

Biopsy Image Analysis Using Hybrid Transfer Learning for Prostate Cancer Gleason Grading

Dr. Sheshang Degadwala¹ Prof Dr Divya Midhun², Dr. Shakir Khan³

¹ Research Scholar ; ² Supervisor; ³ Co-Supervisor

Sheshang13@gmail.com, divya@lincoln.edu.my, sgkhancs@gmail.com

Abstract: Prostate cancer remains a leading cause of cancer-related mortality among men, emphasizing the importance of accurate diagnosis and grading to guide treatment decisions. The Gleason Grading System is the standard for assessing prostate cancer aggressiveness through biopsy images; however, manual evaluation is time-consuming and prone to variability among pathologists. To address these challenges, this study proposes a novel approach using a hybrid transfer learning model based on InceptionResNet. The model leverages the strengths of InceptionResNet's deep learning architecture, integrating pre-trained features with fine-tuning to adapt specifically to prostate biopsy images. Experimental results on a publicly available prostate cancer dataset demonstrate that the proposed model achieves superior accuracy, sensitivity, and specificity compared to existing methods. The findings highlight the model's ability to classify Gleason grades with reduced diagnostic errors, enhancing reliability and efficiency in clinical workflows. This approach holds significant potential for applications in computer-aided diagnosis, providing pathologists with a powerful tool for consistent and precise prostate cancer grading. The use of InceptionResNet in hybrid transfer learning establishes a promising direction for future research in medical image analysis, ultimately contributing to improved patient outcomes.

Keywords: Prostate cancer; Gleason grading; Prostate biopsy; InceptionResNet; Hybrid transfer learning

Introduction

The prostate cancer incidence maintains strong placement as one of the major cancers affecting men worldwide where it continues to be a principal contributor to mortality rates. The precise determination of diagnosis together with grading helps healthcare providers identify proper treatments which yield enhanced medical results. Microscopic analysis by trained pathologists using the Gleason Grading System represents the standard method for determining prostate cancer aggression levels. Trained pathologists perform manual examinations, but this assessment method remains both subject to human interpretation biases and takes considerable time while also showing inconsistent results between examiners. Pathologists face an urgent requirement for automated analysis systems which provide reliable prostate cancer grading while being efficient and highly accurate.

The recent advancements in artificial intelligence (AI) combined with deep learning create new opportunities for medical image analysis through transfer learning solutions that solve the issues of model generalization and data scarcity challenges. Pre-trained large-dataset patterns from models become adjustable through transfer learning so that they work on tasks with minimal annotated data. InceptionResNet stands out among contemporary architectures because it demonstrates outstanding capabilities for recognizing detailed characteristics in advanced image applications. These specific architectures present an opportunity to surpass traditional clinical process limitations while creating better diagnostic tools for prostate diagnosis. **Figure 1** shows biopsy results grouped as per Gleason grades to determine prostate cancer clinical severity. The figure presents benign tissue followed by different Gleason score stages from 6 (3+3) to 10 (5+5). Different colored highlighted areas indicate portions of interest that are being evaluated. Cancer aggressiveness increases along with Gleason score progression because the tissue structure becomes more chaotic. The graphical display stands as an essential tool for analyzing prostate cancer grades during histopathological examination.

