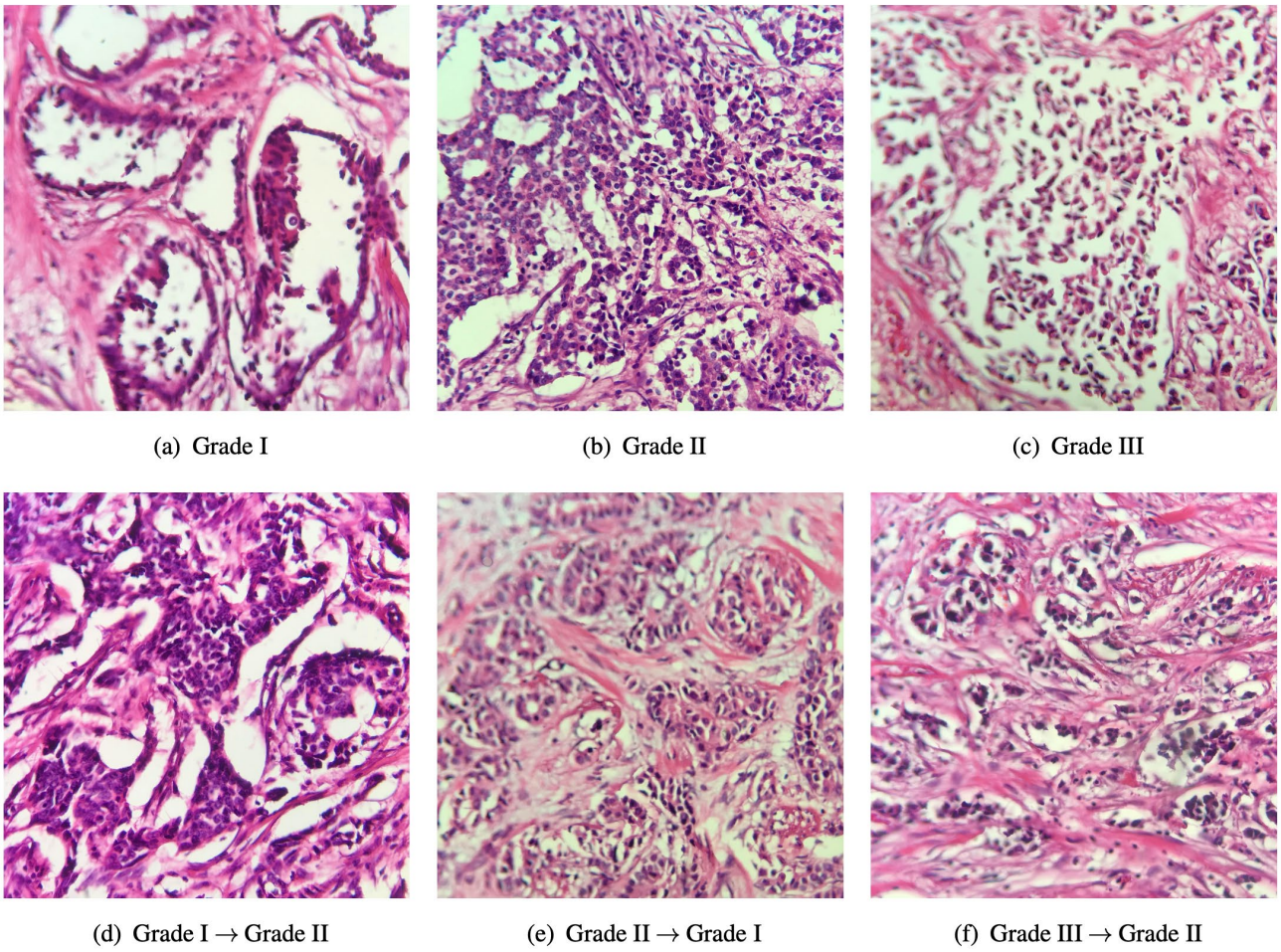


Fig. 9 Visual samples of correct grading (top) and incorrect grading (bottom) by ResNet-50 with feature encoding technique.

These visual examples complement our quantitative results, illustrating both the strengths of our feature encoding approach in capturing diagnostic patterns and the inherent challenges of IDC grading that exist even for human experts. The model performs particularly well on cases with clear, classic features of their respective grades, while misclassifications typically occur in borderline cases with mixed or ambiguous features - precisely the same cases that often generate disagreement among pathologists.

