Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)

Journal Impact Factor® (2023): 3.1
CiteScore® (2023): 7.4
www.biomolbiomed.com | www.blog.bjbms.org

RESEARCH ARTICLE

Kabir et al: Deep learning for PD-L1 in NSCLC

A novel deep learning framework for automatic scoring of PD-L1 expression in non-small cell lung cancer

Saidul Kabir[1], Muhammad E. H. Chowdhury[2*], Rusab Sarmun[1], Semir Vranic[3], Rafif Mahmood Al Saady[3], Inga Rose[4], Zoran Gatalica[4]

[1]Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka, Bangladesh.

[2]Department of Electrical Engineering, Qatar University, Doha, Qatar.

[3]College of Medicine, QU Health, Qatar University, Doha, Qatar.

[4]Reference Medicine, Phoenix, Arizona, United States of America.

[*]**Correspondence to** Muhammad E. H. Chowdhury: mchowdhury@qu.edu.qa

DOI: https://doi.org/10.17305/bb.2025.12056

ABSTRACT

A critical predictive marker for anti-PD-1/PD-L1 therapy is programmed death-ligand 1 (PD-L1) expression, assessed by immunohistochemistry (IHC). This paper explores a novel automated framework using deep learning to accurately evaluate PD-L1 expression from whole slide images

(WSIs) of non-small cell lung cancer (NSCLC), aiming to improve the precision and consistency of Tumor Proportion Score (TPS) evaluation, which is essential for determining patient eligibility for immunotherapy. Automating TPS evaluation can enhance accuracy and consistency while reducing pathologists' workload. The proposed automated framework encompasses three stages: identifying tumor patches, segmenting tumor areas, and detecting cell nuclei within these areas, followed by estimating the TPS based on the ratio of positively stained to total viable tumor cells. This study utilized a Reference Medicine (Phoenix, Arizona) dataset containing 66 NSCLC tissue samples, adopting a hybrid human-machine approach for annotating extensive WSIs. Patches of size 1000x1000 pixels were generated to train classification models such as EfficientNet, Inception, and Vision Transformer models. Additionally, segmentation performance was evaluated across various UNet and DeepLabV3 architectures, and the pre-trained StarDist model was employed for nuclei detection, replacing traditional watershed techniques. PD-L1 expression was categorized into three levels based on TPS: negative expression (TPS < 1%), low expression (TPS 1-49%), and high expression (TPS ≥ 50%). The Vision Transformer-based model excelled in classification, achieving an F1-score of 97.54%, while the modified DeepLabV3+ model led in segmentation, attaining a Dice Similarity Coefficient of 83.47%. The TPS predicted by the framework closely correlated with the pathologist's TPS at 0.9635, and the framework's three-level classification F1-score was 93.89%. The proposed deep learning framework for automatically evaluating the TPS of PD-L1 expression in NSCLC demonstrated promising performance. This framework presents a potential tool that could produce clinically significant results more efficiently and cost-effectively.

Keywords: Programmed death-ligand 1; PD-L1; non-small cell lung cancer; NSCLC; artificial intelligence; AI; deep learning, classification, segmentation.

INTRODUCTION

Lung cancer, with nearly two million new cases each year, stands as the most prevalent cancer globally [1]. Patients with stage IV non-small cell lung cancer (NSCLC) have a survival rate of only 5% [2], [3]. Immune checkpoint inhibitors targeting the PD-1/PD-L1 pathway have significantly improved survival rates for patients diagnosed with NSCLC [11], [12], [13]. The programmed cell death-1 (PD-1) receptor and its ligands PD-L1 and PD-L2 belong to a family of immune checkpoint proteins. These molecules function as co-inhibitory factors for T-cells, effectively dampening immune responses. The interaction between PD-1 and PD-L1 plays a crucial role in regulating the timing of immune system activation [4]. Expression of PD-L1 on tumor cells binds to PD-1 receptors on activated T-cells, enabling tumor cells to evade anticancer immunity [5], [6]. Monoclonal antibodies that block this interaction between PD-1 and PD-L1 can restore the immune system's ability to recognize and destroy cancer cells [7], [8].

Researchers have developed several inhibitors based on this mechanism of action. Existing immune checkpoint inhibitors (ICI), including anti-PD-1 and anti-PD-L1 inhibitors, have shown promising results in clinical trials [9], [10].

The Food Drug Association (FDA) has approved DAKO PD-L1 22C3 PharmDx as a companion diagnostic test for the immunotherapeutic drug pembrolizumab in patients with NSCLC [14]. Evaluating PD-L1 expression is crucial in managing patients, as it helps in identifying patients who are more likely to respond to pembrolizumab. This applies to its use as first or second-line monotherapy, or alongside standard chemotherapy [15].

The tumor proportion score (TPS) is calculated as the percentage of tumor cells (TCs) showing at least partial membranous staining for PD-L1, relative to the total number of TCs [16]. This calculation excludes tumor-associated immune cells (ICs), normal, necrotic, and non-neoplastic cells. The TPS can be defined using the formula $TPS(\%) = \frac{Number\ of\ PD{-}L1\ positive\ TCs}{Total\ number\ of\ viable\ TCs} \times 100.$

Pathologists usually estimate TPS through microscopical examination. For specimens with heterogeneous tumor regions exhibiting varying PD-L1 expression, the TPS is determined by calculating the average percentage of stained tumor cells across multiple divided tumor regions. This approach accounts for the spatial heterogeneity of PD-L1 expression often encountered in tumor samples, providing a more representative assessment of overall PD-L1 status [17].

PD-L1 expression is commonly observed in NSCLC [18], [19], [20], [21] and is predictive of response to ICI. However, the scoring of PD-L1 expression in NSCLC specimens presents significant challenges, particularly in advanced-stage patients [22], [23]. This process necessitates experienced evaluation to ensure accurate tumor classification. Inter-observer variability among pathologists during manual scoring has been reported (kappa score as low as 0.45), potentially leading to inconsistent results. Pathologist-dependent scoring introduces an inherent source of error according to multiple studies, which becomes particularly pronounced in the case of low PD-L1 expression [24], [25]. Moreover, the manual evaluation of PD-L1 expression can be a tedious process vulnerable to subjectivity [16]. This subjectivity stems from the difficulties associated with accurate visual quantification of cellular elements across entire slide sections. The process is further complicated by the subjective nature of stain intensity assessment, which introduces additional variability. These factors collectively contribute to challenges in maintaining reproducibility and inter-observer consistency in PD-L1 scoring. The complexity of this assessment underscores the need for potentially automated solutions to enhance the accuracy and reproducibility of PD-L1 expression evaluation in NSCLC specimens.

Deep learning has been widely integrated within the healthcare sector in recent years, demonstrating its potential to address diagnostic inconsistencies. By leveraging deep learning techniques, medicine can benefit from the ability of these models to identify complex patterns and features within extensive datasets, leading to precise and consistent evaluations [26], [27], [28]. This technological advancement can mitigate the reliance on individual medical practitioners and reduce the variability in subjective interpretations among different observers. The extensive application of deep learning in healthcare underscores its transformative impact on medical diagnostics and treatment [29].

Digital image analysis techniques offer a promising approach to address the limitations in PD-L1 scoring especially in PD-L1 scoring of whole tissue sections. Artificial intelligence (AI) methodologies, particularly those employing deep learning (DL) algorithms, have demonstrated the potential to augment pathologists' capabilities, enhancing diagnostic accuracy, inter-observer concordance, and overall efficiency [30], [31], [32], [33], [34].

Previous investigations have primarily focused on evaluating the correlation between pathologist-derived and automated PD-L1 scores. The findings from these studies indicate that automated

systems demonstrate a level of accuracy comparable to experienced pathologists in PD-L1 expression assessment [35], [36], [37], [38], [39], [40].

