Indian Institute of Technology Kanpur

SURGE INTERNSHIP REPORT

Automated Tumor Proportion Score Analysis

Student Name	Student ID
1. Ajay Singh Panwar	2430150

Guide: Dr. Tushar Sandhan

Submission Date : 12/07/2024

1 Abstract

This study focuses on the automated analysis of tumor proportion score (TPS) for PD-L1 expression in lung squamous cell carcinoma. PD-L1 is a crucial biomarker in predicting responses to immunotherapy, making its accurate assessment vital. We developed a machine learning-based approach to segment positively stained tumor regions and calculate the TPS, aiming to enhance diagnostic accuracy and efficiency.

Our methodology integrates several advanced image processing techniques, including Contrast Limited Adaptive Histogram Equalization (CLAHE), Gaussian blur, Otsu's thresholding, watershed segmentation, and contour detection using the OpenCV library. These techniques ensure precise segmentation, annotation, and quantification of tumor regions. The calculated TPS is derived from the proportion of PD-L1 positive tumor cells relative to the total viable tumor cells, providing a robust metric for clinical decision-making. The results demonstrate that our automated system can significantly reduce the subjectivity and time required for manual scoring, enhancing the reliability of PD-L1 evaluations. This innovation supports faster and more accurate clinical decision-making, potentially improving patient outcomes and advancing personalized medicine in lung cancer treatment.

2 Keywords

PDL1, TPS, IMMUNOTHERAPY, SEGMENTATION.

3 Introduction

Lung cancer, one of the most prevalent and deadly forms of cancer, poses significant challenges in diagnosis and treatment. It is broadly categorized into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC, which includes adenocarcinoma, squamous cell carcinoma, and large cell carcinoma, accounts for approximately 85% of all lung cancer cases. Squamous cell carcinoma, a subtype of NSCLC, is particularly aggressive and is often diagnosed at an advanced stage, complicating treatment options.

One of the essential methods to detect lung cancer is through histopathological analysis of tissue samples. Hematoxylin and eosin (H&E) stained images are commonly used for this purpose. These images provide detailed insights into the cellular architecture and morphology, enabling pathologists to identify cancerous cells. However, manual interpretation of these images is subjective and time-consuming, leading to variability in diagnosis.

In recent years, the integration of biomarkers has enhanced the diagnostic accuracy for lung cancer, especially for squamous cell carcinoma. Programmed cell death ligand-1 (PD-L1) is a significant biomarker used in immunohistochemistry (IHC) assays to predict the response to anti-PD-1/PD-L1 therapy. PD-L1 expression on tumor cells helps them evade the immune system by interacting with PD-1 receptors on T-cells. Inhibitors that block this interaction can restore

the immune response, making PD-L1 a crucial target in cancer immunotherapy.

PD-L1 (Programmed Death-Ligand 1) is a protein that plays a crucial role in suppressing the immune system by binding to its receptor PD-1 on T-cells. This interaction helps tumor cells evade the immune system, making PD-L1 a significant target in cancer immunotherapy.

The Tumor Proportion Score (TPS) is a key metric used by pathologists to assess PD-L1 expression. It is calculated as the percentage of PD-L1 positive tumor cells relative to the total number of viable tumor cells. This assessment, traditionally performed manually, is not only labor-intensive but also prone to inter- and intra-observer variability. Thus, there is a growing need to automate this process to ensure consistency and efficiency.

$$TPS = \left(\frac{Total\ number\ of\ PD-L1\ positive\ TCs}{Total\ number\ of\ viable\ TCs} \right) \times 100$$

Automated systems leveraging advanced image processing and deep learning techniques have shown promise in addressing these challenges. These systems can analyze whole slide images (WSIs) and provide accurate TPS assessments, mimicking the diagnostic workflow of pathologists. By reducing the subjectivity and time required for manual scoring, automated systems can significantly enhance the reliability of PD-L1 evaluations and support clinical decision-making.

4 Objectives and Learning Outcomes

1. To optimize the subjectivity and time consumption inherent in current methods used by pathologists. we will involve extensive work on both data and models. Therefore, we will focus on:

   (a) Implement a machine learning algorithm for the segmentation of tumor regions utilizing thresholding-based segmentation, specifically employing Otsu's thresholding method.

   (b) The primary objective of this project is to segment positively stained tumor regions and predict the mask. This is achieved by applying thresholding, followed by segmentation using the watershed algorithm and contour detection to facilitate mask prediction.

   (c) we implement a machine learning technique for counting positive and total tumor cells. For this purpose, we use contour detection and draw circles around each detected object for annotation.

   (d) Finally, the objective is to calculate the Tumor Proportional Score (TPS) based on the segmented and annotated information.

5 Methodology and Results

5.1 Biology Understanding and Data Pre-processing

Initially, I reviewed several published research papers related to the project theme, identifying key points for image analysis and focusing on the biological mechanisms and significance of PD-L1 expression in tumors.

