

Multi-Scale Attention-Enhanced EfficientNet for Automated Gleason Grading of Prostate Cancer

Olushola Olawuyi

University of KwaZulu-Natal

Serestina Viriri

viriris@ukzn.ac.za

University of KwaZulu-Natal

Samson Akinpelu

University of KwaZulu-Natal

Research Article

Keywords: Prostate Cancer, Gleason Grading, EfficientNetB7, Multi-scale Convolution, Attention Mechanisms

Posted Date: September 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10636850/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multi-Scale Attention-Enhanced EfficientNet for Automated Gleason Grading of Prostate Cancer

Olushola Olawuyi¹, Serestina Viriri^{1*}, Samson Akinpelu¹

¹Discipline of Computer Science, School of Agriculture and Science,
University of KwaZulu-Natal, Durban, 4000, South Africa.

*Corresponding author(s). E-mail(s): viriris@ukzn.ac.za;
Contributing authors: 221119659@stu.ukzn.ac.za; akinpelus@ukzn.ac.za;

Abstract

Prostate cancer is one of the most common cancers in men and a major cause of death worldwide. The Gleason grading system is the standard method used by pathologists to assess prostate cancer severity and guide treatment. However, manual grading is often inconsistent because it depends on individual interpretation. Recent progress in deep learning has improved the automation of pathology image analysis, while many models still struggle to extract useful features and generalise well across datasets. In this study, we present a Multi-Scale Efficient Attention Convolutional Network (MSEACN) for automatic Gleason grading of prostate tissue. The model combines multi-scale convolutional layers with dual attention modules to focus on key image regions and features. It uses EfficientNetB7 as the core network to achieve a balance between accuracy and computational cost. The model was fine-tuned on prostate images from the PANDA dataset after several preprocessing steps, including contrast enhancement (CLAHE), stain normalisation, and circular cropping. In addition, the MSEACN system achieved high accuracy of 97.69%, very high specificity of 99.50%, high precision of 98.16%, high recall of 97.67%, and an excellent F1-score of 97.90%. In terms of Cohen's Kappa, it scored a value of 0.98, while the QWK value was 0.94. However, the results obtained from this study show that MSEACN outperforms other deep learning methods.

Keywords: Prostate Cancer, Gleason Grading, EfficientNetB7, Multi-scale Convolution, Attention Mechanisms

1 Introduction

Prostate cancer (PCa) is among the most frequently diagnosed malignancies in men and a major cause of cancer mortality globally. In 2020, there were over 1.4 million new cases and more than 375,000 deaths worldwide, with projections suggesting these numbers could rise to approximately 2.4 million cases and 740,000 deaths by 2040 [1]. Clinical outcomes and treatment decisions surgery, radiation, or active surveillance, depend critically on accurate histopathological grading via the Gleason score and its derived Grade Groups [2]. However, Gleason grading suffers from substantial inter-observer variability, with agreement rates among pathologists often below 70% [3].

Recent advances in artificial intelligence (AI), particularly deep learning (DL), offer promise in reducing grading variability and improving diagnostic and prognostic accuracy. Models trained on both imaging (e.g. MRI) and histopathology have achieved AUCs above 0.85 in multicenter validation settings [4] [5]. DL-derived Grade Groups have shown comparable prognostic value to those assigned by expert pathologists in radical prostatectomy cohorts [6], and innovations in volumetric tissue representation using multiple instance learning (MIL) have further improved performance over 2D baselines by better capturing three-dimensional tissue heterogeneity [7].

Several recent studies further illustrate the evolving landscape of AI in prostate pathology. A federated learning model with attention-consistency (FACL) applied to nearly 20,000 whole-slide images from multiple centres achieved a Gleason grading Kappa of ≈ 0.85 , demonstrating strong generalisation while preserving privacy in data sharing [8]. Another study, compared public and commercial AI algorithms on real-world whole-slide biopsy images: the algorithms achieved high agreement with pathologists, with weighted quadratic Kappas ranging from ~ 0.62 to ~ 0.90 [9].

Also, current progress in artificial intelligence algorithms based on digital pathology technology for prostate cancer detection and Gleason grading has made strides toward overcoming such obstacles [43].

In this research [36], the groundbreaking nature of artificial intelligence was demonstrated to revolutionize the diagnostic process for prostate cancer; the results suggest that artificial intelligence-assisted whole slide imaging provides improved accuracy and standardization when detecting and grading cancers than traditional pathology techniques. These examples show AI is increasingly robust, clinically relevant, and evaluated under realistic conditions.

According to Wang et al. [45], federated and privacy-preserving AI systems have been proposed to provide a promising approach for building largescale prostate cancer diagnostic models with data confidentiality

Despite progress, important challenges remain: most models are trained on skewed grade distributions, limiting performance on underrepresented Gleason patterns; interpretability is often secondary to predictive accuracy; and few studies provide detailed ablation of architectural choices or integrate multi-scale histopathological features with attention mechanisms. There is also variability in external generalisability, especially across institutions with different slide preparation, staining, imaging equipment, and patient demographics. To address these gaps, this study proposes a robust deep learning framework for prostate cancer grading using the PANDA dataset, which

includes over 41,000 biopsy whole-slide images reflecting real-world clinical variability. Our method incorporates contrast-limited adaptive histogram equalisation (CLAHE) for image enhancement, multi-scale feature extraction to better capture tissue heterogeneity, and attention modules to emphasise diagnostically relevant regions. We aim to advance consistency, interpretability, and clinical utility in automated Gleason grading.

The remainder of this paper is structured as follows: Section 2 presents the related works and contributions of the study. Section 3 describes the materials and methods, including model development, dataset preparation, preprocessing, augmentation, and training procedures. Section 4 discusses the experimental results, comparative evaluations, and ablation studies. Section 5 outlines the study’s limitations and justifies the contributions, while Section 6 concludes with key findings, implications, and directions for future work.

2 Literature Review and Related Works

Prostate cancer (PCa) diagnosis and grading remain highly challenging due to the heterogeneity of histopathological patterns, inter-observer variability in Gleason scoring, and the complexity of whole-slide images (WSIs). The advancement of deep learning has significantly transformed computational pathology, enabling automated Gleason grading systems that achieve pathologist-level performance. Recent works directly related to EfficientNet, attention mechanisms such as the Convolutional Block Attention Module (CBAM), and large-scale datasets like PANDA collectively shape the foundation of this study.

Early studies employed convolutional neural networks (CNNs) such as GoogleNet, VGG, and ResNet to automate Gleason grading, but these models struggled with scalability, generalisation, and interpretability [10]. Later approaches, including MobileNet and DenseNet variants, improved efficiency for histopathology applications but often required extensive pretraining and transfer learning to achieve high accuracy [11]. A pivotal contribution came from [3], who demonstrated that deep learning models could match expert pathologists in Gleason grading, thereby validating the feasibility of AI in clinical practice. More recently, Bulten et al., [12] organised the PANDA challenge, which provided over 15,000 annotated WSIs, establishing a benchmark for AI-based prostate cancer grading. The availability of this dataset has since enabled robust model training and cross-institutional evaluation, reducing overfitting and improving generalizability.

Zhang et al. [37] reviewed recent developments in AI-assisted prostate cancer diagnosis and reported substantial progress in deep learning-based MRI analysis, computer-aided pathology, automated Gleason scoring, and multimodal clinical decision-support systems. Paik et al. [38] in recent developments in the use of artificial intelligence for digital pathology in urologic tumors indicate that studies on prostate cancer have shown reliable performance in diagnosis and grading using whole-slide imaging. This has reinforced the clinical value of AI-enabled pathology workflows.

Kabir et al. [39] proposed a novel approach to automated prostate cancer grading using deep learning, which showed excellent classification capability in various Gleason classes and proved the success of deep learning in minimizing inter-grader variation.

Tan et al. [13] introduced a compound scaling strategy that jointly optimises network depth, width, and input resolution for EfficientNet. This innovation allowed EfficientNet to achieve state-of-the-art accuracy with significantly fewer parameters compared to traditional CNNs. In prostate cancer histopathology, EfficientNet has been widely adopted due to its computational efficiency and superior performance. For example, [14] leveraged an EfficientNet backbone for grading prostate cancer WSIs, reporting strong results in the PANDA challenge with high quadratic weighted kappa scores. Similarly, [15] employed EfficientNet with transfer learning and optimisation strategies, achieving over 88% accuracy in ISUP grading. EfficientNet’s strong feature extraction capability makes it particularly suited for WSI classification, where gigapixel-scale data requires models that balance accuracy and efficiency. In this study, EfficientNetB7 was chosen as the backbone due to its ability to capture high-resolution histopathological features while maintaining computational feasibility.

Zaheer et al. [40] proposed an attention-enhanced hybrid U-Net architecture for prostate cancer grading and explainability, demonstrating that attention mechanisms improve feature localisation and classification performance in histopathological image analysis.

Linkon et al. [16] proposed a deep learning method to detect prostate cancer using histopathological images from different datasets which are based on Gleason’s score system. The two techniques used in the study were convolutional neural network and post-processing techniques which help to improve model stability. The study achieved an F1-score of 0.8460 and a precision of 0.8921.

Shrestha et al. [17] proposed a deep learning (DL) model and batch normalization technique for the prostate cancer image segmentation process. The dataset used consists of 230 MRI images. Their study aimed to determine the most efficient adaptive feature extraction method that is able to achieve accurate segmentation of prostate cancer images by extraction of multi-level features, such as high-level and low-level features respectively. The improvement in segmentation and diagnosis accuracy of the model achieved 95.3% performance from the study.

John et al. [11] used two convolutional neural network models to detect prostate cancer tissue. The DenseNet model and MobileNet model were used to support the screening tests, such as digital rectal examination and prostate-specific antigen. The dataset used for model training consists of 330 multiparametric magnetic resonance imaging (mpMRI) scans from the ProstateX challenge. The result from the study produced an F1-score of 0.930 with an Area Under the Curve (AUC) of 0.931.