In the studies by Taylor et al. [38] and Sha et al. [41], PD-L1 TPS was calculated at the field-of-view level by measuring areas of tumor regions. However, this regional area ratio-based method lacks precision as it fails to align with clinical guidelines, which mandate that TPS must be determined based on tumor cellular counts. Methods of calculating at the cellular level have demonstrated superior results, as subsequent TPS calculations are derived from individual tumor cells [42]. Subsequent research has shown that the open-source program QuPath, used for scoring PD-L1 in NSCLC, has produced promising results [43], [44]. Notably, most studies have utilized watershed-based image processing techniques to identify cell nuclei [36], [39]. While this approach may be effective for clear-cut cases, it tends to struggle with challenging scenarios involving varying stain intensity and coloring.

Huang et al. [45] tested an AI-assisted scoring system for the assessment of PD-L1 expression in NSCLC, using the UNet segmentation model trained, validated, and tested on whole slide images (WSIs). The results showed that the model results correlated strongly with the gold standard TPS and performed comparably to experienced pathologists, though it performed less well in high TPS groups due to false positives, still demonstrating potential for aiding routine diagnosis by pathologists. This study employed a basic UNet architecture to segment positive and negative nuclei. However, this straightforward approach may lead to inaccuracies in more complex cases regarding color intensity. Wu et al. [39] proposed developing a system based on AI using WSIs of the 22c3 assay using a UNet architecture based on residual block to segment tumor areas and automatically calculate TPS of PD-L1 expression. The system showed strong consistency with trained pathologists and improved the efficiency and workload of untrained pathologists, demonstrating high precision in both the 22c3 and SP263 assays. Cheng et al. [46] developed a YOLO-based artificial intelligence (AI) model for the assessment of PD-L1 expression in lung cancer patients, including 1,288 participants. It used the detection model to detect positive and negative nuclei of tumor cells. This AI model's diagnostic results were consistent with pathologists, demonstrating similar performance across different lung cancer subtypes and suggesting that AI-assisted diagnostic methods are promising tools for enhancing clinical pathologist efficiency.

Liu et al. [36] developed a novel Automated Tumor Proportion Scoring System (ATPSS) to compare image analysis results with pathologist scores. The ATPSS employs a three-stage process

that integrates Resnet-UNet-based architectures for detecting tumor regions and nuclei, alongside image processing techniques to identify positive staining. The ATPSS demonstrated a high correlation with pathologist scores, achieving a Mean Absolute Error (MAE) of 8.65 and a Pearson Correlation Coefficient (PCC) of 0.9436. However, image processing-based detection of positive regions may erroneously classify artifacts and stained ICs as positive cancer cells.

Ito et al. [47] developed a model to calculate the TPS of the PD-L1 22C3 assay and evaluate its effectiveness in assisting pathologists. They used UNet architecture to segment nuclei and DeepLab architecture to segment tumor areas. The findings highlight the AI-assisted system's potential to enhance pathologists' accuracy, particularly in challenging cases where their judgments were inconsistent.

In this study, a comparative analysis of the segmentation performance of various UNet architectures was conducted with different encoders, alongside a modified DeepLabV3+ architecture. Additionally, we developed an end-to-end framework for calculating TPS from WSIs, incorporating a classification stage to enhance resilience against artifacts and misidentified positive ICs. For nuclei detection, we employed a deep learning-based approach complemented by image processing techniques, addressing the limitations of the commonly used watershed method, which often fails in cases of overstaining and low intensity.

The major contributions of this study are as follows:

- Development of a novel end-to-end framework for the automated assessment of PD-L1 expression TPS using WSIs from surgical resections.
- Annotation of entire surgical resection WSIs through a combined human-machine approach.
- Comparative analysis of various segmentation networks, including UNet with different encoders and a modified DeepLabV3+ architecture, for tumor region segmentation.
- Deep learning-based cancer cell nuclei detection ensures robust and precise TPS calculation.

MATERIALS AND METHODS

Dataset

This study examined 66 surgically obtained tissue samples from patients with confirmed NSCLC. The specimens were collected at Reference Medicine (Phoenix, AZ) between January 2020 and

October 2022. For each case in the dataset, PD-L1 IHC slides were prepared using the Dako Autostainer Link 48 platform, following the automated staining protocol with the PD-L1 22C3 antibody. TPS was used for PD-L1 assessment following the 'PD-L1 IHC 22C3 pharmDx Interpretation Manual NSCLC [17]. All slides were digitized using the Motic EasyScan Pro slide scanner. **Table 1** shows the baseline characteristics of the NSCLC cohort.

Data processing

In this study, we employed a hybrid machine-human approach for case annotation. From each WSI, a representative tumor area was selected for manual annotation. This approach was adopted to mitigate the considerable challenges associated with annotating entire WSIs, including annotator fatigue and the time-intensive nature of the task. Two independent pathologists conducted annotations on the selected regions from each WSI using QuPath software (Version 0.2.2) [43]. The annotation process involved categorizing the tissue into three distinct classes: (class 0: Non-tumor cell regions, class 1: Tumor cell regions with PD-L1 expression, and class 2: Tumor cell regions without PD-L1 expression). A segmentation model was initially trained on these manually annotated tumor regions obtained from the WSIs. Once trained, these models were utilized to predict tumor regions throughout the entire WSIs. This preliminary machine-generated segmentation provided a basis for the initial annotation of tumor areas, thereby streamlining the subsequent review and annotation process allowing annotations of large WSIs with ease. Then the two pathologists conducted a comprehensive review of the automated annotations to adjust and refine the annotations. This process enhanced the reliability of the training data for further training of segmentation models. This iterative refinement was crucial for training robust models capable of precise tumor segmentation in the WSIs, thereby supporting more effective and efficient pathological assessments. Figure 1 visualizes the annotation approach undertaken in this study.

Patches of size 1000×1000 pixels were generated from the WSIs at the highest magnification (40× optical magnification, 0.267 μm/pixel). The WSIs were divided into five folds for cross-validation, with the patches from each image assigned to their respective fold. Two distinct datasets were created from the extracted image patches. The first dataset was designed to facilitate the classification of each patch by the presence of tumor tissue, while the second was intended for the segmentation of regions into positive and negative tumor areas. For the classification dataset,

patches devoid of tumor cells were labeled as "non-tumor", whereas patches containing any tumor cells were labeled as "tumor". For the segmentation dataset, both the patches and their corresponding masks were generated based on annotations provided by pathologists. Figure 2 shows an example of a classification and segmentation dataset.

Proposed method

We propose a novel end-to-end automated framework for determining TPS in DAB-stained NSCLC WSI. It consists of three key stages: tumor patch classification, tumor area segmentation, and nuclei detection. Initially, a deep learning model identifies and excludes non-tumor patches by discarding patch images containing artifacts or lacking tumor cells. In the subsequent stage, segmentation networks predict pixel-wise classifications (negative or positive) within the tumor patches, delineating the negative and positive regions. Finally, we employ a secondary neural network utilizing the pre-trained "StarDist" model to facilitate cell detection within the annotated tumor regions [48]. Figure 3 demonstrates the flow of the proposed framework.

Tumor patch classification

The first stage for automated tumor patch identification in WSIs of tumor regions was developed using convolutional neural network (CNN) based classification models. CNN-based architectures are highly popular in image classification due to their ability to spatial patterns in local regions and learn abstract features at many levels. Their hierarchical structure and shared weights have enabled them to obtain state-of-the-art results on standard classification datasets and benchmarks [29], [49]. Vision Transformer models leverage self-attention mechanisms which can capture global dependencies and interactions at long distances, effectively representing contexts at both local and global scales and surpassing the limitations of traditional CNN architectures in image classification. In image classification tasks, pre-trained models initialized with ImageNet weights undergo fine-tuning on smaller, task-specific datasets. This fine-tuning process involves replacing the model's final layers and adjusting the weights at a lower learning rate, thereby significantly reducing training time and costs by utilizing previously learned generic features.

In this study, we conducted an extensive evaluation of various deep-learning models for the classification tasks. The performance of the top three performing models, Inception v3 [50], EfficientNet [51], and a Vision Transformer-based model [52], were reported based on their classification performance.

