International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Hybrid vision transformer model for accurate prostate cancer classification in MRI images

Farhan Bin Jashim, Fajle Rabbi Refat [1], Mohammad Hasnatul Karim [1], Farhad Uddin Mahmud [2] and Fariha Ashrafi [3,*]

[1] *Department of Business Administration and Management, International American University, CA 90010, USA.*
[2] *Department of Business Administration and Management in Information System, International American University, CA 90010, USA.*
[3] *Department of Information Technology, Westcliff University, CA 92614, USA.*

International Journal of Science and Research Archive, 2025, 15(02), 1505–1517

Publication history: Received on 08 April 2025; revised on 27 May 2025; accepted on 29 May 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.15.2.1509

Abstract

Prostate cancer remains one of the most prevalent malignancies among men globally, with early diagnosis complicated by its heterogeneous characteristics and the constraints of existing diagnostic approaches. This research introduces an advanced framework that integrates Convolutional Neural Networks (CNNs) with Vision Transformers (ViTs) to enhance the classification of prostate cancer using MRI scans. To mitigate class imbalance and improve generalization, we employed a combination of dual synthetic oversampling strategies along with data augmentation techniques. Our preprocessing workflow was designed to suppress image noise while maintaining edge integrity and enhancing local contrast without introducing artifacts. For robust feature representation, we extracted both Gray-Level Co-occurrence Matrix (GLCM) features and shape descriptors to capture the intricate patterns within the MRI data. Among the tested deep learning models, the ConvNeXt architecture delivered the highest performance. Specifically, using the SMOTE technique, it achieved an F1-score of 97.21% and a Matthews Correlation Coefficient (MCC) of 95.32%, while the application of ADASYN led to further gains, with an F1-score of 98.82% and an MCC of 97.86%. To support real-time clinical use, we also developed a web-based platform capable of analyzing prostate MRI scans interactively. These findings highlight the effectiveness and interpretability of our proposed method in facilitating accurate prostate cancer diagnosis.

Keywords: Prostate Cancer; Deep Learning; Vision Transformer; MRI; Cancer Informatics

1. Introduction

Prostate cancer originates in the prostate gland, a vital component of the male reproductive system [1]. It develops when abnormal genetic changes lead to uncontrolled cell proliferation, eventually forming tumors. These tumors can vary in behavior from indolent, asymptomatic growths to highly aggressive types that metastasize to distant sites such as bones and lymph nodes. According to statistics from the Global Cancer Observatory (2020), prostate cancer accounted for approximately 1.4 million new diagnoses globally, with around 375,000 associated deaths [2]. In the United States alone, the American Cancer Society estimated over 268,000 new cases and nearly 34,000 deaths in 2022. The rising prevalence of prostate cancer underscores the pressing need for improved diagnostic and screening technologies capable of identifying the disease at an early and more treatable stage.

Precise classification of prostate cancer plays a pivotal role in enhancing patient outcomes and prolonging survival [3]. Early-stage detection allows for the implementation of less invasive and more targeted treatment options, including

* Corresponding author: Fariha Ashrafi.

active surveillance, surgical removal of the prostate, or localized radiation therapy [4]. In contrast, delayed identification often results in disease progression to advanced stages, complicating treatment efforts and significantly diminishing survival prospects. Among current imaging modalities, Magnetic Resonance Imaging (MRI) stands out for its exceptional soft tissue contrast and multiparametric imaging capabilities, making it a key tool in the early detection and precise localization of prostate tumors [5].

Although MRI has significantly advanced prostate cancer diagnostics, several limitations continue to hinder its effectiveness. One major challenge is the reliance on radiologist expertise for interpreting prostate MRI scans, which introduces subjectivity into the diagnostic process. Even with structured frameworks like the Prostate Imaging Reporting and Data System (PI-RADS), variability in assessments among clinicians can lead to inconsistent evaluations and possible diagnostic errors. Additionally, manual analysis of MRI images is often labor-intensive, potentially delaying critical treatment decisions [6]. The heterogeneous nature of prostate tumors, reflected in a wide range of histopathological patterns categorized by the Gleason grading system, adds further complexity to accurate diagnosis and prognostication. These issues underscore the urgent demand for automated, rapid, and reliable diagnostic solutions that can enhance radiological assessments and support timely clinical interventions.

In recent years, transfer learning approaches have gained traction for automating medical image analysis, offering improvements in both diagnostic precision and reproducibility. Pretrained Convolutional Neural Networks (CNNs) have been extensively applied to image classification problems; however, their effectiveness can be constrained by the need for large, well-labeled datasets and their limited capacity to model long-range spatial dependencies. This challenge becomes more pronounced in medical imaging, where abnormal cases are often underrepresented, leading to significant class imbalance that can hinder training and generalization [7]. Vision Transformers (ViTs) present a promising alternative by efficiently modeling contextual relationships across the entire image, thereby enhancing generalization capabilities. Their attention-based architecture allows for more effective handling of imbalanced data distributions and offers improved robustness. Furthermore, the inherent interpretability of the self-attention mechanism in ViTs enables clearer understanding of which image regions influence model decisions, contributing to more transparent and explainable AI systems in clinical settings.

This research aims to design a Vision Transformer (ViT)-based framework for the precise classification of prostate cancer from MRI images. To overcome issues related to class imbalance and limited dataset size, we employed two synthetic oversampling techniques: the Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). These methods enhance the representation of underrepresented classes and contribute to a more balanced training process. Our approach incorporates hybrid deep learning architectures that fuse Convolutional Neural Networks (CNNs) with transformer-based attention mechanisms, aiming to improve both model interpretability and generalization capability. Furthermore, we explored multiple feature extraction strategies to optimize classification outcomes. The overarching objective is to deliver an effective and reliable tool for early prostate cancer detection. The complete methodological pipeline is depicted in Figure 1. Key contributions include:

- Introduced an innovative Vision Transformer-based framework that improves both classification accuracy and model interpretability by effectively combining localized feature extraction with global context modeling.
- Implemented advanced preprocessing techniques such as Non-Local Means (NLM) denoising and Contrast-Limited Adaptive Histogram Equalization (CLAHE) to improve image clarity while maintaining essential anatomical structures.
- Enhanced model generalization by applying a dual synthetic oversampling approach in combination with sophisticated data augmentation methods to mitigate class imbalance and dataset limitations.
- Built a web-based platform incorporating the top-performing proposed model to enable real-time prostate MRI analysis, facilitating prompt and informed clinical decision-making.