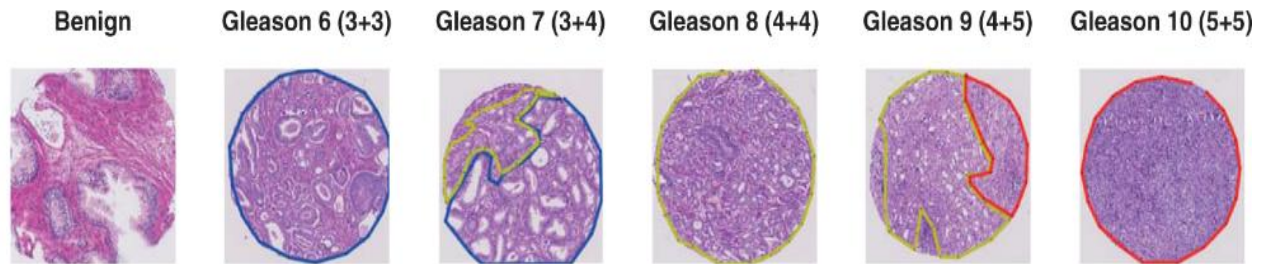


Figure 1. Prostate Cancer Gleason Grading

Researchers built a novel hybrid transfer learning framework using InceptionResNet architecture to enhance Gleason grade diagnostic reliability together with classification accuracy. The proposed model achieved effective adaptation to prostate cancer biopsy images through its combination of pre-trained features with task-specific fine-tuned adjustment that helped address the complex and varied nature of pathological grading. This method's superiority demonstrates through experimental testing on prostate cancer datasets presently available to the public for its future use as an advanced automated diagnostic system. Findings demonstrate how AI-driven solutions revolutionize prostate cancer management through diagnostic error reduction while streamlining clinical decisions to produce better patient outcomes.

Related work

Significant research exists on prostate cancer diagnosis together with Gleason Grading which relies on the Gleason Grading System for clinical practice. The time-consuming nature of subject-based grading combined with its individual variability drives researchers toward investigating automatic computational alternatives. Early investigations utilized statistical models alongside traditional machine learning approaches to classify prostate cancer aggressiveness though these methods struggled with feature extraction and generalization [1]. The combination of deep learning with convolutional neural networks (CNNs) and artificial intelligence (AI)-driven techniques has delivered significant improvements to both accuracy and operational efficiency for prostate cancer detection. Studies by Egevad et al. [1] show that Gleason scoring provides better prognostic value than grade groups therefore maintaining its essential role in diagnosis. The combination of transfer learning in a CNN-based classifier by Sariateş and Özbay [2] resulted in enhanced classification accuracy outcomes. Predicting prostate biopsy needs using machine learning involved the integration of PSA levels with MRI features and hematologic parameters according to Sungur et al. [3] in their work toward AI-based clinical management.

Research has investigated different forms of AI methods for prostate cancer grading improvement. Bottillo and his team researched pathogenetic variants of homologous recombination repair genes to develop early detection methods [4]. PCaseek represents a deep learning algorithm developed by Li et al. [5] to detect tumor DNA through urine tests for prostate cancer diagnosis and Gleason grading. A federated attention-consistent learning model developed by Kong et al. [6] resolved data privacy restrictions in multi-institutional artificial intelligence constructs. Reality-based virtual biopsy systems developed by Harder et al. [7] produce more efficient prostate cancer diagnosis utilizing MRI-targeting protocols. Using biopsy images Alici-Karaca and Akay [8] developed a deep learning system with exceptional performance for prostate cancer identification. The evaluation of deep learning-based Gleason grading methods by Patel et al. [9] showed CNN's central importance in automated diagnosis. Butt et al. [10] resolved Gleason grading inconsistencies through their creation of a multi-label ensemble CNN classifier which produced more reliable outcome predictions.

Recent progress in AI-driven classification programs improved the accuracy of prostate cancer identification devices. The LSTM-Deep Belief Network system proposed by Sethi et al. [11] achieved successful results in analyzing gene expression data for prostate cancer diagnosis. A combination of key radiomics features with a SVM algorithm enhanced classification accuracy according to Varan et al. [12]. An AI-based Gleason grading system achieved validation across multiple institutions by Tolkach et al. [13] which demonstrated its clinical readiness. Bulten et al. [14] took the lead in arranging the PANDA challenge where AI models competed for Gleason grading tasks using substantial multi-center data collections. The authors of Li et al. [15] developed a dual-attention feature autoencoder system to enhance the robustness of multimodal prostate cancer classification operations. A team of researchers led by Malibari [16] engineered an AI-based biomimetic imaging tool that showed excellent results during testing with multiple databases. Liu et al. [17] developed a deep learning system which detects prostate cancer risks within

patients who receive benign biopsy results. Hammouda et al. [18] created a deep learning pipeline that boosting diagnostic accuracy through Professor biopsy specimen digitization for grade group classification.

Table 1 Comparative Study of Different Methodology demonstrates the fast-growing role of artificial intelligence in prostate cancer screening and Gleason grading. The diagnostic approaches being studied right now show high reliability and accuracy, yet hybrid transfer learning evaluated in this research represents an exciting opportunity to enhance prostate cancer detection along with its clinical implementation.