EfficientNet-B0 [51], the foundational model in the EfficientNet series, employs compound scaling to enhance CNNs by adjusting network width, depth, and resolution. Developed via neural architecture search, this method optimizes both accuracy and computational efficiency by uniformly scaling network dimensions with fixed coefficients, ensuring balanced growth and effectiveness.

Inception v3 [50], developed by Google, is an advanced CNN that improves on its predecessors by employing factorized convolutions and expanded inception modules to reduce parameters without sacrificing depth or width. It also introduces label smoothing to prevent overfitting, enhancing its performance in complex image classification tasks. This architecture optimally balances computational efficiency with robust capabilities, making it highly effective for various image-processing applications.

The Vision Transformer (ViT) [52] adapts the transformer architecture, originally designed for natural language processing, to image classification tasks. ViT splits the input image into fixed-size patches, transforms them into token embeddings with added positional embeddings for spatial context, and processes these through several transformer encoder layers. These layers equipped with self-attention mechanisms, allow the model to capture complex relationships across the image, and feed-forward networks that apply nonlinear transformations to the data. The token embeddings are then processed by a classifier head.

Tumor area segmentation

This stage of the framework identifies negative and positive regions in the patches using a segmentation network trained on pathologist annotations. The U-Net architecture [53], developed for precise image segmentation in the biological domain, features a U-shaped design with an encoder that compresses and a decoder that decompresses. The encoder consists of convolutional layers with ReLU activation and max pooling to reduce spatial dimensions while increasing feature depth. The decoder then restores the feature maps to their original spatial dimensions. U-Net's skip connections link encoder and decoder layers, merging high-level and detailed information in the output. A common approach to improving the architecture's performance involves the integration of advanced encoder architectures. Specifically, the U-Net framework for image segmentation is enhanced by incorporating encoders such as DenseNet [54], which utilizes dense connections, and EfficientNet [51], known for its optimized performance in resource-limited settings. These enhancements facilitate feature reuse, alleviate the issue of vanishing gradients, and enforce

feature propagation, thereby achieving more precise segmentation outcomes. This study presents comparative results between the conventional U-Net and the modified U-Net frameworks employing DenseNet and EfficientNet encoders. DeepLabV3 [55] and DeepLabV3+ [56] are advanced models designed for semantic segmentation, aiming to enhance object segmentation at various scales and achieve more precise boundaries. These models are significant enhancements of the DeepLab series, leveraging deep convolutional neural networks for high-resolution image segmentation. DeepLabV3, introduced by Chen et al., integrates an atrous convolution technique to expand filter ranges and capture context at multiple scales without losing resolution. It features an atrous spatial pyramid pooling (ASPP) module that analyzes a convolutional feature layer using filters with varying sampling rates and effective fields-of-view, effectively capturing objects and context at various scales.

We enhanced the DeepLabV3+ network by incorporating Self Organized Operational Neural Networks (Self-ONN) [57], which have been shown to outperform traditional CNNs. CNNs, with their homogeneous, linear structures, do not fully replicate the complexity of biological neural systems. Addressing these limitations, Generalized Operational Perceptrons (GOPs) and Operational Neural Networks (ONNs) introduce heterogeneous and non-linear architectures. GOPs, drawing inspiration from biological mechanisms, are adept at handling complex tasks where traditional models falter. ONNs extend these advancements by incorporating a variety of operational units per neuron, such as nodal and pool operators, which transcend standard linear convolutions. This approach retains the fundamental CNN principles like weight sharing and localized connectivity while expanding the functional capabilities of the network layers. In our modified architecture, all CNN layers in the DeepLabV3 network were replaced with Self-ONN layers. Figure 4 presents the architecture of the Self-ONN-based DeepLabV3 model. Additionally, we conducted comparative analyses between the original and modified networks to highlight the improved performance of our Self-ONN-based architecture.

Nuclei detection and TPS calculation

The objective of this stage of the framework was to accurately identify the nuclei within both positive and negative tumor regions. For this purpose, we employed a pre-trained network, StarDist [48], which is a deep learning-based method designed for object detection and segmentation in biological images. It distinguishes itself from conventional object detection

techniques by employing star-convex polygons for object representation, as opposed to the traditional use of axis-aligned bounding boxes. This technique involves regressing distances from each pixel within an object to its boundary along a set of predefined radial directions. These distance calculations are only relevant for pixels that have been definitively identified as parts of an object, where object probabilities are ascertained through a predictive model. To further refine object representations, non-maximum suppression is utilized to select the polygons that most accurately represent objects, based on the highest computed object probabilities. These probabilities are determined by the normalized Euclidean distances to the nearest background pixel, focusing on polygons nearer to the object's center for more accurate depictions. The framework employs the U-Net architecture, augmented with an additional convolutional layer designed to enhance feature discrimination before the output phases. Object probabilities are derived from a sigmoid-activated convolutional layer, whereas polygon distances are produced directly, scaled by the number of radial directions without subsequent activation. Collectively, this approach offers a refined and computationally efficient alternative to traditional object detection methods, significantly improving the accuracy of complex image segmentation, especially in medical imaging scenarios where precise object delineation is crucial.

This innovative approach facilitates precise and adaptable modeling of the typically irregular and complex shapes observed in biological microscopy images. A pre-trained model was used in this work which required some image preprocessing. Initially, the patch image was deconvoluted to separate stain channels, specifically to isolate the hematoxylin channel. This channel was then converted to grayscale, and a blurring filter was applied to reduce noise. Subsequently, the StarDist model was utilized to detect nuclei within both the positive and negative tumor regions. The procedure is demonstrated in Figure 5. Any cell detected within the region identified as positive during the segmentation stage was classified as a PD-L1 positive cancer cell. Conversely, any nuclei located within the negative region were classified as negative. Following this, the TPS was calculated as the ratio of the total number of stained positive tumor cells to the total number of viable tumor cells.

Training and testing methodology

A five-fold cross-validation approach was utilized to ensure a comprehensive and accurate evaluation of the deep learning models' performance, dividing the 66 WSIs into five folds. In this approach, performance evaluation can be done on the entire dataset where each fold is used as the

test set once, while the remaining folds are used for training and validation. This fold split was consistently applied to both classification and segmentation training. From the WSI, patches measuring 1000 by 1000 pixels at 40x magnification were created, resulting in a total of 120,360 patches.

Classification and segmentation networks were trained respectively on the patches. Training for both classification and segmentation was conducted over 100 epochs, with the best epoch's result saved based on validation set performance. To prepare the input for the models, all patch images were resized to 224x224 pixels for classification and 256x256 for segmentation, which are the image dimensions required for using the ImageNet weights. Preliminary training involved experimenting with various learning rates, ultimately selecting 0.0001 with the Adam optimizer for its optimal results.

All experiments were conducted on a hardware setup consisting of a NVIDIA GeForce RTX 3090 with 32 GB GPU memory, a 36-core Intel Xeon(R) CPU @ 2.30Ghz, 64GB of system memory, Python 3.9.16 and Pytorch version 1.13.

RESULTS

Performance metrics

The performances of classification and segmentation tasks were evaluated using a comprehensive set of metrics. For the classification task, we utilized precision, recall, F1-score, and accuracy as the primary evaluation metrics. In the segmentation task, the assessment was conducted using Intersection over Union (IoU), Dice Similarity Coefficient (DSC), True Positive Rate (TPR), False Positive Rate (FPR), and specificity. Mathematical formulations of the metrics are detailed below:

$$Precision = \frac{T_p}{T_p + F_p} \quad (1)$$

$$Recall/Sensitivity = \frac{T_p}{T_p + F_n} \quad (2)$$

$$Specificity = \frac{T_n}{T_n + F_p} \quad (3)$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

$$Accuracy = \frac{\sum_{c=1}^{N} TP_c}{N_T} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$FNR = \frac{FN}{TP + FN} \quad (8)$$

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

where T_p/TP is True Positive, F_p/FP is False Positive, F_n/FN is False Negative, T_n/TN is True Negative, and N_T is the number of classes.

