mathematics

MDPI

Article

RMD-Net: A Deep Learning Framework for Automated IHC Scoring of Lung Cancer IL-24

Zihao He [1], Dongyao Jia [1,*], Yinan Shi [2,3], Ziqi Li [1], Nengkai Wu [1] and Feng Zeng [1]

[1] School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China; 22115032@bjtu.edu.cn (Z.H.); 21111057@bjtu.edu.cn (Z.L.); 19111054@bjtu.edu.cn (N.W.); 23125099@bjtu.edu.cn (F.Z.)
[2] College of Life Science and Bioengineering, Beijing Jiaotong University, Beijing 100044, China; syn95z@163.com
[3] China Academy of Chinese Medical Sciences, Beijing 100700, China
* Correspondence: dongyaojia1974@163.com

Abstract: Immunohistochemical (IHC) detection is crucial in diagnosing lung cancer. Interleukin-24 (IL-24) is a valuable marker in IHC analysis, aiding in tumor characterization and prognostication. However, current manual scoring methods are labor-intensive, imprecise, and subjective, leading to inconsistencies among observers. Automated scoring methods also have limitations, such as poor segmentation and lack of interpretability. In this paper, we introduce RMD-Net, a novel scoring network framework specifically designed for IL-24 scoring in lung cancer. The framework incorporates a regional attention mechanism and a multi-channel scoring network. Initially, diagnostic region identification and segmentation are accomplished by integrating the diagnostic regional spatial attention module into the fully convolutional network. Subsequently, we employ the Adaptive Multi-Thresholding algorithm to derive expert, strong feature description maps. Finally, the attention-guided IHC images and expert feature description maps are fed into a multi-channel scoring network. Its backbone includes feature fusion layers and scoring layers to ensure the accuracy and interpretability of the final result. To the best of our knowledge, this is the first system that directly employs lung cancer IL-24 IHC images as input and combines both expert-derived features and deep-learning abstract features to produce clinical scores. Our dataset is sourced from the Institute of Life Sciences and Bioengineering at Beijing Jiaotong University. The experimental results demonstrate that the proposed method achieves an IL-24 score precision of 89.25%, an F1 score of 89.00, and an accuracy of 95.94%, outperforming other state-of-the-art methods. This contribution has the potential to advance clinical diagnosis and treatment strategies for lung cancer.

Keywords: lung cancer; immunohistochemistry scoring; Interleukin-24; regional attention; multi-channel model

MSC: 68T07; 68U10; 68U35

check for updates

Academic Editor: Konstantin Kozlov

Received: 18 December 2024
Revised: 22 January 2025
Accepted: 24 January 2025
Published: 27 January 2025

Citation: He, Z.; Jia, D.; Shi, Y.; Li, Z.; Wu, N.; Zeng, F. RMD-Net: A Deep Learning Framework for Automated IHC Scoring of Lung Cancer IL-24. *Mathematics* **2025**, *13*, 417. https://doi.org/10.3390/math13030417

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Lung cancer stands as the foremost cause of global morbidity and mortality, accounting for approximately 1.8 million fatalities in 2020 alone [1]. Exacerbated by the repercussions of the novel coronavirus, the prevalence of lung diseases has been on an upward trajectory [2]. Throughout the treatment process, expeditious differential diagnosis and prognosis evaluation remain paramount [3]. Recently, molecular markers, or biomarkers, have showcased

their potential in identifying traits indicative of tumor behavior and relevant to disease progression and prognosis [4]. IL-24 is an emerging candidate for tumor suppression [5]. It has been found to trigger apoptosis in a variety of human malignant cells via both secretory and non-secretory mechanisms [6]. Studies suggest that a therapy combining IL-24 with chemotherapy and endocrine treatment may present the most efficacious approach to lung cancer management [7]. Consequently, accurate and expedited assessment of IL-24 expression becomes pivotal in devising optimal therapeutic strategies for lung cancer patients. The prevalent method for evaluating IL-24 biomarker expression is IHC staining. This technique identifies specific antigens, including proteins and peptides, by eliciting targeted reactions with tissue samples using specific chromogenic agents [8]. Subsequent quantitative and locational analyses result in an IHC score for the antigen, as determined by a pathologist. Diaminobenzidine (DAB)-based IHC techniques result in a brown coloration (indicative of a positive result) for IL-24, contrasted against a blue counterstain provided by Hematoxylin [9]. The IHC-based scoring and classification of the cancer marker IL-24 have proven instrumental in assessing the severity of specific cancers and in forecasting patient outcomes across numerous cancer types [10]. Pathologists assess pathological alterations in lung cancer cells by identifying IL-24 biomarkers under microscopy. Expert knowledge is used to assign scores to observations, guiding doctors in developing personalized and more effective treatment plans [11]. In essence, the accurate scoring of IL-24 expression is pivotal for the pathological diagnosis and treatment planning of lung cancer [12].

1.1. Manual Scoring Approaches

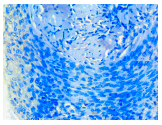
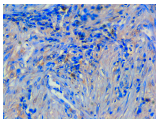
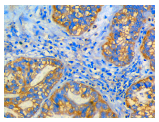
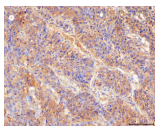
In clinical practice, the evaluation of IL-24 expression in lung cancer tissues proceeds in the following manner: Initially, seasoned pathologists visually inspect lung cancer histology slides stained with Hematoxylin-Eosin-DAB [13] under a microscope to quantify IL-24 expression. Subsequently, they provide scores based on a combination of theoretical knowledge and practical experience. This scoring takes into account factors such as the overall staining intensity and the proportion of neoplastic tissue exhibiting staining [14], as detailed in Table 1. This practice is constrained by the time-consuming and labor-intensive process of manually inspecting glass slides. A minimum of two seasoned pathologists are typically needed for the assessment, and their collective evaluations are then synthesized to produce the final results. However, the manual IHC scoring of IL-24 by expert pathologists is inherently subjective. This often leads to variations between observers and compromises the repeatability of the assessment [15]. Additionally, the inherent semi-quantitative nature of this method, including subjective categorization into strong, medium, and weak staining intensities, complicates efforts to maintain consistency both within and between evaluators. The lack of precision and objectivity further exacerbates this challenge. For instance, the error rate in scoring the HER2 antigen in breast cancer reaches up to 20% [16]. Additionally, artifacts, including high background or inconsistent staining deposition, can potentially skew the results [17]. The quality of slide staining also greatly affects expert judgment. Therefore, the automated scoring method addresses the inherent flaws of manual scoring, enhancing efficiency while ensuring consistency and accuracy. Additionally, its scalability meets oncology's demand for precise and efficient diagnostic tools.

1.2. Digital Scoring Approaches

With advancements in medical image processing technology, several digital processing software, both commercial and open-source, have been introduced to facilitate high-volume IHC analysis and scoring [18,19]. However, when determining scores, users are required to specify the intensity threshold and frequency of stained cells within the image area. Ram et al. [20] utilized two whole-slide image analysis software packages, Visiopharm

and HALO, to assess the expression of membrane biomarkers. However, they required manual specification of thresholds for the nuclear staining channel. These thresholds can be influenced by the imaging quality of the tissue slide, leading to variability in tissue sample scoring based on user decisions. Antigens and antibodies bind to specific targets, creating unique staining patterns in tissue regions [17]. For example, the lung cancer biomarker IL-24 is in the cytoplasm or interstitial fluid [21], while PD-L1 is on the cell membrane [16]. In breast cancer, PR, ER, and Ki-67 are nuclear markers, whereas HER2 and EGFR are found on the cell membrane and in the cytoplasm [22], respectively. Different locations of biomarker expression necessitate varying requirements for digital software-assisted scoring. Cytoplasmic markers like IL-24 exhibit diffuse and widespread staining patterns with blurred and irregular boundaries. Therefore, most digital image processing software focuses on evaluating nuclear and membrane markers and is not suitable for assessing IL-24 expression levels. For example, The ImageJ 1.54f software [23] is proficient in counting positive cells to assess the expression of breast cancer markers like PR, ER, and Ki-67. The study by Alataki et al. [24] evaluated the use of the Cognition Master Professional Suite (CogM) for automated Ki67 assessment in primary breast cancer. The SlidePath tissue image analysis system, which utilizes a Leica camera for its analysis [25], is considerably dependent on the quality of slide staining. Therefore, we focused on the IHC scoring of IL-24 in lung cancer to address the issues of excessive manual intervention and variable image quality in existing digital software, aiming to inspire the development of digital scoring software for cytoplasmic markers.

Table 1. The IL-24 scoring guidelines from «Clinical Cancer Research» [14].

Score	Definition	Image
0	Minimal yellow staining, covering ≤10% of tumor tissue.	
1+	Weak yellow staining, covering 10–30% of tumor tissue.	
2+	Moderate yellow staining, covering 30–50% of tumor tissue.	
3+	Intense yellow staining, covering ≥50% of tumor tissue.	

1.3. Automated Scoring Approaches

IHC images encompass a wealth of phenotypic data that offer insights into the intrinsic mechanisms driving disease progression and patient survival outcomes. [26] Recently, deep learning has emerged as the predominant methodology for the analysis and interpretation of IHC images. The current research on automatic scoring of IHC images is mainly divided into three steps [27]: (1) ROI selection and segmentation; (2) Extraction of features; (3) Building the scoring model. There are several challenges in applying state-of-the-art IHC scoring research to evaluate the expression of IL-24 in lung cancer.