Table 1. Comparative Study of Different Methodology with Datasets and Key Innovation

Category	Methodology	Dataset Used	Key Innovation
Gleason Scoring & Prognostic Analysis[1]	Clinical Gleason scoring analysis	Pathology data	Established superior prognostic value of Gleason scores
Deep Learning-Based Cancer Classification[2,8,10,14]]	Fine-tuned CNN with transfer learning, Efficient deep learning model, multi-label ensemble CNN classifiers, PANDA AI challenge	Public dataset, Biopsy images, Patch-level Gleason grading, multi-center dataset	Enhanced classification accuracy using hybrid CNN, achieved state-of-the-art classification performance, Reduced label inconsistencies in automated grading, Established global AI benchmark for Gleason grading
AI-Assisted Biopsy & Risk Prediction[3,17]	Machine learning with PSA, MRI, & hematologic markers, Deep learning for risk prediction	Patient dataset, Patient biopsy dataset	AI-assisted prediction of biopsy necessity, Identified prostate cancer risk despite benign biopsies
Genetic & Molecular Analysis for Early Detection[4]	Gene variant analysis	Genomic dataset	Identified pathogenic variants in homologous recombination repair genes
Non-Invasive Prostate Cancer Detection[5]	Deep learning-based urinary tumor DNA detection (PCaseek)	Patient urinary samples	Enabled non-invasive prostate cancer diagnosis
Federated Learning & Privacy-Preserving AI[6]	Federated attention-consistent learning	Multi-institutional dataset	Addressed privacy concerns in AI-based cancer grading
AI-Driven Virtual Biopsy & Imaging Analysis[7]	Virtual biopsy for MRI-targeted Gleason grading	Clinical MRI dataset	Optimized biopsy strategy using AI
Radiomics & Feature-Based AI Classification[12]	Fine-tuned SVM with key radiomics features	Clinical imaging dataset	Improved cancer classification using feature engineering
Multi-Institutional AI Validation[13]	Cross-center AI validation study, Dual-attention feature autoencoder	Large-scale dataset	Demonstrated robustness of AI models in multiple settings
Hybrid & Multi-Modal AI Approaches[15,16]	AI-based biomedical image classification	Multimodal dataset, public dataset	Integrated multimodal learning for robust classification, High-performance classification of prostate cancer
Deep Learning for Automated Grading[18]	AI pipeline for grade group classification	Digitized biopsy dataset	Improved precision in automated Gleason grading

Key Contribution

Figure 2 shows how deep neural networks powered by hybrid transfer learning perform the complete prostate cancer Gleason grading workflow. The initial step begins with data reading which obtains the resized training data from the Kaggle PANDA

dataset, the largest public prostate cancer histopathology repository available. Before entering the system, the images undergo pre-processing treatments where resizing happens to associate every image with 224x224x3 pixel dimensions followed by normalization which establishes standard pixel brightness levels for maintaining model input stability. Data augmentation methods transform images to achieve data balance which leads to model generalization improvements. The proposed dataset gets distributed into training and testing parts where training takes 80% while testing makes up the remaining 20% for effective evaluation.