Tumor patch classifier performance

This section presents the performance comparison of three different deep learning architectures EfficientNet, Inception v3, and the Vision Transformer model on the tumor patch classification stage. EfficientNet displayed consistent performance across tumor and non-tumor classes with an overall accuracy of 97.5%. For the non-tumor class, the model achieved an accuracy of 97.5%, precision of 99.02%, sensitivity of 97.16%, F1 score of 98.08%, and specificity of 98.15%. Similarly, for the tumor class, it recorded an accuracy of 97.5%, precision of 94.71%, sensitivity of 98.15%, F1 score of 96.4%, and specificity of 97.16%. The weighted average metrics were 97.5% for accuracy, 97.55% for precision, 97.5% for sensitivity, 97.51% for the F1 score, and 97.81% for specificity. Inception v3 demonstrated a slightly lower overall accuracy of 96.32% compared to EfficientNet. The non-tumor classification results showed an accuracy of 96.32%, precision of 95.91%, sensitivity of 98.63%, F1 score of 97.25%, and specificity of 91.87%. For tumor detection, the model recorded an accuracy of 96.32%, precision of 97.19%, sensitivity of 91.87%, F1 score of 94.46%, and specificity of 98.63%. The weighted average figures for Inception v3 were an accuracy of 96.32%, precision of 96.35%, sensitivity of 96.32%, F1 score of 96.3%, and specificity of 94.18%. The Vision Transformer model showed an overall accuracy equal to EfficientNet, at 97.53%. For the non-tumor class, it achieved an accuracy of 97.53%, precision of 99.12%, sensitivity of 97.12%, F1 score of 98.11%, and specificity of 98.33%. In tumor classification, it scored an accuracy of 97.53%, precision of 94.64%, sensitivity of 98.33%, F1 score of 96.45%, and specificity of 97.12%. The weighted averages were 97.53% for accuracy, 97.59% for precision, 97.53% for sensitivity, 97.54% for the F1 score, and 97.92% for specificity. The Vision Transformer model matched the overall accuracy of EfficientNet, achieving 97.53%, and exhibited superior sensitivity in tumor detection. This model's performance highlights its

potential for applications requiring high sensitivity to avoid missing tumor cases. The confusion matrices of the models are shown in Figure 6.

Tumor area segmentation performance

A comparative evaluation of five segmentation models, including UNet, EfficientNet UNet, DenseNet UNet, DeepLabV3+, and SelfONN-based DeepLabV3+, are shown in this section. The DSC values for various models were analyzed to evaluate their segmentation performance for positive and negative tumors. UNet achieved an overall DSC of 76.28%, demonstrating consistent performance with DSC values of 76.31% for positive tumors and 76.24% for negative tumors. EfficientNet UNet significantly improved overall DSC to 82.33%, with a remarkable 87.566% for positive tumors and 77.092% for negative tumors, indicating strong positive tumor segmentation capabilities. DenseNet UNet showed similar robustness with an overall DSC of 81.89%, achieving 86.57% for positive tumors and 77.21% for negative tumors. DeepLabV3+ exhibited excellent performance in positive tumor segmentation with a DSC of 89.62%, though its performance in negative tumor segmentation was lower, with a DSC of 73.62%, resulting in an overall DSC of 81.62%. The SelfONN-enhanced DeepLabV3+ achieved the highest overall DSC of 83.47%, maintaining high performance in both positive (DSC of 89.58%) and negative tumors (DSC of 77.36%). These results indicate that the SelfONN enhancement particularly enhances the model's robustness, providing the most balanced and effective segmentation performance among the models evaluated. Table 2 shows the segmentation performance of the different models. Figure 7 shows segmentation prediction on patch images.

Automated PD-L1 scoring performance

After the segmentation stage, to calculate the TPS of a WSI, the nuclei within the positive and negative areas need to be identified. The pre-trained model StarDist was used for this purpose. Figure **8** shows examples of detected cell nuclei in patch images. The TPS of a WSI was calculated as:

$$\text{TPS} = \frac{\#\ of\ nuclei\ in\ positive\ region}{\#\ of\ nuclei\ in\ positive\ region\ and\ negative\ region} \tag{10}$$

Automated PD-L1 expression was assessed based on the TPS, which was categorized into three levels: negative (TPS < 1%), low expression (TPS 1-49%), and high expression (TPS ≥50%). According to the ground truth data provided in the dataset, there were 30 cases with negative

expression, 22 with low expression, and 14 with high expression. The automated framework achieved an accuracy of 96.67% for negative cases, 86.36% for low-expression cases, and 100% for high-expression cases. An overall accuracy of 93.94% was attained, accompanied by an F1 score of 93.89%. Figure 9 shows the confusion matrix. The correlation between the ground truth TPS and the TPS predicted by the framework was 0.9635.

DISCUSSION

This study aimed to develop an automated framework for assessing PD-L1 expression in NSCLC using WSIs from surgical resections. The objective was to automate the evaluation of the TPS to enhance clinical decision-making for ICI. Our approach comprised three key stages: tumor patch classification, tumor area segmentation, and nuclei detection.

Initially, the framework identified patches containing tumor cells while discarding those with artifacts or non-tumor cells through a classification stage. This stage demonstrated robust performance across three deep learning models: EfficientNet, Inception v3, and Vision Transformer. Both EfficientNet and Vision Transformer achieved an overall accuracy of 97.5%, surpassing Inception v3, which attained an accuracy of 96.32%. Notably, the Vision Transformer model exhibited superior sensitivity in tumor detection, underscoring its potential for applications requiring high sensitivity to avoid missing tumor cases. This may be particularly important in small biopsy samples or in cytology (e.g., cell blocks). Most existing methodologies do not employ artifact filtering, which can result in false predictions and failure in real-world scenarios. The automated framework proposed in this study processed entire slide images, accounting for artifacts and positive ICs, which should be discarded for more accurate predictions. This consideration is crucial for enhancing prediction accuracy and reliability.

The subsequent stage of the framework involved a segmentation network designed to predict positive and negative tumor regions within a patch. The segmentation performance of the models varied, with the SelfONN-enhanced DeepLabV3+ achieving the highest overall DSC of 83.47%. This model exhibited superior performance in both positive (DSC of 89.58%) and negative tumor areas (DSC of 77.36%), demonstrating its robustness and effectiveness in segmenting complex tumor regions. The incorporation of Self-ONN into the DeepLabV3+ architecture significantly enhanced performance. Previous studies primarily reported the performance of single models, particularly basic UNet models. In contrast, this work presents a comparative analysis of

segmentation performance across various networks, including UNet with high-level encoders capable of capturing diverse and deeper features efficiently, and the modified DeepLabV3+ architecture, which represents a state-of-the-art segmentation network.

The final stage of the framework involved a deep learning network called StarDist, pre-trained to detect cell nuclei in WSIs. The number of nuclei in both positive and negative regions was determined, and the TPS was subsequently calculated for the WSI. The most common method for calculating the number of nuclei, employed by most studies, is the watershed algorithm, which relies solely on image processing. Consequently, it is susceptible to issues such as hard stains and low-intensity cells. The StarDist model overcame these limitations and performed better across various scenarios, making it a more practical and reliable method. The automated framework's performance in calculating the TPS from WSIs showed a strong correlation with the ground truth data, achieving an overall accuracy of 93.94% and an F1 score of 93.89%. The framework performed exceptionally well in identifying high expression cases (TPS $\geq$ 50%) with 100% accuracy, though it exhibited slightly lower accuracy (86.36%) in low expression cases (TPS 1-49%). This discrepancy highlights the ongoing challenge of accurately quantifying low PD-L1 expression levels, which is crucial for patient management and treatment planning.