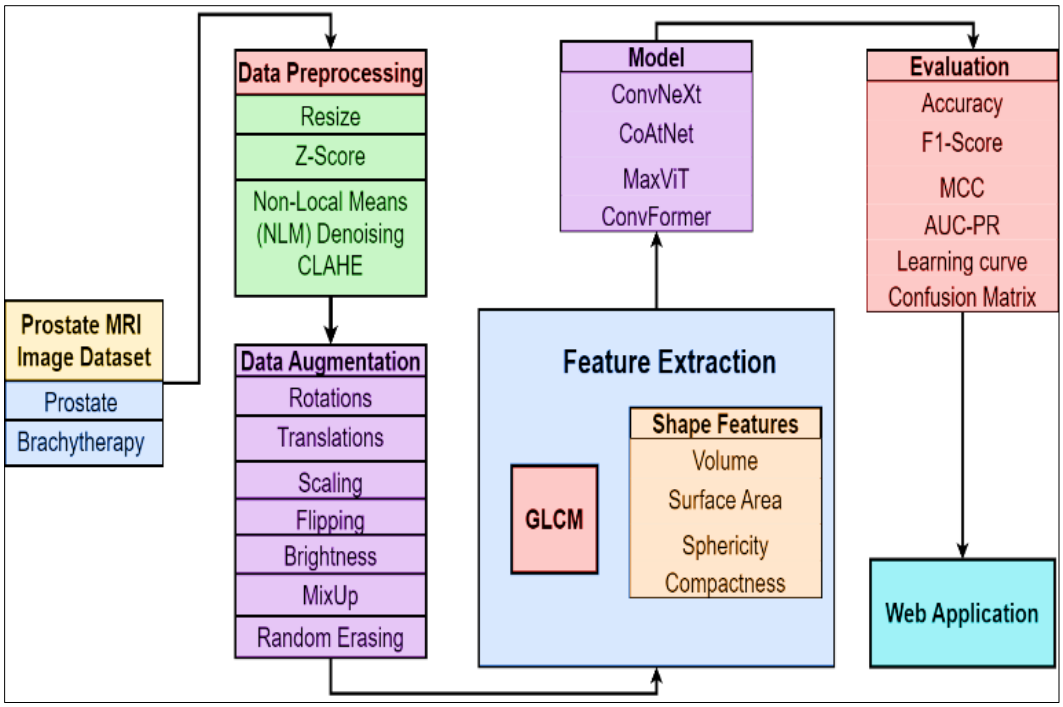

Figure 1 Overview of proposed methodology

The remainder of this paper is organized as follows: Section II reviews existing literature relevant to our research. Section III details the proposed methodology, encompassing preprocessing techniques and the design of the model architecture. Section IV showcases the experimental results and their corresponding analysis. Section V interprets the findings, highlights the study's limitations, and Section VI concludes the work while outlining directions for future investigation.

2. Related Work

Recent research has increasingly leveraged machine learning (ML) and deep learning (DL) techniques to enhance prostate cancer detection and classification. For instance, Ali et al. [8] introduced a two-dimensional CNN model applied to MRI scans, attaining an impressive AUC of 0.9993 following preprocessing to reduce image noise. Nonetheless, the model was designed for binary classification, which limits its utility in multiclass diagnostic scenarios. Moreover, the study's robustness remains uncertain due to the lack of validation across diverse datasets. In another effort, Malibari et al. [9] utilized EfficientNet for feature extraction and fuzzy K-nearest neighbors (FKNN) for classification, reporting an accuracy of 85.09%. Although the model showed promise, its dependence on a particular MRI dataset raises concerns regarding broader applicability. Singhal et al. [10] explored a deep learning approach for grading prostate cancer using Whole Slide Images (WSIs), achieving 89.4% accuracy on internal datasets and 83.1% on external ones. While these results are encouraging, the performance drop during external validation underscores the need for more robust domain adaptation techniques to ensure clinical reliability.

Alzboon et al. [11] employed a random forest algorithm trained on clinical and radiological data from a cohort of 400 patients, achieving a classification accuracy of 92% along with strong sensitivity and specificity metrics. Nevertheless, the absence of deep learning integration indicates room for further performance improvements. In a separate study, Salvi et al. [12] proposed a deep learning-based method utilizing immunohistochemical (IHC) staining and image segmentation, which attained a Dice Score of 90.36%. While effective, the model's dependency on a specific staining protocol could hinder its generalizability to other diagnostic imaging modalities. Gu et al. [13] introduced NAFNet, a deep learning framework designed to predict adverse pathology and biochemical recurrence-free survival (bRFS) from MRI scans. The model achieved an AUC of 0.915 and an accuracy of 85.0%, outperforming ResNet50. However, since the study focused exclusively on pre-treatment MRI data, its effectiveness in post-treatment or longitudinal follow-ups remains uncertain.