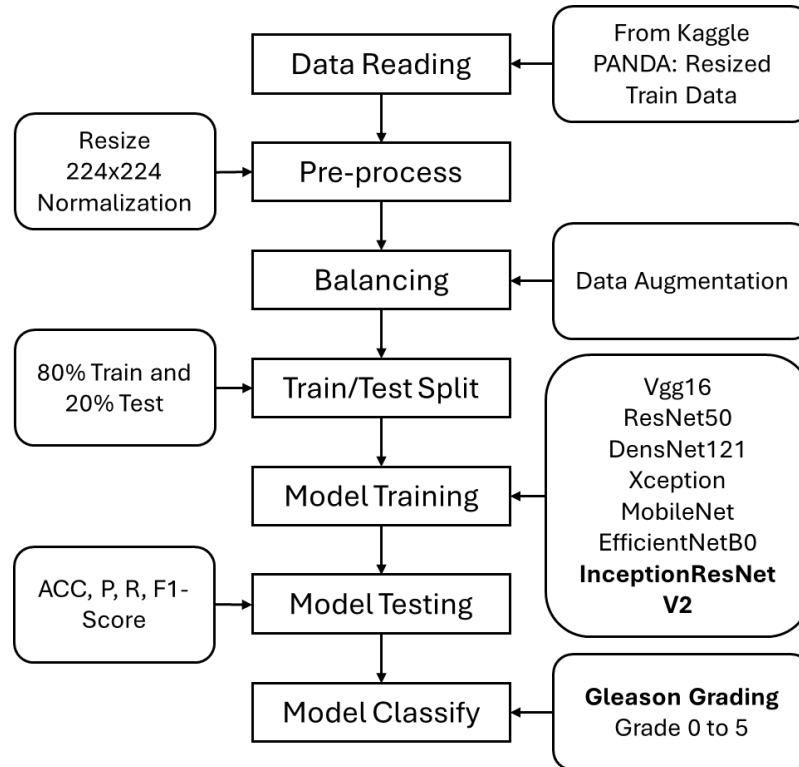


Figure 2. System Flow

Our proposed workflow centers on model training as we found multiple deep learning networks to discover the best solution for prostate cancer Gleason grading. Six distinct deep learning architectures were evaluated for this project: VGG16, ResNet50, DenseNet121, Xception and MobileNet and EfficientNetB0. These networks have distinct advantages for feature extraction and operational efficiency. Through extensive experimentation we chose InceptionResNetV2 implemented using TensorFlow Keras applications because this model effectively joins the Inception module's parallel convolution operations with ResNet50 shortcut connections to provide better gradient propagation alongside improved learning capacity and decreased gradient disappearance. The model's hybrid design allows it to detect complex prostate biopsy structures through advanced pattern analysis while running at an efficient computational rate.

A thorough evaluation of the proposed model employs established metrics including Accuracy (ACC) together with Precision (P) and Recall (R) and F1-score. The selection of these metrics allows for a reliable determination of model effectiveness in subclassifying Gleason grades ranging from 0 to 5. The training process utilizes augmented biopsies to improve both model generalization and accommodation of varying conditions. After optimization the deployed model enables automated prostate cancer grading which assists pathologists to perform accurate and consistent diagnostic tasks. The systematic deep learning pipeline achieves efficient histopathological image analysis and lowers observational discrepancies to provide an expandable and duplicable solution for assisted prostate cancer diagnosis.

Model: "sequential"

Layer (type)	Output Shape	Param #
inception_resnet_v2 (Functional)	(None, 5, 5, 1536)	54,336,736
dropout (Dropout)	(None, 5, 5, 1536)	0
flatten (Flatten)	(None, 38400)	0
batch_normalization_203 (BatchNormalization)	(None, 38400)	153,600
dense (Dense)	(None, 1024)	39,322,624
batch_normalization_204 (BatchNormalization)	(None, 1024)	4,096
activation_203 (Activation)	(None, 1024)	0
dropout_1 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 512)	524,800
activation_204 (Activation)	(None, 512)	0
dense_2 (Dense)	(None, 6)	3,078

Total params: 94,344,934 (359.90 MB)
Trainable params: 39,929,350 (152.32 MB)
Non-trainable params: 54,415,584 (207.58 MB)

Figure 3. InceptionResNetV2 Model Architecture

Method, Experiments and Results

Figure 4 shows the dataset has 250 samples of each isup_grade from 0-5. Figure 5 InceptionResNetV2 training and validation plots are in linear form, so it can say that the model is also stable. Figure 6 shows outstanding performance in its classification of diverse categories with a 93% total accuracy rate. Across the six prostate cancer Gleason classes the InceptionResNetV2 model consistently displays high precision and recall together with F1-scores which demonstrates its reliable performance in this task. Predictions demonstrate high accuracy because the model shows minimal class-specific misidentification.

