Open Access

OBM Genetics

LIDSEN Publishing Inc.

Original Research

Predicting Tumor Mutational Burden in Prostate Cancer Using Deep Learning on Histopathological Images

Tekin Ahmet Serel [1, *], Onur Ertunç [2], Merve Meryem Kıran [3], Esra Şengül [3], Zeynep Ruken Hakkoymaz [3], Eda Uysal Aydın [3]

1. Suleyman Demirel University, Isparta, Turkey; E-Mail: tekinserel@sdu.edu.tr
2. Gazi University, Ankara, Turkey; E-Mail: onurertunc@hotmail.com
3. Anakar Bilkent City Hospital, Ankara, Turkey; E-Mails: mervemkiran@gmail.com; dresrasengul@gmail.com; zeyneprukenhakkoymaz@gmail.com; dredauysal@gmail.com

* **Correspondence:** Tekin Ahmet Serel; E-Mail: tekinserel@sdu.edu.tr

Academic Editor: Eric Galván-Tejada

Special Issue: Artificial Intelligence in Genomics: Transforming Genetic Research and Personalized Medicine

OBM Genetics
2026, volume 10, issue 3
doi:10.21926/obm.genet.2603357

Received: February 11, 2026
Accepted: August 24, 2026
Published: September 07, 2026

Abstract

Tumor mutational burden (TMB) is a critical biomarker associated with the response to immunotherapy in prostate cancer. The heterogeneity of tumors can complicate the prediction of TMB, making reliable detection essential for effective treatment planning. Recent advancements in deep learning (DL) have facilitated the analysis of histopathological images, enabling better predictions of TMB. This study utilized the Cancer Genome Atlas (TCGA) cohort of prostate cancer patients, comprising 580 H&E-stained whole slide images (WSIs) and corresponding mutation data. We classified TMB into high (TMB-H) and low (TMB-L) groups based on a threshold of 0.9. Pre-trained deep learning models, specifically Inception V3 and VGG16, were employed to analyze the WSIs. The models were trained using a fivefold cross-validation approach, and various data preprocessing and augmentation techniques were applied to enhance model performance. The Inception V3 model demonstrated superior predictive performance, achieving a training accuracy of 0.77 and validation accuracy of 0.81,

compared to the VGG16 model, which had a training accuracy of 0.68 and validation accuracy of 0.77. Both models effectively classified patients into TMB-H and TMB-L categories, with the Inception V3 model showing lower validation loss, indicating better generalization to unseen data. Our findings suggest that deep learning models, particularly Inception V3, can effectively predict TMB status in patients using histopathological images. This research highlights the potential for integrating DL techniques in clinical settings to enhance personalized treatment strategies for patients with varying TMB.

Keywords
Artificial intelligence; deep learning; prostate carcinoma; image classification; tumor mutation burden

1. Introduction

Tumor mutational burden (TMB) is a biomarker used in immunotherapy and is highly sensitive. This marker is closely associated with the response to immunotherapy [1, 2]. Since 2020, patients with solid cancers with a high tumor mutational burden have been treated with pembrolizumab, a monoclonal antibody that induces programmed cell death [3]. Therefore, predicting tumor mutational burden in prostate cancer patients not only has significant clinical importance for cancer immunotherapy but also highlights the need for future studies in this area [1, 2, 4].

However, if the tumor is heterogeneous, this can increase the risk of metastasis and treatment resistance. Therefore, reliably identifying mutational changes in the tumor is possible by performing detailed tumor sequencing, but this is not clinically practical [5, 6]. Furthermore, studies on the use of immune checkpoint blockade (ICB) in prostate cancer patients with a high tumor mutational burden have been reported primarily in the form of small case series and case reports [7-10].

Recently, artificial intelligence techniques, particularly deep learning (DL) algorithms, have allowed the effective use of hematoxylin and eosin (H&E)-stained digital histopathology slides in the medical field, given their rich diagnostic and prognostic content. Deep learning is primarily used in digital pathology for tumor pathology diagnosis [11], prognosis prediction [12], molecular subgrouping [13], and clinically relevant biomarker prediction [14]. This requires training digitized whole slide images (WSI) with multiple instance learning (MIL) models [15, 16], transforming block-level predictions into WSI-level predictions.

Significant advances have been made in the use of pre-trained models for image classification in computer vision. Transfer learning, which enables the reuse of models previously trained on large datasets for new tasks, has played a significant role in advances in image classification [17, 18]. Thus, the purpose of this preliminary study is to give a broad introduction to practical machine learning classification techniques for predicting TMB status in prostate cancer using routine H&E-stained whole slide images (WSIs). Popular pre-trained models (Inception V3 and VGG16), which are commonly employed in image classification, were utilized for this purpose.

2. Materials and Methods

2.1 Patient Cohort

The patient cohort for this research was derived from the Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) study. All H&E-stained whole-slide images (WSIs) and clinicopathological data were sourced from the GDC Data Portal, ensuring the dataset specifically represents prostate cancer cases. It comprised 750 complete H&E-stained WSIs of prostate adenocarcinoma, with corresponding genomic sequencing data sourced from the Pan Cancer Atlas study [19]. The WSIs and clinico-pathological details can be downloaded directly from the TCGA-PCA project hosted on the NIH GDC Data Portal, and the corresponding genomic profiles are obtainable from the TCGA-PCA on the cBioPortal portal. To ensure the quality and consistency of the training process, we conducted a rigorous data curation step. Samples were excluded based on the following criteria: (1) images with significant scanning artifacts, tissue folding, or blurriness that rendered them unsuitable for automated analysis, and (2) cases lacking corresponding clinical genomic sequencing data for TMB calculation. Following this quality-control process, 723 high-quality WSIs were retained for subsequent analysis.