Zhao et al. [14] introduced a deep learning-based model aimed at identifying clinically significant prostate cancer (csPCa) using biparametric MRI (bpMRI). When integrated with the Prostate Imaging Reporting and Data System (PI-

RADS), the model exhibited strong specificity. However, its performance declined in one external validation cohort, indicating a need for further enhancements to ensure broader generalizability across diverse clinical datasets. In another study, Bygari et al. [15] designed a multistage deep learning pipeline for prostate cancer grading, utilizing a UNet architecture for image segmentation and combining Xception, ResNet-50, and EfficientNetB7 in an ensemble for grading. Trained on the Prostate Cancer Grade Assessment Challenge dataset, the model achieved a notable accuracy of 92.38%. Despite its success on a large dataset, additional external validation is necessary to assess its practical clinical utility. Additionally, Saiful et al. [16], [17] presented a highly optimized VGG-16 CNN framework for brain tumor classification, reaching an accuracy of 99.5% on a dataset of 6,328 MRI images spanning three tumor types. This approach outperformed earlier models in terms of precision and robustness.

A common limitation among many existing models is their reliance on small, homogeneous datasets, which restrict their applicability across varied patient populations and clinical environments. These models also tend to underperform when exposed to domain shifts, as discrepancies in data distributions often lead to noticeable drops in accuracy. Moreover, comprehensive evaluation metrics such as Matthews Correlation Coefficient (MCC) and Precision-Recall AUC (PR AUC) are frequently neglected, resulting in inflated assessments of model effectiveness. To overcome these challenges, advanced architectures like Vision Transformers (ViTs) are essential, as they excel at capturing global contextual patterns and adapting to diverse datasets. Furthermore, deploying ViT-based systems in real-time web applications can significantly improve diagnostic efficiency and streamline clinical decision-making.

3. Methodology

3.1. Data Description

This study utilized a publicly accessible dataset comprising 647 MRI scans. The data included 447 scans categorized as Prostate Cases which encompassed both benign prostatic hyperplasia and prostate cancer and 200 scans from Brachytherapy Cases involving patients receiving radiation treatment for prostate cancer. To ensure balanced representation across subsets, the dataset was partitioned into 80% for training, 5% for validation, and 15% for testing. Specifically, the training set included 358 Prostate Cases and 160 Brachytherapy Cases, while the validation set consisted of 23 Prostate Cases and 10 Brachytherapy Cases. The remaining 67 Prostate Cases and 30 Brachytherapy Cases were allocated to the testing set [18]. Figure 2 provides representative MRI samples from each category, illustrating the anatomical differences and treatment-specific characteristics.

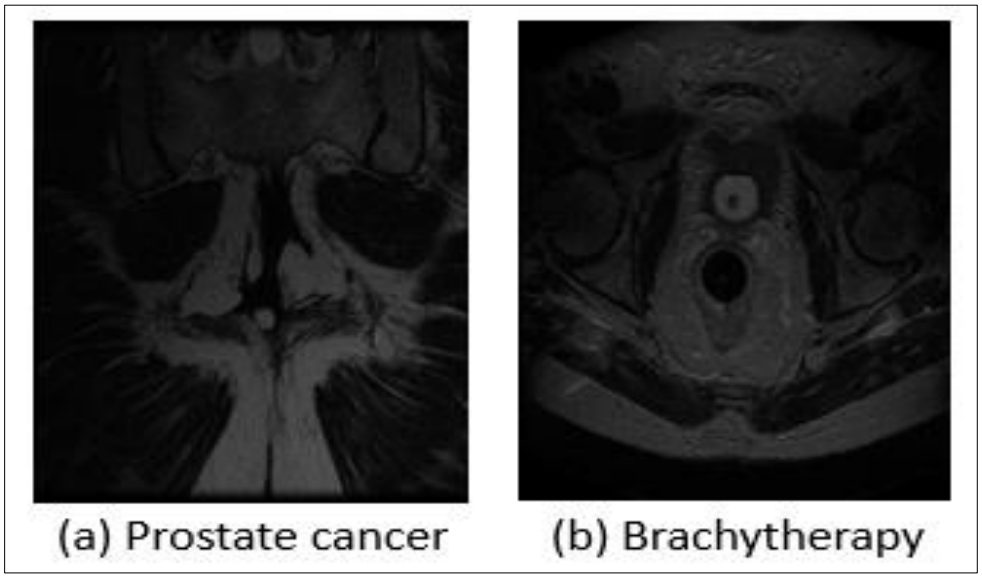

Figure 2 Sample PCa and BC images from the dataset

3.2. Data Pre-processing

To ensure uniformity across the dataset, all MRI images were resized to 224×224 pixels using bilinear interpolation [19]. Z-Score Normalization was then applied to standardize pixel intensity values, minimizing inconsistencies introduced by different imaging equipment and scanning protocols [20]. To enhance image quality, Non-Local Means (NLM) denoising was used to suppress background noise while retaining critical anatomical structures. Contrast-

Limited Adaptive Histogram Equalization (CLAHE) further improved local contrast without introducing noise amplification. To address the issue of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) were independently utilized, generating two distinct balanced training sets. This dual approach enabled comprehensive evaluation of oversampling strategies and facilitated effective learning from both uniformly and adaptively balanced data [21]. To increase the variability and robustness of the training dataset, we applied a range of augmentation techniques. These included rotations of ±15 degrees to simulate different orientations, translations of ±10% to represent spatial shifts, and scaling within the range of 0.9 to 1.1 to account for size variations. Horizontal and vertical flipping was also performed to introduce mirrored perspectives. Additionally, brightness adjustments of ±10% were used to emulate varying illumination conditions [22]. MixUp augmentation was incorporated to blend image-label pairs, promoting generalization. Lastly, Random Erasing was employed to randomly mask regions within the images, helping the model better handle occlusions and missing data.

