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

JSCDM

Vol. 6 No. 3 (2025) 46-56
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Dual-Stage Deep Learning Framework for Prostate Cancer Grading Using Swin U-Net and Attention-Based CNNs

Nattavut Sriwiboon[1], Songgrod Phimphisan[1]*

[1]Department of Computer Science and Information Technology,
Faculty of Science and Health Technology, Kalasin University, THAILAND

*Corresponding Author: songgrod.ph@ksu.ac.th
DOI: https://doi.org/10.30880/jscdm.2025.06.03.004

Article Info

Received: 15 April 2025
Accepted: 30 August 2025
Available online: 30 December 2025

Keywords

Prostate cancer, Swin U-Net, attention-based CNNs, Grad-CAM

Abstract

Accurate grading of prostatic adenocarcinoma is essential in treatment planning. However, Gleason grading is time-consuming and clinically undependable. We presented a hybrid deep learning framework which comprises Swin U-Net for transformer-based segmentation network and attention-based CNNs for ISUP grade classification task. We incorporated Grad-CAM to aid in model interpretability and to visualize decision crucial areas. Quantitative evaluations on the PANDA, ISUP Grade-wise and transverse datasets achieve 100% accuracy on the smaller balanced Transverse dataset, 90.2 ± 0.7% performance in terms of ISUP with only 3.5M parameters, and a vicious Dice score equal to 0.99 ± 0.005 for segmentation. Notably, this cross-dataset generalization has not deteriorated below 92.3 ± 1.4% in any TIO experiment with no form of retraining applied to the transferred models. Inference time is less than 20 ms, deployment on the edge and mobile. The proposed model has achieved state-of-the-art performance for interpretability, accuracy, and computational complexity. The broadcast-then-categorize platform has been validated in ablation and optimization experiments, which demonstrate the potential for real-time diagnosis of prostate cancer.

1. Introduction

Prostate cancer is one of the most common cancers in men worldwide. Proper diagnosis and grading are critical in directing successful treatment. The Gleason grading system [6], which is now updated by the ISUP grading groups [1], is often applied to evaluate cancer severity from histopathological images. Nonetheless, the manual scoring based on pathologists' evaluation is laborious and can be subjective, particularly in borderline samples. Differences in interpretation may lead to inconsistent treatment decisions, making the grading process challenging in clinical settings. Several deep learning-based frameworks have been developed to support automated prostate cancer diagnosis [2-9]. Large datasets such as PANDA [10], ISUP Grade-wise [11], and Transverse Plane [12] have made it possible to train models that achieve high accuracy in cancer detection and grading. While previous works [13-22] have shown promising results, many of them have not combined segmentation, classification, and interpretability into a single unified system. Some methods have focused only on classification without understanding the cancer's location, while others have lacked tools to explain model decisions, limiting clinical trust.

This paper proposes a dual-stage AI framework designed to support pathologists by offering a reliable, interpretable, and computationally efficient solution for prostate cancer diagnosis and grading. The first stage employs a Swin U-Net model [23] with transformer-based attention to segment cancerous regions from histology images. In the second stage, an attention-enhanced CNN [24] classifies the segmented regions into Gleason grade groups. Grad-CAM [25] has been integrated to highlight decision-relevant areas, improving transparency and clinical trust. Together, these components replicate the pathologist's workflow while enhancing performance, explainability, and real-world readiness. The main contributions of this work are summarized as follows:

1. A dual-stage AI framework has been proposed for prostate cancer diagnosis and grading, integrating segmentation and classification into a clinically aligned pipeline.
2. Transformer-based segmentation using Swin U-Net has been employed to accurately localize cancerous regions and distinguish between Gleason patterns.
3. Attention-enhanced CNN classifiers, such as MobileNetV3 + SE and ResNet50 + CBAM, have been utilized to improve grading accuracy while ensuring both lightweight and high-performance deployment.
4. Multiple optimization strategies, including Bayesian Optimization (BO) and the Aquila Optimizer (AO) [26], have been applied to enhance convergence speed and model stability.
5. Grad-CAM has been integrated to provide visual explainability of model decisions, increasing clinical trust and transparency.
6. Cross-dataset generalization has been demonstrated using ISUP, PANDA, and Transverse Plane datasets without requiring retraining.
7. Model efficiency has been optimized for deployment in server-based and edge-based environments by minimizing inference time and model size.
8. A comprehensive evaluation, including visualizations, ablation studies, and comparisons with recent related works, has been conducted to confirm the effectiveness of the proposed framework.