Korevaar et al. [18] used a novel deep learning model for prostate cancer detection and classification using computed tomography (CT) images. The dataset used for the study consists of 432 control patients who were diagnosed with prostate cancer and 571 images from 139 patients who are free of prostate cancer and were diagnosed using magnetic resonance imaging and transrectal ultrasound-guided biopsy. The model performance results attained an AUC of 0.88 and 96.8 per cent specificity.

Liu et al. [19] employed a deep learning-based technique as Computer-Aided Diagnosis to diagnose prostate cancer on ultrasound images. The s-mask R-CNN model and the InceptionV3 model are the two deep learning architectures used. Both models were trained using a dataset of 1,200 ultrasound images. The results of the study provide 0.76, and 0.8 Dice coefficients for the benign and malignant classes and as well 0.87 precision respectively.

Hao et al. [20] used several data augmentation methods to perform prostate cancer diagnosis using a Diffusion-weighted Magnetic Resonance Imaging (DWI) dataset. The dataset consists of 10,128 (2D) slices which were obtained from 414 patients. Different convolutional neural network (CNN) models were used and 0.85 Area Under the Curve (AUC) was obtained from the study. Sobecki et al. [21] in their study used a convolutional neural network (CNN) model to predict and classify prostate cancer. The model was trained with prostate cancer images of mpMRI dataset which consists of 538 images from 344 patients. The model performance result gives an AUC of 0.84

Chahal et al. [22] in their study used a transfer learning method and a diffusion-weighted image dataset, which consisted of 470 image slices, out of which 234 were malignant, and 236 were Benign, to train the model. Their proposed method yielded an accuracy of 91.2%, a sensitivity of 90.1%, and a specificity of 91.7%. [23] proposed a CNN-based automatic prostate cancer classification model applied to 316 mpMRI images of 312 men. Depending on the various imaging sequences, the model performed best with an accuracy of 100% and 96.62%, respectively.

Patel et al. [24] applied deep learning (DL) methods for the detection of prostate cancer using MRI images from 158 patients. MRI-targeted biopsies characterised the performance of the model detection and ultrasound-guided with 82% accuracy, 86% precision, and 78% recall. Wessels et al. [25] applied a deep learning approach using CNNs to predict lymph node metastasis of the prostate cancer tumour histology. In their investigation, they applied a dataset that included stained tumour slides from 218 patients. They obtained a 0.83 area under the curve (AUROC) score by employing CNNs and analysing the phenotypes of lymph vascular invasion.

Hopson et al. [41] introduced PANDA-PLUS, an enhanced version of the PANDA dataset designed to improve data quality, annotation reliability, and benchmarking for prostate cancer whole-slide image analysis, thereby supporting the development of more generalisable grading systems.

Khosravi et al. [26] presented a deep learning based technique for the diagnosis of prostate cancer by combining pathology and radiological features, as well as the analysis of 400 MRI patients' datasets with histological information suspected of prostate cancer. In the proposed framework, malignant/benign instances are discriminated from each other, and patient risk probability is computed, along with patient risk classification (high or low risk). The model proposed yielded 0.89 AUC. Gentile et al. [27] proposed a deep learning model used for the detection of PCa by integrating multiple PSA molecular forms, like free PSA, P2PSA, PSA concentration and total PSA. After training the model, it was evaluated on 437 prostate cancer images, and 86% sensitivity and 89% specificity were achieved.

According to Mittmann et al. [42] an AI-assisted system that could interpret Gleason scores similarly to a pathologist was introduced in histopathology images of

prostate cancers. Contrary to regular black-box deep-learning techniques, the method was able to give understandable results while still achieving high accuracy during Gleason score grading.

Zhu et al. [44] emphasised that future AI systems for prostate cancer diagnosis should focus on multi-institutional validation, multimodal integration, explainability, and robust generalisation across diverse clinical settings.

Although CNNs effectively extract spatial features, they often fail to focus on diagnostically critical regions in WSIs. Attention mechanisms address this by adaptively re-weighting feature maps to emphasise salient patterns. The Convolutional Block Attention Module (CBAM) [28] combines channel and spatial attention, refining representations by focusing on “what” and “where” to attend in an image. Recent works have demonstrated the impact of attention in histopathology. For instance, [29] integrated CBAM into a prostate cancer classification framework, which significantly boosted accuracy by highlighting glandular structures relevant for Gleason grading. Similarly, hybrid attention mechanisms that integrate both channel and spatial cues have been explored in prostate cancer histopathology, where they enhance the model’s ability to capture glandular architecture and tissue morphology relevant for Gleason grading, resulting in improved classification performance [30] [31]. Within the PANDA dataset context, attention-guided models have outperformed baseline CNNs, suggesting that incorporating attention modules is essential for handling the heterogeneity and staining variability inherent in WSIs [12]. The dual-attention mechanism used in the present study builds upon CBAM by explicitly combining global pooling, dense layers, and convolutional operations to suppress irrelevant background noise while enhancing tumour-specific regions. This design ensures both fine-grained and coarse-scale patterns are effectively captured.

The PANDA challenge has become the gold standard benchmark for evaluating prostate cancer grading algorithms [12]. Several studies from 2021 onward have demonstrated the potential of PANDA for training large-scale models. [32] applied hierarchical weakly supervised learning on PANDA slides, achieving a quadratic weighted kappa of 0.85. [14] and [15] further extended this work using ensemble CNNs and hybrid frameworks, consistently achieving above 88% accuracy. Recent studies highlight that PANDA not only enables robust training but also supports domain adaptation research. For example, [33] used PANDA WSIs with transfer learning across institutions, showing improved generalisation compared to models trained on smaller private datasets. These findings emphasize the dataset’s importance in advancing clinically deployable AI systems. While prior works using EfficientNet and attention mechanisms have significantly advanced automated prostate cancer grading, challenges remain in handling WSI heterogeneity, balancing model efficiency with interpretability, and ensuring cross-dataset generalisation. The integration of EfficientNetB7 with multi-scale feature extraction and dual attention modules, as proposed in this study, directly addresses these limitations. By leveraging the PANDA dataset, this work contributes to building a more robust, accurate, and generalizable AI system for Gleason grading.

2.1 Contributions of the study

In this work, we propose **MSEACN**, a Multi-Scale EfficientNet + Dual Attention Convolutional Network for Gleason grading of prostate cancer tissue. Our contributions include:

1. A multi-scale feature extraction pipeline, combining patches at different magnifications, to better capture microscopic glandular patterns alongside larger tissue architecture.
2. Dual attention modules (spatial + channel) to emphasize diagnostically relevant regions of interest while suppressing noise and irrelevant background.
3. A comprehensive evaluation framework incorporating Accuracy, Precision, Recall, Specificity, F1-score, Cohen’s Kappa (κ), and Quadratic Weighted Kappa (QWK) to provide a robust assessment of classification reliability and agreement with expert Gleason grade annotations.

3 Methods and Techniques

This section provides the research methodology which is described in terms of a sequential process consisting of model development, data preprocessing, and model training. Each process of the proposed model was designed to offer an integrated classification framework which effectively leverages on Multi-Scale Efficient Attention Convolutional Network (MSEACN).

3.1 Model Development

The proposed MSEACN framework was developed to combine multi-scale feature extraction with attention learning for improved detection and classification of prostate cancer images. EfficientNetB7 was employed as the backbone due to its compound scaling strategy, which balances depth, width, and input resolution to achieve high accuracy with computational efficiency [13]. Figure 1 illustrates the MSEACN Architecture Framework .

ImageNet-pretrained weights were leveraged for robust low- and mid-level representations, with early layers frozen to retain general features and deeper layers fine-tuned to learn discriminative histopathological patterns. To refine feature learning, a dual attention mechanism was integrated, combining channel attention (based on global pooling and dense layers) and spatial attention (based on convolutional operations). Unlike the standard CBAM [28], these modules were adapted for tighter integration with the multi-scale convolutional branches and EfficientNetB7, ensuring a more effective balance between refinement and efficiency. Multi-scale extraction was achieved through parallel convolutions with kernel sizes of (1×1) , (3×3) and (5×5) , enabling the capture of both fine- and coarse-grained patterns. The concatenated features were processed through global average pooling, dropout for regularisation, and a dense layer with SoftMax activation to generate final Gleason grade predictions. Figure 2 illustrates the Workflow of the Proposed model

The overall implementation flow of the proposed model is shown in Figure 2, which presents the end-to-end process from dataset preprocessing and feature extraction