3.3. Feature Extraction

3.3.1. Gray-Level Co-occurrence Matrix (GLCM)

This method is used to characterize the spatial relationships between voxel intensity values, effectively capturing the textural complexity found in prostate MRI scans [23]. By analyzing the frequency of voxel intensity pairs that occur in specific spatial arrangements, it uncovers structural patterns that may indicate pathological tissue changes. For an image with M gray levels, the Gray-Level Co-occurrence Matrix (GLCM) is constructed, as shown in Equation (1). In this context, N(i, j) refers to the count of voxel pairs with intensity levels i and j, occurring at a predetermined distance and angle. The matrix is then normalized by the total number of these voxel pairs, N, ensuring that the resulting features are scale-invariant and consistent across different image dimensions.

$$C(i,j) = \frac{N(i,j)}{N} \tag{1}$$

3.3.2. Several statistical features are derived to assess textural properties

The feature Energy (E), defined in Equation (2), assesses the uniformity of the voxel pair distribution within the GLCM, with higher values indicating more homogeneous textures. Entropy (H), shown in Equation (3), captures the degree of randomness or complexity in the intensity distribution—higher entropy reflects greater heterogeneity. Contrast (C), as presented in Equation (4), measures the difference in intensity between a voxel and its neighboring voxels, offering insight into edge strength and local variation. Lastly, Homogeneity (H⇳) in Equation (5) evaluates how closely the elements in the GLCM are distributed toward its diagonal, with higher values indicating more uniform or smooth textures [24].

$$E = \sum_{i=1}^{M}\sum_{j=1}^{M}[C(i,j)]^2 \tag{2}$$

$$\mathcal{H} = -\sum_{i=1}^{M}\sum_{j=1}^{M}C(i,j)\log\big(C(i,j)\big) \tag{3}$$

$$\mathcal{C} = \sum_{i=1}^{M}\sum_{j=1}^{M}(i-j)^2C(i,j) \tag{4}$$

$$\mathcal{H}_m = \sum_{i=1}^{M}\sum_{j=1}^{M}\frac{C(i,j)}{1+|i-j|} \tag{5}$$

3.3.3. Shape Features

This approach captures the geometric characteristics of the prostate, providing valuable insights into its structural integrity. Accurate segmentation of the prostate boundaries allows for the extraction of key morphological features such as Volume (V), Surface Area (S), Sphericity (Ψ), and Compactness (C), which can help identify potential abnormalities [25]. Volume, as defined in Equation (6), refers to the total count of voxels (N_v) identified as prostate tissue, serving as a measure of gland size. Surface area, calculated using Equation (7), quantifies the extent of the prostate's boundary. Sphericity, outlined in Equation (8), indicates how closely the prostate's shape approximates that of a perfect sphere,

while Compactness, described in Equation (9), assesses the regularity and smoothness of the shape. By integrating these shape-based descriptors with texture features derived from the Gray Level Co-occurrence Matrix (GLCM), the model achieves a comprehensive understanding of both internal tissue variations and the prostate's overall anatomical form.

$$V = \sum_{v=1}^{N_v} 1 \tag{6}$$

$$S = \sum_{s=1}^{N_s} A(s) \tag{7}$$

$$\Psi = \frac{\pi^{1/3}(6V)^{2/3}}{S} \tag{8}$$

$$C = \frac{S}{V^{2/3}} \tag{9}$$

3.4. Model Training

3.4.1. ConvNeXt

ConvNeXt is a modern architecture inspired by transformers that enhances traditional convolutional neural networks (CNNs) by integrating larger convolutional kernels and incorporating layer normalization [26]. It also eliminates unnecessary components, such as fully connected layers. These improvements facilitate more efficient processing of high-resolution medical images, like prostate MRI scans, without sacrificing predictive performance. In ConvNeXt, the convolution operation involves an input feature map (X), a convolutional kernel (W), and a bias term (b). The notation (z_i) denotes the logit corresponding to class (i), while (n) represents the total number of output classes. By combining the robustness of CNNs with streamlined design principles, ConvNeXt is particularly effective at extracting detailed and localized features from complex medical images. The use of larger kernels increases the receptive field, allowing the network to capture both fine-grained details and broader contextual patterns within the image simultaneously.

$$\mathrm{P}(\mathrm{y_i}|\mathrm{X}) = \frac{\mathrm{e}^{\mathrm{z_i}}}{\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{e}^{\mathrm{z_j}}} \tag{10}$$

$$\mathrm{Y} = \mathrm{Conv}(\mathrm{X}, \mathrm{W}) + \mathrm{b} \tag{11}$$

3.4.2. CoAtNet

This hybrid model architecture combines convolutional layers with transformer-based self-attention mechanisms, taking advantage of both approaches. Convolutional layers effectively extract localized features, such as textures and edges, while self-attention modules analyze interactions between different spatial positions. This enables the model to learn long-range dependencies and understand global contextual information. As described in Equation (12), the self-attention mechanism uses query (Q), key (K), and value (V) matrices, where ($\mathrm{d_k}$) represents the dimensionality of the key vectors. By fusing local feature extraction with global reasoning, this model enhances its ability to interpret complex medical images. This makes it particularly well-suited for identifying subtle abnormalities and recognizing broader anatomical patterns that are crucial in prostate cancer diagnosis.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{12}$$

3.4.3. MaxViT

MaxViT improves multi-scale feature learning by dividing an image into grids and applying self-attention mechanisms both within and across these regions. This method enhances image classification accuracy by capturing features at various spatial scales. The architecture combines convolutional operations with multi-axis attention derived from transformer models, utilizing a grid-based attention mechanism to efficiently extract both local and global patterns. As illustrated in Equation (13), local spatial features are captured using convolutional filters (F) along with a bias term (c), which allows the model to concentrate on fine-grained details critical for identifying subtle variations in prostate MRI

scans. After the convolution process, MaxViT employs multi-axis attention (as shown in Equation 14) by utilizing query, key, and value matrices, where (d_m) represents the dimensionality of the key vectors. This attention mechanism enables the modeling of long-range dependencies across the image, allowing the network to integrate fine details with broader anatomical context, thereby enhancing the robustness of cancer detection.