(1) ROI Selection and Segmentation: Before processing IHC images, selecting and segmenting specific biomarker expression regions is crucial [28]. However, current advanced methods often overlook the segmentation step, suffer from segmentation accuracy

issues, and lack robustness. For example, Zidane et al. [29] introduced a groundbreaking graph neural network (GNN) model designed to predict HER2 status directly from whole-slide images of standard H&E stained slides. Feng et al. [30] proposed a recognition and registration system for IHC scoring of Ki-67 molecules in breast cancer tissue images. However, positive cell segmentation was conducted on whole tissue sections. Feng et al. [16] proposed a new automatic image analysis model for scoring and grading ER, PR, Ki-67, and HER-2 immunohistochemical images of breast cancer. Similarly, the model processes whole slide tissue images without considering the different expression locations of various molecules. The above methods carry an inherent bias in model predictions, as regions irrelevant to diagnosis might influence the accuracy of IHC scoring. Additionally, a notable limitation is computational inefficiency, as these models require processing every region within an image. Xu et al. [31] introduced an FCN, trained using instance-level labels derived from image-level annotations, for the segmentation of metastatic areas in lung cancer images. Khameneh et al. [32] utilized an enhanced U-Net model for segmenting cell membranes to predict the HER2 IHC score in lung cancer. Zhang et al. [33] proposed an automatic IHC scoring model for P53 molecules in tumors, consisting of an ROI extraction module and a feature fusion scoring network. However, excessive prior knowledge was introduced to exclude irrelevant regions, potentially reducing robustness when addressing complex backgrounds and heterogeneous tissues. Lung cancer tissue slide images contain diagnostic areas (tumor tissue and cytoplasmic regions infiltrated by inflammatory cells) and non-diagnostic areas such as normal lung tissue, scar tissue, and background areas. Segmenting the cytoplasmic regions specific to IL-24 expression in lung cancer tumor tissue and ensuring alignment with pathologist standards is crucial for enhancing subsequent model evaluation. To address the challenges in fully or semi-supervised histopathology image segmentation, Jin et al. [34] proposed an inter- and intra-uncertainty regularization-based semi-supervised segmentation framework. Inspired by this, we introduced a regional attention mechanism in the segmentation network to focus more on the diagnostic area and suppress the background and irrelevant areas.

(2) Extraction of Features: Feature extraction is crucial in the IHC scoring process, directly impacting the accuracy, efficiency, and interpretability of the results. Currently, most advanced methods focus solely on artificial features or deep abstract features. For instance, Guo et al. [35] proposed a scoring method for immunohistochemical staining on Ki67, which considered only the staining intensity and staining area of the positive regions. Kabakçı et al. [22] proposed an automatic scoring system for CerbB2/HER2 molecules in breast cancer tissue, where the reference feature was based on the membrane intensity histogram, including membrane staining intensity and integrity. Liu et al. [36] proposed an IHC score for PD-L1 in lung squamous cell carcinoma, where the pathological evaluation of its expression was based on two semi-quantitative metrics: the number of PD-L1-positive tumor cells and the number of total viable tumor cells. In the above studies, the primary consideration is based on expert experience, followed by simple thresholding. The selection and definition of expert features are subjective, impacting the consistency and repeatability of the scoring results. Additionally, these methods exhibit poor ability to extract global features from complex images. The application of deep networks has improved the processing efficiency of IHC images. For example, Baxi et al. [37] provide an exhaustive overview of contemporary deep-learning methodologies employed in IHC image analysis. Durán-Díaz et al. [38] proposed an unsupervised robust IHC scoring method that achieves image classification through clustering regression. Xue et al. [39] trained a classification method based on the entire IHC image to achieve the IHC scoring of biomarkers in colon cancer. Ghoshal et al. [40] constructed HBNet to extract deep image features and classify immunohistochemistry images. The inherent "black-

box" nature of deep learning models renders their decision-making mechanisms elusive. Ensuring transparency and reliability in the decision-making process is paramount for computer-assisted diagnosis (CAD) systems. At the same time, the quality of deep abstract features is affected by data distribution. Class imbalance in the training data can cause the model to perform poorly with minority categories, impacting the comprehensiveness and accuracy of the scoring. In summary, the feature extraction techniques employed in the aforementioned studies tend to be somewhat unidimensional and lacking in diversity. This often results in models that are less interpretable and robust. In this paper, we delve into methods for extracting both expert-driven features and deep-learning features from tumor immunohistochemistry images. By achieving multi-modal data learning, we aim to harness the interpretability and reliability of traditional feature engineering while also leveraging the automated feature extraction and recognition capabilities inherent to deep learning.

(3) Building the Scoring Model: The feature vector is fed into the scoring model, and the IHC scoring result is output after processing. A scientific and well-designed model architecture ensures that the model possesses sufficient complexity and expressive ability to accurately identify and evaluate biomarkers in IHC images. Currently, linear classifiers and deep learning models are widely used [41]. Hameed et al. [42] used a support vector machine to classify each nucleus in tissue images as positive or negative based on staining intensity and then determined the tissue score according to the J-score scheme. Additionally, Rmili et al. [43] performed a color space assessment, selected appropriate color channels for threshold processing, and implemented auxiliary scoring of neuroendocrine tumors. Mahanta et al. [44] proposed the IHC-net, which integrates three machine learning models (Naive Bayes, Support Vector Machine, and Random Forest) and uses the majority voting principle to implement the Allred score for breast cancer. In the above methods, IHC images are only processed and scored through simple linear superposition or classifiers. This approach cannot effectively capture the nonlinear relationships in IHC images, nor can it adaptively learn features from the original images. It is more sensitive to noise and interference. As a result, the robustness and accuracy of these methods are not ideal. Anaam et al. [45] proposed an end-to-end deep learning framework using a deep active learning strategy for cell detection and scoring in HEp-2 specimen images. Molnar et al. [46] proposed the IHCScoreGAN deep network, which inputs breast cancer IHC images and outputs the Ki-67 score results. This is an end-to-end, unsupervised processing method. Wen et al. [47] focused solely on the DAB staining features of the attention area, fed these features into a deep network and obtained the IHC scoring results within a few seconds. End-to-end IHC image scoring based solely on deep networks carries the risk of poor interpretability and overfitting. Springenberg et al. [48] proposed that the classification prediction performance, interpretability, and robustness of the tissue pathology model should be comprehensively considered. They pointed out that insufficient robustness is a serious problem that hinders clinical application. Therefore, this paper plans to design a multi-channel scoring network that can effectively process and fuse multi-modal features to provide more accurate and comprehensive scoring results.

1.4. Objectives

To address the above challenges, as summarized in Table 2, the main features and limitations of various IHC image scoring methods are outlined. Additionally, we found that most studies focused on other biomarkers (ER, PR, Ki-67, and HER2) or cancer types (breast cancer, colon cancer, etc.). Therefore, we posed the following question: Is it possible to develop a system that automatically provides accurate scores for IL-24 IHC pathology images in lung cancer? This system aims to address the labor-intensive, time-consuming, and subjective aspects of manual scoring. It avoids the drawbacks of digital scoring

software that requires excessive user intervention and is not well-suited for cytoplasmic markers (IL-24). Additionally, it addresses the limitations of current automatic scoring methods, which include the interference of non-diagnostic areas when the entire tissue section image is used, increased computational load, reliance on a single feature extraction technique, and suboptimal scoring accuracy and interpretability.

Table 2. Overview of immunohistochemical image processing methods.

Method	Approach	Key Features	Limitations
Taylor et al. [49] (2016)	Manual	Manual and Automated Scoring Thresholds	Labor-intensive
Hameed et al. [50] (2015)	Manual	Maximal separation techniques	Time-consuming
Irshad et al. [51] (2017)	Manual	Crowdsourcing scoring of IHC images	Subjective variability
Ram et al. [20] (2021)	Digital software	Two whole-slide image analysis software packages	Excessive user intervention
Alataki et al. [24] (2021)	Digital software	Cognition Master Professional Suite (CogM) for automated Ki67 assessment	Not well-suited for cytoplasmic markers
Lu et al. [25] (2022)	Digital software	SlidePath tissue image analysis system	Partial automation
Sode et al. [52] (2023)	Digital software	HER2 Digital image analysis software	Interference from non-diagnostic areas
Zilenaite et al. [53] (2020)	Digital software	The HALO AI tissue classifier	Suboptimal scoring accuracy
Zidane et al. [29] (2023)	Automated	Groundbreaking graph neural network	Suboptimal scoring accuracy
Feng et al. [30] (2020)	Automated	An accurate registration algorithm for IHC images and their corresponding HE images	Lack of interpretability
KA et al. [22] (2021)	Automated	Membrane Intensity Histogram + Classifier	Not well-suited for cytoplasmic markers
Ghoshal et al. [40] (2021)	Automated	The Hybrid Bayesian Neural Network (HBNet)	Reliance on a single feature
Mahanta et al. [44] (2021)	Automated	A fully convolutional neural network for automated nuclear segmentation and ensemble classification	Lack of interpretability
Wen et al. [47] (2024)	Automated	Region segmentation + generation of nuclear mask	Lack of interpretability

In addition, IL-24 IHC images exhibit distinctive features, including specific cytoplasmic expression patterns in lung cancer tissues, as well as rich information such as

overlapping color signals, cell clustering, and irregular histomorphology, as demonstrated in Figure 1. Developing a system capable of effectively analyzing these complex images holds significant potential for advancing research and clinical applications.