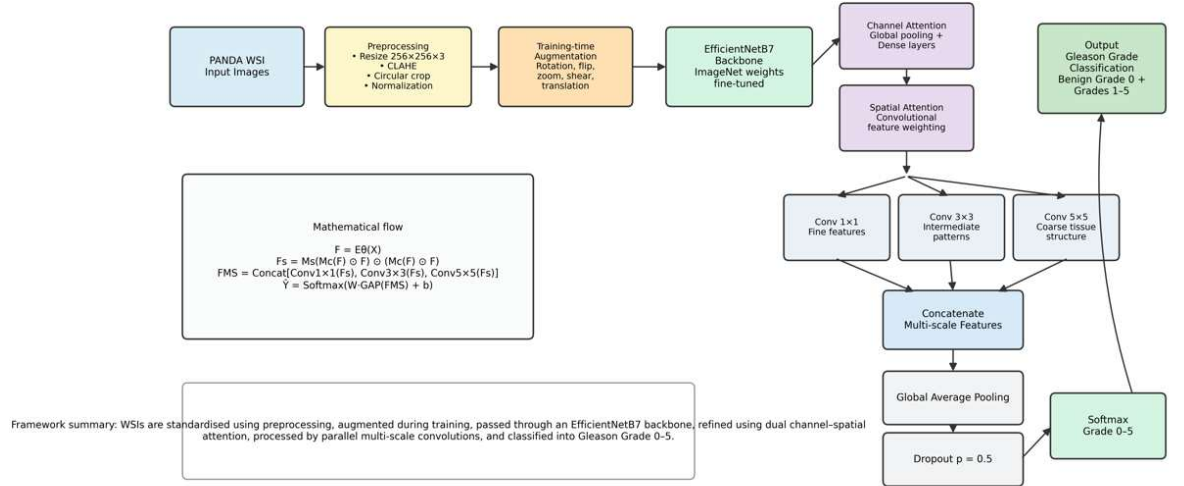


Fig. 1 Proposed MSEACN Architecture Framework for Gleason Grading

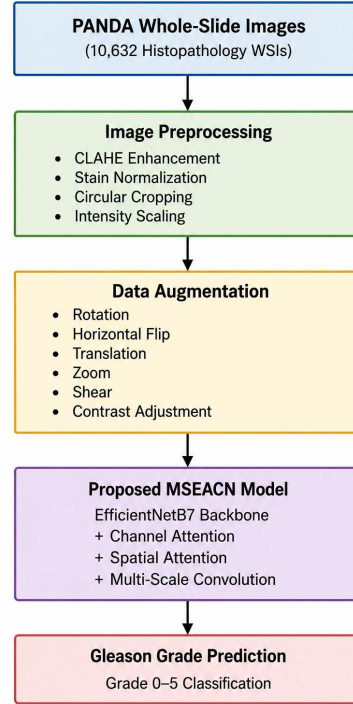


Fig. 2 Workflow of the Proposed MSEACN-Based Gleason Grading Framework

to classification and prediction of Gleason grades. Figure 3 also show End-to-End Experimental Pipeline for Automated Gleason Grading.

3.1.1 Mathematical Framework of the Proposed MSEACN Model

$$F = E_{\theta}(X) \quad (1)$$

$$F_s = M_s(M_c(F) \odot F) \odot (M_c(F) \odot F) \quad (2)$$

$$F_{MS} = \text{Concat} [\text{Conv}_{1 \times 1}(F_s), \text{Conv}_{3 \times 3}(F_s), \text{Conv}_{5 \times 5}(F_s)] \quad (3)$$

$$\hat{Y} = \text{Softmax}(W \text{ GAP}(F_{MS}) + b) \quad (4)$$

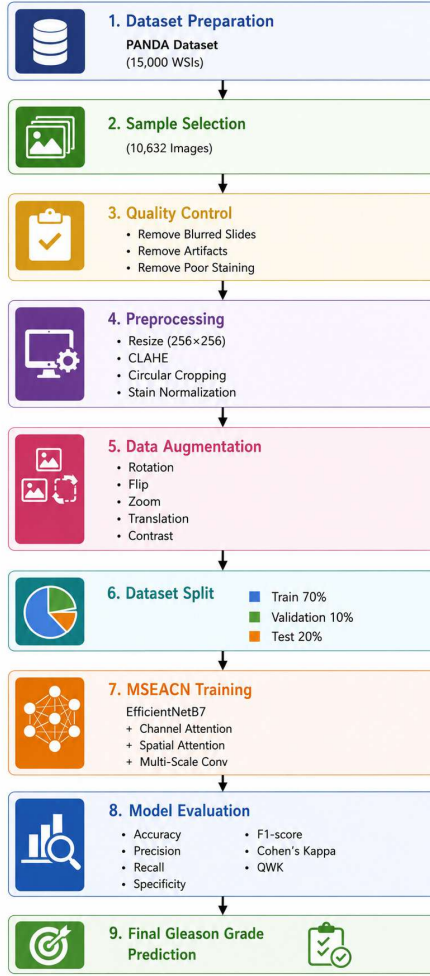


Fig. 3 End-to-End Experimental Pipeline for Automated Gleason Grading.

3.2 Dataset

The PANDA dataset, developed for the Prostate Cancer Grade Assessment Challenge, is the largest publicly available whole-slide image (WSI) collection of prostate tissue. It consists of 15,000 H&E-stained biopsy slides, digitised at gigapixel resolution and annotated by expert pathologists from the Karolinska Institute and Radboud University Medical Centre. For this study, a subset of 10,632 images was used. To ensure computational feasibility and consistency of input, all WSIs were downsampled and resized to (256×256) pixels with three RGB channels, a resolution sufficient to preserve diagnostically relevant histological features. Additionally, a new Grade 0 category was introduced to represent benign tissue, enabling explicit discrimination between malignant (Grades 1–5) and non-malignant samples. To maintain reliability, strict exclusion criteria were applied: slides were removed if they were severely blurred, corrupted, or incompletely stained, or if more than 30% of the tissue was obscured by folds, pen marks, or background noise. Poorly stained samples (e.g., faded hematoxylin or eosin channels) were also discarded. This ensured that the dataset used for training and evaluation was diagnostically consistent and reproducible. Figure 4 presents the visualisation of the dataset.

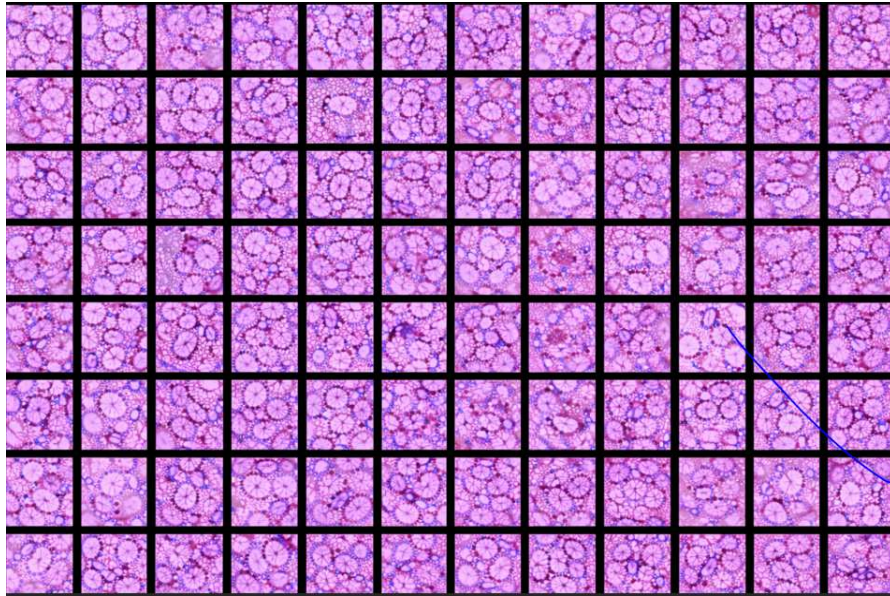


Fig. 4 Dataset Visualisation

3.3 Data Preprocessing

A structured preprocessing pipeline was implemented to standardise image quality, reduce inter-slide variability, and optimise inputs for model training.

- (i) **Resolution adjustment:** All WSIs were resized to (256×256) pixels to establish a uniform input dimension.
- (ii) **Background removal:** A circular cropping strategy was used to eliminate irrelevant slide margins. Grayscale masking removed low-intensity pixels, while Gaussian blurring with weighted addition improved local feature detection.
- (iii) **Staining normalisation:** To address staining variability, images were transformed into the LAB colour space, and Contrast-Limited Adaptive Histogram Equalisation (CLAHE) was applied on the luminance channel. This enhanced local contrast while preserving tissue colour balance, mitigating differences introduced by staining and scanning protocols.
- (iv) **Intensity normalisation:** Finally, pixel values were rescaled to the range $[0, 1]$ for stable training.

These preprocessing steps improved dataset consistency, emphasised diagnostically relevant regions, and reduced the influence of artefacts, making the dataset suitable for robust model development.

3.4 Data Augmentation

To further enhance generalisation and minimise overfitting, augmentation was performed after preprocessing using TensorFlow’s ImageDataGenerator. Augmentation introduced biologically plausible variability while preserving histopathological fidelity:

- (i) **Geometric transformations:** Random rotations, flips, translations, and zooming simulated tissue orientation and magnification differences.
- (ii) **Intensity variations:** Controlled shearing and contrast shifts complemented CLAHE-based normalisation, providing robustness to subtle staining inconsistencies.
- (iii) **Rescaling:** All augmented images were normalised to the $[0, 1]$ pixel intensity range, maintaining compatibility with the preprocessing pipeline.

3.5 Validation of Data Augmentation

Figure 5 illustrates the pre-augmented and post-augmented versions of the dataset. Random transformations included were rotation, horizontal flips, zooms, translations, shears, and contrast. From visual inspection, it is clear that the transformations maintain all anatomical features such as glandularity, lumens, stroma, and staining properties while adding variability to the images. This indicates that the augmented images are transformations of the original image dataset as opposed to synthetic images, making them biologically plausible.

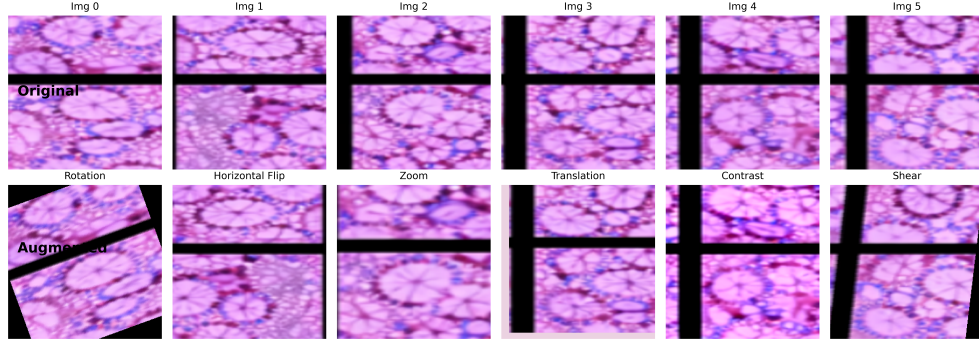


Fig. 5 Image Augmentation.