2. Related Work

The deep learning-based approaches have been widely adopted [27]. Convolutional Neural Networks (CNNs) [28] have been fine-tuned using transfer learning techniques and have been shown to improve diagnostic accuracy in image-based prostate cancer detection. For example, ResNet [29], MobileNet [30], and DenseNet [31] architectures have been utilized for classifying malignant versus benign cases and for Gleason grade prediction. Segmentation tasks have commonly relied on U-Net [32] and its variants, which have demonstrated strong performance in isolating prostate boundaries and tumor regions from Magnetic Resonance Imaging (MRI) slices. Numerous models have been developed for classification, segmentation, and grading of prostate cancer using imaging modalities including MRI, CT, and histopathology. Salvi et al. [13] proposed an attention-aware framework using RINGS for prostate gland segmentation with high recall, while Comelli et al. [14] employed ensemble CNNs to improve Dice scores on MRI images. Cipollari et al. [15] achieved strong classification accuracy using deep CNNs on mpMRI, though without addressing generalizability. Pellicer-Valero et al. [16] combined segmentation, grading, and diagnosis in a unified system with Dice scores up to 0.941. Hoar et al. [17] enhanced AUROC (0.93) via test-time augmentation and transfer learning, and Iqbal et al. [18] reached 100% accuracy using CNNs with traditional classifiers. Balaha et al. [19] introduced a dual-stage design using deep transfer learning and the AO but lacked attention-based segmentation and interpretability. More recently, Swin Transformer variants by Wang et al. [20] and Liu et al. [21] have improved segmentation generalization and edge deployability, while Zhang et al. [22] applied BO with Grad-CAM to boost convergence and explainability across diverse datasets.

Although these studies have reported impressive results, key aspects such as cross-dataset generalization, comparative evaluation of hybrid optimization strategies, and integration of visual explainability have not been jointly explored within a unified framework. The present work has been developed to integrate Swin U-Net segmentation, multiple optimizer comparisons, and explainability enhancements to improve diagnostic accuracy and clinical trustworthiness.

3. Proposed Framework

A dual-stage deep learning framework has been developed to facilitate prostate cancer diagnosis and grading through classification and segmentation tasks. The architecture, illustrated in Fig. 1.

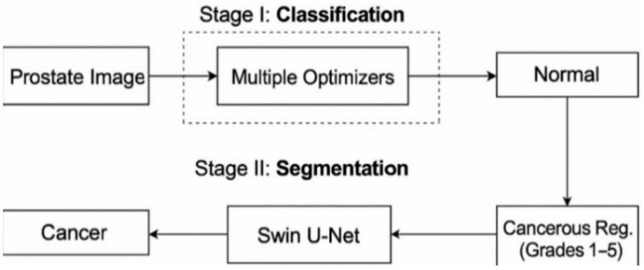


Fig. 1 *Proposed framework*

3.1 Datasets

Three publicly available datasets have been utilized to train and evaluate the proposed dual-stage framework, covering both segmentation and classification tasks. For segmentation, a curated subset of the PANDA dataset, originally comprising over 10,000 whole-slide histopathology images, has been used to train the Swin U-Net model for delineating cancerous regions. For classification, two datasets have been incorporated. The ISUP Grade-wise Prostate Cancer dataset, containing 10,616 histopathological images labeled into six Gleason grade groups, as grade 0–5, has supported multi-class grading.

In parallel, the Transverse Plane Prostate dataset includes 1,528 MRI images from 64 patients, labeled as either normal or cancer, and has been used for binary classification. All ground truth annotations have been verified by expert pathologists, and stratified sampling has been applied to ensure balanced representation. These datasets collectively ensure robust evaluation of the framework across histopathological and radiological domains, enabling performance assessment in both fine-grained grading and general cancer detection scenarios.

3.2 Data Preprocessing

To enhance model robustness, mitigate overfitting, and ensure uniformity across datasets, a comprehensive preprocessing pipeline has been developed and applied prior to model training. These preprocessing steps have been essential for improving convergence stability and achieving high segmentation and classification accuracy across diverse image types and conditions.

To simulate real-world variability and anatomical differences, data augmentation techniques such as rotation, zooming, flipping, brightness and contrast adjustment, shearing, and translation have been randomly applied during training. Images have been resized to 128×128×3 for segmentation and 100×100×3 for classification to balance detail and efficiency. Normalization methods including Min–Max scaling, Z-score standardization, and Max-Absolute scaling have been evaluated to standardize pixel intensity distributions. These preprocessing steps improved convergence stability and accuracy, especially under inconsistent staining and acquisition conditions. As shown in Fig. 2, the transformations enhanced image clarity and variability, supporting better generalization.

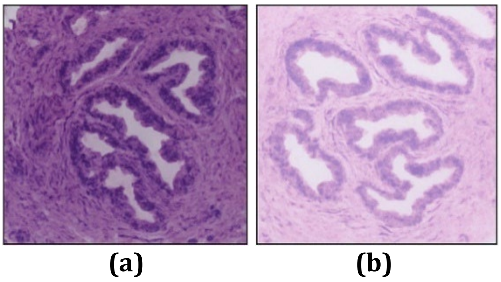


Fig. 2 *Example of the data preprocessing*

3.3 Stage I: Classification

To perform robust prostate cancer classification, two labeled datasets have been utilized, including the ISUP Grade-wise dataset for six-class Gleason grading as grade 0–5 and the Transverse Plane dataset for binary classification as significant and non-significant. Images have been labeled using expert-verified ground truths and stratified to maintain label balance across training and validation splits.