Beyond classification performance, our method demonstrated remarkable computational efficiency. To ensure consistency, all experiments were conducted on the Google Colab platform using an NVIDIA L4 GPU. As illustrated in Fig. 10, training time was reduced by approximately 50-67% compared to baseline approaches. This efficiency gain is attributable to two factors: (1) the compact nature of our encoded features (288×288×1) compared to baseline resized WSIs (224×224×3), and (2) our optimized training pipeline that pre-computes encoded features, eliminating redundant processing during training iterations.

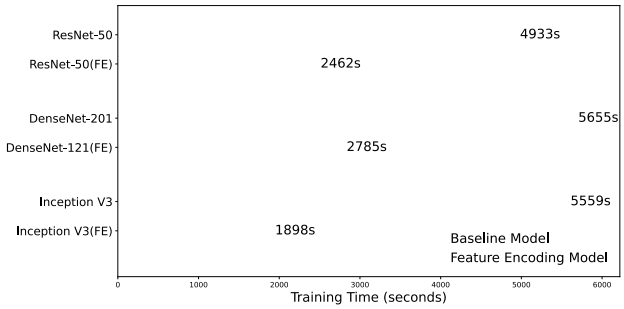


Fig. 10 Training time comparison (100 epochs): baseline model vs feature encoding models.

4.5 Discussion

We established model effectiveness through multiple complementary approaches. First, we implemented k-fold cross-validation (k=5) to ensure robust performance estimation across different data partitions, mitigating potential selection bias. Second, we verified the biological plausibility of our approach by confirming that the autoencoder preserves critical histological features essential for pathological diagnosis (as demonstrated in Fig. 7). Third, we quantitatively validated our feature extraction method by demonstrating high PSNR (34.86dB) and SSIM (0.90) values, confirming that encoded representations retain diagnostically relevant information. Finally, the consistent performance improvement across multiple classification architectures (ResNet-50, DenseNet-121, and Inception V3) provides strong evidence for the validity of our feature encoding approach,

demonstrating that the benefits are independent of the specific classification architecture employed.

These results collectively demonstrate that our feature encoding approach effectively captures the essential diagnostic information from high-resolution WSIs while significantly reducing computational demands. The superior performance across all metrics, combined with substantially improved efficiency, positions this method as a viable solution for practical clinical implementation in IDC grading tasks.

While our best model achieved 71.43% accuracy, which is lower than some recent CNN-based approaches reporting >80% accuracy, our method offers several significant advantages that justify its adoption. First, most higher-accuracy methods in literature were evaluated on different datasets with potentially less challenging cases or different grading criteria, making direct accuracy comparisons problematic. Second, our approach provides substantial computational efficiency gains, reducing training time by 50-67% compared to baseline methods (Fig. 10), which is crucial for real-world clinical implementation. Third, our method maintains interpretability by preserving spatial relationships between tissue regions, unlike black-box end-to-end approaches. Finally, our feature encoding technique demonstrates remarkable robustness to the unique challenges of the Databiox dataset, which contains significant inter-grade similarities that reflect real-world diagnostic challenges faced by pathologists. This robustness to challenging cases, combined with computational efficiency and interpretability, makes our approach preferable for practical clinical applications despite the moderate accuracy improvement over baseline methods on this specific dataset.

5. Conclusion

This study introduced a novel deep learning-based feature encoding approach for grading IDC from high-resolution whole-slide images. Unlike traditional methods, which often compromise diagnostic accuracy due to downsampling or computational constraints, our approach effectively balances computational efficiency and diagnostic fidelity by utilizing a patch-based autoencoder framework. The encoder extracted meaningful latent representations from tissue patches, which were aggregated to preserve spatial relationships and improve classification accuracy. Experimental results demonstrated that our method achieved an accuracy of 71.43% and an AUC-ROC score of 0.8940, outperforming conventional resizing-based approaches.

Our findings are consistent with prior studies highlighting the importance of preserving histopathological details in WSI analysis while addressing computational challenges. Specifically, our framework overcomes the limitations of the traditional CNN network using downsampling methods, which often lose critical diagnostic features, by employing a high-fidelity patch encoding and dimensionality reduction approach combined with effective feature aggregation. Moreover, our method meets the clinical demand for efficient and reliable IDC grading tools by significantly reducing training time (by approximately 50-67%) while maintaining interpretability—both crucial factors for real-world clinical implementation.