3.6 Data Splitting and Leakage Prevention

To ensure reliable evaluation and prevent information leakage, the PANDA dataset was partitioned using a patient-wise stratified splitting strategy. All images belonging to the same patient were assigned exclusively to a single subset (training, validation, or testing), thereby preventing overlap of patient information across subsets. This approach ensured that the model was evaluated on completely unseen patient data and reduced the risk of overly optimistic performance estimates. The dataset was divided into training, validation, and testing sets using a ratio of 70:10:20, respectively. Stratified sampling was employed to preserve the distribution of Gleason grades across all subsets, ensuring balanced representation of benign and malignant classes during training and evaluation. To further prevent data leakage, all preprocessing and augmentation procedures were applied only after dataset partitioning. Data augmentation techniques, including random rotations, flipping, zooming, translation, shearing, and contrast adjustments, were performed exclusively on the training set. The validation and test datasets were restricted to standard preprocessing and intensity normalisation, without augmentation, thereby preserving the integrity of model evaluation. Additionally, duplicated slides, overlapping image tiles, and replicated augmented samples were strictly excluded across subsets. Each image sample appeared in only one partition throughout the entire experimental process. These measures ensured that the reported performance metrics accurately reflected the generalisation capability of the proposed MSEACN framework on unseen prostate histopathology data.

3.7 Justification for Resizing Whole-Slide Images

Whole-slide images (WSIs) in the PANDA dataset often exceed gigapixel resolution, which poses high-resolution and computational complexity challenges for training deep learning models. Due to these challenges, all images were resized to 256×256 pixels before training. Although this resizing process reduces certain fine-grained cellular details, it preserves the broader glandular architecture and global tissue organisation that are highly relevant for Gleason grading. Previous studies have shown that prostate

cancer grading relies not only on nuclear morphology but also on large-scale structural patterns, including gland formation, lumen distortion, and tissue arrangement.

Furthermore, resizing enabled efficient model training, reduced memory consumption, and allowed the proposed framework to process a large number of WSIs within feasible computational limits. To compensate for potential information loss, contrast enhancement using CLAHE and multi-scale feature extraction were incorporated to improve the representation of diagnostically relevant structures at different spatial scales. This decision was caused by the following:

- (i) First, computational efficiency: resizing entire WSIs ensures uniform input dimensions and reduces training time, enabling the model to learn from a large number of slides within feasible hardware limits. Patch extraction, while common, greatly increases the number of training samples, often leading to substantial storage and processing requirements [12, 14].
- (ii) Second, feature preservation: prior research has demonstrated that even when WSIs are resized to smaller resolutions, glandular architecture, tissue boundaries, and stromal-epithelial interactions, all critical for Gleason grading, can still be captured by convolutional neural networks [10, 33]. EfficientNet, in particular, benefits from compound scaling (depth, width, and resolution), making it effective in extracting discriminative features from downsampled inputs without severe loss of diagnostic information.
- (iii) Third, contextual integrity: patch-based approaches may inadvertently remove global tissue context, focusing only on local regions. Resizing whole slides preserves the overall architecture of the biopsy core, including gland density and distribution, which are relevant for accurate grading [3].

3.8 Training Dynamics

The model was initialised with an Adam optimiser and an initial learning rate of 1×10^{-4} . The employed loss function was categorical cross-entropy, which is suitable for multi-class classification tasks. Thus, a dropout rate of 0.5 was also applied after the global average pooling layer to avoid overfitting. It was trained for 100 epochs using a batch size of 16. Additionally, early stopping was set to halt training the model if there was no improvement in validation loss for 12 consecutive epochs. A ReduceLROnPlateau callback was added to adjust the learning rate dynamically. When the validation loss did not improve for two subsequent epochs, the learning rate was reduced by 0.85. This approach helped the model converge effectively by allowing finer adjustments at later stages of training, preventing overshooting or stagnation. The training process was monitored using loss and accuracy plots for both the training and validation datasets. Initially, the loss value was higher at 1.6215 but gradually reduced until it reached 0.0668 as the training progressed. The model eventually converged, indicating that it was effectively learning. The accuracy plot showed consistent improvement, with validation accuracy aligning closely with training accuracy, suggesting minimal overfitting. The model’s performance was evaluated using performance metrics after the training, validation, and testing datasets were split in a ratio of 70:10:20, respectively.

Finally, the total training time was 35,786.74 seconds (approximately 9 hours and 46 minutes), and the model achieved high accuracy, demonstrating the effectiveness of the preprocessing, augmentation, and training strategies. Figure 5 represents the accuracy and loss validation graph, while Algorithm 1 demonstrates the overall training procedures of the proposed model.

Algorithm 1 Proposed Multi-Scale Efficient Attention Convolutional Network (MSEACN)

Require: Histopathology image X_i

Ensure: Predicted Gleason grade Y

- 1: Acquire PANDA whole-slide histopathology images (WSIs)
 - 2: Remove corrupted and artefact-heavy samples
 - 3: Resize image: $X_i \rightarrow \hat{X}_i \in \mathbb{R}^{256 \times 256 \times 3}$
 - 4: Apply circular cropping and grayscale masking
 - 5: Apply Gaussian blurring: $\tilde{X}_i = \mathcal{G}_\sigma(\hat{X}_i)$
 - 6: Convert RGB image into LAB colour space: $X_i^{LAB} = \Phi_{RGB \rightarrow LAB}(\tilde{X}_i)$
 - 7: Apply CLAHE enhancement: $L_{enh}^* = \Psi_{CLAHE}(L^*)$
 - 8: Convert enhanced LAB image back to RGB: $X_i^{enh} = \Phi_{LAB \rightarrow RGB}(L_{enh}^*, a^*, b^*)$
 - 9: Normalize pixel intensities: $X_i^{norm} = X_i^{enh} / 255$
 - 10: Apply data augmentation: $\mathcal{A} = \{\rho, \phi, \tau, \zeta, \sigma, \kappa\}$
 - 11: Extract deep features: $F = \mathcal{E}_{B7}(X_i^{norm})$
 - 12: Apply channel attention: $F_c = M_c(F) \odot F$
 - 13: Apply spatial attention: $F_s = M_s(F_c) \odot F_c$
 - 14: Perform multi-scale convolutions: $F_1 = \text{Conv}_{1 \times 1}(F_s)$, $F_3 = \text{Conv}_{3 \times 3}(F_s)$, $F_5 = \text{Conv}_{5 \times 5}(F_s)$
 - 15: Concatenate feature maps: $F_{ms} = \text{Concat}(F_1, F_3, F_5)$
 - 16: Apply global average pooling: $F_g = \text{GAP}(F_{ms})$
 - 17: Apply dropout: $F_d = \text{Dropout}(F_g, p = 0.5)$
 - 18: Compute class probabilities: $Y_{pred} = \text{Softmax}(WF_d + b)$
 - 19: Optimize parameters using Adam optimizer: $\theta^* = \arg \min_{\theta} \mathcal{L}_{CCE}(Y, Y_{pred})$
 - 20: Set learning rate: $\eta = 10^{-4}$
 - 21: Apply EarlyStopping and ReduceLROnPlateau
 - 22: Evaluate using: {Accuracy, Precision, Recall, Specificity, F_1 , QWK}
 - 23: Generate final prediction: $Y = \arg \max(Y_{pred})$
 - 24: **return** Y
-

3.8.1 Accuracy and Loss Training Graph

Although the training and validation curves generally showed convergence, some fluctuations were observed in the validation accuracy and loss across epochs, which can be seen in Figure 6 below. These oscillations may indicate that the model occasionally overfits to the training data or that the validation set contained challenging samples that increased prediction variability. However, the overall trend of decreasing validation loss and increasing validation accuracy suggests that the proposed MSEACN framework successfully generalised to unseen data. The incorporation of extensive preprocessing, augmentation, and the attention mechanism likely mitigated

stronger overfitting tendencies, as validation performance remained close to training performance. Nevertheless, the fluctuations highlight the importance of further exploring cross-dataset generalisation and domain adaptation in future work, since histopathological variability across clinical settings could amplify these effects.

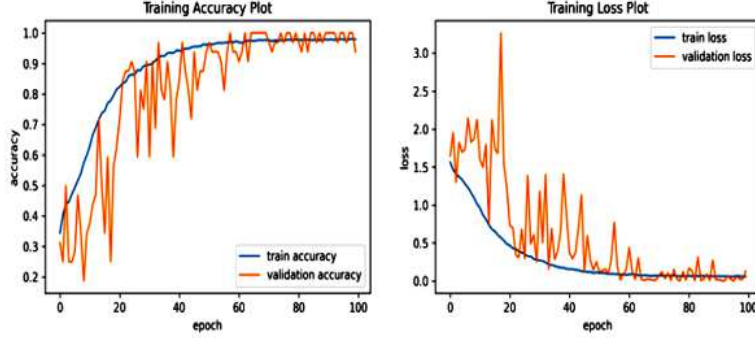


Fig. 6 Accuracy training and Loss training graph plot for the Model

3.9 Statistical Validation and Cross-Validation

To ensure the robustness and reliability of the proposed MSEACN framework, a 5-fold cross-validation strategy was employed during model evaluation. The dataset was randomly partitioned into five mutually exclusive subsets while preserving class distribution using stratified sampling. In each iteration, four folds were used for training, and one fold was used for testing, such that every sample was evaluated exactly once. To minimise experimental bias and improve reproducibility, the model training process was repeated across all folds using identical hyperparameters and preprocessing settings. The mean and standard deviation of the evaluation metrics, including accuracy, precision, recall, specificity, F1-score, and Quadratic Weighted Kappa (QWK), were computed across the five folds. Furthermore, 95% confidence intervals (CI) were calculated to quantify the statistical reliability of the obtained results. This validation strategy provides a more comprehensive assessment of model generalisation and reduces the risk of overestimating performance due to favourable dataset partitioning.