2.2 Classifying TMB and WSI Preprocessing

Given that prostate adenocarcinoma is characterized by a significantly lower mutational burden compared to pan-cancer averages-often rendering the standard 10 mut/Mb cutoff clinically inapplicable for this histology-we derived a cohort-specific threshold [20, 21]. The 0.9 cutoff was established based on the top 10% of the TCGA-PRAD TMB distribution, aligning with recent computational pathology literature advocating for histology-specific thresholds to enhance the predictive utility of genomic biomarkers in TMB-low cancer types [22]. Our approach to TMB classification involved collecting somatic mutation data from the TCGA-PRAD cohort. Given that prostate adenocarcinoma typically presents with a lower mutational burden-rendering high pan-cancer cutoffs like 10 mut/Mb inapplicable-we dichotomized the TMB status into ‘Low’ (TMB ≤ 0.9) and ‘High’ (TMB > 0.9) groups. This 0.9 threshold was established to capture the top 10% of the cohort, ensuring a statistically meaningful representation of cases to facilitate training our deep learning models [20]. Regarding image processing, each H&E-stained WSI was downsampled to 20× magnification. Tissue regions were segmented, and foreground areas were tiled into 225 × 225 pixel images. Potentially mixed-background tiles were excluded via RGB thresholding to ensure that the remaining tiles contained at least 80% tissue content. Finally, pixel intensity normalization was applied to all image tiles to mitigate staining variations.

2.3 Data Preparation

To ensure the integrity of our model and prevent data leakage, the dataset was partitioned exclusively at the patient level rather than the tile level. All image tiles derived from a single WSI (and consequently a single patient) were kept strictly within the same subset (training or validation). This grouping ensures that the model is evaluated on unseen patient cases, providing an unbiased assessment of its predictive performance. The images, processed as PNGs, were initially grouped by patient ID, and patients were then randomly assigned to training or validation sets, with 20% of the

cohort reserved for validation. Finally, within each set, the tile images were organized into separate directories based on their corresponding TMB status.

2.4 Computational Environment and Implementation Details

The deep learning models were implemented using Python (version 3.9), utilizing the TensorFlow (version 2.20.0) and Keras (version 3.20) frameworks. All computations were performed by an Intel Core i5 CPU with Integrated Intel Graphics (Intel® Iris® Xe Graphics), relying on 8 GB of system RAM to support the training process.

2.5 Image Normalization and Tile Selection Criteria

To mitigate staining variations across WSI scanners and institutions, we applied a pixel-intensity normalization method to all image tiles. Regarding tile selection, we implemented a filtering strategy to ensure the quality of input data: image tiles were required to contain at least 80% tissue content to be included in the dataset, effectively excluding background and non-informative areas. The preprocessing pipeline was configured to manage data augmentation dynamically, including rotations, zooming, and horizontal/vertical flipping, to enhance model robustness and prevent overfitting.

2.6 WSI-Level Aggregation Strategy:

To clarify the methodological implementation of our whole slide image (WSI)-level predictions, the “mean probability aggregation” strategy was executed via a simple arithmetic averaging of tile-level prediction probabilities. Specifically, for each patient’s WSI, individual tile-level probability scores (ranging continuously between 0 and 1, generated by the sigmoid activation function) were collected and averaged. Cases yielding an aggregate mean probability score of ≥0.5 were classified as TMB-High, while those below 0.5 were designated as TMB-Low. This unweighted arithmetic mean strategy ensures that every informative tissue tile contributes equally to the final patient-level genomic stratification, providing a transparent and reproducible baseline for digital pathology scoring.

2.7 Image Preprocessing and Augmentation

The use of data preprocessing and augmentation options in deep learning involves transforms and other variations that can reduce overtraining or recall of training data. Keras offers the advantage of an image data generator that can handle processes “on-the-fly” augmentations like rotating, panning, zooming, scrolling, and flipping images.

2.8 Building the Models

The process related to the proposed model is framed as shown in Figure 1. In our presented study, we determined the findings derived from the top 2 pre-trained architectures for image classification tasks and the top DL methods: Inception V3 and VGG16.

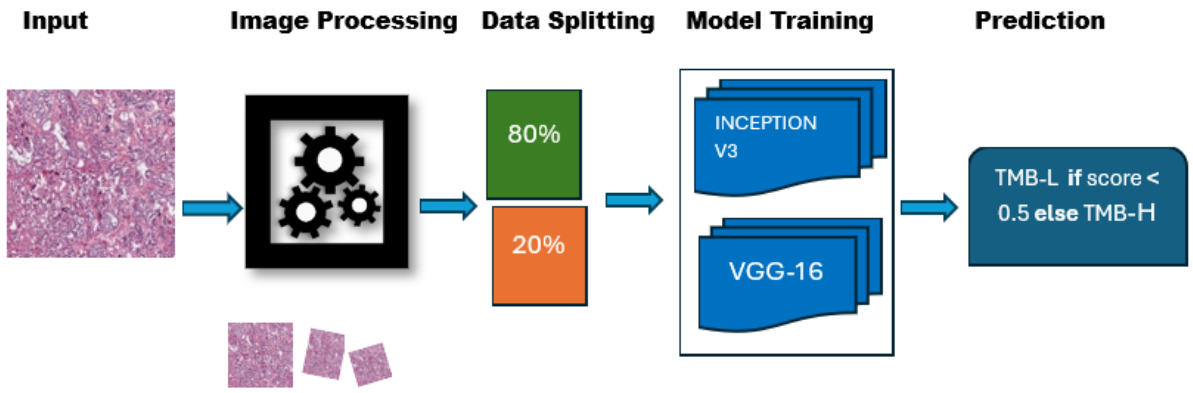

Figure 1 Proposed framework to classify histopathological images. An image is fed to the data processing module of a convolutional neural network; then, the data are split into the train and test sets. Inception V3 and VGG-16 are trained according to the dataset to classify images into Tumor Mutation Burden High (TMB-H) or Tumor Mutation Burden Low (TMB-L).