Despite its contributions, this study has limitations. The dataset size and diversity were constrained, potentially limiting the generalizability of the model to other cancer types or populations. In addition, compared to classification results from expertly curated datasets (accuracy > 80%), this model's accuracy still has significant opportunities for improvement, particularly in distinguishing between Grade II and Grade III cases. Future research should expand dataset diversity, explore other deep encoding components (such as VGG [29]) within our framework, investigate advanced architectures (e.g., Vision Transformers [32]), and develop interpretable methods for visualizing aggregated features. Moreover, extending the application of this feature encoding framework to other domains, such as radiology or multi-modal medical imaging, could further validate its utility.

This study provides a robust foundation for automated IDC grading, demonstrating the potential of deep learning-based feature encoding to enhance diagnostic consistency, reduce pathologists' workload, and improve patient care. Future efforts should focus on addressing identified limitations and expanding the applicability of this approach to broader clinical and research contexts.

Acknowledgements

We would like to thank Hamidreza Bolhasani for providing the grading dataset for breast-invasive ductal carcinomas. This research project was financially supported by Mahasarakham University.

References

[1] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine,* vol. 25, no. 8, pp. 1301-1309, 2019/08/01 2019, doi: 10.1038/s41591-019-0508-1.

[2] C. van Dooijeweert, P. J. van Diest, and I. O. Ellis, "Grading of invasive breast carcinoma: the way forward," *Virchows Archiv,* vol. 480, no. 1, pp. 33-43, 2022/01/01 2022, doi: 10.1007/s00428-021-03141-2.

[3] E. Gray, A. Donten, K. Payne, and P. S. Hall, "Survival estimates stratified by the Nottingham Prognostic Index for early breast cancer: a systematic review and meta-analysis of observational studies," *Systematic Reviews,* vol. 7, no. 1, p. 142, 2018/09/15 2018, doi: 10.1186/s13643-018-0803-9.

[4] N. K. C. Pratiwi, Y. N. Fu'adah, N. Ibrahim, S. Rizal, and S. Saidah, "Invasive Ductal Carcinoma (IDC) Classification Based on Breast Histopathology Images Using Convolutional Neural Network," Singapore, 2021: Springer Singapore, in Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics, pp. 477-486.

[5] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis,* vol. 67, p. 101813, 2021/01/01/ 2021, doi: https://doi.org/10.1016/j.media.2020.101813.

[6] A. Korzynska, L. Roszkowiak, J. Zak, and K. Siemion, "A review of current systems for annotation of cell and tissue images in digital pathology," *Biocybernetics and Biomedical Engineering,* vol. 41, no. 4, pp. 1436-1453, 2021/10/01/ 2021, doi: https://doi.org/10.1016/j.bbe.2021.04.012.

[7] F. Aeffner *et al.*, "Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association," *Journal of Pathology Informatics,* vol. 10, no. 1, p. 9, 2019/01/01/ 2019, doi: https://doi.org/10.4103/jpi.jpi_82_18.

[8] D. Komura and S. Ishikawa, "Machine Learning Methods for Histopathological Image Analysis," *Computational and Structural Biotechnology Journal,* vol. 16, pp. 34-42, 2018/01/01/ 2018, doi: https://doi.org/10.1016/j.csbj.2018.01.001.

[9] I. Hirra *et al.*, "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling," *IEEE Access,* vol. 9, pp. 24273-24287, 2021, doi: 10.1109/ACCESS.2021.3056516.

[10] M. Shaban *et al.*, "Context-Aware Convolutional Neural Network for Grading of Colorectal Cancer Histology Images," *IEEE Transactions on Medical Imaging,* vol. 39, no. 7, pp. 2395-2405, 2020, doi: 10.1109/TMI.2020.2971006.

[11] A. Cruz-Roa *et al.*, *Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks* (SPIE Medical Imaging). SPIE, 2014.

[12] N. Bayramoglu and J. Heikkilä, "Transfer Learning for Cell Nuclei Classification in Histopathology Images," Cham, 2016: Springer International Publishing, in Computer Vision – ECCV 2016 Workshops, pp. 532-539.

[13] S. Dash, S. Padhy, P. Suman, and R. Kumar Das, "Enhancing Lung Cancer Diagnosis through CT Scan Image Analysis using Mask-EffNet," *Engineering Access,* vol. 11, no. 1, pp. 92-107, 12/20 2024. [Online]. Available: https://ph02.tci-thaijo.org/index.php/mijet/article/view/254068.