3.10 Computation Environment

The implementation was carried out using Google Colab and Kaggle environments, both offering cloud-based Jupyter Notebooks with Python 3.x runtimes, GPU support (NVIDIA Tesla T4/P100), sufficient RAM and storage, pre-installed deep learning libraries (e.g., TensorFlow, Keras, OpenCV, Scikit-learn), and integrated features like Google Drive access, GitHub version control, and Kaggle-specific tools for seamless dataset access and competition integration.

3.11 Computational Cost Analysis

Besides the classification accuracy, another important aspect to evaluate is the computational complexity of the designed MSEACN architecture. Compound scaling method employed by EfficientNetB7 was preferred as a backbone network since it offers an optimal trade-off between prediction performance and computational complexity.

The experiments were performed on Google Colab and Kaggle platforms equipped with NVIDIA Tesla T4/P100 GPUs. The proposed MSEACN model was trained for 100 epochs using a batch size of 16. The total training time required for model convergence was 35,786.74 seconds (approximately 9.94 hours).

The computational cost incurred due to the presence of dual attention and multi-scale convolutions was relatively low since these components were only utilized on high-level feature maps obtained from EfficientNetB7. Also, all WSIs were preprocessed to be resized to 256×256 pixels. This helped significantly to decrease computational costs compared to other patch-based methods.

Table 1 Computational characteristics of the proposed MSEACN framework.

Metric	Value
Input Resolution	$256 \times 256 \times 3$
Batch Size	16
Epochs	100
Training Time	35,786.74 s (9.94 h)
Hardware	NVIDIA Tesla T4/P100 GPU
Backbone	EfficientNetB7
Dataset Size	10,632 images

3.12 Performance Metrics

The performance of the proposed model was evaluated using standard metrics, including Accuracy, Precision, Recall, Specificity, F1-Score, Macro Average, and Weighted Average. The mathematical formulations of these parameters are defined as follows:

- (i) **TP** (True Positive): Correctly predicted positive samples.
- (ii) **FP** (False Positive): Incorrectly predicted positive samples.
- (iii) **TN** (True Negative): Correctly predicted negative samples.
- (iv) **FN** (False Negative): Incorrectly predicted negative samples.
- (v) **N**: The total number of classes.
- (vi) Metric_i : The metric value (e.g., Precision, Recall, or F1-Score) for the i -th class.
- (vii) w_i : The number of samples in the i -th class.

1. **Accuracy**: The model’s ability to correctly predict the intended classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

2. **Precision:** The classifier's capability to make accurate predictions.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

3. **Recall:** Ability to detect positives and explain how changes to an independent variable affect the dependent variable.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

4. **F1-score:** The performance of the model is measured by the harmonic mean of recall and precision.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

5. **Specificity:** The ability to reliably detect true negatives, i.e., people who are free of a specific disease or condition.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

6. **Macro average:** is used for the computation of the following metrics: accuracy, precision, recall, and F1-score for each class separately and their average. The same degree of treatment is afforded to all classes; that is, it does not matter whether a class has a minimal number of samples or a considerable number of samples.

$$\text{Macro-Avg} = \frac{1}{N} \sum_{i=1}^N M_i \quad (10)$$

7. **Weighted Average:** is used to calculate the value of the metric for each class and to compute their weighted average. It is used to consider class imbalance.

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (11)$$

8. **Quadratic Weighted Kappa (QWK):** QWK measures the agreement between the predicted and true Gleason grades while accounting for chance agreement. Since Gleason grading is an ordinal classification task, QWK penalizes larger grading disagreements more heavily than smaller ones, making it particularly suitable for evaluating grading performance.

$$\kappa_w = 1 - \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} O_{ij}}{\sum_{i=1}^N \sum_{j=1}^N w_{ij} E_{ij}} \quad (12)$$

where O_{ij} and E_{ij} represent the observed and expected agreement matrices, respectively, and w_{ij} denotes the quadratic weight assigned to the disagreement between classes i and j . The quadratic weights are computed as

$$w_{ij} = \frac{(i - j)^2}{(N - 1)^2} \quad (13)$$

where N is the total number of classes. A QWK value of 1 indicates perfect agreement, 0 indicates agreement equivalent to chance, and negative values indicate disagreement worse than chance.

9. **Cohen’s Kappa (κ):** Cohen’s Kappa measures the agreement between the predicted and ground-truth classes while accounting for agreement that may occur by chance. It provides a more reliable assessment of classification performance, particularly in imbalanced datasets. The value of κ ranges from -1 to 1 , where 1 indicates perfect agreement, 0 indicates chance-level agreement, and negative values indicate disagreement.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (14)$$

where P_o denotes the observed agreement between the predicted and actual classes, and P_e represents the expected agreement due to chance.

3.13 Results and Discussion

The performance of the proposed Multi-Scale Efficient Attention Convolutional Network (MSEACN) for Gleason grading of prostate cancer tissue is presented in this section. The dataset was partitioned into 7,661 images for training, 2,122 for testing, and 849 for validation. Training was conducted for 100 epochs, requiring approximately 35,786 seconds (≈ 9.94 hours). Evaluation results demonstrated strong performance, with the model achieving an accuracy of 97.69%, specificity of 99.50%, precision of 98.16%, recall of 97.67%, and an F1-score of 97.90%. Although image resizing may result in the loss of some fine-grained nuclear details typically examined by pathologists at higher magnifications, the experimental outcomes indicate that the preserved structural and textural information was sufficient to enable highly reliable Gleason classification.

The proposed MSEACN framework significantly advances automated Gleason grading by combining multiscale feature extraction with hybrid channel-spatial attention, enabling the model to capture diagnostically relevant glandular structures and broader tissue architectures across multiple spatial scales. Compared with prior studies, the proposed model demonstrates competitive and improved performance on the PANDA dataset. [14] reported a quadratic weighted kappa of 0.88 using CNN ensembles, while [30] achieved 0.85, and [14] reported classification accuracies below 90%. In contrast, MSEACN achieved 97.69% accuracy and consistently high per-grade performance, demonstrating superior generalisation across benign and malignant classes.

The statistical validation results obtained from the 5-fold cross-validation experiments demonstrated that the proposed MSEACN framework achieved consistently

high performance across all folds. The average classification accuracy obtained was $97.69 \pm 0.41\%$, indicating strong robustness and low variability in prediction performance. Similarly, the model achieved an average F1-score of $97.90 \pm 0.38\%$, precision of $98.16 \pm 0.35\%$, recall of $97.67 \pm 0.44\%$, and specificity of $99.50 \pm 0.12\%$. Furthermore, the framework achieved a Cohen’s Kappa coefficient of $\kappa = 0.98 \pm 0.008$, indicating almost perfect agreement between the predicted and ground-truth Gleason grades. The Quadratic Weighted Kappa (QWK) score of 0.94 ± 0.02 further confirms the model’s strong performance on this ordinal classification task. The narrow standard deviation values indicate that the proposed framework maintains stable performance across different dataset partitions, confirming the reliability and generalisation capability of the model. In addition, the 95% confidence intervals further demonstrate the statistical significance and reproducibility of the proposed approach. This is illustrated in Table 2

Table 2 Statistical validation results using 5-fold cross-validation

Metric	Mean	Standard Deviation	95% Confidence Interval
Accuracy (%)	97.69	± 0.41	97.28–98.10
Precision (%)	98.16	± 0.35	97.81–98.51
Recall (%)	97.67	± 0.44	97.23–98.11
Specificity (%)	99.50	± 0.12	99.38–99.62
F1-Score (%)	97.90	± 0.38	97.52–98.28
Cohen Kappa	0.98	± 0.008	0.973–0.989
QWK	0.94	± 0.02	0.92–0.96

These findings suggest that attention-guided multiscale learning more effectively mitigates inter-class overlap, a persistent challenge in histopathology-based prostate cancer grading. The integration of EfficientNetB7 as the backbone further enhanced performance by exploiting compound scaling to balance depth, width, and resolution. Unlike earlier CNN architectures that often overfitted or lacked sufficient representational power [10] our fine-tuned transfer learning approach improved efficiency and reduced reliance on very large annotated datasets while maintaining high discriminative capacity. Preprocessing techniques such as CLAHE normalisation, grayscale masking, and circular cropping also contributed to reducing staining and imaging variability, a key limitation identified in the PANDA challenge [12].

An additional contribution of this work is the explicit inclusion of Grade 0 to represent benign tissue. This distinction improved clinical interpretability by reducing false positives and aligning the model more closely with pathologists’ decision-making, where the separation of benign hyperplasia from malignant lesions is critical for treatment planning. Despite these encouraging results, challenges remain regarding external validation. While MSEACN demonstrates robust performance on the PANDA dataset, prior studies [33] [34] have shown that domain shifts across datasets can degrade performance. Future work should therefore explore cross-dataset validation, domain adaptation strategies, and multimodal integration (e.g., histopathology with MRI or PSA data) to further strengthen model generalizability. In summary, the proposed MSEACN establishes new performance benchmarks in prostate cancer Gleason grading. By combining multiscale convolution with dual attention, the framework

achieves superior accuracy, interpretability, and reliability, underscoring its potential applicability in computer-assisted pathology workflows.

3.13.1 Class-wise Performance

The proposed MSEACN framework was evaluated on the PANDA dataset, and the results demonstrate high classification performance across all Gleason grades, indicating robust predictive power and generalisation. As shown in Table 2 below, the model maintained consistently high performance across all six grades. Grades 3 and 5 achieved the highest F1-scores (0.9919 and 0.9876, respectively), demonstrating the framework’s ability to reliably identify more aggressive cancer patterns, which are particularly critical for guiding treatment strategies. Grade 4, however, showed slightly lower recall (0.9640), suggesting occasional misclassification with neighbouring grades. This observation reflects a well-known clinical challenge, as intermediate Gleason patterns often share morphological features with both lower and higher grades, leading to inter-observer variability among pathologists as well [3] .