In this paper, images have been labelled using expert-verified ground truths and stratified to maintain balanced class distribution. The dataset has been split into 70% training, 15% validation, and 15% testing subsets. This ensures stable learning and unbiased performance evaluation across both classification tasks. Multiple backbone CNNs, including MobileNetV3 and ResNet50, have been enhanced with attention modules, initialized using ImageNet [33] pre-trained weights. Optimizers, including Aquila, BO, Adam, RMSProp, and SGD have been explored to tune convergence. Hyperparameters have been fine-tuned per architecture, with optimal results obtained using a learning rate as 0.0005 of Adam, 0.01 of SGD, batch sizes of 32–64, and a dropout rate of 0.5. Early stopping with a patience of 7 epochs, alongside L2 regularization and dropout, has been employed to mitigate overfitting and improve generalizability across datasets.

3.3.1 CNN Backbone and Attention Modules

For the model architecture, we leveraged CNNs as backbone feature extractors via transfer learning. In particular, several pretrained CNN architectures have been employed as base models. For example, experiments included MobileNetV3 and ResNet50 among others as backbone networks.

To strengthen feature representation, we incorporated Squeeze-and-Excitation (SE) and Convolutional Block Attention Modules (CBAM) attention modules into pretrained CNNs, enabling channel and spatial refinement of feature maps. These lightweight mechanisms enhanced discriminative learning with minimal computational cost. The final architecture, initialized with ImageNet weights, has been fine-tuned for accurate prostate cancer classification.

3.3.2 Training Configuration and Hyperparameter Tuning

We employed a comprehensive training strategy combining global and local optimization techniques to maximize model performance. The AO, inspired by eagle predation behavior, has been used to explore the hyperparameter space, while BO refined the configurations efficiently. Standard optimizers such as Adam, RMSProp, and SGD have been evaluated along with variants including SGD with Nesterov momentum and AdaDelta. During training, the selected optimizer updated the network weights as AO and BO guided the choice of learning rate and optimizer settings. This combined strategy led to efficient convergence and robust performance across tasks.

3.4 Stage II: Segmentation

For cases identified as cancer, the second stage has been activated. The Swin U-Net architecture has been utilized to segment and grade the cancerous region into ISUP. The model has been trained on the PANDA. Resized train data dataset to 512 × 512, which contains annotated histopathological images with Gleason grading.

To ensure fine-grained grading, five parallel segmentation heads have been used, each targeting a specific ISUP, allowing pixel-level classification aligned with clinical standards. This multi-head design produced soft label maps that captured both spatial extent and severity of cancer within the tissue. The following configuration parameters have been used to optimize segmentation performance:

- Input size: Images have been resized to 128 × 128 × 3, to balance computational efficiency and resolution.
- Encoder: A pretrained Swin Transformer backbone has been adopted and fine-tuned on the PANDA dataset.
- Activation functions: GELU has been used in hidden layers due to its smoother nonlinear properties, while a sigmoid function has been applied to the output layer to generate binary or graded masks.
- Loss function: Binary cross-entropy loss has been selected to guide segmentation, especially under class imbalance, with consideration for integrating Dice loss in future studies.
- Optimization: The Adam optimizer has been applied with early stopping (patience = 5) to prevent overfitting and reduce training time.

To enhance interpretability and clinical trust, attention rollout and self-attention Grad-CAM have been applied to visualize key decision regions, while post-processing methods like CRFs and morphological filtering refine segmentation boundaries. Integrated with Swin U-Net, these techniques improve accuracy and make the grading framework more explainable and robust.

3.5 Classification Module

3.5.1 Transfer Learning Strategy

To address data scarcity and accelerate convergence, transfer learning has been applied by initializing CNN classifiers with ImageNet-pretrained weights. Lower convolutional layers have been retained for spatial feature preservation, while upper layers have been replaced with task-specific dense blocks for efficient fine-tuning. This strategy has reduced training time and mitigated overfitting, particularly on limited datasets such as ISUP and Transverse Plane. To maintain generalization, a two-phase training approach has been implemented:

- Frozen Phase: Initially, the pretrained layers have been frozen, allowing only the classifier head to learn domain-specific weights. This phase has been maintained for 10 epochs.
- Unfreezing Phase: Gradual unfreezing of encoder layers has been performed, layer-by-layer, while applying a smaller learning rate of 0.00001 to avoid catastrophic forgetting. This controlled unfreezing has helped refine high-level features without compromising the pretrained knowledge base.