[14] H. U. Khan, B. Raza, A. Waheed, and H. Shah, "MSF-Model: Multi-Scale Feature Fusion-Based Domain Adaptive Model for Breast Cancer Classification of Histopathology Images," *IEEE Access,* vol. 10, pp. 122530-122547, 2022, doi: 10.1109/ACCESS.2022.3223870.

[15] X. Wang *et al.*, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Medical Image Analysis,* vol. 81, p. 102559, 2022/10/01/ 2022, doi: https://doi.org/10.1016/j.media.2022.102559.

[16] K. Sirinukunwattana, D. R. J. Snead, and N. M. Rajpoot, "A Stochastic Polygons Model for Glandular Structures in Colon Histology Images," *IEEE Transactions on Medical Imaging,* vol. 34, no. 11, pp. 2366-2378, 2015, doi: 10.1109/TMI.2015.2433900.

[17] S. Calderaro, G. L. Bosco, F. Vella, and R. Rizzo, "Breast Cancer Histologic Grade Identification by Graph Neural Network Embeddings," Cham, 2023: Springer Nature Switzerland, in Bioinformatics and Biomedical Engineering, pp. 283-296.

[18] Z. Gao, Z. Lu, J. Wang, S. Ying, and J. Shi, "A Convolutional Neural Network and Graph Convolutional Network Based Framework for Classification of Breast Histopathological Images," *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 7, pp. 3163-3173, 2022, doi: 10.1109/JBHI.2022.3153671.

[19] K. Wang, F. Zheng, L. Cheng, H. N. Dai, Q. Dou, and J. Qin, "Breast Cancer Classification From Digital Pathology Images via Connectivity-Aware Graph Transformer," *IEEE Transactions on Medical Imaging,* vol. 43, no. 8, pp. 2854-2865, 2024, doi: 10.1109/TMI.2024.3381239.

[20] W. Voon *et al.*, "Performance analysis of seven Convolutional Neural Networks (CNNs) with transfer learning for Invasive Ductal Carcinoma (IDC) grading in breast histopathological images," *Scientific Reports,* vol. 12, no. 1, p. 19200, 2022/11/10 2022, doi: 10.1038/s41598-022-21848-3.

[21] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," presented at the Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2019. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18-23 June 2018 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[24] A. Abdelli, R. Saouli, K. Djemal, and I. Youkana, "Combined Datasets For Breast Cancer Grading Based On Multi-CNN Architectures," in *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 9-12 Nov. 2020 2020, pp. 1-7, doi: 10.1109/IPTA50016.2020.9286653.

[25] T. Hattiya, K. Dittakan, and S. Musikasuwan, "Diabetic Retinopathy Detection using Convolutional Neural Network: A Comparative Study on Different Architectures: doi: 10.14456/mijet.2021.8," *Engineering Access,* vol. 7, no. 1, pp. 50-60, 01/09 2021. [Online]. Available: https://ph02.tci-thaijo.org/index.php/mijet/article/view/10.14456.mijet.2021.8.

[26] S. Sharma, S. Kumar, M. Sharma, and A. Kalkal, "An ensemble of deep CNNs for automatic grading of breast cancer in digital pathology images," *Neural Computing and Applications,* vol. 36, no. 11, pp. 5673-5693, 2024/04/01 2024, doi: 10.1007/s00521-023-09368-1.

[27] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21-26 July 2017 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[29] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," p. arXiv:1409.1556doi: 10.48550/arXiv.1409.1556.

[30] H. Bolhasani, E. Amjadi, M. Tabatabaeian, and S. J. Jassbi, "A histopathological image dataset for grading breast invasive ductal carcinomas," *Informatics in Medicine Unlocked,* vol. 19, p. 100341, 2020/01/01/ 2020, doi: https://doi.org/10.1016/j.imu.2020.100341.

[31] E. Kumaraswamy, S. Kumar, and S. Sharma, "Performance Analysis of Feature Extraction and Deep Learning Approaches on Whole and Segmented Histopathological Images for Cancer Grade Classification," in *2024 IEEE International Conference*

on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 14-16 March 2024 2024, vol. 2, pp. 1-7, doi: 10.1109/IATMSI60426.2024.10502483.

[32] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," p. arXiv:2010.11929doi: 10.48550/arXiv.2010.11929.