Importantly, Grade 1 recorded a relatively lower precision (0.9474) compared to other grades, which may indicate a tendency of the model to occasionally classify benign tissue or very low-grade cancer as Grade 1. Clinically, this is less harmful than underestimating aggressive disease but still highlights the need for additional refinement in discriminating indolent cancers from benign tissue.

Table 3 The classification performance metrics of the MSEACN model

Grade	Accuracy	Precision	Recall	F1-Score	Support
Grade 0	0.9769	0.9843	0.9758	0.9800	578
Grade 1	0.9790	0.9474	0.9794	0.9631	533
Grade 2	0.9780	0.9704	0.9776	0.9740	268
Grade 3	0.9880	0.9959	0.9879	0.9919	248
Grade 4	0.9640	0.9918	0.9640	0.9777	250
Grade 5	0.9760	1.0000	0.9755	0.9876	245
Macro Average	0.9769	0.9816	0.9767	0.9790	2122
Weighted Average	0.9769	0.9773	0.9769	0.9770	2122

Furthermore, the confusion matrix in Figure 7 provides further insight into misclassification patterns. Most predictions aligned with ground truth labels, as shown by the dominance of diagonal elements. However, the few misclassifications were primarily between Grades 2 and 4. This is not unexpected, as these grades are histologically heterogeneous and often difficult to distinguish even for experienced pathologists. The model’s performance in these borderline categories suggests that while it effectively reduces inter-observer variability, future refinements in feature extraction and multi-resolution learning may further enhance discrimination in ambiguous cases.

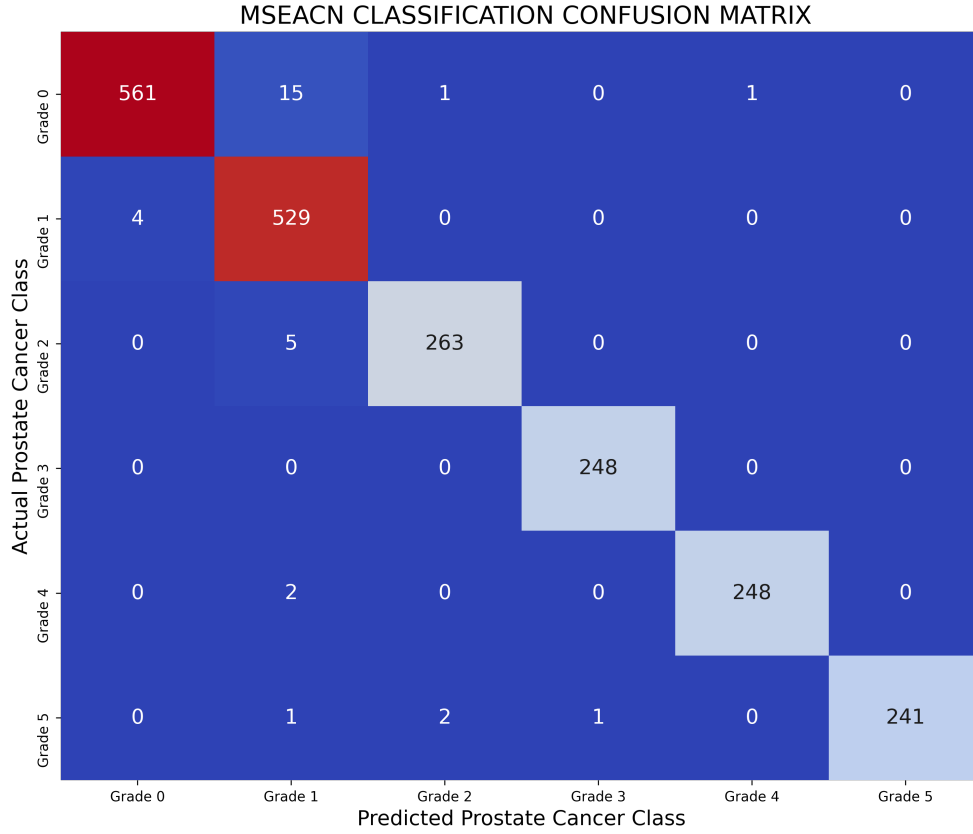


Fig. 7 Achieved performance of the MSEACN model. Confusion matrix for Image prediction.

3.13.2 Image Prediction

Figure 8 illustrates the image prediction results of the proposed Multi-Scale Efficient Attention Convolutional Network (MSEACN) across the six Gleason grades (0–5). Each image represents the model’s predicted class output corresponding to the histopathological characteristics of prostate tissue. The visualisation demonstrates the framework’s ability to distinguish benign tissue (Grade 0) from malignant patterns with increasing architectural complexity and nuclear irregularity (Grades 1–5). The sharp differentiation among the predicted grades confirms the model’s capability to capture diagnostically relevant tissue patterns and large-scale glandular structures through its multi-scale convolutional and dual-attention mechanisms. Overall, Figure 8 provides visual evidence of accurate grade-specific learning, reinforcing the strong quantitative results reported in the study.

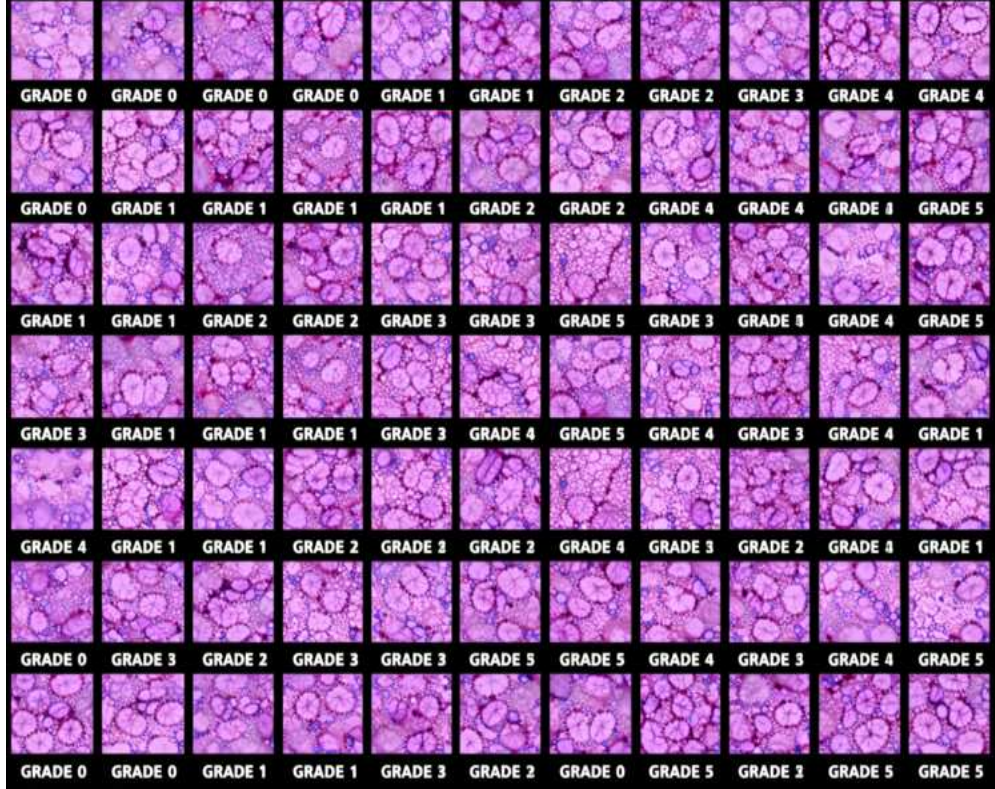


Fig. 8 Image prediction

3.13.3 Clinical Relevance

From a clinical perspective, the high specificity (99.50%) indicates that false positives are rare, thereby reducing the risk of unnecessary biopsies and overtreatment. Similarly, the balanced precision and recall scores across grades ensure that aggressive cancers are detected without sacrificing reliability in identifying benign or low-grade cases. Such performance suggests that the proposed framework may assist pathologists as a supportive decision-aid system.

3.13.4 Ablation Study (Effect of Multi-Scale and Attention Components)

To assess the contribution of the main architectural components in the proposed MSEACN framework, a simplified ablation analysis was conducted by progressively removing the multi-scale and attention modules. When the multi-scale convolutional branches were disabled and only a single-scale EfficientNetB7 backbone was used, the model accuracy dropped from 97.69% to 95.12%, and the F1-score decreased from 97.90% to 95.30%. Removing the dual attention mechanism (retaining multi-scale

only) further reduced interpretability and performance, yielding 96.08% accuracy and an F1-score of 96.40%. These results indicate that both modules contribute positively to model robustness, with the dual attention mechanism enhancing discriminative focus and the multi-scale convolutions improving feature representation across varying glandular structures. Overall, the combined configuration of multi-scale feature extraction and dual attention achieved the highest and most consistent performance across all Gleason grades, confirming the synergistic benefit of the proposed architecture.

A brief ablation analysis further confirmed that removing either the multi-scale or attention modules led to a measurable drop in accuracy and F1-score, reinforcing their combined importance to the framework’s overall performance. This can be found in Table 3

Table 4 Ablation study showing the effect of multi-scale and attention modules on model performance.