Although the study achieved promising results, it comes with certain limitations. While it performed well in negative cases, it encountered difficulties detecting unusually shaped tumor cell nuclei, which are uncommon. Additionally, cases with high PD-L1 expression exhibited heterogeneous staining, leading to some inaccuracies. The study was conducted with only 66 specimens; although these were large surgical specimens, increasing the number of cases could introduce greater variability in cell morphology and stain patterns, enhancing the model's generalizability. Moreover, while the human-machine collaborative annotation was effective in this study, more extensive and detailed human annotations could further improve model performance due to the subjective nature of the problem.

Scopes of future work can focus on further refining the segmentation and classification models, particularly in improving the accuracy of low-expression PD-L1 cases. Additionally, the inclusion of a more diverse range of NSCLC subtypes to expand the dataset and PD-L1 staining patterns will be essential for generalizing the model's applicability. The incorporation of multimodal data, such as genomic and clinical information, could also enhance the predictive power of the framework.

CONCLUSION

The automated framework developed in this study shows significant promise in the field of digital pathology, offering a valuable tool for the accurate and efficient assessment of PD-L1 expression in NSCLC. By harnessing the power of deep learning, this framework provides a reliable, scalable method for automating PD-L1 TPS evaluation, which is a critical factor in determining eligibility for ICI. This study developed a robust deep learning-based model that achieved high accuracy across several essential tasks, including tumor patch classification, segmentation, and cancer cell nuclei detection. The framework demonstrated impressive performance in distinguishing between tumor and non-tumor regions and accurately identifying PD-L1-positive cancer cell nuclei, a key component in TPS calculation. The ability of this automated system to consistently and accurately quantify PD-L1 expression highlights its potential as a powerful tool for pathologists, helping to reduce the subjectivity and variability that often arise in manual assessments. The findings of this research underscore the broader potential of AI-driven solutions in improving diagnostic accuracy, streamlining workflows in pathology labs, and facilitating more precise patient selection for treatment with ICI. By integrating such automated systems into clinical practice, there is a significant opportunity to enhance both the speed and consistency of cancer diagnosis and treatment decisions. As the healthcare sector continues to adopt AI-based technologies, this framework could play a critical role in optimizing personalized treatment strategies, improving patient outcomes, and potentially reducing the overall cost and burden of cancer care. Future work could further refine the model, extend its applicability to other cancer types, and integrate it into routine clinical workflows, reinforcing the growing role of AI in advancing precision medicine.

Conflicts of interest: Authors declare no conflicts of interest.
Funding: Authors received no specific funding for this work.
Data availability: The processed dataset used in this study can be made available upon a reasonable request to the corresponding author.
Submitted: 14 January 2025
Accepted: 1 March 2025
Published online: 3 March 2025

REFERENCES

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, 2021.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 68, no. 6, pp. 394–424, 2018.

[3] N. C. I. (US). S. Program, S. P. (National C. I. (US)), and N. C. for H. S. (US), *Cancer statistics review*. US Department of Health and Human Services, Public Health Service, National …, 1973.

[4] P. A. Ott, F. S. Hodi, H. L. Kaufman, J. M. Wigginton, and J. D. Wolchok, "Combination immunotherapy: a road map," *J Immunother Cancer*, vol. 5, pp. 1–15, 2017.

[5] Y. Iwai, J. Hamanishi, K. Chamoto, and T. Honjo, "Cancer immunotherapies targeting the PD-1 signaling pathway," *J Biomed Sci*, vol. 24, pp. 1–11, 2017.

[6] W. Zou, J. D. Wolchok, and L. Chen, "PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations," *Sci Transl Med*, vol. 8, no. 328, pp. 328rv4-328rv4, 2016.

[7] K. Shien, V. A. Papadimitrakopoulou, and I. I. Wistuba, "Predictive biomarkers of response to PD-1/PD-L1 immune checkpoint inhibitors in non–small cell lung cancer," *Lung Cancer*, vol. 99, pp. 79–87, 2016.

[8] Y. Iwai, M. Ishida, Y. Tanaka, T. Okazaki, T. Honjo, and N. Minato, "Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade," *Proceedings of the National Academy of Sciences*, vol. 99, no. 19, pp. 12293–12297, 2002.

[9] C. K. Lee *et al.*, "Clinical and molecular characteristics associated with survival among patients treated with checkpoint inhibitors for advanced non–small cell lung carcinoma: a systematic review and meta-analysis," *JAMA Oncol*, vol. 4, no. 2, pp. 210–216, 2018.

[10] A. R. Almutairi *et al.*, "Comparative efficacy and safety of immunotherapies targeting the PD-1/PD-L1 pathway for previously treated advanced non-small cell lung cancer: A Bayesian network meta-analysis," *Crit Rev Oncol Hematol*, vol. 142, pp. 16–25, 2019.

[11] L. Gandhi *et al.*, "Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer," *New England journal of medicine*, vol. 378, no. 22, pp. 2078–2092, 2018.

[12] S. Gettinger *et al.*, "Nivolumab monotherapy for first-line treatment of advanced non–small-cell lung cancer," *Journal of Clinical Oncology*, vol. 34, no. 25, pp. 2980–2987, 2016.

[13] E. B. Garon *et al.*, "Pembrolizumab for the treatment of non–small-cell lung cancer," *New England Journal of Medicine*, vol. 372, no. 21, pp. 2018–2028, 2015.

[14] A. A. Davis and V. G. Patel, "The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors," *J Immunother Cancer*, vol. 7, no. 1, p. 278, 2019.

[15] E. Munari *et al.*, "PD-L1 Assays 22C3 and SP263 are Not Interchangeable in Non–Small Cell Lung Cancer When Considering Clinically Relevant Cutoffs: An Interclone Evaluation by Differently Trained Pathologists," *Am J Surg Pathol*, vol. 42, no. 10, pp. 1384–1389, 2018.

[16] S. Vranic and Z. Gatalica, "PD-L1 testing by immunohistochemistry in immuno-oncology," *Biomolecules and Biomedicine*, vol. 23, no. 1, p. 15, 2023.

[17] D. N. America, "PD-L1 IHC 22C3 pharmDx interpretation manual," *2018*, 2018.

[18] M. Reck *et al.*, "Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer," *New England Journal of Medicine*, vol. 375, no. 19, pp. 1823–1833, 2016.

[19] A. D'incecco *et al.*, "PD-1 and PD-L1 expression in molecularly selected non-small-cell lung cancer patients," *Br J Cancer*, vol. 112, no. 1, pp. 95–102, 2015.

[20] Z. Gatalica *et al.*, "Programmed cell death 1 (PD-1) and its ligand (PD-L1) in common cancers and their correlation with molecular cancer type," *Cancer epidemiology, biomarkers & prevention*, vol. 23, no. 12, pp. 2965–2970, 2014.

[21] K. Pawelczyk *et al.*, "Role of PD-L1 expression in non-small cell lung cancer and their prognostic significance according to clinicopathological factors and diagnostic markers," *Int J Mol Sci*, vol. 20, no. 4, p. 824, 2019.

[22] E. Vigliar *et al.*, "The Reproducibility of the Immunohistochemical PD-L1 Testing in Non-Small-Cell Lung Cancer: A Multicentric Italian Experience," *Biomed Res Int*, vol. 2019, no. 1, p. 6832909, 2019.

[23] H. Yu, T. A. Boyle, C. Zhou, D. L. Rimm, and F. R. Hirsch, "PD-L1 expression in lung cancer," *Journal of Thoracic Oncology*, vol. 11, no. 7, pp. 964–975, 2016.