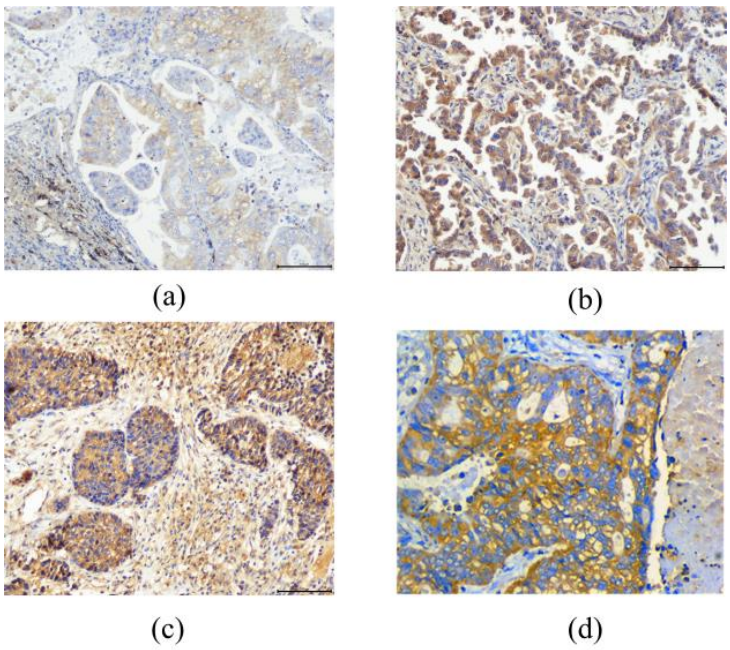


Figure 1. Variations in IL-24 immunohistochemical staining in lung cancer samples: (**a**) disrupted tissue morphology, (**b**) diffuse and weak staining, (**c**) cellular aggregation and crowding, and (**d**) uneven distribution of staining intensities.

1.5. Contributions

In this paper, we propose a scoring network framework, denoted as RMD-Net, designed for lung cancer IHC IL-24. The framework incorporates regional attention and a multi-channel scoring network. As illustrated in Figure 2, the framework first employs a regional segmentation network, diagnostic regional spatial attention fully convolutional network(FCN-DRSA), to achieve localized segmentation of diagnostic regions(the cytoplasmic region within the tumor tissue). Subsequently, we employ the DAB-Luminance Adaptive Multi-Thresholding algorithm (DAB-LAMT) to generate expert-level feature description maps from the IHC image. Ultimately, the attention-guided IHC images and expert feature description maps are fed into a multi-channel scoring network to facilitate the automated scoring of lung cancer IHC IL-24. In summary, the contributions of this work are as follows:

(1) The "FCN-DRSA", a novel diagnosis-related region segmentation network, effectively integrates the DRSA module into the foundational structure of the FCN, efficiently emphasizing the essential diagnostic cytoplasmic region while minimizing irrelevant cell nuclei and background noise;

(2) The multi-channel scoring network is designed with a backbone comprising feature fusion layers and scoring layers. The former processes the received multi-modal feature map, while the latter is responsible for outputting the IHC score of lung cancer. This design ensures interpretability throughout the process and enhances the ultimate scoring accuracy.

This study introduced a regional attention mechanism and an expert feature description map, combining expert knowledge with deep abstract features through a multi-channel scoring network to ultimately output the scoring results. This method not only improves the accuracy of the scoring but also enhances the interpretability and transparency of the decision-making process. Based on our extensive literature review and analysis of related studies, especially lung cancer and IL-24 markers, we believe that this study is a ground-

breaking attempt. Our forthcoming experimental results will underscore the high precision of our proposed scoring network and its robust alignment with scores rendered by seasoned pathologists. Our work will greatly enrich the current research ideas and objectives for automatic IHC scoring. Although much work remains to be completed, in future clinical applications, it will promote the personalization and precision of tumor treatment.

The structure of this paper is outlined as follows: In Section 2, we delve into the specifics of the proposed methodology. Experimental findings and comparative analyses are presented in Section 3. Finally, Section 4 provides a summary of the paper and then transitions into discussing the constraints of our proposed approach and prospective directions for future research.

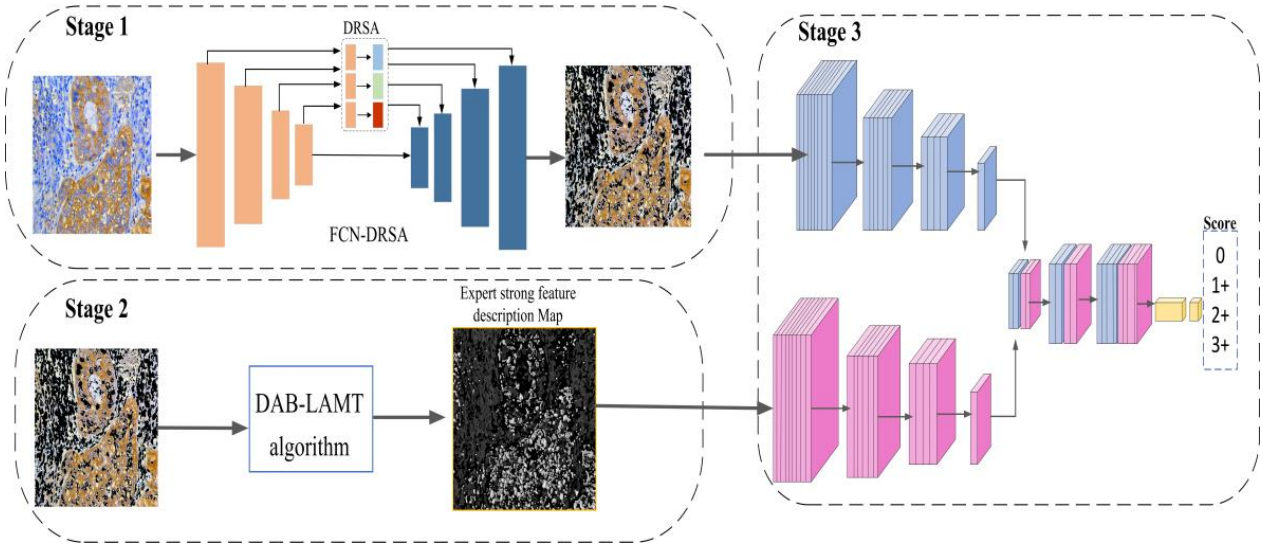


Figure 2. The overview of our proposed IL-24 scoring framework. An input lung cancer IHC image undergoes initial processing using the FCN-DRSA algorithm to segment the diagnostic area. Then, the attention-guided IHC images are processed using the DAB-LAMT algorithm to derive the expert strong feature description maps. These multi-modal features are subsequently fed into the multi-channel scoring network to predict the IL-24 IHC score.

2. Materials and Methods

2.1. Materials

2.1.1. Slide Preparation and Image Acquisition

This study is based on 128 lung cancer tissue pathology sections, including 67 squamous cell carcinomas and 61 adenocarcinomas, all from the Institute of Institute of Life Sciences and Bioengineering, Beijing Jiaotong University. The patients and their next of kin have signed informed consent forms, and the relevant research has passed the strict review of the Medical Ethics Committee of Beijing Jiaotong University. All slides were prepared and maintained by pathology experts, and a standardized immunohistochemical staining process was performed. The process involved fixing paraffin-embedded tissue sections with formalin, localizing IL-24 using DAB colorimetric agents, and finally performing hematoxylin contrast staining. The slides were then dehydrated and sealed. The staining quality is high, and the color contrast is clear.

Digital slide images were acquired using a Leica DM2000 microscope (40× magnification), a DFK 33G274 industrial camera (The Imaging Source, Bremen, Germany), and a ProScan H101F 3D (ProScan, Taipei, Taiwan) motorized stage (see Figure 3). The system allows for automatic focus and scanning. The resulting dataset contained 2496 images, which were scored by three pathologists.

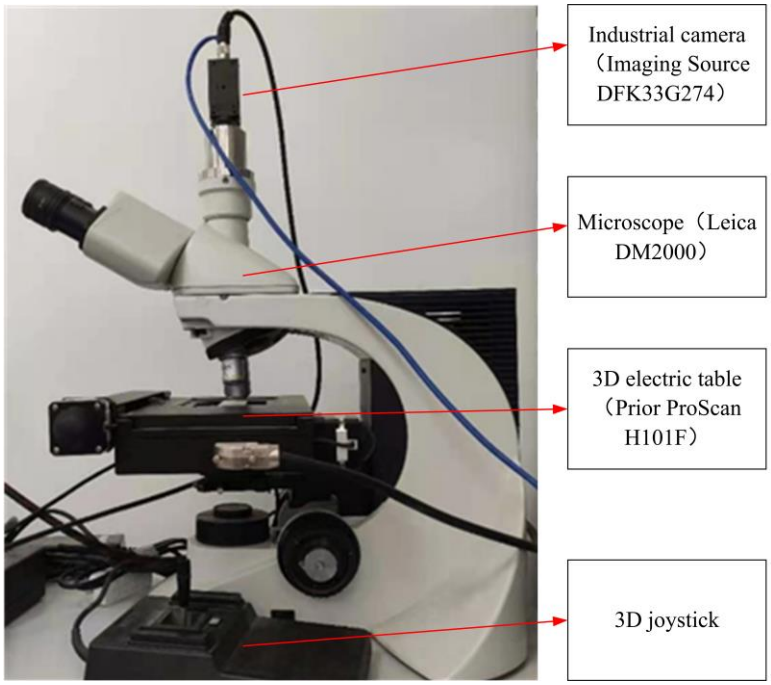


Figure 3. Image acquisition system. The system is capable of auto-focusing and automatic scanning of slides according to a "Z"-shaped trajectory. Finally, the IL-24 IHC images are stored in the database.