We obtained microscopic images of tumors stained with PDL1 andH&E directly from doctors. Additionally, we conducted image preprocessing on these images. This preprocessing involved careful cropping of the images to suit our specific purposes and manual marking of positive strained regions within the tumor region.

5.2 Tumor Segmentation

For the segmentation of tumor regions in a grayscale image, we employ a comprehensive sequence of advanced image processing techniques. Initially, the image undergoes enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE), which is specifically designed to enhance the contrast and improve the visualization of subtle details within the image. This enhancement is crucial as it allows for better differentiation of the tumor region from the surrounding tissue.

Following the enhancement, a Gaussian blur is applied to the image. This step is essential for achieving smoother transitions and reducing noise, which can otherwise interfere with the accuracy of the subsequent segmentation process. The Gaussian blur helps in creating a more uniform appearance of the image, making the tumor boundaries more distinct.

Once the image is enhanced and smoothed, Otsu's thresholding method is utilized to segment the tumor region. This method automatically determines the optimal threshold value based on the image's histogram, effectively separating the tumor areas from the background. Otsu's thresholding is particularly advantageous as it adapts to the specific characteristics of the image, ensuring a more precise segmentation.

To further refine the segmentation, morphological operations are employed. These operations are crucial for eliminating small objects and noise that may have been incorrectly identified as part of the tumor during the initial segmentation. By using techniques such as erosion and dilation, the binary image is cleaned up, resulting in a more accurate and robust delineation of the tumor regions.

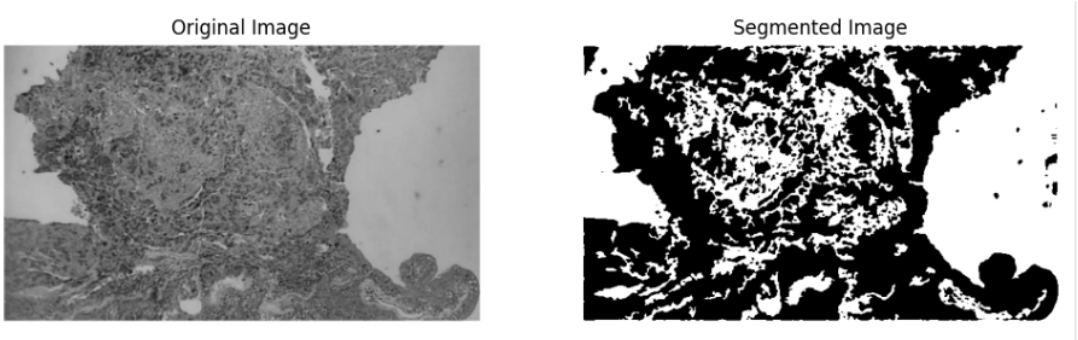


Figure 1

5.3 Mask Predection

We implement a comprehensive multi-step approach to segment positively stained tumor regions from a given image. The process begins by loading the image and converting it to grayscale. The first critical step involves the use of Otsu's thresholding method to create a binary image. This method automatically determines the optimal threshold value based on the histogram of pixel intensities, effectively highlighting tumor regions against a dark background.

Following this, we compute the distance transform, which measures the distance of each pixel from the nearest zero-valued pixel in the binary image. This distance map is crucial for identifying local maxima, which serve as markers in the subsequent watershed segmentation algorithm. These markers guide the watershed algorithm in delineating distinct tumor regions, thereby refining the initial segmentation and ensuring that the boundaries of tumor regions are accurately defined. Contours within each segmented region are then detected to provide precise outlines of the tumor areas. To highlight and annotate individual tumor regions, bounding circles are drawn around the detected contours. This step not only enhances the visualization of tumor areas but also aids in subsequent quantitative analysis.

To further improve the quality of the segmented image, we apply Contrast Limited Adaptive Histogram Equalization (CLAHE). This technique enhances the contrast of the image. Additionally, a Gaussian blur is applied to smooth transitions and reduce noise, which helps in achieving a cleaner segmentation result and Finally, additional thresholding and noise reduction techniques are employed to produce a refined segmentation. These steps eliminate any remaining artifacts and ensure that only the most relevant tumor regions are highlighted.