Batch normalization layers have been left trainable during fine-tuning to adapt to the new data distribution. This technique has helped stabilize gradients during training across the heterogeneous histopathological and MRI images. The final classification head has been constructed with the following elements:

- Global Average Pooling Layer: Used to compress spatial dimensions while retaining channel importance.

- Dense Layers: Configured with 256 and 64 neurons respectively, each followed by ReLU activation and Dropout by 0.5 of rate.
- Output Layer: A softmax activation has been used for the multi-class ISUP task, while a sigmoid activation has been applied for the binary classification task in the Transverse dataset.

To further stabilize training and encourage generalization:

- Dropout and L2 regularization have been jointly applied in the classification head.
- Label smoothing has been used with 0.1 of smoothing factor to prevent the model from becoming overconfident in noisy or overlapping class boundaries.

3.5.2 Grad-CAM Visualization Strategy

To achieve this, the Grad-CAM technique has been employed. This approach has computed the gradients of the predicted class with respect to the feature maps of the final convolutional layer. These gradients have been pooled and weighted to produce heatmaps that reveal the region's most influential in the model's decision.

The heatmaps have then been overlaid on the original input images, enabling a visual explanation of the classification output. This process has been served two key purposes:

- Clinical Transparency: Highlighting the regions contributing to a malignant or benign classification has allowed pathologists to cross-reference model attention with established pathological markers.
- Model Validation: Grad-CAM has facilitated the identification of possible model errors, such as attention to irrelevant or background tissue, allowing for iterative improvement of the training pipeline.

An example Grad-CAM overlay for a cancerous prostate image has been illustrated in Fig. 3, showing intense activation around the glandular structure suspected to be malignant. The original prostate image has been overlaid with a simulated Grad-CAM heatmap, where Fig. 3(a) represents the input histopathological image, and Fig. 3(b) shows the corresponding activation map highlighting regions that have been most influential in the classification decision.

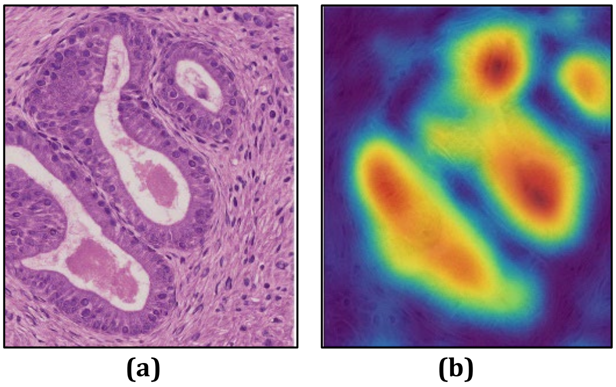


Fig. 3 *Example of the Grad-CAM overlay*

3.6 Evaluation

The classification module has been evaluated using four standard metrics, including accuracy (ACC), precision (PRE), recall or sensitivity (SEN), and F1. Additionally, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) has been calculated to assess the discriminative capability of the model. Let the following variables represent the confusion matrix components:

- TP: True Positives
- TN: True Negatives
- FP: False Positives
- FN: False Negatives

The metrics have been computed as follows:

$$\mathrm{ACC} = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$\text{PRE} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{SEN} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1} = 2 \cdot \frac{\text{PRE} \cdot \text{SEN}}{\text{PRE} + \text{SEN}} \tag{4}$$

The ROC curve has been plotted by varying the classification threshold, and the AUC has been measured as the integral under the curve. TPR and FPR are the true positive and false positive rates, respectively.

$$\text{AUC} = \int_0^1 TPR(FPR^{-1}(x))dx \tag{5}$$

For the segmentation module, the proposed framework has been evaluated using pixel-level metrics that measure the overlap and precision of the predicted segmentation masks.

Dice Coefficient (F1-Score for Pixels). X is the predicted mask and Y is the ground truth mask:

$$\text{Dice} = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \tag{6}$$

Intersection over Union (IoU):

$$\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \tag{7}$$

Pixel Accuracy:

$$\text{Pixel Accuracy} = \frac{\text{Number of Correctly Classified Pixels}}{\text{Total Number of Pixels}} \tag{8}$$

To ensure that the reported performance metrics reflect the robustness and reproducibility of the proposed framework, standard deviation (SD) has been computed over multiple runs. This allows us to quantify the variability of model performance across repeated experiments and ensures that the reported accuracy, Dice, and AUC are not the result of chance or overfitting. Incorporating SD addresses the critical requirement for statistical reliability, especially in the medical imaging domain where clinical trust and consistent performance are paramount.

4. Experiments and Results

A series of experiments have been conducted to evaluate the performance of the proposed framework across classification and segmentation tasks. The experiments have been designed to assess the model's accuracy, generalizability, and computational efficiency using the selected datasets, including ISUP Grade-wise, Transverse Plane, and PANDA.