2.1.2. Image Processing and Labeling

In IHC staining image processing, color normalization is a key step to standardize the staining intensity and color performance between different slices, ensuring consistency and comparability of images in subsequent analysis. In this paper, we adopt the Vahadane method based on the non-negative matrix factorization (NMF) proposed by Vahadane et al. [54]. Specifically, the image matrix I is decomposed into the product of the staining matrix W and the concentration matrix H, as shown in Equation (1).

$$I = W * H \tag{1}$$

The color matrix W_{target} of the target image is adjusted to be consistent with the staining matrix $W_{reference}$ of the reference image. Then, the normalized image is reconstructed using the adjusted staining matrix and the concentration matrix H_{target} of the target image, as shown in Equation (2). The normalized example image is shown in Figure 4b.

$$I_{normalized} = W_{reference} * H_{target} \tag{2}$$

The annotation of tissue images consists of two parts: diagnostic region annotation and IHC scoring. The diagnostic region annotation adopts a weak supervision strategy, which only requires annotating the general outline of the diagnostic region without marking all pixels. Labelme v5.5.0 software is used for this purpose, as shown in Figure 4c.

The IHC scoring of lung cancer tissue sections is performed by pathologists based on their professional knowledge and experience to observe the staining intensity and staining area of IL-24. The results are then obtained according to the scoring guidelines in Table 1. This process requires the participation of at least three experts, and the Kappa coefficient statistical method is used to evaluate the consistency of the scores among multiple experts for all samples and any section. When the Kappa coefficient is higher than 0.80, it indicates that the scores are highly consistent. Finally, the final score results of each section are summarized in the scoring table.

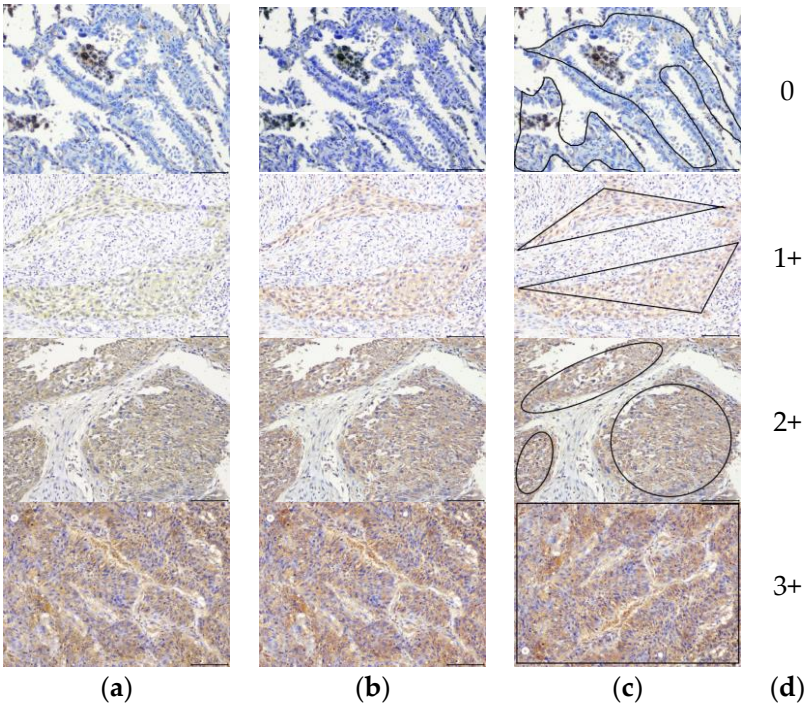


Figure 4. Processing and labeling of lung cancer IHC images. (**a**) Original images; (**b**) color normalized images; (**c**) diagnostic area labeling; (**d**) labeling the scoring results.

2.2. Methods

In this paper, to address the problems of low efficiency, human interference, and excessive subjectivity of manual scoring and digital-assisted software, as well as the non-diagnostic area interference, suboptimal scoring accuracy, and interpretability issues in current automated scoring methods, we designed an automated IHC scoring network framework tailored for lung cancer IL-24, named RMD-Net. As depicted in Figure 2, the framework consists of three stages:

(1) Diagnostic area attention: The FCN network integrates the Diagnostic Region Spatial Attention (DRSA) module to perform preliminary processing and diagnostic region segmentation on the input lung cancer IHC images;

(2) Feature Extraction Stage: The DAB-brightness adaptive multi-threshold algorithm (DAB-LAMT) is used to extract expert-level feature description maps, which primarily express the characteristics of positive staining intensity and area;

(3) Scoring Stage: The attention-guided IHC images and expert feature description maps are fed into a multi-channel scoring network, which consists of a feature fusion layer and a scoring layer and finally generates the IL-24 IHC scoring results.

2.2.1. Diagnostic Region Attention Based on FCN-DRSA

Lung cancer IHC images have complex staining patterns, tissue structures, and high overlap. Traditional image segmentation methods such as threshold segmentation, edge detection, and clustering have certain limitations. These methods require manual setup, are sensitive to noise and illumination changes, and have limited processing capabilities for complex structures. Compared with traditional methods, commonly used segmentation networks such as U-Net, FCN, and DeepLab can more accurately capture complex features and details in images. However, they perform poorly in small target areas and boundary areas. Therefore, we designed a DRSA module that introduces a spatial attention mechanism. By generating an attention map, the model can better focus on the diagnostic area, suppressing the influence of background or non-diagnostic areas on the segmentation

results and improving the accuracy and robustness of segmentation. It is worth noting that the combination of average pooling and maximum pooling in DRSA can better extract features of different scales and perform feature fusion through skip connections to enhance the segmentation accuracy of boundary areas. The FCN-DRSA proposed in this paper meets the requirements of pathologists.

The architecture of FCN-DRSA is illustrated in Figure 5. The network backbone comprises an encoding pathway, a decoding pathway, and a regional attention pathway. The encoding path inputs lung cancer IHC images, which are passed through multiple convolution layers, batch normalization layers, and ReLU activation functions to gradually increase the number of feature channels. Low-level features such as edges and textures, as well as high-level semantic information, are extracted. The size of the convolution kernel is 3×3, the stride is 1, and the number of channels is 32, 64, 128, and 256, respectively. Then, a maximum pooling layer (size 2×2, stride 2) is used for downsampling to gradually reduce the spatial size of the feature map.

The decoder path is responsible for feature reconstruction, gradually restoring the resolution of the feature map through the deconvolution layer, cropping layer, and fusion layer (add), and finally generating a segmentation map with the same resolution as the input image. The deconvolution layer is used to restore the spatial information of the feature map, with a kernel size of 2×2, a stride of 2, and the number of channels set to 128, 64, 32, and 3. The fusion layer combines the feature attention map output by DRSA with the upsampled feature map through a skip connection, retaining more detailed information while focusing on the diagnostic area. The cropping layer ensures that the size of the feature map before fusion is consistent.

The regional attention path consists of DRSA modules. Through global average pooling, maximum pooling, and convolution operations, a weight matrix is generated to weight the feature map of the encoding path, ensuring that feature extraction focuses on the diagnostic area. Then, feature fusion is performed with the decoding path through skip connections.

The specific process begins by inputting the lung cancer IL-24 IHC image $I(h, w)$. The encoding pathway processes it to derive the feature vector $F(h/n, w/n, c)$, where n is the feature scaling factor and c denotes the number of feature channels. This vector is then introduced to the DRSA module. As illustrated in Figure 6, both AvgPool and MaxPool are applied to each feature channel, resulting in two feature maps $H \times W \times 1$. These maps are then concatenated channel-wise. Following a 7×7 convolution layer and a Sigmoid activation function, the spatial weight coefficient matrix MS is derived. By multiplying MS with the feature vector F, we obtain the spatial attention feature matrix L. The matrix L is merged with the original feature vector to produce the output matrix $W(a \times b \times num_class)$,$num_class \in \{0, 1\}$. In pixel-level classification, "1" indicates a pixel relevant to diagnosis and '0' for irrelevant pixels. The process concludes with decoding, resulting in a predicted segmentation mask map Y, that aligns with the input size and exclusively emphasizes the diagnostic area. The network's training loss is determined using the dice loss function. In processing medical images, the Dice loss function is particularly suitable for segmentation tasks. It effectively reduces false positives and false negatives by maximizing the overlap between the real mask and the predicted mask. Additionally, considering the imbalance in the number and area of diagnostic and non-diagnostic regions in lung cancer tissue images, the segmentation effect can be poor. The Dice loss function directly optimizes the overlapping area, ensuring that the contribution of each pixel to the loss is relatively balanced. This approach prevents deviations due to class imbalance, thereby improving the accuracy of segmentation.