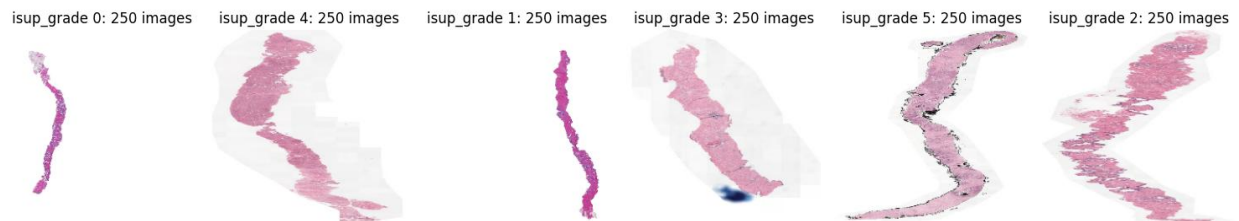


Figure 4. Dataset Reading

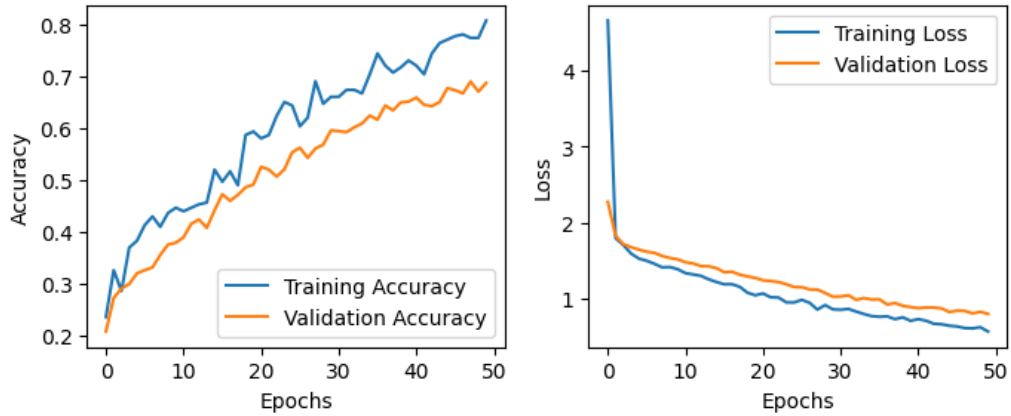


Figure 5. Training Model Plots

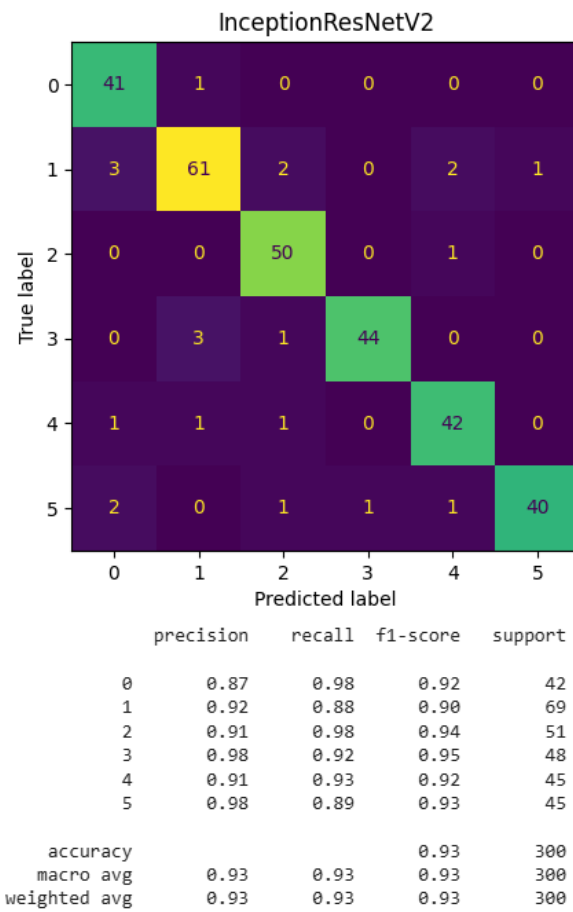


Figure 6. Evaluation Model Parameters

Table 2. Comparative Analysis of Transfer Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Vgg16	26%	26%	26%	26%
ResNet50	33%	34%	34%	33%
DensNet121	31%	30%	31%	31%
Xception	26%	26%	27%	26%
MobileNet	32%	32%	33%	32%

InceptionResNetV2	93%	93%	93%	93%
-------------------	-----	-----	-----	-----

Discussions

The biopsy image analysis **Table 2** with transfer learning models to evaluate prostate cancer Gleason grades produced varying performance outcomes for different architectural frameworks. The analysis confirmed InceptionResNetV2 as the best-performing model with 93% accuracy and consistent 93% precision and recall and F1-score levels. Recent research shows that advanced deep learning architectures and transfer learning excel when applied to complex medical image classification purposes including prostate cancer Gleason grade diagnosis.