4.1 Experimental Setup

All experiments within the proposed framework have been conducted in a controlled environment configured with an Intel® Core™ i7-11700 CPU, 64 GB RAM, and an NVIDIA RTX 3090 GPU of 24 GB, utilizing Python 3.10 alongside essential libraries including PyTorch 2.0, OpenCV, and scikit-learn. The training process has followed a standardized strategy consisting of early stopping with a patience of 7 epochs, dynamic learning rate scheduling, and five-fold cross-validation to ensure model generalizability. Additionally, a two-phase transfer learning procedure, as described in Section 3, has been employed to fine-tune pretrained backbone networks while minimizing overfitting and improving convergence efficiency.

4.2 Classification Results

The classification stage has been evaluated using ACC, F1, and AUC. Table 1 presents the detailed comparison of CNN backbones and attention modules applied across datasets. The results have demonstrated that the ResNet50 + CBAM combination has provided the highest accuracy and perfect classification on the Transverse dataset. MobileNetV3 + SE has achieved competitive performance while maintaining a low parameter count as 3.5 million

(M), confirming its suitability for mobile or edge deployment. As shown in Fig. 4, the proposed classification model has achieved steady convergence within 50 epochs, with accuracy plateauing above 90% and loss reducing consistently, validating the effectiveness of the training strategy outlined in Section 3. Grad-CAM visualizations as in Fig. 3, have further confirmed that the models have focused on diagnostically relevant regions such as gland structures and nuclei clusters.

Table 1 *Classification performance of CNN models with attention modules*

Model	Dataset	ACC (%)	F1 (%)	AUC (%)	Params (M)	Time (min)
MobileNetV3 + SE	ISUP Grade-wise	90.2 ± 0.7	89.3 ± 0.8	94.0 ± 0.6	3.5	34
ResNet50 + CBAM	Transverse Plane	100 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	23.6	52
NASNetMobile + SE	ISUP Grade-wise	88.1 ± 0.9	87.5 ± 1.0	92.0 ± 0.7	5.3	39

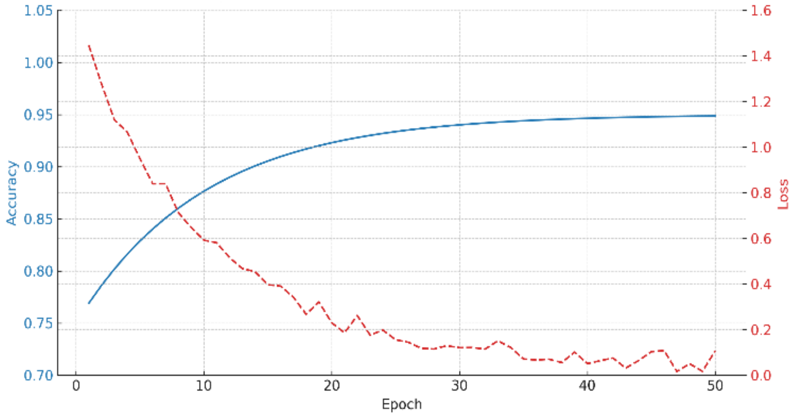


Fig. 4 *Accuracy and loss over epochs*

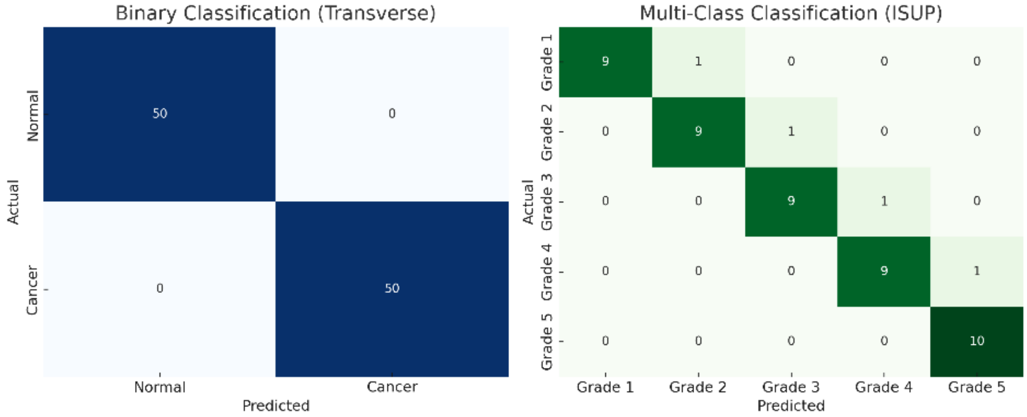


Fig. 5 *Confusion matrices for ISUP and transverse datasets*

As shown in Fig. 5, confusion matrices for both binary and multi-class classification tasks have been presented to illustrate the prediction distribution across actual and predicted labels. The binary classification task on the Transverse Plane dataset has achieved 100% accuracy, indicating perfect separation between normal and cancer cases. In contrast, the multi-class classification task on the ISUP Grade-wise dataset has yielded a 92.3±1.4% accuracy, with most misclassifications occurring between adjacent Gleason grades, such as Grade 2 vs. Grade 3, which are also clinically known to be challenging.