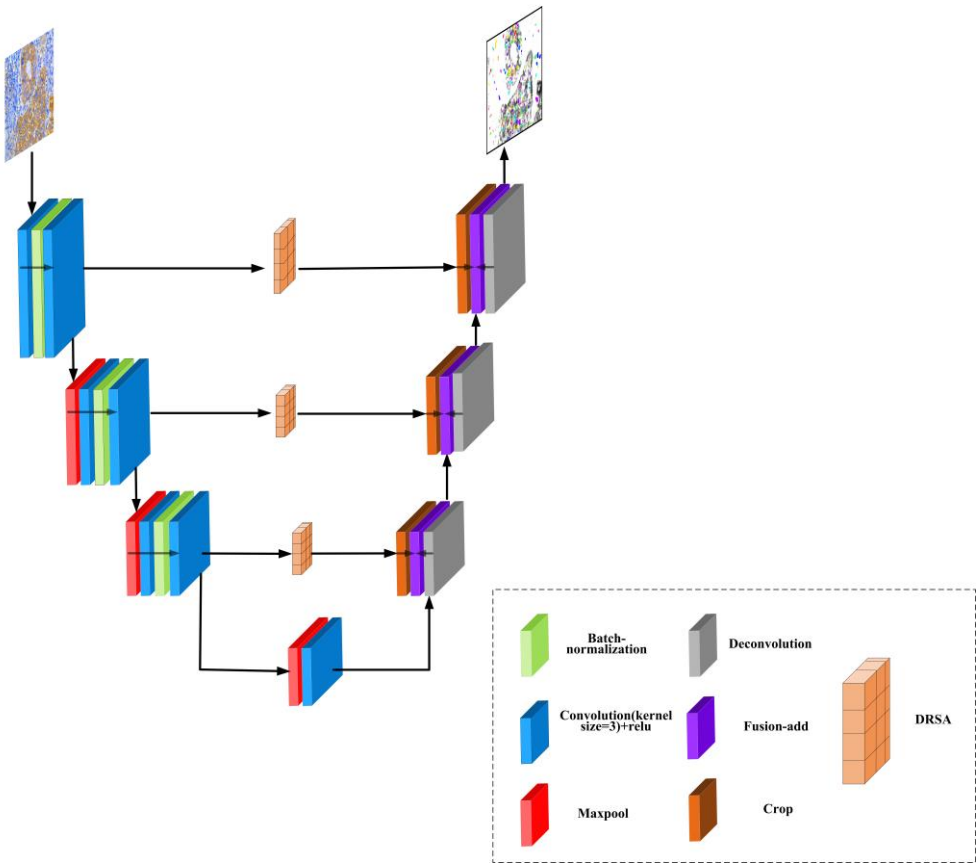


Figure 5. Network structure diagram of the FCN-DRSA for diagnostic area segmentation. (The network backbone is structured with the encoding pathway on the left, the decoding pathway on the right, and the regional attention pathway in the middle. The features flow follows the direction of the arrows).

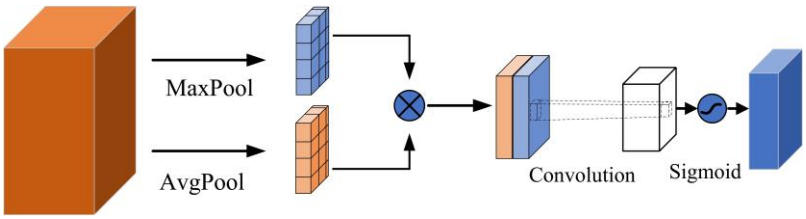


Figure 6. Structure diagram of the diagnostic regional spatial attention mechanism (DRSA) module.

As shown in Equation (3), y represents the label mask map of the input image. ε as a smoothing term, it ensures that the denominator and numerator are not zero, thus avoiding numerical instability. Additionally, the Dice loss function calculates the gradient based on regional overlap, making its gradient calculation more stable compared to pixel-level loss functions such as cross-entropy loss.

$$L_{dice} = \frac{2 \times (\sum_{i=1}^{h}\sum_{j=1}^{w} Y_{ij} y_{ij} + \varepsilon)}{(\sum_{i=1}^{h}\sum_{j=1}^{w} (Y_{ij} + y_{ij} + \varepsilon)} \quad (3)$$

Finally, we broadcast the predicted segmentation mask image Y obtained by the segmentation network, which only focuses on the diagnostic area with the original image $I(h, w)$ to obtain the final output image $G(h, w)$. The non-diagnostic areas and background

are eliminated, retaining only the ROI. This will enable subsequent models to focus on foreground diagnostic objects, which will improve scoring accuracy and significantly reduce computational costs.

2.2.2. Expert Strong Feature Description Map Extraction

During tumor tissue section preparation, the HE-DAB staining method was employed, with DAB stain specifically manifesting in brown when reacting with IL-24. Pathologists subsequently score the IL-24 IHC based on staining intensity and area. Existing techniques resort to manually setting a threshold based on the brightness value of the image to classify staining intensity (strong, medium, weak) subsequent to DAB dye separation. This approach is subjective. Image luminance can be influenced by various factors, including lighting conditions, parameter configurations, and dye deposition levels. These factors can result in erratic data with limited continuity and structure. To address this, our study introduces a DAB-LAMT algorithm-based approach to produce feature description images that predominantly capture DAB staining intensity and area, offering a more objective measure of IL-24 expression. The main process is as follows: (1) Obtain the initial brightness matrix of the IHC image. (2) Obtain the DAB staining matrix based on the color deconvolution algorithm. (3) Divide the ROI and perform adaptive threshold processing to obtain the brightness description feature matrix. (4) Perform weighted fusion of the DAB staining matrix and the brightness description matrix according to different staining areas to obtain the expert strong feature description map. The specific description is as follows:

We transform the original RGB image into a three-channel dye component image $I = [I_{DAB,}I_{H,}I_E]$ using the color deconvolution algorithm [55], as described in Equation (4). I_{OD} is the optical density converted image calculated according to Lambert–Beers law. M denotes the color matrix detailed in Equation (5), constructed from individual staining matrices. Then, we take the DAB stain channel I_{DAB} from I.

$$I = M^{-1} I_{OD} \tag{4}$$

$$M = \begin{bmatrix} 0.268 & 0.570 & 0.776 \\ 0.650 & 0.704 & 0.286 \\ 0.000 & 0.000 & 0.000 \end{bmatrix} \tag{5}$$

The detailed process of the DAB-LAMT algorithm is shown in Algorithm 1.

Algorithm 1. DAB-LAMT algorithm

Compute the initial brightness matrix I_l of the IHC image, as illustrated in Equation (6);
Calculate I_{DAB} with the color deconvolution algorithm;
The IHC image is first transformed into a grayscale representation, after which it's evenly partitioned into K regions;
Repeat:
    for *each region* **do:**
        Find the pixel coordinates of the center of the area:(x_0, y_0);
        Calculate the Gaussian weight of any pixel (x, y):$w(x, y)$, as illustrated in Equation (7);
        Compute the Gaussian-weighted mean brightness, I_L, of the region, as illustrated in Equation (8);
Until: the luminance description matrix I_{Lum} of the original image is obtained;
The expert strong feature description image, $I_{Expert\ Map}(x, y)$, is derived by performing a weighted fusion of the brightness descriptor matrix I_{Lum} and the DAB stain channel matrix I_{DAB}, as illustrated in Equation (9).

$$I_l = 0.299 \times I_R + 0.587 \times I_G + 0.114 \times I_B \tag{6}$$

$$w(x,y) = \exp(-((x-x_0)^2 + (y-y_0)^2)/(2\sigma^2)) \tag{7}$$

where σ denotes the standard deviation of the Gaussian function. A smaller value of σ captures more detailed image features. For this study, σ is set to 0.75.

$$I_L = sum(w(x,y) \times I_l(x,y))/sum(w(x,y)) \tag{8}$$

A particular pixel within this region, represented by (x,y), has a luminance value of $I_l(x,y)$.

$$I_{Expert\ Map}(x,y) = \begin{cases} 0.6 * I_{DAB} + 0.4 * I_{Lum}\ (x,y) \in Positive \quad Area \\ 0.4 * I_{DAB} + 0.6 * I_{Lum}\ (x,y) \in Negative \quad Area \end{cases} \tag{9}$$

For positive regions with high IL-24 expression, the DAB staining channel matrix has a higher weight of 0.6 due to its significant contribution, while the brightness description matrix is weighted at 0.4. Conversely, for negative areas that are mainly blue or white with higher brightness values, the weight of the brightness description matrix is set at 0.6. Since these areas have low IL-24 expression, the DAB staining channel matrix is given a lower weight of 0.4 [15]. Both matrices' values are normalized between 0 and 255.

The DAB-LAMT algorithm obtains the distribution of IL-24 staining through color deconvolution and optical density conversion and then quantifies the characteristics of the IL-24 staining area through linear superposition. Additionally, the algorithm quantifies the staining intensity characteristics of IL-24 by setting sub-regions and calculating the Gaussian weighted average brightness value in the region. Notably, this algorithm avoids the subjectivity of manually setting thresholds during feature extraction. At the same time, regionalization processing can reduce the problem of poor continuity and smoothness of features caused by uneven staining or different lighting conditions, thereby achieving a more objective IL-24 expression measurement.