[24] H. Brunnström *et al.*, "PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability," *Modern Pathology*, vol. 30, no. 10, pp. 1411–1421, 2017.
[25] S. Lantuejoul, D. Damotte, V. Hofman, and J. Adam, "Programmed death ligand 1 immunohistochemistry in non-small cell lung carcinoma," *J Thorac Dis*, vol. 11, no. Suppl 1, p. S89, 2019.
[26] S. Kabir *et al.*, "The utility of a deep learning-based approach in Her-2/neu assessment in breast cancer," *Expert Syst Appl*, vol. 238, p. 122051, 2024.
[27] M. E. H. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *Ieee Access*, vol. 8, pp. 132665–132676, 2020.
[28] N. Coudray *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nat Med*, vol. 24, no. 10, pp. 1559–1567, 2018.
[29] T. Rahman *et al.*, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, 2020.
[30] B. E. Bejnordi, G. Litjens, and J. A. W. M. van der Laak, "Machine learning compared with pathologist assessment—Reply," *JAMA*, vol. 319, no. 16, p. 1726, 2018.
[31] P.-H. C. Chen *et al.*, "An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis," *Nat Med*, vol. 25, no. 9, pp. 1453–1457, 2019.
[32] T. Mahmood, M. Arsalan, M. Owais, M. B. Lee, and K. R. Park, "Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs," *J Clin Med*, vol. 9, no. 3, p. 749, 2020.
[33] A. R. Zlotta and A. Sadeghian, "Re: artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study," *Eur Urol*, vol. 78, no. 2, pp. 290–291, 2020.
[34] J. Saltz *et al.*, "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images," *Cell Rep*, vol. 23, no. 1, pp. 181–193, 2018.
[35] A. Kapil *et al.*, "Deep semi supervised generative learning for automated tumor proportion scoring on NSCLC tissue needle biopsies," *Sci Rep*, vol. 8, no. 1, p. 17343, 2018.
[36] J. Liu *et al.*, "Automated tumor proportion score analysis for PD-L1 (22C3) expression in lung squamous cell carcinoma," *Sci Rep*, vol. 11, no. 1, p. 15907, 2021.
[37] B. Pan *et al.*, "Automated tumor proportion scoring for PD-L1 expression based on multistage ensemble strategy in non-small cell lung cancer," *J Transl Med*, vol. 19, no. 1, p. 249, 2021.
[38] C. R. Taylor *et al.*, "A multi-institutional study to evaluate automated whole slide scoring of immunohistochemistry for assessment of Programmed Death-Ligand 1 (PD-L1) expression in non–small cell lung cancer," *Applied Immunohistochemistry & Molecular Morphology*, vol. 27, no. 4, pp. 263–269, 2019.
[39] J. Wu *et al.*, "Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer," *Modern Pathology*, vol. 35, no. 3, pp. 403–411, 2022.
[40] L. M. Hondelink *et al.*, "Development and validation of a supervised deep learning algorithm for automated whole-slide programmed death-ligand 1 tumour proportion score assessment in non-small cell lung cancer," *Histopathology*, vol. 80, no. 4, pp. 635–647, 2022.
[41] L. Sha *et al.*, "Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images," *J Pathol Inform*, vol. 10, no. 1, p. 24, 2019.
[42] E. Vigliar *et al.*, "PD-L1 expression on routine samples of non-small cell lung cancer: results and critical issues from a 1-year experience of a centralised laboratory," *J Clin Pathol*, vol. 72, no. 6, pp. 412–417, 2019.
[43] M. P. Humphries, P. Maxwell, and M. Salto-Tellez, "QuPath: The global impact of an open source digital pathology system," *Comput Struct Biotechnol J*, vol. 19, pp. 852–859, 2021.
[44] M. P. Humphries *et al.*, "Improving the diagnostic accuracy of the PD-L1 test with image analysis and multiplex hybridization," *Cancers (Basel)*, vol. 12, no. 5, p. 1114, 2020.
[45] Z. Huang *et al.*, "A new AI-assisted scoring system for PD-L1 expression in NSCLC," *Comput Methods Programs Biomed*, vol. 221, p. 106829, 2022.
[46] G. Cheng *et al.*, "Artificial intelligence-assisted score analysis for predicting the expression of the immunotherapy biomarker PD-L1 in lung cancer," *Front Immunol*, vol. 13, p. 893198, 2022.
[47] H. Ito *et al.*, "A Deep Learning–Based Assay for Programmed Death Ligand 1 Immunohistochemistry Scoring in Non–Small Cell Lung Carcinoma: Does it Help Pathologists Score?," *Modern Pathology*, vol. 37, no. 6, p. 100485, 2024.
[48] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, Springer, 2018, pp. 265–273.

[49] M. M. Hossen *et al.*, "GCDN-Net: Garbage classifier deep neural network for recyclable urban waste management," *Waste Management*, vol. 174, pp. 439–450, 2024.
[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
[51] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
[52] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 2015, pp. 234–241.
[54] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
[55] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation. arXiv," *arXiv preprint arXiv:1706.05587*, vol. 5, 2017.
[56] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
[57] S. Kiranyaz, J. Malik, H. Ben Abdallah, T. Ince, A. Iosifidis, and M. Gabbouj, "Self-organized operational neural networks with generative neurons," *Neural Networks*, vol. 140, pp. 294–308, 2021.

TABLES AND FIGURES WITH LEGENDS

Table 1: Baseline characteristics of the NSCLC patient cohort.

Characteristic		Dataset (N = 66)
Age, years	Mean Range	66 50-82
Sex	Men Women	36 (54.5%) 30 (45.5%)
Specimen site	Primary (lung) Metastatic (lymph nodes)	64 (97%) 2 (3%)
Tumor type	Adenocarcinoma Squamous cell carcinoma Other subtypes of NSCLC	46 (69.7%) 13 (19.7%) 7 (10.6%)
TPS	<1% 1-49% ≥50%	30 (45.5%) 22 (33.3%) 14 (21.2%)

TPS – Tumor proportion score; NSCLC – Non-small cell lung cancer

Table 2: Segmentation performance comparison of different architectures.

Model	Accuracy (%)	IoU (%)	DSC (%)	IoU (Positive Tumor) (%)	DSC (Positive Tumor) (%)	IoU (Negative Tumor) (%)	DSC (Negative Tumor) (%)
UNet	67.97	69.07	76.28	71.12	76.31	67.02	76.24
EfficientNet UNet	70.782	76.73	82.33	84.72	87.56	68.74	77.09
DenseNet UNet	70.88	76.44	81.89	83.6	86.57	69.28	77.21
DeepLabV3+	71.58	75.27	81.62	86.89	89.62	63.66	73.62
DeepLabV3+ (SelfONN)	71.25	77.69	83.47	86.79	89.58	68.59	77.36

IoU – Intersection over Union; DSC – Dice Similarity Coefficient

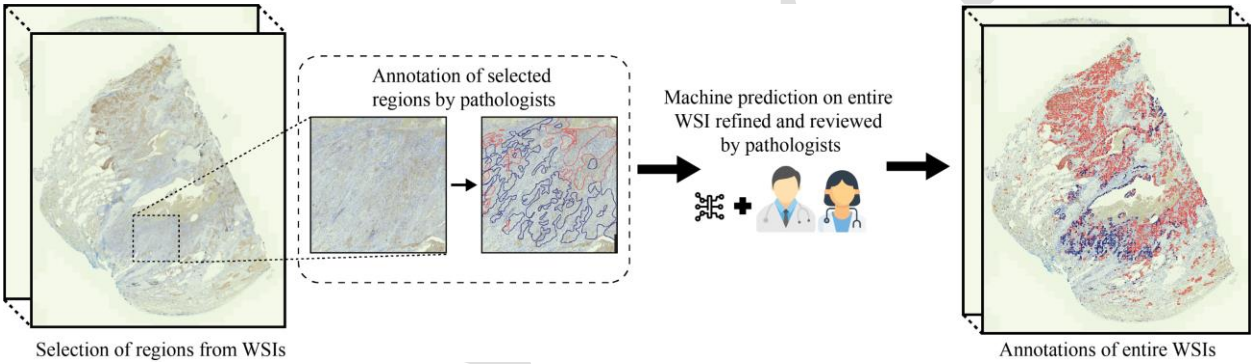

Figure 1: **Annotation procedure based on human-machine collaborative approach.**