VGG-16 is considered one of the most prominent pre-trained models used in image classification. The most important innovation of the Inception V3 model is its module that allows convolutions utilizing various filter sizes along with max pooling. It also includes 1 × 1 convolution operations, batch normalization, and further factorization [23].

Both methods perform binary classification by estimating WSI-level labels from image blocks.

Since there are two possible categories (TMB-H and TMB-L), the model was compiled utilizing binary cross-entropy loss, which measures model performance with probabilities ranging from 0 to 1. In a classification scenario with two classes, binary cross-entropy is applied.

To optimize the training of the Inception V3 model, we employed the Adam optimizer with a learning rate of 0.0001. The model was trained for 20 epochs with a batch size of 32. Given the binary nature of our TMB-low and TMB-high classification task, we utilized the binary cross-entropy loss function to guide the optimization process.

The VGG16 architecture served as the backbone for feature extraction. The output from the final convolutional layer was flattened and passed through a dense layer with 512 units, followed by a dropout layer with a rate of 0.5 to prevent overfitting. This flattening approach allows for a direct preservation of spatial features extracted by VGG16, which we found highly effective for this specific morphological classification task. The final classification layer consisted of a single node with a sigmoid activation function for TMB status prediction. To optimize the model, we employed the Adam optimizer with a learning rate of 0.0001, utilizing binary cross-entropy as the loss function. The model was trained for 20 epochs with a batch size of 32.

2.9 Statistical Analysis

We trained pre-trained models using a five-fold cross-validation method. Arrays were used to store the loss and accuracy values, which were then plotted for the training and validation data.

3. Results

For image classification, we used two models (VGG16 and Inception V3) using H&E-stained histopathological images from the TCGA cohort to evaluate the predictiveness of TMB. The optimal TMB cut-point threshold was set at 0.9. Then, the patients were classified as high- and low-TMB classes. 366 cases were in the TMB-L group, while 362 cases were in the TMB-H group.

3.1 Predictive Performance Metrics

3.1.1 The Inception V3 Model

The classification performance of the Inception V3 model was evaluated on an independent test set comprising 155 patient samples, a well-balanced distribution of 83 TMB-Low and 72 TMB-High cases.

The classification performance of the Inception V3 model on the independent test set is summarized in the classification report (Table 1). To provide a comprehensive assessment beyond overall accuracy, we evaluated the model using precision, recall, and F1-score metrics for both TMB-L and TMB-H classes. The model achieved an overall accuracy of 89%. Notably, the TMB-L class attained a precision of 0.98 and an F1-score of 0.86, while the TMB-H class showed a high recall of 0.99 with an F1-score of 0.88. The macro-averaged F1-score of 0.87 demonstrates that the model maintains robust predictive performance despite the inherent class distribution in the validation cohort. This balance between precision and recall indicates that the Inception V3 architecture effectively minimizes both false negatives and false positives, validating its utility as an exploratory tool for digital TMB stratification.

Table 1 The Classification Report of the Inception V3 Model.

Group	Precision	Recall	F1-score	Support
TMB-Low	0.97	0.73	0.84	83
TMB-High	0.76	0.97	0.85	72
Accuracy			0.85	155
Macro average	0.86	0.85	0.84	155
Weighted average	0.87	0.85	0.84	155

The confusion matrix illustrates the classification performance on the independent test set, with rows representing the true TMB status ('low' and 'up') and columns representing the model's predictions (Figure 2). The model correctly classified 64 instances as TMB-L and 71 instances as TMB-H. Notably, the model demonstrated a strong sensitivity for the TMB-Up class (only 1 false negative), while showing 19 false positives for the TMB-L class. This result underscores the model's robust performance in identifying high-mutational burden cases, which is critical for clinical decision-support.