2.2.3. The Multi-Channel Scoring Network

Traditional scoring networks typically utilize a single channel, processing the raw IHC image to yield an end-to-end scoring result. This approach often results in suboptimal accuracy and interpretability. Multi-channel architectures have found applications in fields such as sentiment analysis [56], high dynamic range imaging, and sentence relationship classification. To strike a balance between interpretability and objectivity in the scoring results, we propose a multi-channel IHC scoring model. It includes the input layer, fusion layer, and output layer, illustrated in Figure 7.

The input layer comprises two channels to receive the attention-guided IHC images and expert strong feature description maps. Given that their dimensions are $128 \times 128 \times 3$ and $128 \times 128 \times 1$ respectively, we employ convolution and pooling operations for preprocessing. This ensures the provision of precise and consistent inputs for subsequent fusion layers and output layers.

The fusion layer accepts two input feature maps of size $32 \times 32 \times 60$ and applies convolution and pooling operations, resulting in the generation of two feature maps of size $8 \times 8 \times 8$. Subsequently, the features from different channels are integrated through feature splicing (Concat) to provide a more comprehensive and enhanced feature representation for the subsequent output layer. This integration aims to capture a broader range of information and improve the overall effectiveness of the feature representation.

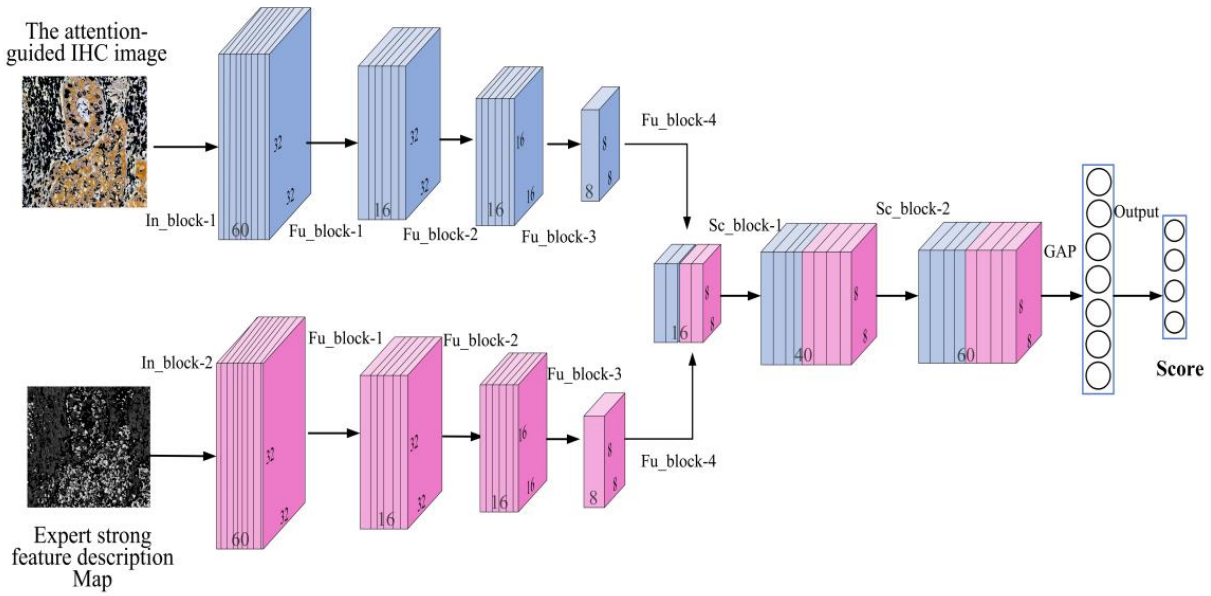


Figure 7. Structure diagram of the multi-channel scoring network.

The scoring layer consists of two scoring blocks. Taking into account the comprehensiveness and complexity of the fused hybrid features, we have optimized the design of the scoring block, as shown in Figure 8. (1) Multi-class small convolution kernels: three 1×1 kernels, a 3×3 kernel, and a 5×5 kernel are employed. This design increases the receptive field while reducing computational requirements. (2) Group normalization (GN) operation: Instead of traditional batch normalization (BN), the scoring block utilizes the GN operation. This choice minimizes the model's reliance on batch size and hardware requirements. (3) Feature fusion: The input and the output of the 3×3 kernel are added together to achieve the addition of residuals. Additionally, the outputs of the 3×3 and 5×5 kernels are concatenated to form the final output. This ensures that the scoring layer fully considers feature maps of different sizes, enhancing its ability to capture both local and global information.

In the output layer, global average pooling (GAP) is employed instead of a fully connected layer. This choice effectively addresses the issue of large parameter quantities. We configure the output layer with 4 channels to generate 4 classification results: 0+, 1+, 2+, and 3+. The Softmax function can normalize the actual values in the 1×4 vector to [0, 1], with a probability sum of 1.

The detailed architectural parameters of the multi-channel scoring network for IHC are delineated in Table 3. The loss function for IHC-Score prediction is defined as:

$$L_{score} = -\sum_{i=1}^{m}\sum_{k=0}^{3} y_k^{(i)} * \log(\hat{y}_k^{(i)}) \tag{10}$$

where m denotes the number of IHC image samples, k represents the IHC score type, $y_k^{(i)}$ is the actual label of the corresponding category, and $\hat{y}_k^{(i)}$ signifies the IHC score predicted by the Softmax function. The dual-channel scoring network combines the abstract features extracted by deep learning with the specific features defined by experts, making the scoring results not only highly accurate and robust but also well interpretable. Consequently, the overall performance of the model is superior.

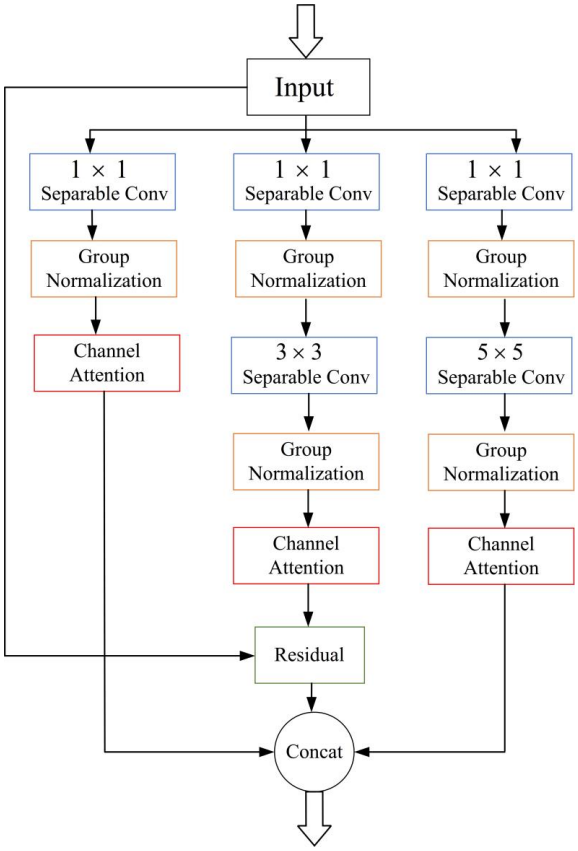


Figure 8. The scoring block designed for lung cancer IHC images.

Table 3. The architecture of multi-channel scoring network.

Layer					Input Size	Output Size
		Size/Stride	**Filters**	**Padding**		
Input layer	In_block-1	Conv 5 × 5/1 MaxPool 2 × 2/4	60	Same	128 × 128 × 3	32 × 32 × 60
	In_block-2	Conv 5 × 5/1 MaxPool 2 × 2/4	60	Same	128 × 128 × 1	32 × 32 × 60
Fusion layer	Fu_block-1	Conv 3 × 3/1	16	Same	32 × 32 × 60	32 × 32 × 16
	Fu_block-2	MaxPool 2 × 2/2			32 × 32 × 16	16 × 16 × 16
	Fu_block-3	Conv 3 × 3/1	8	Same	16 × 16 × 16	16 × 16 × 8
	Fu_block-4	MaxPool 2 × 2/2			16 × 16 × 8	8 × 8 × 8
		Concat			8 × 8 × 8 8 × 8 × 8	8 × 8 × 16
Scoring layer	Sc_block-1	Conv 1 × 1/1 Conv 1 × 1/1 Conv 1 × 1/1 Conv 3 × 3/1 Conv 5 × 5/1	8 16 32 16 16	Same	8 × 8 × 16	8 × 8 × 40
	Sc_block-2	Conv 1 × 1/1 Conv 1 × 1/1 Conv 1 × 1/1 Conv 3 × 3/1 Conv 5 × 5/1	8 16 32 40 16	Same	8 × 8 × 40	8 × 8 × 60
Output layer	GAP	8 × 8			8 × 8 × 60	1 × 60
	Output	Classifier			1 × 60	4

3. Results and Discussions

3.1. Implementation Details

The FCN-DRSA network structure incorporates the FCN-8S as the backbone, with the addition of the DRSA module between the encoder and decoder. The network is fine-tuned with a learning rate of 1×10^{-4}, a Batch Size of 16, and it is trained for 100 epochs. We utilize the L_{dice} as the loss function and employ Adam to optimize the segmentation network.