Other candidate models displayed distinctively reduced performance levels. ResNet50 delivered 33% accuracy which outperformed DenseNet121 and MobileNet whose results came in at 31% and 32% accuracy respectively. Xception together with VGG16 provided the least successful results among all tested models because both models achieved only 26% accuracy levels. The experimental data suggests InceptionResNetV2 excels at detecting complex biopsy image features that guide precise Gleason grading assessments.

The significant differences in performance stem from structural variations combined with extraction strength and image pattern recognition capabilities in medical dimensions. The combination of InceptionResNetV2's Inception modules and its residual connections creates an architecture which captures spatial hierarchies effectively and sustains gradients during computations to produce top results in this application task. The performance of VGG16 and Xception becomes limited when analyzing this medical dataset because of inadequate architectural depth and generalization limitations.

Medical imaging applications depend strongly on choosing optimal models for implementation. The remarkable performance gains of InceptionResNetV2 demonstrate its potential for precise Gleason grade assessment in prostate cancer which benefits diagnostic decision protocols. Additional research should focus on optimizing performance through parameter adjustments along with dataset augmentation methods and ensemble techniques to enhance different architectural approaches.

Conclusions

This research tackles automated Gleason grading of prostate cancer through biopsy image analysis which remains essential for both diagnostic and treatment planning assessment. An objective solution to grading problems becomes possible through hybrid transfer learning techniques which boost diagnostic precision.

A total of six different pre-trained network models including VGG16, ResNet50, DenseNet121, Xception, MobileNet and InceptionResNetV2 were used through applied transfer learning to process biopsy image data. Models implemented the hybrid transfer learning method to reach peak performance through adaptation of existing models to Gleason grading data.

The model InceptionResNetV2 surpassed all other tested models by reaching an accuracy of 93% together with precision, recall, and F1-score of 93% when analyzing biopsy images due to its exceptional processing of complex texture patterns and spatial information in biomedical images. The performance of ResNet50 alongside DenseNet121 and MobileNet reached 30-33% accuracy although VGG16 and Xception exhibited only 26% accuracy across the studied models.

This study faces two main restrictions: insufficient hyperparameter tuning and absence of ensemble approaches that potentially could enhance final output results. The models achieved only limited success because of the restrictive nature of working with the dataset's small and universe collection. The lack of interpretability mechanisms to explain model predictions represents a significant barrier to clinical adoption of this research.

Enhancing both model generalization and robustness requires data collection efforts to concentrate on more comprehensive datasets. The implementation of advanced optimization methods together with ensemble techniques will optimize performance capability. The application of reinforcement learning mechanisms needs investigation to enhance effective clinical decisions in unpredictable patient scenarios. The successful implementation of interpretive AI methods into prediction platforms creates understandable outcomes that make them acceptable to medical practitioners.

References

1. L. Egevad *et al.*, "Gleason scores provide more accurate prognostic information than grade groups," *Pathology*, p. 104334, Jan. 2025. <https://doi.org/10.1016/j.pathol.2024.12.633>
2. M. Sariateş and E. Özbay, "A Classifier Model Using Fine-Tuned Convolutional Neural Network and Transfer Learning Approaches for Prostate Cancer Detection," *Applied Sciences (Switzerland)*, vol. 15, no. 1, 2025. <https://doi.org/10.3390/app15010225>