$$Z = \text{Conv2D}(A, F) + c \tag{13}$$

$$\text{GridAttention}(Q, V, K) = \text{Concat}_s\text{maxQVTdmW} \tag{14}$$

3.4.4. ConvFormer

The architecture utilizes compound scaling to systematically adjust the network's width, depth, and input resolution, ensuring optimal performance across various image sizes. This scaling mechanism, illustrated in Equation (15), involves applying convolutional filters (C_i) and a bias term (c) to the input feature map (M). This allows the model to extract localized spatial features that are crucial for identifying subtle irregularities in prostate tissue. ConvFormer enhances its functionality with a multi-head self-attention mechanism that captures global dependencies by computing multiple attention heads—each focusing on different regions of the input. These outputs are then aggregated through an output projection matrix (W_O). After the attention stage, the model processes the features through a feed-forward network (FFN), which includes weight matrices (W_1) and (W_2), along with corresponding biases (b_1) and (b_2), as defined in Equation (17). This step further refines the representations learned. To ensure stability during training, layer normalization is applied. Equation (18) employs the mean (μ), standard deviation (σ), and learnable parameters (γ) and (β) for scaling and shifting the normalized data.

$$M' = \sum_{i=1}^{K} (M * C_i) + c \tag{15}$$

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{16}$$

$$\text{LayerNorm}(x) = \frac{x - \mu}{\sigma} \cdot \gamma + \beta \tag{17}$$

3.5. Training Procedure

The model was trained using the AdamW optimizer, which is a variant of the Adam optimizer [27]. AdamW decouples weight decay from gradient updates, enhancing regularization and promoting more stable convergence during training. A learning rate scheduler was included to adaptively adjust the learning rate based on the model's performance. Specifically, it monitored the validation loss and reduced the learning rate when a plateau was detected [28]. This allowed the network to fine-tune its parameters more effectively. To prevent overfitting, early stopping was implemented. Training was terminated if the validation loss did not improve over a predefined number of epochs. This strategy helps maintain the model's generalization by avoiding excessive adaptation to training noise. The training configuration used a batch size of 32 and was conducted over 50 epochs [23].

3.6. Evaluation

To evaluate the model’s effectiveness, we employed several metrics: accuracy, F1-score, Matthews Correlation Coefficient (MCC), and the Area Under the Precision-Recall Curve (AUC-PR). Accuracy reflects the overall proportion of correctly classified instances, while the F1-score offers a balanced measure of precision and recall, making it particularly useful in scenarios with class imbalance [28], [29]. MCC provides a comprehensive assessment by incorporating true and false positives and negatives, yielding values between -1 and +1, with higher values indicating better binary classification performance. AUC-PR emphasizes performance on the positive class and captures the trade-off between precision and recall in imbalanced datasets [7], [30]. Additionally, confusion matrices were used to visualize the alignment between predicted and actual labels, helping identify specific misclassification trends[31]. Learning curves were also analyzed to track training and validation loss across epochs, providing insights into the model’s generalization capability and helping to identify signs of overfitting or underfitting [32].

4. Results analysis

4.1. Comparative Analysis of Performance

Table 1 summarizes the classification performance of the proposed models using SMOTE and ADASYN as oversampling techniques. Among the evaluated architectures, ConvNeXt consistently outperformed the others across both balancing strategies. With SMOTE, ConvNeXt achieved a peak accuracy of 98.75%, an F1-score of 97.21%, and an MCC of 95.32%, reflecting its strong overall classification capability. The model also attained an AUC-PR of 99.25%, highlighting its effectiveness in handling class imbalance. CoAtNet and MaxViT followed, with accuracies of 96.42% and 96.07%, respectively, though their F1-scores and MCCs were slightly lower. ConvFormer, while still reasonably effective, yielded the lowest performance under SMOTE, recording 94.40% accuracy, an F1-score of 92.90%, and an AUC-PR of 94.20%.

When switching to ADASYN, performance improved across all models. ConvNeXt again led with an impressive accuracy of 99.48%, an F1-score of 98.82%, and the highest AUC-PR at 99.86%. MaxViT also showed significant gains, achieving an F1-score of 96.66% and accuracy of 97.88%. Although CoAtNet and ConvFormer improved as well, they remained behind ConvNeXt, with accuracy scores of 97.34% and 96.93%, respectively.

Table 1 Results of our experimental classifiers

Type	Model	Accuracy	F1	MCC	AUC-PR
	ConvNeXt	98.75%	97.21%	95.32%	99.25%
SM	CoAtNet	96.42%	94.20%	95.65%	97.42%
	MaxViT	96.07%	94.45%	94.80%	96.12%
	ConvFormer	94.40%	92.90%	91.50%	94.20%
	ConvNeXt	99.48%	98.82%	97.86%	99.86%
AD	CoAtNet	97.88%	96.66%	94.67%	98.61%
	MaxViT	97.34%	96.48%	93.70%	97.41%
	ConvFormer	96.93%	95.78%	90.42%	95.81%

4.2. Performance Validation

As shown in the confusion matrix in Figure 3, the ConvNeXt model trained using the ADASYN oversampling method demonstrates high classification accuracy, correctly identifying 66 out of 67 prostate cancer (PCa) cases. It also performs strongly in classifying benign conditions (BC), with 30 out of 31 cases correctly predicted. Minor misclassifications occurred, including one PCa instance labeled as BC and one BC case identified as PCa. Despite these isolated errors, the model exhibits strong overall performance in differentiating between malignant and benign cases. Additionally, Figure 4 illustrates the learning curves for the ConvNeXt model, showing a steady and parallel decline in training and validation loss. The minimal gap between the two curves suggests effective learning and strong generalization, with no signs of overfitting during the training process.