The multi-channel scoring network is trained using a five-fold cross-validation strategy. This strategy involves updating the test set five times for each type of IHC image while the remaining data are used as the training set. The initialization parameters of all models were pre-trained using ImageNet. The experimental setup includes the following configurations: the initial learning rate is set to 0.0001, automatically adjusted using a learning rate scheduler; the momentum is set to 0.9; the optimizer is set to the Adam function; num_groups is set to 4. We set L_{score} as the loss function, and the number of epochs is 40. The training process was performed on a system equipped with an i9-9900K CPU, NVIDIA GeForce RTX 2080Ti x2, and 64 GB of memory, running on the Windows 10 operating system. The software versions used were Python 3.6 and PyTorch 0.4. The distribution of the training and test sets for each scoring category is shown in Figure 9.

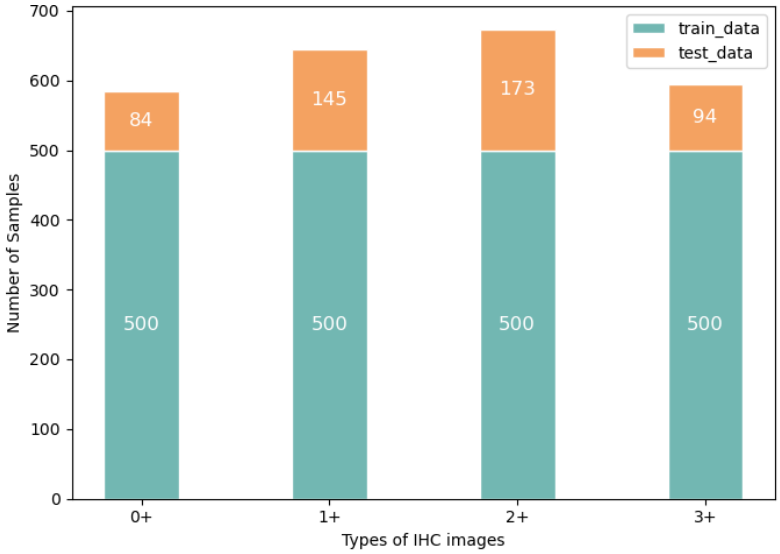


Figure 9. Distribution of experimental datasets, which is divided into the training set and the test set according to a certain proportion.

3.2. Diagnostic Region Segmentation Results

Figure 10 shows examples of diagnostic region segmentation. We display IHC images of four scoring categories in Figure 10a, which exhibit different staining intensities, areas, and tissue structures of IL-24. Key areas with strong staining or complex boundary areas are selected for an enlarged display to observe the segmentation effect more clearly. Figure 10b shows the rough segmentation of the region in the IHC image by FCN-8S. It is evident that the boundary of the segmentation result is fuzzy, and the regional continuity is poor, indicating insufficient attention to the diagnostic region. Figure 10c shows the diagnostic region attention heat map generated by introducing the DRSA module based on FCN-8S. It can be observed that the attention to the diagnostic region is enhanced by the spatial attention mechanism. The color area represents the area with high attention, while the darker surrounding area represents the positive staining area. Figure 10d shows the segmentation result of the FCN-DRSA network proposed in this paper, which is generated by superimposing the diagnostic region attention heat map on the original image. It can be

seen that the diagnostic region segmentation boundary is clear and continuous, with better-preserved details. The segmentation result demonstrates high accuracy and robustness for complex structures and uneven staining. Finally, the diagnostic region is effectively focused on, and the non-diagnostic region is suppressed.

We employ pixel-level classification accuracy metrics Dice Coefficient (DC) and Pixel Accuracy (PA) to quantitatively assess the performance of segmentation, as demonstrated in Equations (11) and (12).

$$DC(PS, RS) = \frac{2 * |PS \cap RS|}{|PS| + |RS|} \tag{11}$$

$$PA = \frac{TDA + TNA}{TA} \tag{12}$$

where PS represents the predicted segmentation results, and RS symbolizes the actual segmentation mask. TDA denotes the number of pixels predicted and accurately labeled as diagnostic areas, while TNA signifies the number of pixels correctly identified as non-diagnostic areas. Lastly, TA stands for the total number of image pixels. Frames Per Second (FPS) refers to the number of samples that the model can perform inference per second.

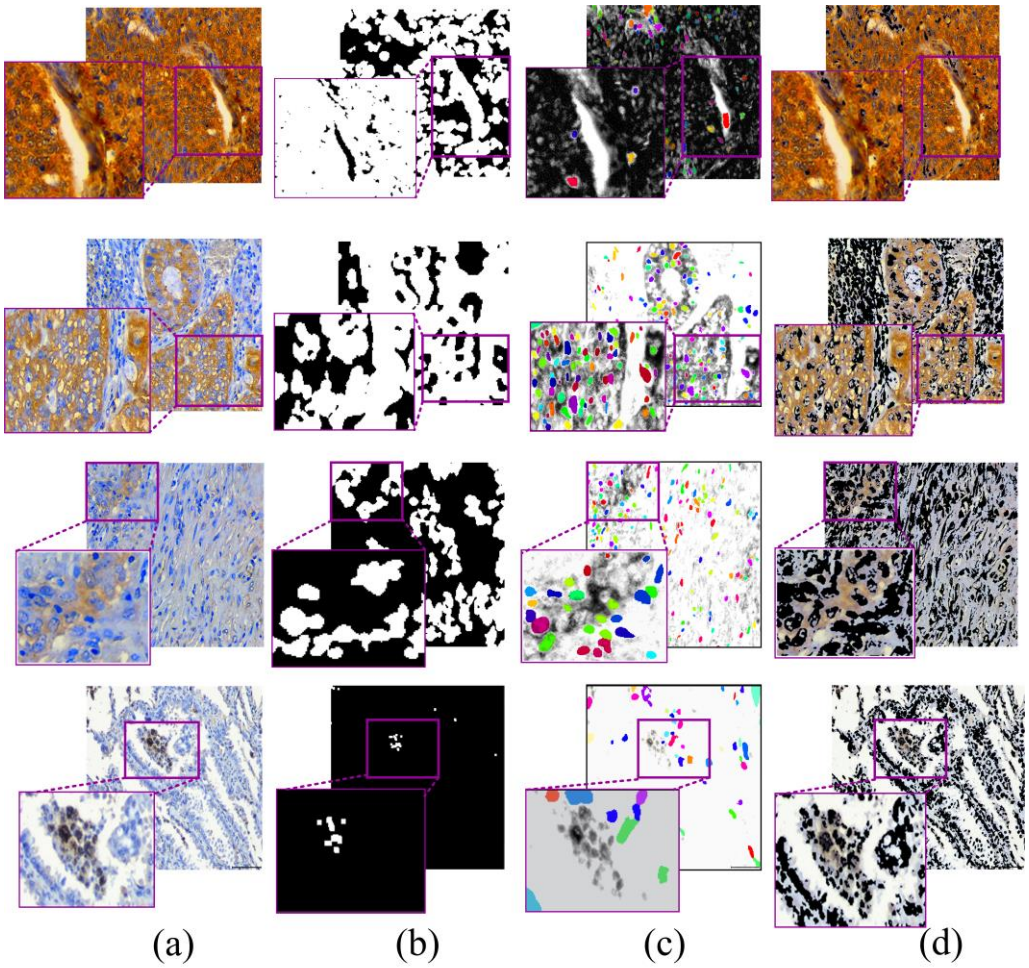


Figure 10. Diagnostic region segmentation example. ((**a**) Original IHC Image displaying categories in descending order: 3+, 2+, 1+, and 0. (**b**) Positive area representation derived from FCN-8S segmentation. (**c**) Intermediate rendering showcasing the introduction of the attention mechanism. (**d**) Final output focusing exclusively on the diagnostic area).

We compared our proposed method, FCN-DRSA, with representative and advanced methods in the field of image segmentation, including SFCN-OPI [57], KiNet [58], SLIC-SVM [32], U-Net [59], and DeepLabv3 [60]. Among them, SLIC-SVM combines traditional machine learning and superpixel technology. U-Net and DeepLabv3 are classic semantic segmentation models that achieve multi-scale feature fusion through skip connections, similar to the model proposed in this paper. SFCN-OPI guides region segmentation by introducing prior knowledge and is suitable for focusing on diagnostic regions. KiNet transforms the segmentation problem into a structural regression problem and designs two branches for ROI segmentation and classification. By comparing these methods, we can comprehensively evaluate the advantages and disadvantages of our proposed FCN-DRSA method and verify its effectiveness in the diagnostic region segmentation task.

As shown in Figure 11 and Table 4, U-Net and DeepLabv3 are commonly used for image semantic segmentation tasks; however, they lack a regional attention mechanism, resulting in fuzzy segmentation of diagnostic areas and reduced accuracy. In contrast, SLIC-SVM leverages an SVM classifier to predict superpixel labels for diagnostic region segmentation. While this approach extracts features at the superpixel level and employs an SVM classifier, it may lead to spatial discontinuities with lower DC and PA values.