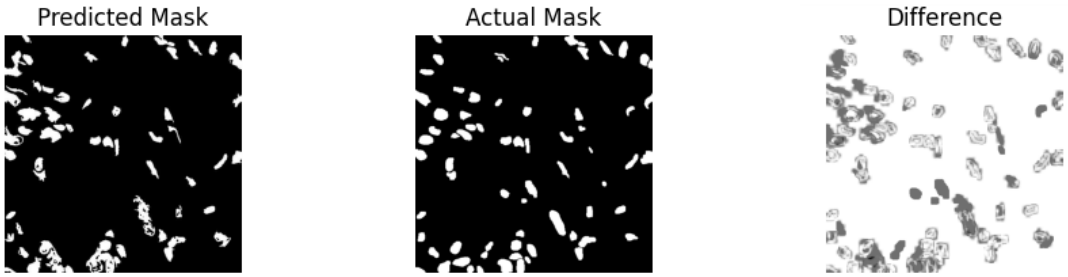
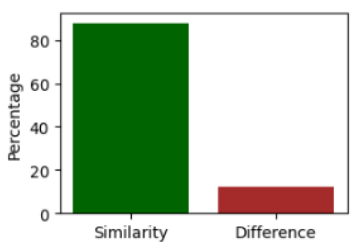
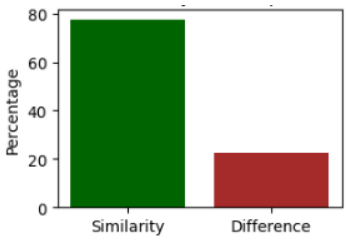