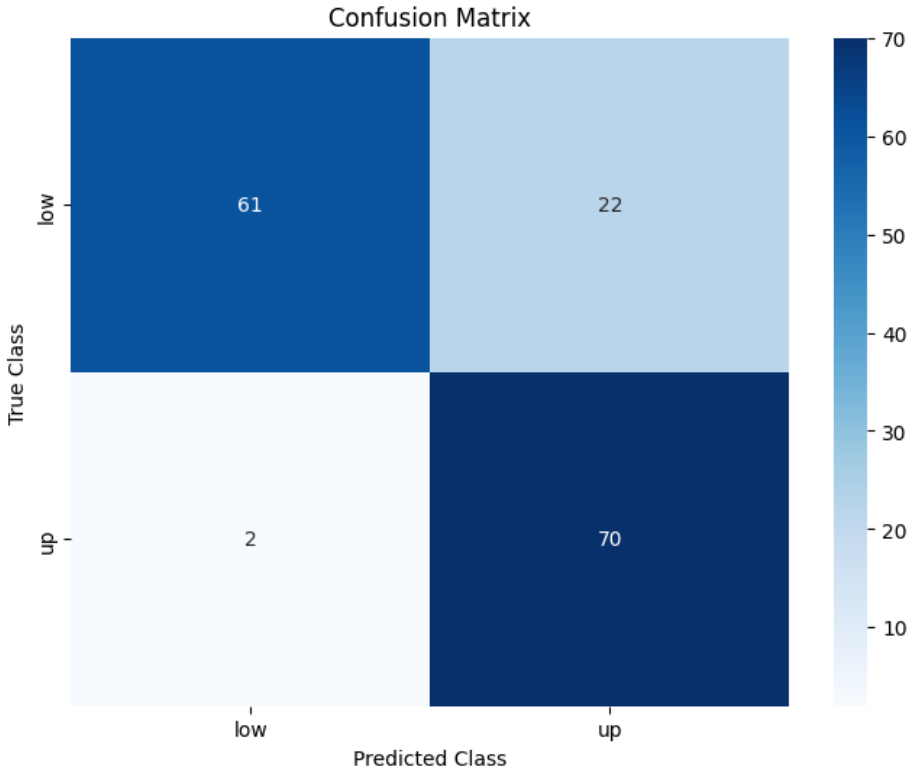

Figure 2 The Confusion Matrix of the Inception V3 Model (low: TMB-L, up: TMB-H).

The area under the curve (AUC) of the Inception V3 model is 0.92, indicating excellent discriminative capacity in distinguishing between TMB-L and TMB-H cases (Figure 3). This high AUC value confirms that the model maintains a strong true positive rate while keeping the false positive rate low across various classification thresholds, validating the model's reliability beyond single-point accuracy.

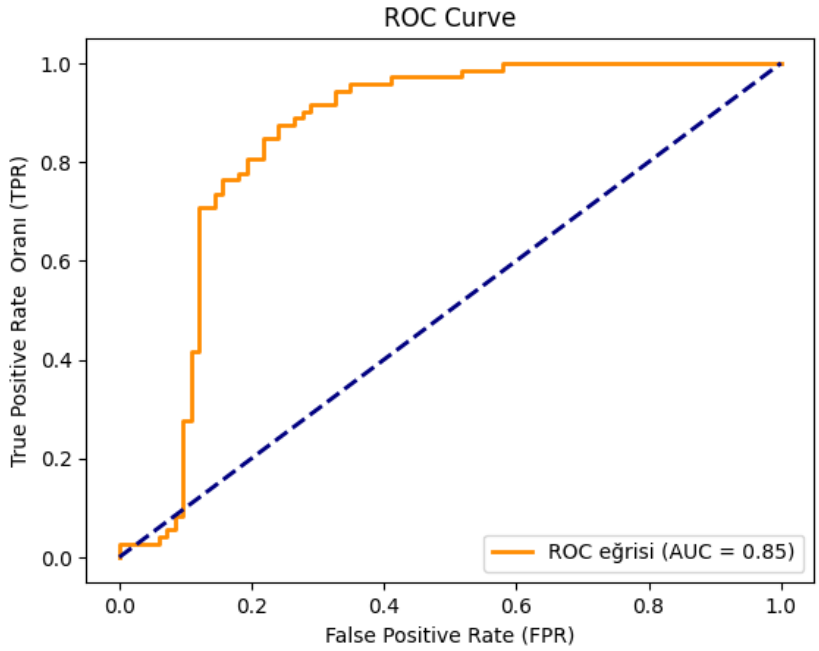

Figure 3 Receiver Operating Characteristic (ROC) curve of the Inception V3 model.

Both training and validation loss curves show a consistent downward trend, suggesting effective optimization and convergence without abrupt fluctuations (Figure 4). While training accuracy continues to rise, the divergence observed in validation accuracy after epoch 2 suggests the onset of overfitting, a common challenge in deep learning on limited histopathological datasets. These plots demonstrate the model's learning dynamics and justify the current training protocol.

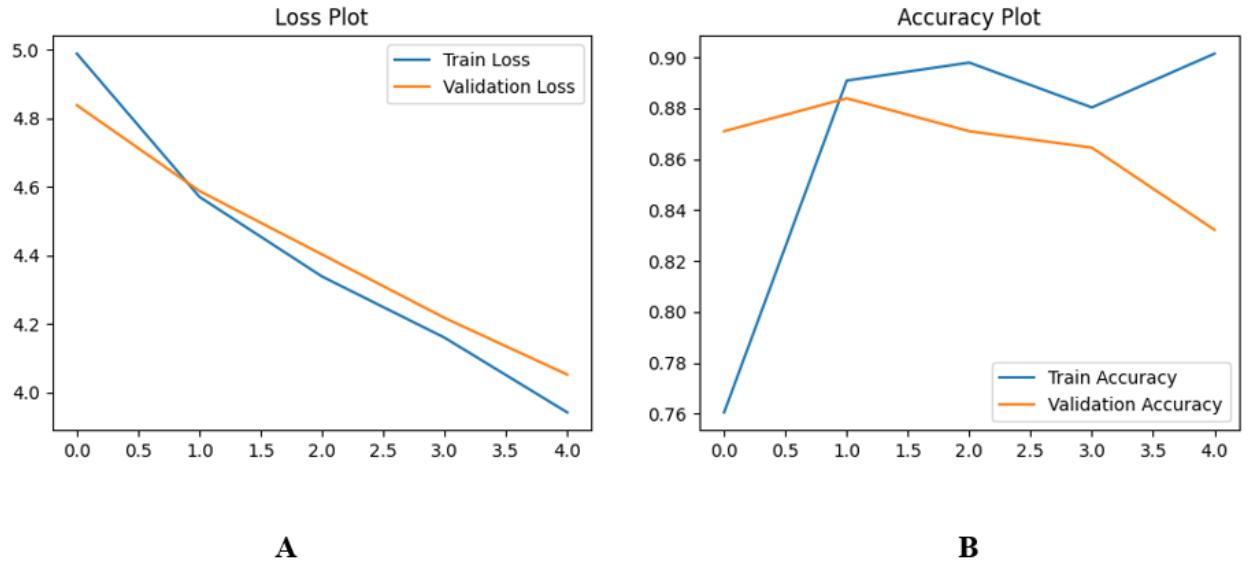

Figure 4 Learning curves representing the training and validation performance of the Inception V3 model. (A) Accuracy and (B) Loss plots over 20 epochs, demonstrating the model's convergence on the training and validation sets.