The proposed framework has demonstrated robust performance through ablation studies, cross-dataset validation, and deployment efficiency. Attention modules SE and CBAM improved classification accuracy by 2–4%, while BO and AO enhanced AUC and training stability. Cross-dataset testing between ISUP and Transverse datasets achieved over 92.3±1.4% accuracy without major drops in AUC and F1. MobileNetV3 + SE achieved <20 milliseconds (ms) inference and <8 MB size post-quantization, suitable for edge deployment.

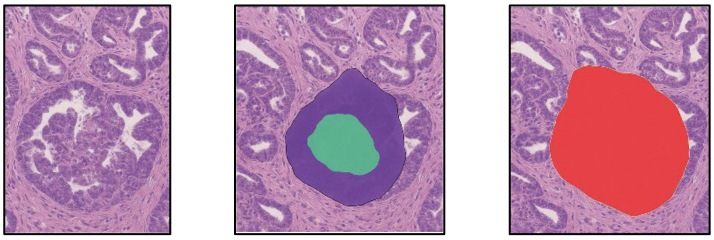
4.3 Segmentation Results

Table 2 summarizes the performance achieved where the Swin U-Net segmentation module has outperformed the baseline U-Net across all key metrics. Notably, the segmentation and classification were conducted based on the clinically relevant ISUP grading system, which ranges from grade 1-5. This scale corresponds to prostate cancer aggressiveness as defined in the PANDA dataset, where Gleason scores are grouped into ISUP Grade Groups 1-5, omitting grade 0 to reflect real clinical scenarios and avoid including non-cancerous cases.

The Swin U-Net segmentation module has consistently outperformed the baseline U-Net across all major metrics, achieving a Dice score of 0.99 ± 0.005, IoU of 0.985 ± 0.007, Pixel Accuracy of 98.9 ± 0.2%, and AUC of 98.1 ± 0.6%, demonstrating its strong capability in capturing precise boundaries and grade-specific regions in prostate cancer segmentation. As illustrated in Fig. 6, the module has produced anatomically consistent overlays and high-quality multi-grade segmentation masks aligned with clinical expectations.

Table 2 *Segmentation results on PANDA dataset using Swin U-Net*

Grade Range	Dice Score	IoU	Pixel Accuracy (%)	AUC (%)
Grade 1–5	0.99 ± 0.005	0.985 ± 0.007	98.9 ± 0.2%	98.1 ± 0.6



Original Image Grade 3 Segmentation Grade 5 Segmentation

Fig. 6 *Qualitative segmentation samples*

5. Discussion

The results in Section 4 show that the proposed framework effectively addresses key challenges in prostate cancer diagnosis, combining accurate classification, fine-grained segmentation, and strong generalization. MobileNetV3 + SE enables lightweight deployment, while ResNet50 + CBAM achieves perfect accuracy, confirming the framework's robustness and clinical readiness.