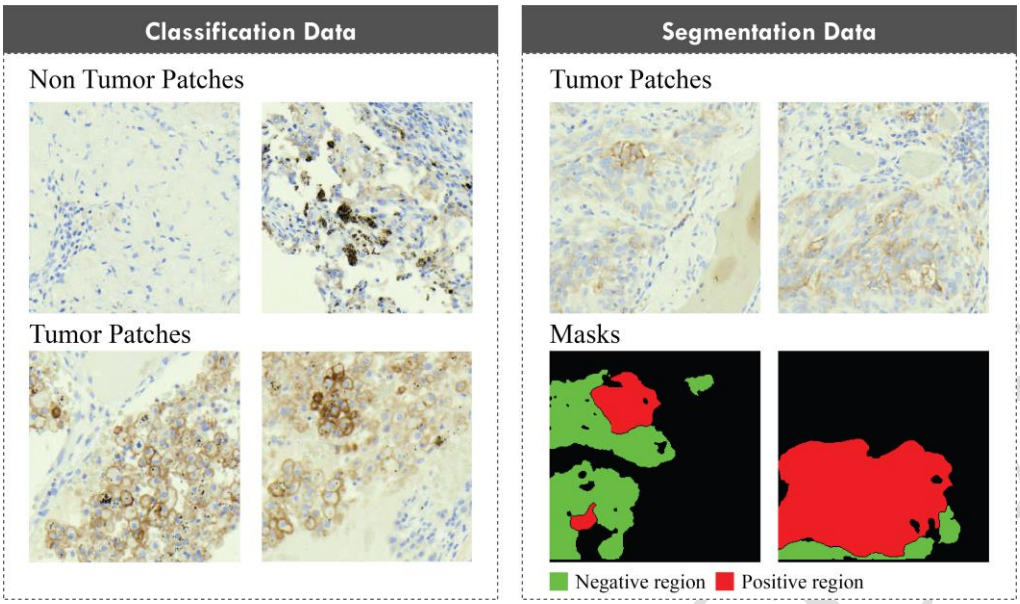

Figure 2: Example of patches in classification and segmentation dataset.

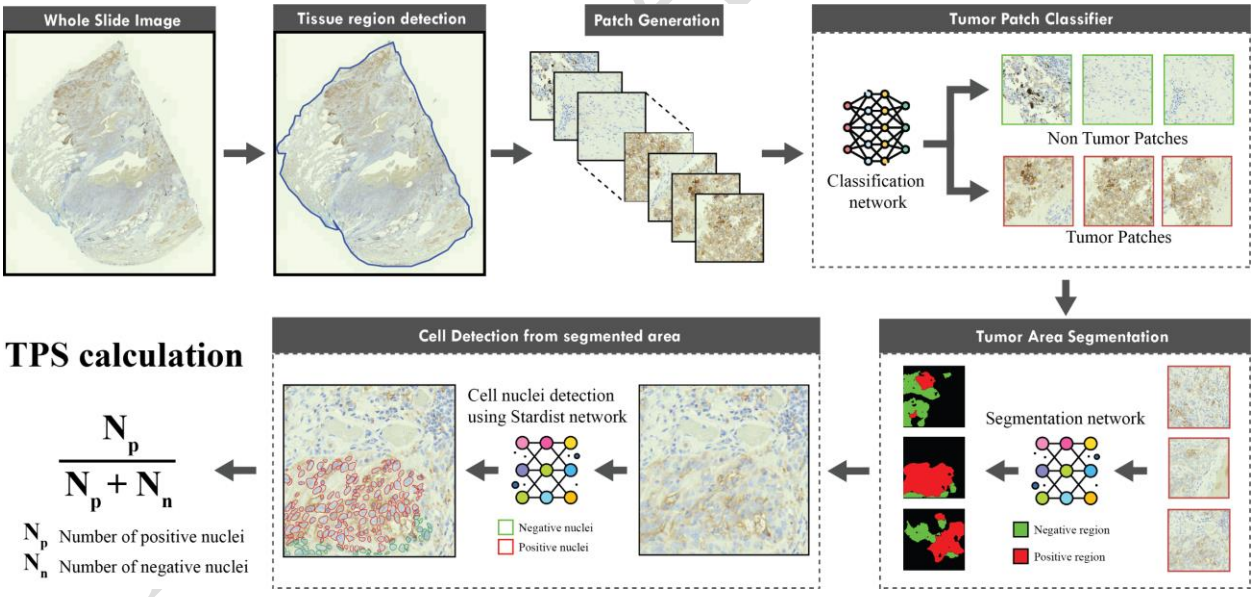

Figure 3: A proposed automated framework for TPS calculation of PD-L1 expression.