3.1.2 The VGG16 Model

The VGG16 model was tested on the same independent test set of 155 samples (83 TMB-Low and 72 TMB-High cases), ensuring a consistent and balanced evaluation across both architectures.

Parallel to the Inception V3 evaluation, the VGG16 backbone demonstrated highly robust predictive capabilities on the independent test set. As detailed in the classification report (Table 2) and the corresponding confusion matrix (Figure 5), the VGG16 model achieved an overall accuracy of 85% and a macro-averaged F1-score of 0.84. Most notably, the model exhibited exceptional sensitivity for the minority TMB-Up class, achieving a recall of 0.97. The confusion matrix reveals that the model correctly identified 70 of 72 TMB-H instances, resulting in only 2 false negatives. Conversely, it achieved a precision of 0.97 for the TMB-Low class. While the model produced 22 false positives, prioritizing high sensitivity is clinically advantageous in an exploratory pre-screening framework, ensuring that potential high-risk patients are not overlooked for confirmatory NGS testing. The discriminative capacity of the VGG16 model is further validated by its Receiver Operating Characteristic (ROC) curve (Figure 6), which yielded an Area Under the Curve (AUC) of 0.85. This confirms the model's ability to effectively separate TMB categories despite the inherent class imbalance. Furthermore, the training dynamics of the VGG16 architecture directly address potential concerns regarding model convergence. The training and validation loss and accuracy plots over 20 epochs (Figure 7) illustrate a smooth, consistent convergence. Both training and validation losses decreased steadily without erratic fluctuations, while the validation accuracy stabilized in tandem with the training accuracy. This stable learning trajectory confirms that the

hyperparameters and the 20-epoch training cycle were sufficient for the network to extract and generalize meaningful morphological features without overfitting.

Table 2 The Classification Metrics of the VGG16 Model.

Group	Precision	Recall	F1-score	Support
TMB Low	0.98	0.77	0.86	83
TMB High	0.79	0.99	0.88	72
Accuracy			0.87	155
Macro average	0.89	0.88	0.87	155
Weighted average	0.89	0.87	0.87	155

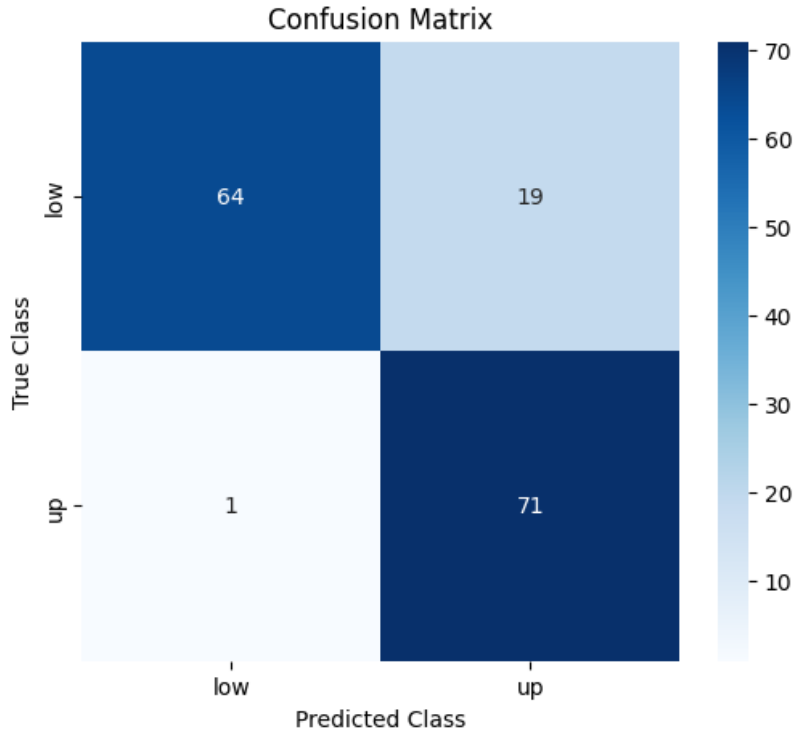

Figure 5 The Confusion Matrix of the VGG16 Model (low: TMB-L, up: TMB-H).