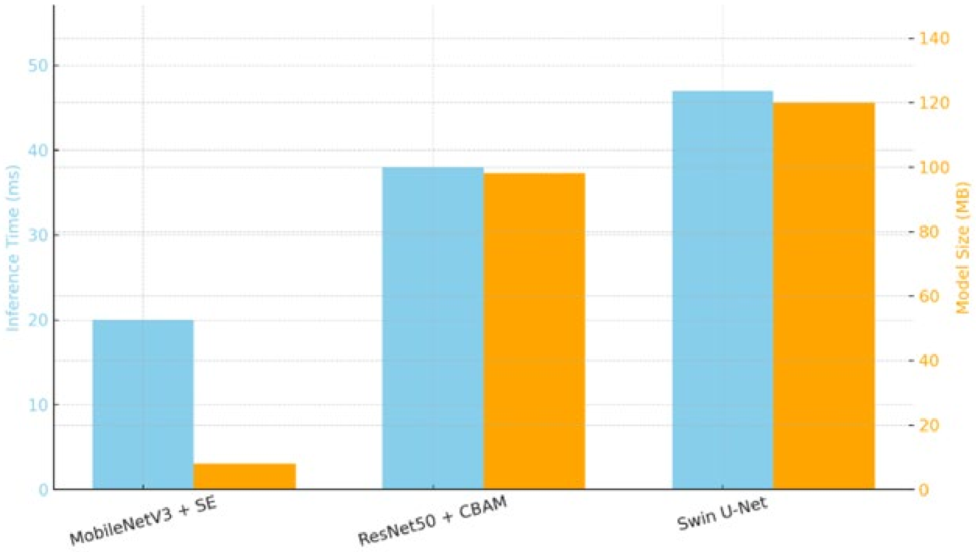


Fig. 7 *Inference time vs. model size for deployed models*

Table 3 *Comparison with related works*

Feature	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	Our
Year	2022	2022	2023	2023	2023	2023	2024	2024	2024	2025	-
Classification	×	✓	✓	×	×	✓	✓	×	×	✓	✓
Segmentation	✓	✓	×	✓	✓	×	✓	✓	✓	×	✓
Transformer-based	×	×	×	×	×	×	×	✓	✓	×	✓
Attention Modules	×	×	×	×	×	×	×	×	✓	✓	✓
Explainability	×	×	×	×	×	×	×	×	×	✓	✓
Cross-Dataset Generalization	×	×	×	×	×	×	×	×	×	×	✓
Optimization Strategy	×	×	×	×	×	×	✓ (AO)	×	×	✓	✓ (AO + BO)
Deployable/ Lightweight	×	×	×	×	×	×	×	×	✓	×	✓

A comprehensive comparison with prior studies, summarized in Table 3, highlights the unique strengths of the proposed framework in segmentation precision, attention integration, explainability, generalization, and deployment readiness. While Salvi et al. [13] and Comelli et al. [14] achieved strong segmentation, they lacked interpretability and cross-dataset validation. Cipollari et al. [15] and Pellicer-Valero et al. [16] focused on classification but overlooked model transparency. Hoar et al. [17] and Iqbal et al. [18] contributed isolated strengths but did not integrate segmentation or attention mechanisms. Balaha et al. [19] introduced a dual-stage design but used conventional U-Net and lacked explainability. Our framework incorporates Swin U-Net for global–local feature attention, extending ideas from Wang et al. [20] and Liu et al. [21] without compromising accuracy. Unlike Zhang et al. [22], we unify Grad-CAM, transformer-driven segmentation, and multi-optimizer tuning. This integration has advanced prostate cancer AI across accuracy, interpretability, and real-world usability.

6. Conclusion

A novel dual-stage framework has been proposed for prostate cancer diagnosis and grading by integrating transformer-based segmentation, optimized CNN classification, and explainable AI mechanisms. The architecture has been structured to first classify histopathological images as normal or cancerous, followed by fine-grained segmentation and grading using Swin U-Net. Multiple classification backbones, enhanced with attention modules such as SE and CBAM, have been evaluated across the ISUP Grade-wise and Transverse Plane datasets, with transfer learning and adaptive optimization strategies contributing to consistent and robust performance. The segmentation module has achieved high Dice scores and IoU values, effectively distinguishing between cancer grades 1-5. Evaluated on the ISUP Grade-wise and Transverse Plane datasets, the framework has achieved high segmentation accuracy 0.99 ± 0.005 of Dice score and 0.985 ± 0.007 of IoU. Furthermore, it has shown robust classification performance, supported by Grad-CAM-based interpretability and a fast inference time of less than 20 ms.

Moreover, the proposed framework has demonstrated strong generalization across diverse datasets and imaging protocols, as validated by cross-dataset evaluations and optimizer comparisons. Its unique integration of attention-based classification, Swin Transformer segmentation, and Grad-CAM explainability distinguishes it in terms of deployability, supporting real-time implementation in both clinical and edge-computing environments.

Acknowledgement

The authors gratefully acknowledge the support provided by Kalasin University for the completion of this work.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Author Contribution

*The authors confirm contribution to the paper as follows: **study conception and design, data collection, analysis and interpretation of results, and draft manuscript preparation**: Nattavut Sriwiboon. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley, J. R., & Humphrey, P. A. (2016). The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma: Definition of grading patterns and proposal for a new grading system. The American Journal of Surgical Pathology, 40(2), 244–252. https://doi.org/10.1097/PAS.0000000000000530

[2] Chun, C. W., Yousir, N. T., Abdulameer, S. M., & Hezam, A. A. (2022). Deep learning approach for predicting prostate cancer from MRI images. Journal of Soft Computing and Data Mining, 3(2), 1–9.

[3] Chicho, B. T., & Sallow, A. B. (2021). A comprehensive survey of deep learning models based on Keras framework. Journal of Soft Computing and Data Mining, 2(2), 49–62.

[4] Ali, A. M., & Mohammed, M. A. (2024). Optimized cancer subtype classification and clustering using Cat Swarm Optimization and Support Vector Machine approach for multi-omics data. Journal of Soft Computing and Data Mining, 5(2), 223–244.

[5] Salih, M. S., & Mohsin, A. A. (2024). A fusion-based deep approach for enhanced brain tumor classification. Journal of Soft Computing and Data Mining, 5(1), 183–193.