Figure 2

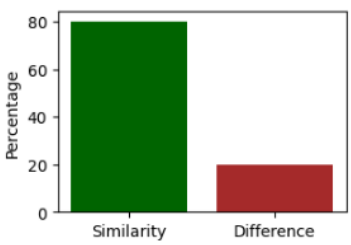
Figure 3



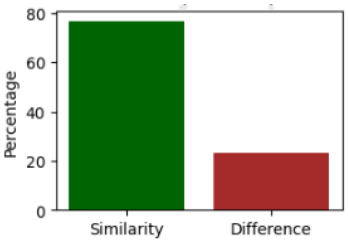
Accuracy: 87.93



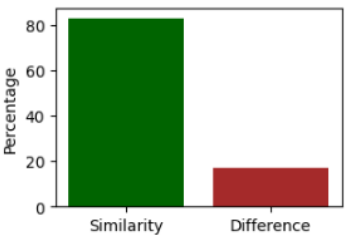
Accuracy: 77.69



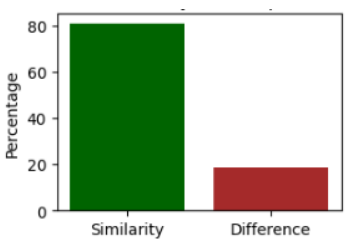
Accuracy: 80.24



Accuracy: 76.85



Accuracy: 83.13



Accuracy: 81.20

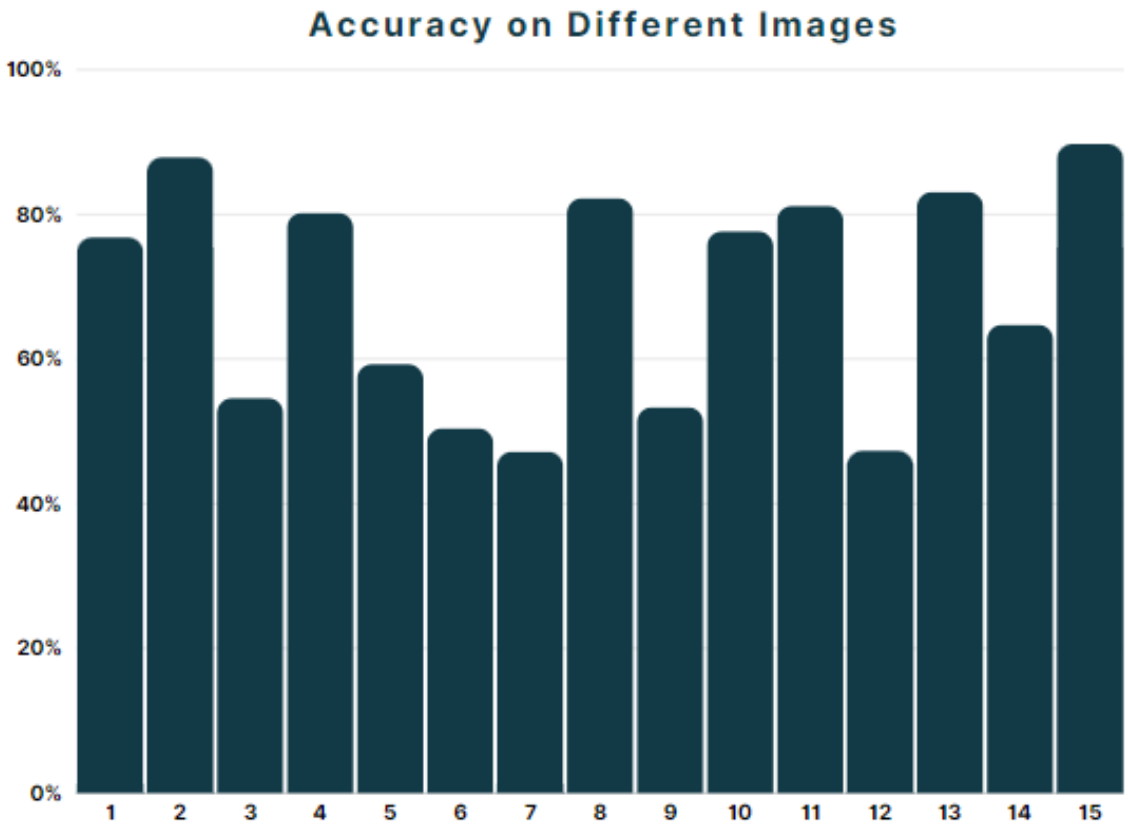


Figure 4

5.4 Cell Counting

We perform bounding box detection on microscopic images of tumor cells to accurately localize and characterize tumor regions. Initially, contours are detected within each segmented region using the OpenCV library. This involves identifying the edges of tumor areas within the image, which serve as the basis for further analysis.

Once the contours are detected, bounding rectangles are drawn around these contours to enclose and annotate each identified tumor area. The use of bounding boxes provides a clear and precise delineation of tumor regions, facilitating easy identification and subsequent analysis. This approach combines advanced thresholding techniques with sophisticated contour detection methods to achieve highly accurate localization of tumors. Thresholding helps to distinguish tumor regions from the surrounding tissue by converting the image into a binary format, where the tumor areas are highlighted. Contour detection then traces the boundaries of these highlighted regions, allowing for the accurate drawing of bounding rectangles.

The integration of these techniques ensures that each tumor region is effectively enclosed and annotated, providing a robust foundation for further quantitative analysis. This method enhances the ability to accurately characterize tumors in medical images, supporting improved diagnostic and treatment planning processes.

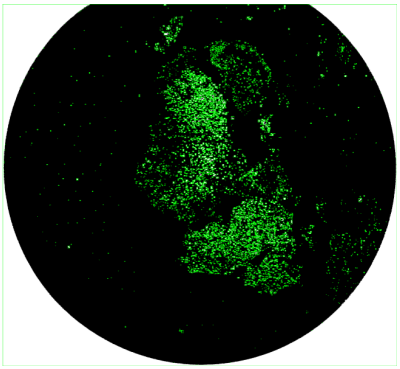


Figure 5

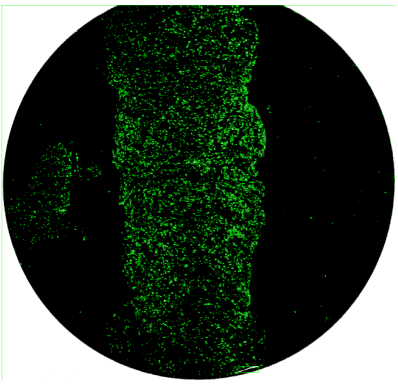


Figure 6

5.5 Tps Calculation

we calculate the Tumor Proportional Score (TPS) based on the segmented and annotated information. This step involves quantifying the proportion of tumor cells that exhibit specific characteristics, such as PD-L1 expression, relative to the total number of viable tumor cells identified in the segmented regions.

- Number of cells in positive stained region: 1971
- Total number of tumor cells: 5960

$$\text{TPS} = \left(\frac{1971}{5960}\right) \times 100 = 33.07$$

6 Discussion

The objective of this project was to develop a machine learning method for calculating the Tumor Proportion Score (TPS) to predict the benefit of immunotherapy for patients. Throughout the various stages of our project, we have achieved significant milestones that contribute to this overarching goal.

The results of our project have significant implications for the field of medical imaging and cancer treatment. The integration of machine learning in TPS calculation streamlines the process, enabling faster decision-making in clinical settings. This innovation not only minimizes human error and variability but also facilitates large-scale analysis, making it feasible to evaluate numerous patient samples in a fraction of the time required by traditional methods. Consequently, the timely and precise treatment adjustments enabled by our system have the potential to significantly improve patient outcomes and advance the field of personalized medicine.

One of the primary limitations of our study was the small dataset size, which impacted the effectiveness for mask prediction. A larger and more diverse dataset would be necessary to fully leverage the capabilities further improve the accuracy of our segmentation and TPS calculation.

7 Conclusion

In conclusion, our project demonstrates the effectiveness of a machine learning-based approach to calculating the Tumor Proportion Score (TPS) for predicting the benefit of immunotherapy in cancer patients. Through a comprehensive methodology that includes various segmentation techniques, image enhancement, and mask prediction, we have achieved significant advancements in the automation and accuracy of TPS calculation. Despite some limitations, the results of our study highlight the potential of this approach to improve personalized treatment plans and advance the field of personalized medicine. Future research should focus on addressing these limitations and further refining the methodology to achieve even greater accuracy and efficiency.