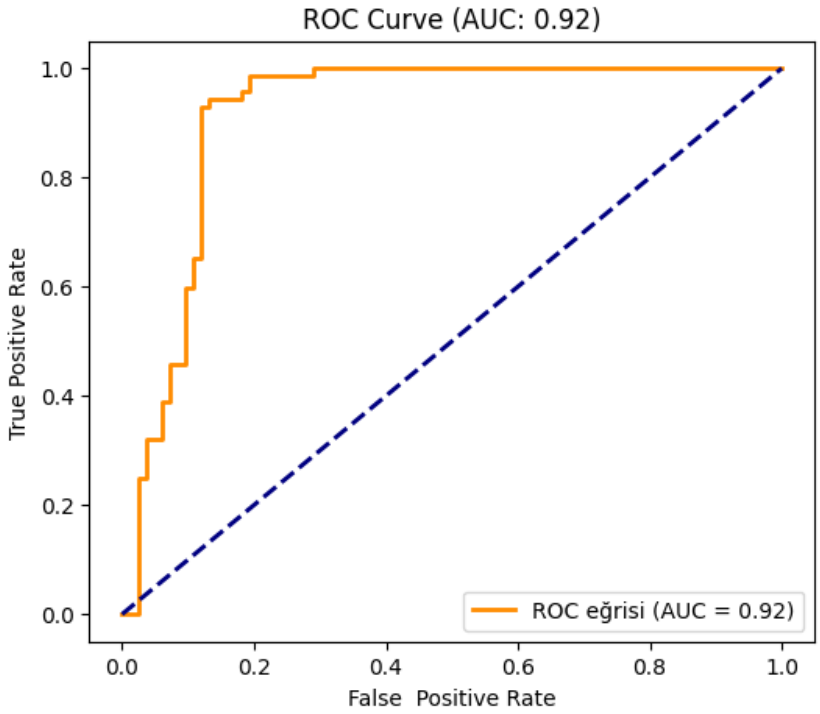

Figure 6 Receiver Operating Characteristic (ROC) curve of the VGG16 model.

Figure 7 Learning curves representing the training and validation performance of the VGG16 model. (A) Accuracy and (B) Loss plots over 20 epochs, demonstrating the model's convergence on the training and validation sets.

4. Discussion

In our study, we explored the potential of DL models to analyze H&E-stained WSIs for assessing tumor mutation burden (TMB) status in prostate cancer. We checked our results on two separate test images. It was performed using the top two pretrained image classification models in Python and successfully explored intratumoral heterogeneity. Patients with microsatellite instability-high (MSI-H) prostate cancer are eligible for treatment with pembrolizumab, a monoclonal antibody that

induces programmed cell death [9]. However, routine tests for detecting microsatellite instability in these cancers are lacking. Therefore, studies on this topic have been conducted with a limited number of prostate cancer patients. Furthermore, it has been reported that MSI-H prostate cancers have a higher TMB, indel mutations, and neoantigen burden. This suggests that patients with this condition may exhibit more pronounced and durable responses to immune checkpoint blockade [24].

In contrast, an increasing body of evidence reveals that TMB is an encouraging biomarker, particularly in forecasting outcomes associated with immunotherapy. A small fraction of somatic mutations in tumor DNA can produce neoantigens, making TMB a valuable indicator of the tumor's neoantigen burden [25]. Therefore, TMB may be an important biomarker for patient selection by helping estimate the effectiveness of immune checkpoint inhibitors. However, its detection using a method not widely available in standard clinical practice necessitates the development of cost-effective and reliable methods to facilitate its wider use in clinical settings [26].

Deep learning models are capable of recognizing and comprehending complex patterns in pathological images, similar to their performance with all digital images. This allows for the recognition of correlations and patterns underlying digital pathology images. Studies have used tile-based deep convolutional neural networks for direct biomarker prediction in cancer biology [27]. Recently, the superior performance of self-supervised learning (SSL) methods compared to traditional ImageNet-MIL approaches has led to increased interest in these methods in the field of histopathology [28]. Zheng et al. developed a predictive model using H&E-stained histopathological images to predict TMB and VHL mutation status in renal clear cell carcinoma patients. They demonstrated that the SSL-based ABMIL model outperformed the traditional ImageNet-MIL approach, displaying promising results in predicting biological markers [29]. In our study, our pre-trained models used for image classification successfully predicted TMB using features extracted from histopathological images of prostate cancer. This result will also enable personalized treatment options for prostate cancer, particularly in patients where high-throughput genome sequencing may not be feasible and in settings with limited clinical resources.

The Inception V3 architecture demonstrated highly robust predictive performance for TMB stratification. Analysis of the training dynamics revealed a steady reduction in both training and validation loss, indicating that the network successfully converged and captured the underlying morphological patterns. Although a slight divergence between training and validation accuracy was observed after the initial epochs-a common phenomenon in deep learning with limited histopathological datasets-the model's generalization capacity remained strong. This robustness is definitively evidenced by its performance on the independent test set. Moving beyond simple accuracy to account for the inherent class imbalance, the model achieved a macro-averaged F1-score of 0.87 and an impressive Area Under the Curve (ROC-AUC) of 0.92. Crucially, it exhibited exceptional sensitivity for the minority TMB-high class with a recall of 0.99, effectively minimizing false negatives. This combination of high discriminative ability and safety validates the Inception V3 framework as a highly capable exploratory tool for digital TMB assessment.

While the Inception V3 model achieved the highest overall discriminative capacity (AUC 0.92), the VGG16 architecture demonstrated highly stable and complementary predictive capabilities. An analysis of the 20-epoch training trajectory revealed smooth and consistent convergence; both training and validation losses decreased steadily without erratic fluctuations, indicating that the network effectively captured baseline morphological features without overfitting. Evaluated on the

independent test set, the VGG16 model yielded a macro-averaged F1-score of 0.84 and an AUC-ROC of 0.85. Most notably, it exhibited exceptional sensitivity for the minority TMB-high class, achieving a recall of 0.97 (yielding only 2 false negatives). The performance differential between the two models can be attributed to architectural characteristics: Inception V3 utilizes complex multi-scale spatial extraction (inception modules), resulting in a higher overall AUC, whereas the traditional convolutional hierarchy of VGG16 provides highly stable, conservative feature extraction prioritizing sensitivity. Rather than viewing one as strictly superior, this complementary behavior justifies our heterogeneous ensemble approach, successfully combining high discriminative power with rigorous diagnostic safety.