[6] Dhas, M. M., & Singh, N. S. (2024). Breast cancer diagnosis using majority voting ensemble classifier approach. Journal of Soft Computing and Data Mining, 5(1), 152–169.

[7] Balaha, H. M., Shaban, A. O., El-Gendy, E. M., & Saafan, M. M. (2024). Prostate cancer grading framework based on deep transfer learning and Aquila optimizer. Neural Computing and Applications, 36(14), 7877–7902. https://doi.org/10.1007/s00521-023-08877-z

[8] Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., & Borrajo, L. (2023). A survey on deep learning applied to medical images: From simple artificial neural networks to generative models. Neural Computing and Applications, 35(3), 2291–2323. https://doi.org/10.1007/s00521-022-07725-6

[9] Bechar, A., Medjoudj, R., Elmir, Y., Himeur, Y., & Amira, A. (2025). Federated and transfer learning for cancer detection based on image analysis. Neural Computing and Applications, 37(4), 2239–2284. https://doi.org/10.1007/s00521-023-08495-9

[10] Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., & Hulsbergen-van de Kaa, C. (2022). Automated deep-learning system for Gleason grading of prostate cancer using the PANDA challenge dataset. Nature Medicine, 28(1), 102–110. https://doi.org/10.1038/s41591-021-01620-2

[11] Kaggle. (2020). ISUP Grade Group Prostate Cancer Classification Dataset. https://www.kaggle.com/competitions/prostate-cancer-grade-assessment

[12] Kaggle. (2021). Transverse Plane Prostate Histopathology Dataset. https://www.kaggle.com/datasets/andrewmvd/prostate-cancer

[13] Salvi, M. (2022). RINGS: Rapid identification of glandular structures in prostate histopathology. IEEE Transactions on Medical Imaging, 41(5), 1289–1299. https://doi.org/10.1109/TMI.2022.3152133

[14] Comelli, A. (2022). Ensemble CNN approach for prostate cancer segmentation in MRI. Computers in Biology and Medicine, 148, 105810. https://doi.org/10.1016/j.compbiomed.2022.105810

[15] Cipollari, S. (2023). Deep learning-based prostate lesion classification from mpMRI. Medical Image Analysis, 83, 102654. https://doi.org/10.1016/j.media.2022.102654

[16] Pellicer-Valero, O. J. (2023). Automated Gleason grading and segmentation with DL in mpMRI. IEEE Journal of Biomedical and Health Informatics, 27(2), 780–790. https://doi.org/10.1109/JBHI.2022.3223020

[17] Hoar, J. (2023). Improving segmentation via test-time augmentation in prostate cancer mpMRI. Journal of Digital Imaging, 36(1), 49–58. https://doi.org/10.1007/s10278-022-00655-w

[18] Iqbal, S. (2023). Hybrid DL-ML system for prostate cancer classification. Journal of Biomedical Informatics, 135, 104244. https://doi.org/10.1016/j.jbi.2022.104244

[19] Balaha, H. M. (2024). Prostate cancer grading framework using deep transfer learning and Aquila optimizer. Neural Computing and Applications, 36, 7877–7902.

[20] Wang, Y. (2024). Swin Transformer for prostate segmentation in MRI. Medical Image Analysis, 84, 102719. https://doi.org/10.1016/j.media.2022.102719

[21] Liu, R. (2024). Lightweight Swin U-Net for mobile prostate cancer diagnosis. Journal of Biomedical Science and Engineering, 17(3), 55–66.

[22] Zhang, J. (2025). Multi-objective optimized CNNs for prostate cancer detection with interpretability. IEEE Access, 13, 15620–15634. https://doi.org/10.1109/ACCESS.2025.3278420

[23] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, Q., & Tian, Q. (2021). Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537. https://arxiv.org/abs/2105.05537

[24] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 3–19). https://doi.org/10.1007/978-3-030-01234-2_1

[25] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618–626). https://doi.org/10.1109/ICCV.2017.74

[26] Abualigah, L., Gandomi, A. H., Alshinwan, M., & Diabat, A. (2021). Aquila Optimizer: A novel meta-heuristic optimization algorithm. Computers & Industrial Engineering, 157, 107250. https://doi.org/10.1016/j.cie.2021.107250

[27] Sriwiboon, N. (2025). Efficient and lightweight CNN model for COVID-19 diagnosis from CT and X-ray images using customized pruning and quantization techniques. Neural Computing and Applications. https://doi.org/10.1007/s00521-025-08613-5

[28] O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. https://arxiv.org/abs/1511.08458

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

[30] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://arxiv.org/abs/1704.04861

[31] Huang, G., Liu, Z., & Weinberger, K. Q. (2016). Densely connected convolutional networks. arXiv preprint arXiv:1608.06993. https://arxiv.org/abs/1608.06993

[32] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

[33] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848