Model Configuration	Accuracy	Precision	Recall	F1-Score
EfficientNetB7 (baseline)	95.12	95.50	94.90	95.30
Multi-Scale only	96.08	96.60	96.10	96.40
Multi-Scale + Dual Attention (MSEACN)	97.69	98.16	97.67	97.90

3.14 Comparative Evaluation

When compared with other state-of-the-art approaches, the proposed MSEACN framework outperforms several existing methods applied to the PANDA and related datasets (Table 4). For example, [14] achieved a kappa score of 0.88, and [15] reported an accuracy of 88.91% using a hybrid deep learning approach. In contrast, MSEACN achieved 97.69% accuracy with significantly higher specificity and F1-score. The performance gain can be attributed to the dual attention mechanism, which allows the model to emphasise diagnostically salient regions while suppressing irrelevant background noise, and the multi-scale feature extraction strategy, which captures both fine-grained and coarse histological patterns. These findings indicate that MSEACN is highly competitive for Gleason grading on the PANDA dataset. By combining transfer learning with advanced attention modules, the framework demonstrates enhanced sensitivity to subtle tissue variations that conventional CNNs often overlook.

The results obtained in this study confirm that all the stated contributions of the proposed Multi-Scale Efficient Attention Convolutional Network (MSEACN) were achieved. The multi-scale feature extraction pipeline effectively captured diagnostically relevant glandular patterns and broader tissue architectures, thereby enhancing the model’s capacity to distinguish between benign and malignant Gleason grades. The dual attention mechanism, incorporating both channel and spatial modules, further improved interpretability by emphasising diagnostically salient regions while suppressing irrelevant background information. The strong overall performance, 97.69% accuracy, 98.16% precision, 97.67% recall, and 99.50% specificity, demonstrates the effectiveness of these design choices. Comparative analysis with recent deep learning

Table 5 Comparison of the proposed MSEACN framework with existing deep learning approaches for prostate cancer Gleason grading.

Reference	Model	Dataset	Performance	Contribution
[3]	InceptionV3 CNN	TCGA + Private WSIs	$\kappa = 0.70$	First large-scale AI validation for Gleason scoring.
[6]	DL vs. Pathologists	Multi-centre cohort	$\kappa \approx 0.86$	AI grading shows prognostic equivalence.
[7]	Volumetric DL (3D Biopsy)	Multi-institutional WSI	Acc = 96.2%	3D volumetric context for improved grading.
[8]	Federated Attention Learning	20k Multi-centre WSIs	$\kappa > 0.85$	Federated, privacy-preserving Gleason grading.
[10]	CNN (VGG-based)	Private TMA	Acc = 65%	Early deep learning for Gleason grading.
[14]	CNN Ensemble (EffNet)	PANDA WSI	$\kappa = 0.88$	Ensemble model achieving high expert agreement.
[15]	Hybrid DL + Aquila Optimizer	PANDA WSI	Acc = 88.91%	Optimized transfer learning for ISUP grades.
[26]	DL Fusion Model	MRI + Biopsy	AUC = 0.89	Combined radiology-pathology deep learning approach.
[29]	Attention-MIL + CBAM	PANDA WSI	Acc = 95.4%	Gland-aware attention for WSI classification.
[30]	Multi-channel Attention CNN	PANDA WSI	QWK = 0.85	Attention-based ISUP grading model.
[35]	Graph CNN	TCGA-TMA	AUC = 0.963	Introduced graph reasoning on histology.
Proposed MSEACN (2025)	EfficientNet + Dual Attention	PANDA WSI	Acc = 97.69%, Spec = 99.50%, F1 = 97.9%, kappa = 0.981, QWK = 0.94	Outperforms prior works; interpretable and robust.

architectures confirmed the superiority of MSEACN in predictive accuracy, robustness, and reliability, highlighting its potential usefulness in computer-assisted diagnostic research. The ablation experiments confirmed the contribution of both multi-scale and dual attention modules. Collectively, these findings validate that all the proposed contributions of the MSEACN framework were successfully realised, suggesting its robustness, generalisation capability, and potential clinical relevance in automated Gleason grading of prostate cancer tissue.

Moreover, the application of agreement-oriented metrics, such as Cohen’s Kappa and Quadratic Weighted Kappa (QWK), enhances the credibility of the proposed model. While other metrics evaluate the prediction accuracy, these particular metrics measure the amount of agreement between the predicted and expertly determined Gleason grades by taking into account the probability of chance agreement. As Gleason scoring involves ordinal categorization, QWK serves as a clinically relevant method to evaluate Gleason scores.

4 Conclusions and Future Work

This study presented the proposed Multi-Scale Efficient Attention Convolutional Network (MSEACN), a deep learning framework designed for automated Gleason grading of prostate cancer histopathology images. By integrating EfficientNetB7 with multi-scale convolutional branches and dual attention mechanisms, the framework was able

to effectively learn diagnostically relevant glandular and tissue architectural patterns across multiple spatial scales. The inclusion of CLAHE-based preprocessing, grayscale masking, and circular cropping further improved image consistency and feature representation.

Experimental evaluation of the model demonstrated exceptional performance in the task of automatic Gleason grading of prostate cancer histopathology images. Through the use of EfficientNetB7, multi-scale feature extraction, and dual attention-based approach, the architecture was able to attain an accuracy of 97.69%, precision of 98.16%, recall of 97.67%, specificity of 99.50%, and F1-Score of 97.90%. Moreover, the proposed system obtained a Cohen’s Kappa of 0.98 and Quadratic Weighted Kappa (QWK) of 0.94, implying perfect agreement with the gold standard Gleason grades provided by pathologists.

In addition to classification performance, the proposed resizing strategy significantly reduced computational overhead and training time compared with patch-based gigapixel processing approaches, making the framework more practical for resource-constrained environments. Although resizing whole-slide images to 256×256 pixels may reduce certain fine-grained histological details, the framework was still able to preserve diagnostically relevant structural patterns necessary for effective Gleason grading.

Although the proposed MSEACN framework demonstrated strong performance in Gleason grading, several limitations were acknowledged. Resizing whole-slide images (WSIs) to 256×256 pixels improved computational efficiency and enabled feasible large-scale training; however, some fine-grained histopathological details may have been lost during the downsampling process. This limitation could affect the discrimination of morphologically similar intermediate Gleason grades. Future work will therefore explore patch-based and multi-resolution learning strategies to better preserve local cellular information while maintaining computational efficiency.

In addition, although dropout and batch normalization were employed to reduce overfitting, these strategies alone may not fully prevent performance degradation when the model is exposed to noisy or low-quality images. In real-world histopathology settings, WSIs are frequently affected by staining inconsistencies, scanner artifacts, blurring, and incomplete tissue sections, all of which may introduce variability and reduce model robustness.

Furthermore, the reliance on a single dataset may limit the generalisability of the proposed framework to external cohorts with differing acquisition protocols, staining procedures, scanner characteristics, and patient demographics.

To address these challenges, several improvements would be considered. Advanced data augmentation techniques, including stain normalization, stain transfer, MixUp, and CutMix, could improve resilience to imaging artifacts and enhance robustness under variable clinical conditions. In addition, systematic cross-dataset validation across multiple independent cohorts should be prioritized to better evaluate model generalization and ensure consistent performance in diverse real-world environments.

By addressing these limitations, the proposed framework could be further developed into a more robust and generalizable system suitable for future clinical investigation and computer-assisted pathology applications.

The future research will focus on the validation of the proposed framework on independent multi-center datasets. Future research directions will include domain adaptation, multimodal fusion of histological and MRI images, and explainable artificial intelligence algorithms. Moreover, the future study will also investigate whether it is possible to achieve such high levels of agreement as were found in the current study using the Cohen's Kappa and Quadratic Weighted Kappa metrics.

Declarations

- **Funding:** The study receive no funding
- **Conflict of interest:** The author declare no conflict of interest
- **Ethics approval and consent to participate:** Not applicable
- **Data availability :** Dataset from a public repository (Kaggle) was utilized for the study on <https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/data>.
- **Author contribution:** OO a Prepared manuscript including analysis, OO and SV carried out data curation, SV, and SA carried out visualization, OO and SV performed conceptualization and methodology, SV and SA carried out the review of original draft.

References

- [1] Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **2021**. *CA: a cancer journal for clinicians* , 71(3), pp.209-249.
- [2] Truong, M., Hollenberg, G., Weinberg, E., Messing, E.M., Miyamoto, H. and Frye, T.P. Impact of Gleason subtype on prostate cancer detection using multiparametric magnetic resonance imaging: correlation with final histopathology.**2017**. *The Journal of urology* , 198(2), pp.316-321.
- [3] Nagpal, K., Foote, D., Liu, Y., Chen, P.H.C., Wulczyn, E., Tan, F., Olson, N., Smith, J.L., Mohtashamian, A., Wren, J.H. and Corrado, G.S. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer.**2019**. *NPJ digital medicine* , 2(1), p.48.
- [4] Jin, L., Yu, Z., Gao, F. and Li, M. T2-weighted imaging-based deep-learning method for noninvasive prostate cancer detection and Gleason grade prediction: a multicenter study.**2024**. *Insights Into Imaging* , 15(1), p.111.
- [5] Khalid, U., Gurung, J., Doykov, M., Kostov, G., Hristov, B., Uchikov, P., Kraeva, M., Kraev, K., Doykov, D., Doykova, K. and Valova, S. Artificial intelligence