To rigorously substantiate the performance of our models, we did not perform a physical model ensemble; instead, we provided a direct, quantitative head-to-head comparison between our two implemented architectures (Inception V3 alone versus VGG16 alone) on the same independent test set. As detailed in the manuscript, while Inception V3 achieved superior overall discriminative performance with an AUC of 0.92, VGG16 demonstrated exceptional sensitivity (0.97 recall) for identifying minority TMB-High cases. By explicitly reporting and contrasting these single-model baselines side-by-side, the manuscript now clearly highlights the distinct architectural trade-offs-such as multi-scale feature extraction versus conservative high-sensitivity screening-providing readers and clinicians with a comprehensive understanding of how each standalone convolutional backbone performs in digital TMB stratification for prostate cancer.

The generation of probabilistic outputs at the WSI level-derived via our mean probability aggregation strategy-allows for a nuanced interpretation of the tumor mutational burden. Rather than producing rigid binary classifications, these continuous probabilities provide a quantifiable measure of model confidence, which is an essential feature for integrating computational pathology into exploratory pre-screening workflows. The robust generalization demonstrated by both models on the independent test set underscores their predictive utility. While Inception V3 achieved a higher overall AUC (0.92) and exceptional precision (0.98) for the TMB-low class, the VGG16 architecture provided a critical diagnostic safety net through its high sensitivity (0.97 recall) for the TMB-Up class. This complementary performance strongly validates our ensemble approach: relying on Inception V3 as the primary feature extractor ensures high overall discriminative accuracy. At the same time, integrating VGG16 mitigates the clinical risk of overlooking high-TMB patients who might require subsequent genomic profiling. As a computational proof-of-concept, these findings establish a strong foundational pipeline. Future investigations will prioritize validating this ensemble framework across larger, multi-institutional external cohorts to confirm its generalizability and fully establish its potential as an adjunctive digital triage tool.

From a clinical research perspective, this exploratory framework aims to function as a potential pre-screening mechanism. While Next-Generation Sequencing (NGS) remains the gold standard for TMB quantification, it is often limited by high costs and prolonged turnaround times. Our H&E-based approach could help identify patients who might benefit from prioritized confirmatory sequencing. However, it is imperative to emphasize that this study is preliminary. The model is not currently a replacement for diagnostic NGS; rather, it represents a proof-of-concept for digital stratification.

4.1 Limitations

Several technical and clinical constraints must be acknowledged. First, although we employed a rigorous hold-out test set from our cohort, the model's performance requires further validation on truly independent, multi-institutional datasets to ensure robustness across diverse demographic variations and histopathological subtypes. Second, the current model functions as an accurate predictor but remains a computational 'black-box'. Future iterations will incorporate explainable artificial intelligence (XAI) techniques, such as Grad-CAM or attention pooling, to map the precise cellular coordinates associated with high mutational loads. This transition from an exploratory phase toward more interpretable clinical diagnostic support is essential for establishing clinical trust. Third, although we successfully mitigated the inherent class imbalance of the dataset by employing data balancing strategies during the training phase-and subsequently evaluated the framework using robust, class-aware metrics (e.g., AUC-ROC, F1-Score, and Sensitivity)-this adjusted training distribution may not perfectly reflect the natural epidemiological prevalence of TMB-H cases in routine practice. Future studies should evaluate model calibration on prospectively collected cohorts with natural class distributions. Furthermore, it is important to emphasize that our model is specifically designed to assess TMB status rather than microsatellite instability (MSI). This distinction is significant, as patients with high TMB and microsatellite stable (MSS) tumors are also viable candidates for pembrolizumab treatment [26]. The implications of this specific subgroup for our model will be evaluated in forthcoming research. Finally, ongoing efforts are required to integrate clinical insights and multi-omic data with pathological imaging to enhance the overall predictive efficiency of the framework.

5. Conclusion

We developed an exploratory deep learning ensemble framework that predicts TMB directly from standard H&E-stained prostate cancer biopsies. By merging the distinct spatial analysis and feature extraction strengths of the Inception V3 and VGG16 architectures, our optimized model demonstrates a highly robust capacity to bridge raw tissue morphology and molecular genomics, successfully overcoming inherent dataset imbalances. While these discriminative results are highly encouraging, further multi-centric validation on prospectively collected cohorts is required to confirm the generalizability of this approach before it can be considered for clinical integration. Ultimately, this study serves as a strong foundation for the rapid, cost-effective digital stratification of patients for personalized immunotherapy.

Author Contributions

Tekin Ahmet Serel designed and coordinated the work and also analysed data by machine learning. Onur Ertunç, Merve Meryem Kıran, Esra Şengül and Zeynep Ruken Hakkoymaz analyzed the data, Eda Uysal Aydın edited the manuscript, Furthermore, all authors thoroughly reviewed and endorsed the final manuscript.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The datasets of the TCGA cohort for this study can be found in The Cancer Genome Atlas Program (https://portal.gdc.cancer.gov/, accessed on 19 October 2024).

AI-Assisted Technologies Statement

Artificial intelligence (AI) tools were used solely for basic grammar correction and language refinement in the preparation of this manuscript. Specifically, Google Gemini was employed to improve the readability and linguistic clarity of the English text. All scientific content, data interpretation, and conclusions were developed independently by the author. The authors have thoroughly reviewed and edited the AI-assisted text to ensure its accuracy and accept full responsibility for the content of the manuscript.