3. M. Sungur, A. Aykaç, M. E. Aydin, O. Celik, and C. Kaya, "Machine Learning-Based Prediction of Prostate Biopsy Necessity Using PSA, MRI, and Hematologic Parameters," *Journal of Clinical Medicine*, vol. 14, no. 1, 2025. <https://doi.org/10.3390/jcm14010183>
4. I. Bottillo *et al.*, "Early Detection of the Pathogenetic Variants of Homologous Recombination Repair Genes in Prostate Cancer: Critical Analysis and Experimental Design," *Biology*, vol. 14, no. 2, p. 117, Jan. 2025. <https://doi.org/10.3390/biology14020117>
5. G. Li *et al.*, "PCaseek: ultraspecific urinary tumor DNA detection using deep learning for prostate cancer diagnosis and Gleason grading," *Cell Discovery*, vol. 10, no. 1, pp. 1–3, 2024. <https://doi.org/10.1038/s41421-024-00710-y>
6. F. Kong *et al.*, "Federated attention consistent learning models for prostate cancer diagnosis and Gleason grading," *Computational and Structural Biotechnology Journal*, vol. 23, no. January, pp. 1439–1449, 2024. <https://doi.org/10.1016/j.csbj.2024.03.028>
7. C. Harder *et al.*, "Enhancing Prostate Cancer Diagnosis: Artificial intelligence-Driven Virtual Biopsy for Optimal Magnetic Resonance Imaging-Targeted Biopsy Approach and Gleason Grading Strategy," *Modern Pathology*, vol. 37, no. 10, p. 100564, 2024. <https://doi.org/10.1016/j.modpat.2024.100564>
8. D. Alici-Karaca and B. Akay, "An efficient deep learning model for prostate cancer diagnosis," *IEEE Access*, vol. 12, no. October, pp. 150776–150792, 2024. <https://doi.org/10.1109/ACCESS.2024.3472209>
9. M. Patel, P. Sanghani, and N. Shah, "Prostate Cancer Gleason Grading: A Review on Deep Learning Approaches for Recognizing," *ITM Web of Conferences*, vol. 65, p. 03013, 2024. <https://doi.org/10.1051/itmconf/20246503013>
10. M. A. Butt, M. F. Kaleem, M. Bilal, and M. S. Hanif, "Using multi-label ensemble CNN classifiers to mitigate labelling inconsistencies in patch-level Gleason grading," *PLoS ONE*, vol. 19, no. 7 July, pp. 1–24, 2024. <https://doi.org/10.1371/journal.pone.0304847>
11. B. K. Sethi, D. Singh, S. K. Rout, and S. K. Panda, "Long Short-Term Memory-Deep Belief Network-Based Gene Expression Data Analysis for Prostate Cancer Detection and Classification," *IEEE Access*, vol. 12, no. January, pp. 1508–1524, 2024. <https://doi.org/10.1109/ACCESS.2023.3346925>
12. M. Varan, J. Azimjonov, and B. Macal, "Enhancing Prostate Cancer Classification by Leveraging Key Radiomics Features and Using the Fine-Tuned Linear SVM Algorithm," *IEEE Access*, vol. 11, no. August, pp. 88025–88039, 2023. <https://doi.org/10.1109/ACCESS.2023.3306515>
13. Y. Tolkach *et al.*, "An international multi-institutional validation study of the algorithm for prostate cancer detection and Gleason grading," *npj Precision Oncology*, vol. 7, no. 1, 2023. <https://doi.org/10.1038/s41698-023-00424-6>
14. W. Bulten *et al.*, "Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge," *Nature Medicine*, vol. 28, no. 1, pp. 154–163, 2022. <https://doi.org/10.1038/s41591-021-01620-2>
15. B. Li, R. Oka, P. Xuan, Y. Yoshimura, and T. Nakaguchi, "Robust multi-modal prostate cancer classification via feature autoencoder and dual attention," *Informatics in Medicine Unlocked*, vol. 30, no. March, p. 100923, 2022. <https://doi.org/10.1016/j.imu.2022.100923>
16. A. A. Malibari, R. Alshahrani, F. N. Al-Wesabi, S. B. H. Hassine, M. A. Alkhonaini, and A. M. Hilal, "Artificial Intelligence Based Prostate Cancer Classification Model Using Biomedical Images," *Computers, Materials and Continua*, vol. 72, no. 2, pp. 3799–3813, 2022. <https://doi.org/10.32604/cmc.2022.026131>
17. B. Liu *et al.*, "Using deep learning to detect patients at risk for prostate cancer despite benign biopsies," *iScience*, vol. 25, no. 7, 2022. <https://doi.org/10.1016/j.isci.2022.104663>
18. K. Hammouda *et al.*, "A deep learning pipeline for grade groups classification using digitized prostate biopsy specimens," *Sensors*, vol. 21, no. 20, pp. 1–14, 2021. <https://doi.org/10.3390/s21206708>