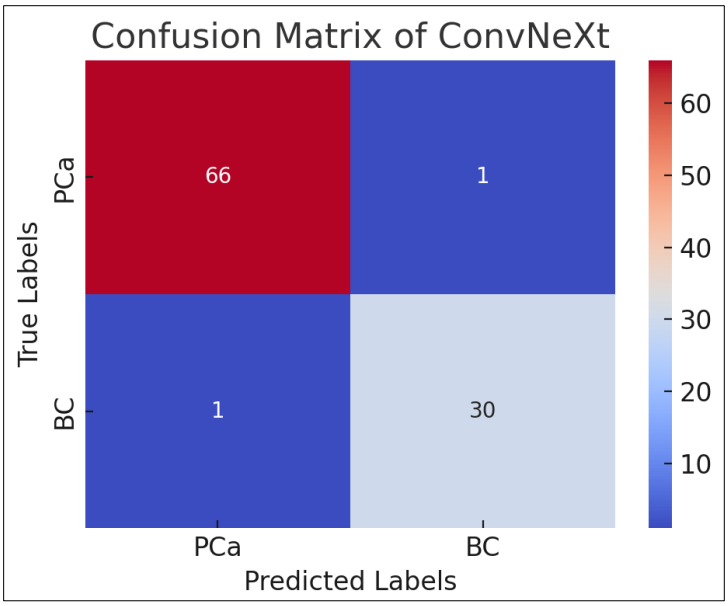

Figure 3 Confusion matrix of the ConvNeXt model

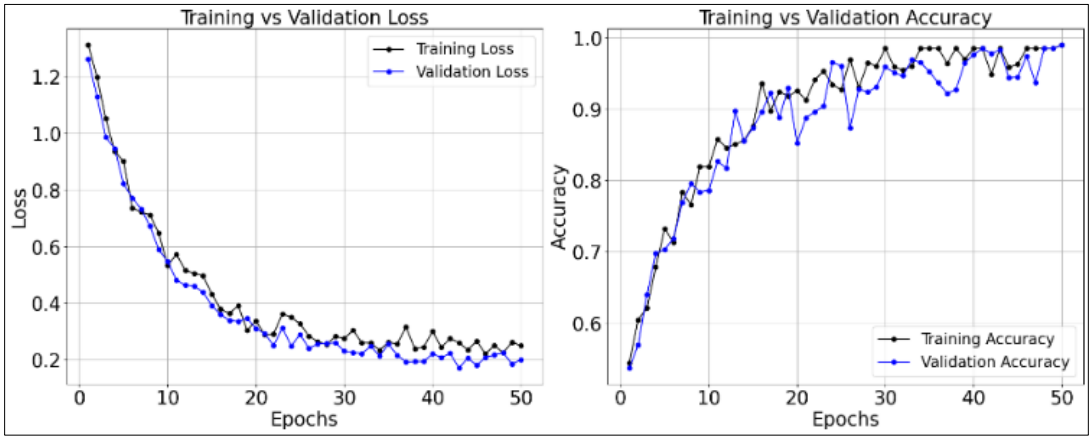

Figure 4 Learning curve of the ConvNeXt model

4.3. Web Application

Figure 5 illustrates a web-based application designed for the classification of prostate cancer using MRI images. The interface adopts a clean and professional aesthetic, featuring a white background complemented by blue accents to convey clarity and clinical precision. At the top of the application, a navigation bar includes links to essential sections such as “Input,” “Output,” and “About,” alongside a prominent title and icon that reinforce the platform's identity and purpose. On the left side of the interface, users are provided with a straightforward image upload module labeled “Upload Image,” accompanied by a cloud icon to signify the action clearly. Once an MRI scan is uploaded, it is visually rendered in the “MRI Image” display panel, allowing users to verify the input before proceeding with analysis. Positioned at the center is a large, easily accessible “CLASSIFY” button, streamlining the user workflow for initiating the model’s prediction. To the right of the image panel, the application displays the classification result. In this instance, the diagnosis is “Prostate Cancer” with a high confidence score of 98.8%, indicating the model’s strong certainty in its prediction. This immediate feedback provides actionable insights for radiologists and clinicians, potentially aiding in timely intervention and treatment planning. Below the classification output, a confusion matrix visualizes the model’s performance in detail. The matrix shows that out of the test samples, 66 prostate cancer cases were correctly identified, while 30 benign cases were accurately classified. Only one case was misclassified in each category, reflecting the model’s excellent diagnostic precision. This matrix is a vital component, offering a quick overview of true positives, true negatives, and rare misclassifications.

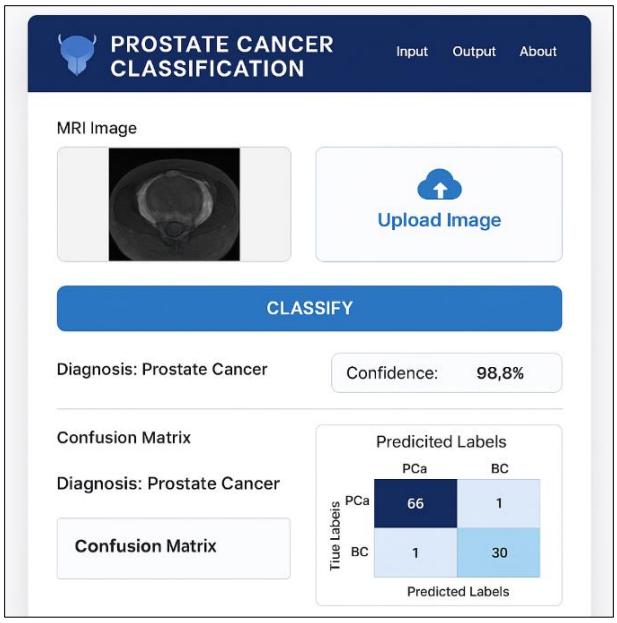

Figure 5 Web Application

5. Discussion

Among the evaluated models, ConvNeXt demonstrated superior performance in classifying prostate cancer cases, attributed to its refined convolutional structure that effectively captures detailed, localized patterns. Its transformer-inspired enhancements facilitate robust feature extraction while maintaining computational efficiency, making it particularly adept at distinguishing subtle variations within prostate MRI images. In terms of data balancing, models trained with ADASYN oversampling outperformed those using SMOTE, as ADASYN focuses on generating synthetic examples for harder-to-learn minority cases. This approach improved model generalization, especially in detecting less common prostate cancer variants. In contrast, SMOTE's uniform sample generation was less effective in addressing this complexity. Additionally, integrating GLCM-based texture features with geometric shape descriptors significantly strengthened the model's discriminative capability by providing a more holistic representation of both internal tissue structure and anatomical form. The deployment of the model in a web-based application offers clinical advantages, enabling early diagnosis, individualized treatment strategies, and reduced workload for radiologists.