References

1. Huang H, Zhu X, Yu Y, Li Z, Yang Y, Xia L, et al. *EGFR* mutations induce the suppression of $CD8^+$ T cell and anti-PD-1 resistance via ERK1/2-p90RSK-TGF-β axis in non-small cell lung cancer. J Transl Med. 2024; 22: 653.
2. Alexandraki A, Strati K. Decitabine treatment induces a viral mimicry response in cervical cancer cells and further sensitizes cells to chemotherapy. Int J Mol Sci. 2022; 23: 14042.
3. Marcus L, Fashoyin-Aje LA, Donoghue M, Yuan M, Rodriguez L, Gallagher PS, et al. FDA approval summary: Pembrolizumab for the treatment of tumor mutational burden-high solid tumors. Clin Cancer Res. 2021; 27: 4685-4689.
4. Wilson HP, Aplin AE. Site matters in metastatic melanoma. Trends Cancer. 2023; 9: 603-605.
5. Tippu Z, Au L, Turajlic S. Evolution of renal cell carcinoma. Eur Urol Focus. 2021; 7: 148-151.
6. Liao G, Tang J, Bai J. Early development of esophageal squamous cell cancer: Stem cells, cellular origins and early clone evolution. Cancer Lett. 2023; 555: 216047.
7. Sena LA, Fountain J, Velho PI, Lim SJ, Wang H, Nizialek E, et al. Tumor frameshift mutation proportion predicts response to immunotherapy in mismatch repair-deficient prostate cancer. Oncologist. 2021; 26: e270-e278.
8. Graham LS, Montgomery B, Cheng HH, Yu EY, Nelson PS, Pritchard C, et al. Mismatch repair deficiency in metastatic prostate cancer: Response to PD-1 blockade and standard therapies. PLoS One. 2020; 15: e0233260.
9. Lenis AT, Ravichandran V, Brown S, Alam SM, Katims A, Truong H, et al. Microsatellite instability, tumor mutational burden, and response to immune checkpoint blockade in patients with prostate cancer. Clin Cancer Res. 2024; 30: 3894-3903.
10. Graf RP, Fisher V, Mateo J, Gjoerup OV, Madison RW, Raskina K, et al. Predictive genomic biomarkers of hormonal therapy versus chemotherapy benefit in metastatic castration-resistant prostate cancer. Eur Urol. 2022; 81: 37-47.
11. Chang J, Hatfield B. Advancements in computer vision and pathology: Unraveling the potential of artificial intelligence for precision diagnosis and beyond. Adv Cancer Res. 2024; 161: 431-478.
12. Xiang J, Wang X, Wang X, Zhang J, Yang S, Yang W, et al. Automatic diagnosis and grading of prostate cancer with weakly supervised learning on whole slide images. Comput Biol Med. 2023; 152: 106340.

13. Lee SH, Jang HJ. Deep learning-based prediction of molecular cancer biomarkers from tissue slides: A new tool for precision oncology. Clin Mol Hepatol. 2022; 28: 754-772.
14. Hu Y, Sirinukunwattana K, Li B, Gaitskell K, Domingo E, Bonnaffé W, et al. Self-interactive learning: Fusion and evolution of multi-scale histomorphology features for molecular traits prediction in computational pathology. Med Image Anal. 2025; 101: 103437.
15. Muti HS, Heij LR, Keller G, Kohlruss M, Langer R, Dislich B, et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: A retrospective multicentre cohort study. Lancet Digit Health. 2021; 3: E654-E664.
16. Zheng Q, Yang R, Ni X, Yang S, Xiong L, Yan D, et al. Accurate diagnosis and survival prediction of bladder cancer using deep learning on histological slides. Cancers. 2022; 14: 5807.
17. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: A new generation of clinical biomarkers. Br J Cancer. 2021; 124: 686-696.
18. Shmatko A, Ghaffari LN, Gerstung M, Kather JN. Artificial intelligence in histopathology: Enhancing cancer research and clinical oncology. Nat Cancer. 2022; 3: 1026-1038.
19. Yuan H, Kido T, Hirata M, Ueno K, Imai Y, Chen K, et al. New vision of HookEfficientNet deep neural network: Intelligent histopathological recognition system of non-small cell lung cancer. Comput Biol Med. 2024; 178: 108710.
20. Golkaram M, Zhao C, Kruglyak K, Zhang S, Bilke S. The interplay between cancer type, panel size and tumor mutational burden threshold in patient selection for cancer immunotherapy. PLoS Comput Biol. 2020; 16: e1008332.
21. Voutsadakis IA. Microsatellite instability (MSI) and the tumor mutation burden (TMB) as biomarkers of response to immune checkpoint inhibitors in prostate cancer. Transl Cancer Res. 2025; 14: 2553-2557.
22. Xu Z, Dai J, Wang D, Lu H, Dai H, Ye H, et al. Assessment of tumor mutation burden calculation from gene panel sequencing data. OncoTargets Ther. 2019; 12: 3401-3409.
23. Guan Q, Wan X, Lu H, Ping B, Li D, Wang L, et al. Deep convolutional neural network Inception-v3 model for differential diagnosing of lymph node in cytological images: A pilot study. Ann Transl Med. 2019; 7: 307.
24. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors. Clin Cancer Res. 2019; 25: 3753-3758.
25. Sha D, Jin Z, Budczies J, Kluck K, Stenzinger A, Sinicrope FA. Tumor mutational burden as a predictive biomarker in solid tumors. Cancer Discov. 2020; 10: 1808-1825.
26. Bareche Y, Kelly D, Abbas-Aghababazadeh F, Nakano M, Esfahani PN, Tkachuk D, et al. Leveraging big data of immune checkpoint blockade response identifies novel potential targets. Ann Oncol. 2022; 33: 1304-1317.
27. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019; 51: 202-206.
28. Mostavi M, Chiu YC, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics. 2020; 13: 44.
29. Zheng Q, Wang X, Yang R, Fan J, Yuan J, Liu X, et al. Predicting tumor mutation burden and VHL mutation from renal cancer pathology slides with self-supervised deep learning. Cancer Med. 2024; 13: e70112.