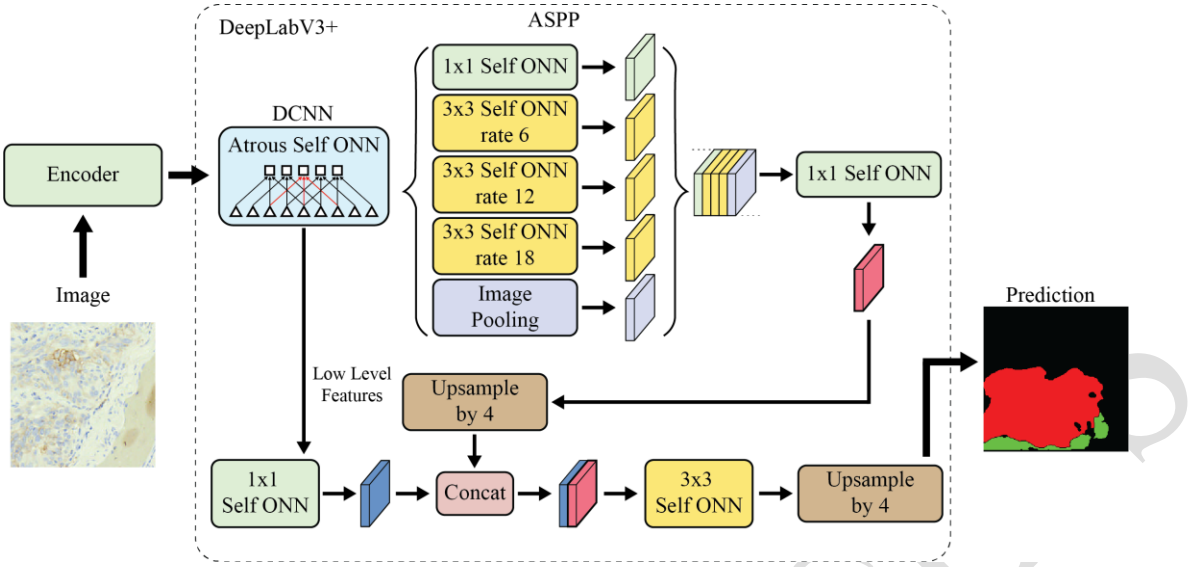

Figure 4: **Architecture of modified DeepLabV3+ network.**

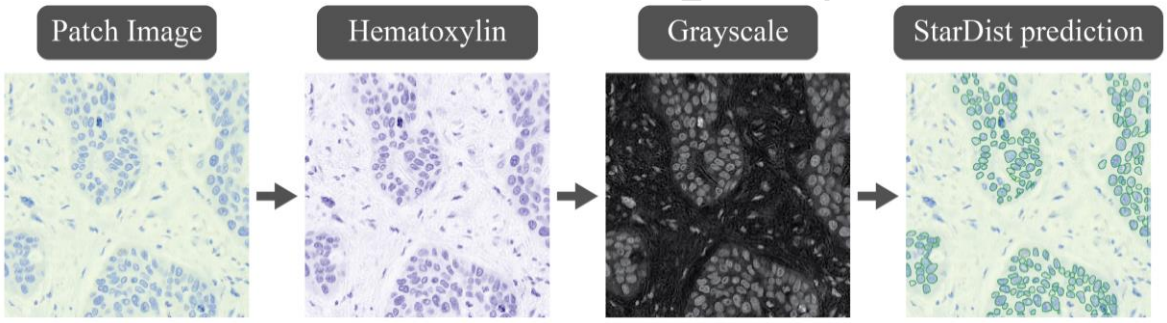

Figure 5: **Detection of cell nuclei using the StarDist network.**

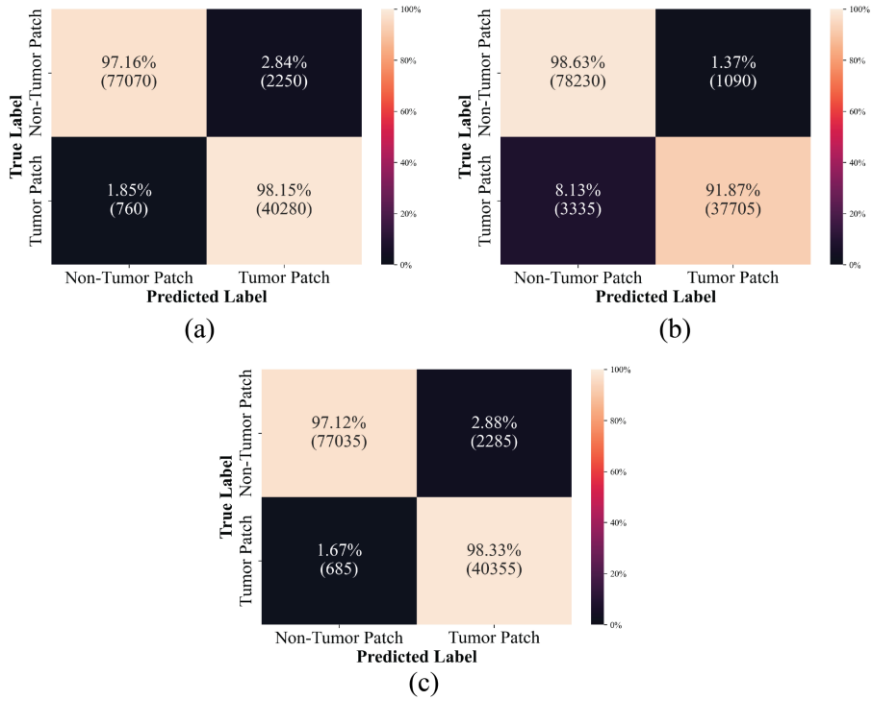

Figure 6: Confusion matrix of (a) EfficientnetB0 (b) Inception v3 and (c) Vision Transformer model.

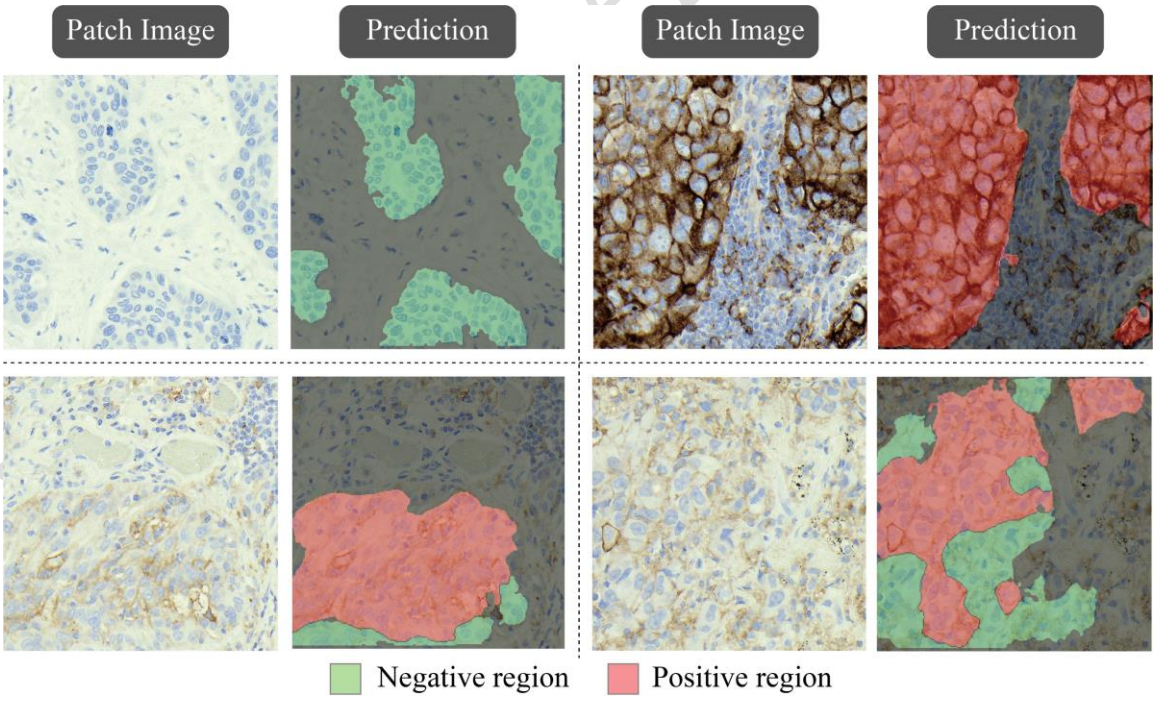

Figure 7: Visualization of the segmentation model performance.

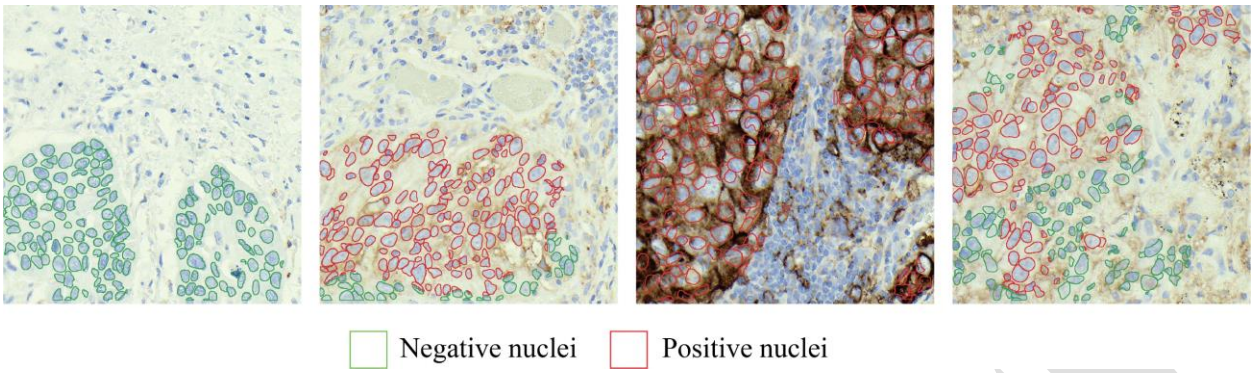

Figure 8: Example of cell nuclei detection using the StarDist model.

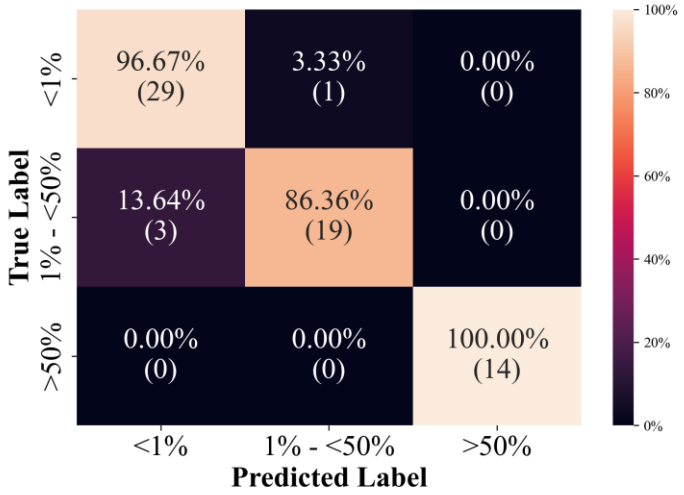

Figure 9: Confusion matrix of the classification of TPS level.