However, several limitations were identified. The relatively small dataset may affect the model's ability to generalize across broader and more diverse patient populations. Reliance on synthetic oversampling methods may introduce artificial biases that do not reflect real-world distributions. Moreover, the computational requirements of hybrid CNN-transformer models pose challenges for deployment in low-resource settings. These constraints underscore the importance of expanding the dataset, conducting external validations, and optimizing model efficiency in future studies.

6. Conclusion

In this study, a Vision Transformer-based framework was developed for the classification of prostate cancer using MRI scans. Among the evaluated models, ConvNeXt demonstrated superior performance, offering improved accuracy and generalization compared to existing approaches. Key contributions include addressing class imbalance through ADASYN-based oversampling, employing advanced feature extraction techniques, and enhancing the classification of less common prostate cancer subtypes. The deployment of a web-based diagnostic tool further highlights the clinical applicability of the proposed system, offering radiologists a reliable and efficient aid for early detection and cancer staging.

Despite its strengths, the approach presents certain limitations, including the computational cost associated with Vision Transformers and the potential for artificial noise introduced by synthetic oversampling techniques. To advance the clinical utility of such systems, future work should focus on integrating multimodal medical data, incorporating Explainable AI (XAI) to enhance interpretability, and optimizing model efficiency for deployment in real-world, resource-constrained environments.

Compliance with ethical standards

Disclosure of conflict of interest

There is not conflict of interests.

Statement of ethical approval

The present research work does not contain any studies performed on animals/humans subjects by any of the authors'.

References

[1] Gogola S, Rejzer M, Poppiti R. Prostate gland anatomy and hormonal factors contributing to cancer development. Therapy Resistance in Prostate Cancer: Mechanisms and Insights. 2024 Jan 1;1–26.

[2] De Silva F, Alcorn J. A Tale of Two Cancers: A Current Concise Overview of Breast and Prostate Cancer. Cancers 2022, Vol 14, Page 2954 [Internet]. 2022 Jun 15 [cited 2025 Feb 2];14(12):2954. Available from: https://www.mdpi.com/2072-6694/14/12/2954/htm.

[3] Wildeboer RR, van Sloun RJG, Wijkstra H, Mischi M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. Comput Methods Programs Biomed. 2020 Jun 1;189:105316.

[4] Haque R, Al Sakib A, Hossain MF, Islam F, Ibne Aziz F, Ahmed MR, et al. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. BioMedInformatics 2024, Vol 4, Pages 966-991 [Internet]. 2024 Apr 1 [cited 2025 May 13];4(2):966–91. Available from: https://www.mdpi.com/2673-7426/4/2/54/htm.

[5] Huang X, Li Z, Zhang M, Gao S. Fusing hand-crafted and deep-learning features in a convolutional neural network model to identify prostate cancer in pathology images. Front Oncol. 2022 Sep 27;12:994950.

[6] Ahmed MR, Haque R, Rahman SMA, Afridi S, Abir MFF, Hossain MF, et al. Towards Automated Detection of Tomato Leaf Diseases. Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024. 2024;387–92.

[7] Al Noman A, Fardin H, Chhabra G, Sultana S, Haque R, Ahmed MR, et al. Monkeypox Lesion Classification: A Transfer Learning Approach for Early Diagnosis and Intervention. Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024. 2024;247–54.

[8] Ali AM, Mohammed AA. Improving classification accuracy for prostate cancer using noise removal filter and deep learning technique. Multimed Tools Appl [Internet]. 2022 Mar 1 [cited 2025 May 13];81(6):8653–69. Available from: https://link.springer.com/article/10.1007/s11042-022-12102-z.

[9] Huang TL, Lu NH, Huang YH, Twan WH, Yeh LR, Liu KY, et al. Transfer learning with CNNs for efficient prostate cancer and BPH detection in transrectal ultrasound images. Scientific Reports 2023 13:1 [Internet]. 2023 Dec 9 [cited 2025 May 13];13(1):1–9. Available from: https://www.nature.com/articles/s41598-023-49159-1.

[10] Singhal N, Soni S, Bonthu S, Chattopadhyay N, Samanta P, Joshi U, et al. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. Scientific Reports 2022 12:1 [Internet]. 2022 Mar 1 [cited 2025 May 13];12(1):1–11. Available from: https://www.nature.com/articles/s41598-022-07217-0.

[11] Manalu DR, Sitompul OS, Mawengkang H, Zarlis M. Model Classification of Fire Weather Index using the SVM-FF Method on Forest Fire in North Sumatra, Indonesia. International Journal of Advanced Computer Science and Applications. 2023;14(8):329–37.

[12] Salvi M, Manini C, López JI, Fenoglio D, Molinari F. Deep learning approach for accurate prostate cancer identification and stratification using combined immunostaining of cytokeratin, p63, and racemase. Computerized Medical Imaging and Graphics. 2023 Oct 1;109:102288.

[13] Gu W jie, Liu Z, Yang Y jie, Zhang X zhi, Chen L yu, Wan F ning, et al. A deep learning model, NAFNet, predicts adverse pathology and recurrence in prostate cancer using MRIs. npj Precision Oncology 2023 7:1 [Internet]. 2023 Dec 11 [cited 2025 Feb 2];7(1):1–10. Available from: https://www.nature.com/articles/s41698-023-00481-x.