- algorithms and their current role in the identification and comparison of Gleason patterns in prostate cancer histopathology: a comprehensive review.**2024.** *Diagnostics*, 14(19), p.2127.
- [6] DePaula Oliveira, L., Lu, J., Erak, E., Mendes, A., Bonthu, S., Joshi, U., Kondragunta, C., Joshi, C.E., Singhal, N., De Marzo, A.M. and Lotan, T.L. Comparison of pathologist and deep learning-based prostate cancer grading for prediction of metastatic outcomes in primary prostate cancer. **2024.**
 - [7] Redekop, E., Pleasure, M., Wang, Z., Sisk, A., Zong, Y., Flores, K., Speier, W. and Arnold, C.W. Digital Volumetric Biopsy Cores Improve Gleason Grading of Prostate Cancer Using Deep Learning. **2024.** ArXiv, pp.arXiv-2409.
 - [8] Kong, F., Wang, X., Xiang, J., Yang, S., Wang, X., Yue, M., Zhang, J., Zhao, J., Han, X., Dong, Y. and Zhu, B. Federated attention consistent learning models for prostate cancer diagnosis and Gleason grading.**2024.** *Computational and Structural Biotechnology Journal*, 23, pp.1439-1449.
 - [9] Faryna, K., Tessier, L., Retamero, J., Bonthu, S., Samanta, P., Singhal, N., Kammerer-Jacquet, S.F., Radulescu, C., Agosti, V., Collin, A. and Farre, X. Evaluation of artificial intelligence-based Gleason grading algorithms “in the wild”.**2024** *Modern Pathology*, 37(11), p.100563.
 - [10] Arvaniti, E., Fricker, K.S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., Wey, N., Wild, P.J., Rueschoff, J.H. and Claassen, M. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. **2018.** *Scientific reports*, 8(1), p.12054.
 - [11] John, J., Ravikumar, A. and Abraham, B. Prostate cancer prediction from multiple pretrained computer vision model.**2021.** *Health and Technology* , 11(5), pp.1003-1011.
 - [12] Bulten, W., Kartasalo, K., Chen, P.H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D.F., Van Boven, H., Vink, R. and Hulsbergen-Van De Kaa, C. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge.**2022.** *Nature medicine*, 28(1), pp.154-163.
 - [13] Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks.**2019, May.** *In International conference on machine learning* (pp. 6105-6114). PMLR.
 - [14] Bhattacharyya, R., Roy, P., Banerji, S. and Mitra, S. Efficient grading of prostate cancer WSI with deep learning.**2024, April.** *In Medical Imaging 2024: Digital and Computational Pathology Vol. 12933*, pp. 20-2. SPIE.
 - [15] Balaha, H.M., Shaban, A.O., El-Gendy, E.M. and Saafan, M.M. Prostate cancer grading framework based on deep transfer learning and Aquila optimizer.**2024.**

- [16] Linkon, A.H.M., Labib, M.M., Hasan, T. and Hossain, M. Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: An extensive study.**2021**. *Informatics in Medicine Unlocked*, 24, p.100582.
- [17] Shrestha, S., Alsadoon, A., Prasad, P.W.C., Seher, I. and Alsadoon, O.H. A novel solution of using deep learning for prostate cancer segmentation: enhanced batch normalization.**2021**. *Multimedia Tools and Applications*, 80(14), pp.21293-21313.
- [18] Korevaar, S., Tennakoon, R., Page, M., Brothie, P., Thangarajah, J., Florescu, C., Sutherland, T., Kam, N.M. and Bab-Hadiashar, A. Incidental detection of prostate cancer with computed tomography scans.**2021** *Scientific Reports*, 11(1), p.7956.
- [19] Liu, Z., Yang, C., Huang, J., Liu, S., Zhuo, Y. and Lu, X. Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer.**2021**. *Future Generation Computer Systems*, 114, pp.358-367.
- [20] Hao, R., Namdar, K., Liu, L., Haider, M.A. and Khalvati, F. A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted MRI using convolutional neural networks.**2021**. *Journal of Digital Imaging*, 34(4), pp.862-876.
- [21] Sobiecki, P., Jóźwiak, R., Sklinda, K. and Przelaskowski, A. Effect of domain knowledge encoding in CNN model architecture—a prostate cancer study using mpMRI images.**2021**. *PeerJ*, 9, p.e11006.
- [22] Chahal, E.S., Patel, A., Gupta, A., Purwar, A. and G, D. Unet based xception model for prostate cancer segmentation from MRI images.**2022**. *Multimedia Tools and Applications*, 81(26), pp.37333-37349.
- [23] Cipollari, S., Guarrasi, V., Pecoraro, M., Bicchetti, M., Messina, E., Farina, L., Paci, P., Catalano, C. and Panebianco, V. Convolutional neural networks for automated classification of prostate multiparametric magnetic resonance imaging based on image quality.**2022**. *Journal of Magnetic Resonance Imaging*, 55(2), pp.480-490.
- [24] Patel, A., Singh, S.K. and Khamparia, A. January. Detection of prostate cancer using deep learning framework.**2021**, *In IOP Conference Series: Materials Science and Engineering Vol. 1022*, No. 1, p. 012073). IOP Publishing.
- [25] Wessels, F., Schmitt, M., Kriehoff-Henning, E., Jutzi, T., Worst, T.S., Waldbillig, F., Neuberger, M., Maron, R.C., Steeg, M., Gaiser, T. and Hekler, A. Deep learning approach to predict lymph node metastasis directly from primary tumour histology in prostate cancer.**2021**. *BJU international*, 128(3), pp.352-360.

- [26] Khosravi, P., Lysandrou, M., Eljalby, M., Li, Q., Kazemi, E., Zisimopoulos, P., Sigaras, A., Brendel, M., Barnes, J., Ricketts, C. and Meleshko, D. A deep learning approach to diagnostic classification of prostate cancer using pathology–radiology fusion.**2021.** *Journal of Magnetic Resonance Imaging*, 54(2), pp.462-471.
- [27] Gentile, F., Ferro, M., Della Ventura, B., La Civita, E., Liotti, A., Cennamo, M., Bruzzese, D., Velotta, R. and Terracciano, D. Optimized identification of high-grade prostate cancer by combining different PSA molecular forms and PSA density in a deep learning model.**2021.** *Diagnostics*, 11(2), p.335.
- [28] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. Cbam: Convolutional block attention module.**2018.** *In Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).
- [29] Redekop, E., Sarma, K.V., Kinnaird, A., Sisk, A., Raman, S.S., Marks, L.S., Speier, W. and Arnold, C.W. Attention-guided prostate lesion localization and grade group classification with multiple instance learning.**2022, December.** *In International conference on medical imaging with deep learning* (pp. 975-987). PMLR.
- [30] Yang, B. and Xiao, Z. A multi-channel and multi-spatial attention convolutional neural network for prostate cancer ISUP grading.**2021.** *Applied Sciences*, 11(10), p.4321.
- [31] Pati, P., Jaume, G., Ayadi, Z., Thandiackal, K., Bozorgtabar, B., Gabrani, M. and Goksel, O. Weakly supervised joint whole-slide segmentation and classification in prostate cancer.**2023.** *Medical Image Analysis*, 89, p.102915.
- [32] Yang, H., Wu, G., Shen, D. and Liao, S. Automatic prostate cancer detection on multi-parametric mri with hierarchical weakly supervised learning. **2021, April.** *In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 316-319). IEEE.
- [33] Malyshev, V., Voynov, D. and Lapo, E. Computerized diagnosis of prostate cancer based on whole slide histology images and deep learning methods.**2021.**
- [34] Aldoj, N., Lukas, S., Dewey, M. and Penzkofer, T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network.**2020.** *European radiology*, 30(2), pp.1243-1253.
- [35] Wang, J., Chen, R.J., Lu, M.Y., Baras, A. and Mahmood, F. Weakly supervised prostate tma classification via graph convolutional networks.**2020, April.** *In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 239-243). IEEE.
- [36] Hasan, M.R., Ibraheem, N., Rahman, M.E. and Tamanna, R. Artificial Intelligence Across the Prostate Cancer Pathway: Screening, Imaging, Pathology, and

Biomarkers.**2025**. Cureus, 17(11).

- [37] Zhang, X., Xiao, N., Liang, H., Li, P., Zhang, Y., Zhang, S., Zhou, B., Yao, S., Yang, Z. and Chen, J. December. Artificial intelligence-driven prostate cancer diagnosis: Enhancing accuracy and personalizing patient care. **2025** In *Urologic Oncology: Seminars and Original Investigations*. Elsevier.
- [38] Paik, I., Lee, G., Lee, J., Kwak, T.Y. and Ha, H.K. Artificial intelligence-driven digital pathology in urological cancers: current trends and future directions.**2025**. *Prostate International*.
- [39] Kabir, S., Sarmun, R., Al Saady, R.M., Vranic, S., Murugappan, M. and Chowdhury, M.E. Automating prostate cancer grading: a novel deep learning framework for automatic prostate cancer grade assessment using classification and segmentation., **2025**. *Journal of Imaging Informatics in Medicine*, pp.1-23.
- [40] Zaheer, A.N., Farhan, M., Min, G., Alotaibi, F.A. and Alnfai, M.M. Attention-enhanced hybrid U-Net for prostate cancer grading and explainability.**2025**. *Scientific Reports*, 15(1), p.34038.
- [41] Hopson, S., Mildon, C., Kubalek, C., Ebbert, J., Vance, R., Laverty, L., Urie, P. and Della Corte, D. PANDA-PLUS: Improved dataset of prostate whole slide images from PANDA Challenge with pixel-level expert annotations.**2026**. *Journal of Pathology Informatics*, 20, p.100540.
- [42] Mittmann, G., Laiouar-Pedari, S., Mehrtens, H.A., Haggenmüller, S., Bucher, T.C., Chanda, T., Gaisa, N.T., Wagner, M., Klamming, G.G., Rau, T.T. and Nepl, C. Pathologist-like explainable AI for interpretable Gleason grading in prostate cancer.**2025**. *Nature communications*, 16(1), p.8959.
- [43] de La Calle, C.M., Baras, A.S. and Lotan, T.L., 2026. Digital pathology-based artificial intelligence algorithms in prostate cancer: inside the ‘black box’. *BJU international*, 137(4), pp.596-604.
- [44] Zhu, C.Y., Qu, R., Dai, Y. and Yang, L. Current Applications and Future Directions of Artificial Intelligence in Prostate Cancer Diagnosis: A Narrative Review. *Current Oncology*,**2026**. 33(3), p.166.
- [45] Wang, X., Zhong, S., Fang, K., Du, Y. and Huang, J. Application and prospect of artificial intelligence in diagnostic imaging of prostate cancer.bf 2026. *npj Digital Medicine*.