[14] Zhao L, Bao J, Qiao X, Jin P, Ji Y, Li Z, et al. Predicting clinically significant prostate cancer with a deep learning approach: a multicentre retrospective study. Eur J Nucl Med Mol Imaging [Internet]. 2023 Feb 1 [cited 2025 Feb 2];50(3):727–41. Available from: https://link.springer.com/article/10.1007/s00259-022-06036-9.

[15] Bygari R, Rithesh K, Ambesange S, Koolagudi SG. Prostate Cancer Grading Using Multistage Deep Neural Networks. Lecture Notes in Electrical Engineering [Internet]. 2023 [cited 2025 Feb 2];946:271–83. Available from: https://link.springer.com/chapter/10.1007/978-981-19-5868-7_21.

[16] Saiful M, Haider S, Rahman SMA, Reza N, Reza AW, Arefin MS. MRI-Based Brain Tumor Classification Using Various Deep Learning Convolutional Networks and CNN. Lecture Notes in Networks and Systems [Internet]. 2023 [cited 2025 Feb 2];729 LNNS:177–88. Available from: https://link.springer.com/chapter/10.1007/978-3-031-36246-0_17.

[17] Hasib Fardin, Hasan Md Imran, Hamdadur Rahman, Anamul Haque Sakib, Md Ismail Hossain Siddiqui. Robust and explainable poultry disease classification via MaxViT with attention-guided visualization. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1848–59. Available from: https://journalijsra.com/node/1054.

[18] Mohammad Rasel Mahmud, Al Shahriar Uddin Khondakar Pranta, Anamul Haque Sakib, Abdullah Al Sakib, Md Ismail Hossain Siddiqui. Robust feature selection for improved sleep stage classification. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1790–7. Available from: https://journalijsra.com/node/1049.

[19] Hasan J, Hasan K, Al Noman A, Hasan S, Sultana S, Arafat MA, et al. Transforming Leukemia Classification: A Comprehensive Study on Deep Learning Models for Enhanced Diagnostic Accuracy. PEEIACON 2024 - International Conference on Power, Electrical, Electronics and Industrial Applications. 2024;266–71.

[20] Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Amira Hossain, Hasib Fardin, Al Shahriar Uddin Khondakar Pranta. Custom CNN for acoustic emission classification in gas pipelines. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1760–8. Available from: https://journalijsra.com/node/1046.

[21] Md Ariful Islam, Mohammad Rasel Mahmud, Anamul Haque Sakib, Md Ismail Hossain Siddiqui, Hasib Fardin. Time domain feature analysis for gas pipeline fault detection using LSTM. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1769–77. Available from: https://journalijsra.com/node/1047.

[22] Mohammad Rasel Mahmud, Hasib Fardin, Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Abdullah Al Sakib. Hybrid deep learning for interpretable lung cancer recognition across computed tomography and histopathological imaging modalities. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1798–810. Available from: https://journalijsra.com/node/1050.

[23] Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Sanjida Akter, Jesika Debnath, Mohammad Rasel Mahmud. Comparative analysis of traditional machine learning Vs deep learning for sleep stage classification. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1778–89. Available from: https://journalijsra.com/node/1048.

[24] Sanjida Akter, Mohammad Rasel Mahmud, Md Ariful Islam, Md Ismail Hossain Siddiqui, Anamul Haque Sakib. Efficient and interpretable monkeypox detection using vision transformers with explainable visualizations. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1811–22. Available from: https://journalijsra.com/node/1051.

[25] Anamul Haque Sakib, Md Ismail Hossain Siddiqui, Sanjida Akter, Abdullah Al Sakib, Mohammad Rasel Mahmud. LEVit-Skin: A balanced and interpretable transformer-CNN model for multi-class skin cancer diagnosis. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1860–73. Available from: https://journalijsra.com/node/1055.

[26] Hamdadur Rahman, Hasan Md Imran, Amira Hossain, Md Ismail Hossain Siddiqui, Anamul Haque Sakib. Explainable vision transformers for real time chili and onion leaf disease identification and diagnosis. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1823–33. Available from: https://journalijsra.com/node/1052.

[27] Siddiqui IH, Al Sakib A, Sakib AH, Fardin H, Debnath J. Dual-branch CrossViT for ovarian cancer diagnosis: Integrating and explainable AI for real-time clinical applications. Article in International Journal of Science and

Research Archive [Internet]. 2025 [cited 2025 May 14];2025(01):1834–47. Available from: https://doi.org/10.30574/ijsra.2025.15.1.1164.

[28] Al-Sakib A, Limon ZH, Sakib A, Pranto MN, Islam MA, Sultana S, et al. Robust Phishing URL Classification Using FastText Character Embeddings and Hybrid Deep Learning. 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things, RAAICON 2024 - Proceedings. 2024;53–8.

[29] Haque R, Khan MA, Rahman H, Khan S, Siddiqui MIH, Limon ZH, et al. Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. Comput Biol Med [Internet]. 2025 Jun 1 [cited 2025 May 13];191:110166. Available from: https://www.sciencedirect.com/science/article/pii/S0010482525005177.

[30] Masum A Al, Limon ZH, Islam MA, Rahman MS, Khan M, Afridi SS, et al. Web Application-Based Enhanced Esophageal Disease Diagnosis in Low-Resource Settings. 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) [Internet]. 2024 Nov 28 [cited 2025 May 13];153–8. Available from: https://ieeexplore.ieee.org/document/10962580.

[31] Hosen MD, Bin Mohiuddin A, Sarker N, Sakib MS, Al Sakib A, Dip RH, et al. Parasitology Unveiled: Revolutionizing Microorganism Classification Through Deep Learning. Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024. 2024;1163–8.

[32] Noman A Al, Hossain A, Sakib A, Debnath J, Fardin H, Sakib A Al, et al. ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence. Computers 2025, Vol 14, Page 171 [Internet]. 2025 May 2 [cited 2025 May 13];14(5):171. Available from: https://www.mdpi.com/2073-431X/14/5/171/htm.