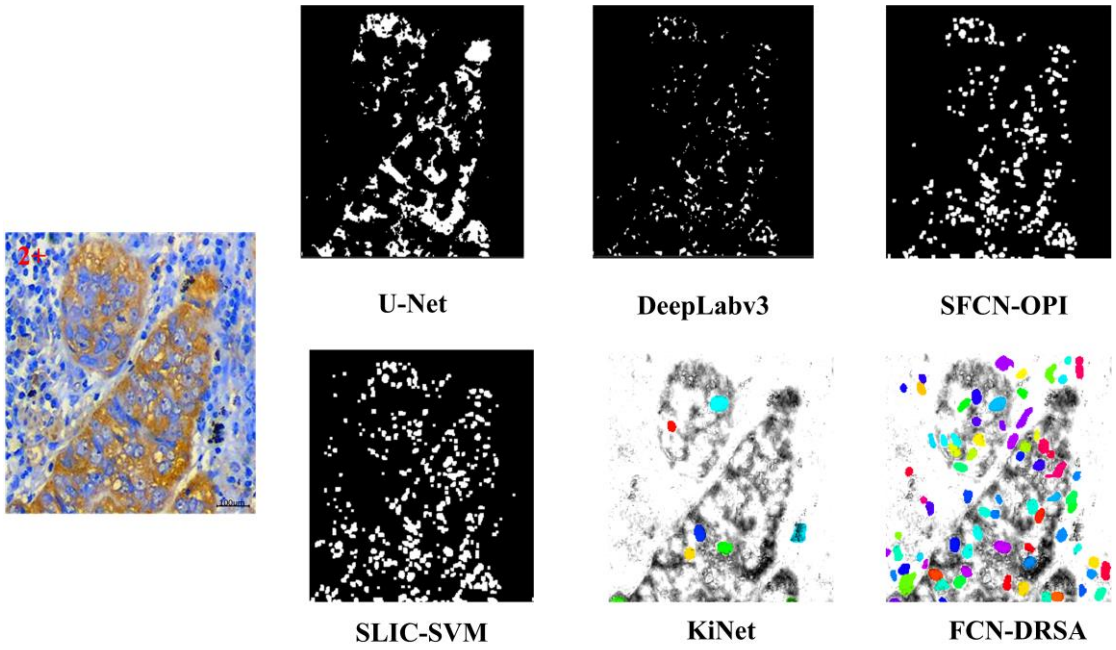


Figure 11. Results of segmentation obtained with SegNet, FCN-8s, and DeepLabv3.

Table 4. Quantitative comparison of methods for diagnostic region segmentation.

Model	DC	PA	FPS
U-Net [59]	0.7934	0.7847	35
DeepLabv3 [60]	0.8054	0.8272	30
SLIC-SVM [32]	0.8301	0.8798	28
SFCN-OPI [57]	0.8579	0.9057	25
KiNet [58]	0.8875	0.898	27
FCN-DRSA	0.9021	0.9471	23

On the other hand, SFCN-OPI adopts direct pixel classification for segmentation, often focusing solely on individual pixel characteristics while neglecting global features. In contrast, KiNet is a deeply structured regression model with two branches designed

to perform both ROI segmentation and classification, achieving notably high DC and PA values.

Our FCN-DRSA shows higher Dice and PA values in IHC image diagnostic region segmentation, promising significantly improved accuracy in subsequent scoring. However, its FPS value is low, indicating that the processing speed is slow. This is mainly because the DRS module needs to perform multiple convolution and weighting operations on the feature map, increasing the computational complexity and processing time. Considering that accurate segmentation of the image has a key impact on the final IHC scoring results, the decrease in processing speed is acceptable. In the future, we will consider using a higher-performance GPU and optimized computing library to improve the operating efficiency of the model, thereby balancing processing speed and accuracy.

3.3. Expert Strong Feature Description Map

We present experimental renderings of IL-24 IHC images for lung cancer, diagnosed as 2+, incorporating areas of partial magnification for detailed view. Figure 12 exhibits discernible disparities in the staining intensity across different regions; the diagnostic area is the most luminous, the non-diagnostic area is predominantly dimmer, and the background area is the darkest. To represent staining intensity, a violin plot is illustrated in Figure 13. It reveals that the IL-24 staining intensity values in the diagnostic area predominantly range between 130 and 200, while in the non-diagnostic area, they are between 50 and 70. The width and area of the box in the violin plot intuitively showcase the staining area characteristics of IL-24, offering an immediate visual insight into the distribution and density of the staining intensity. The mask image created, based on the DAB-Luminance method proposed herein, mitigates the image complexity while preserving the intrinsic robust features identified by experts.

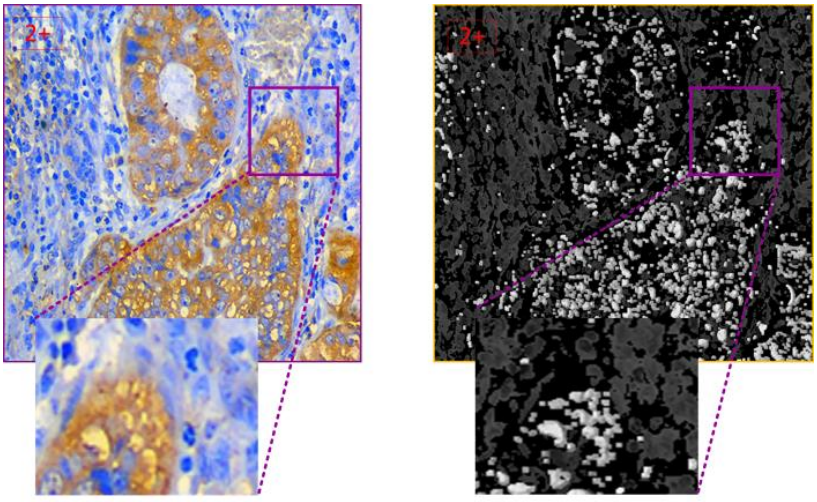


Figure 12. Example of expert strong feature extraction based on DAB-LAMT algorithm. (**a**) original images; (**b**) expert strong feature description map.

3.4. Lung Cancer IHC Scoring Results

To evaluate the effectiveness of the method within the framework presented in this paper, we designed six groups of ablation experiments (see Table 5). Each group of experiment settings differed, including different inputs, the presence or absence of a diagnostic region segmentation method, and the use of distinct scoring networks.

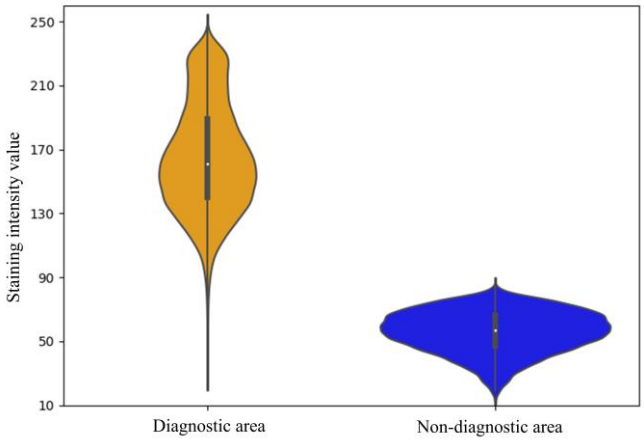


Figure 13. Distribution of staining intensity in diagnostic and non-diagnostic areas.

In Group 1, we input the original IHC images and the expert strong feature description maps obtained by the DAB-LAMT algorithm into the multi-channel scoring network without further processing of diagnostic region segmentation. The comparison between the results of Group 1 and Group 3 is employed to assess the effectiveness of FCN-DRSA. In Group 2, the feature maps after diagnostic region segmentation and expert strong feature extraction are input into a single-channel scoring network (VGG19), whose structural parameters are consistent with the backbone parameters of the dual-channel scoring network. The comparison between the results of Group 2 and Group 3 serves to validate the efficacy of the Multi-channel scoring network introduced in this paper. Furthermore, in Groups 4 and 5, ResNet18 was selected as the scoring network, and the three-channel original IHC images and the single-channel expert strong feature description maps were input, respectively. The comparison between the results of Group 3, Group 4, and Group 5 is conducted to confirm the effectiveness of the proposed expert strong feature description maps. Finally, Groups 3, 4, and 6 present comparative results between the method proposed in this paper and prevalent deep network models.

Table 5. Ablation experiment setup table.

Group		Input		Method	Scoring Network		
		The Original IHC Image	**Expert Strong Feature Description Map**	**FCN-DRSA**	**The Multi-Channel Network**	**ResNet18**	**VGG19**
1	MD-Net	√	√		√		
2	RM-Net	√	√	√			√
3	RMD-Net(Our)	√	√	√	√		
4	ResNet18-RGB	√				√	
5	ResNet18-Expert Map		√			√	
6	VGG19	√					√

The training loss and accuracy curves for the six experimental groups outlined in the table can be observed in Figures 14 and 15. Notably, the training curve of the VGG19 experiment exhibits significant fluctuations, indicating that the model's learning process is unstable and struggles to adapt to the data's characteristic distribution. In contrast, the RM-Net experiment displays a sluggish convergence speed, reaching the fitting state after 25 epochs, which suggests limited learning capacity. ResNet18-RGB demonstrates limited learning capacity for the original image, as evidenced by its slower convergence speed and greater fluctuations. In contrast, ResNet18-Expert Map, when provided with the expert feature map as input, surpasses the RGB input in learning speed, final accuracy, and loss

value. Meanwhile, the MD-Net stabilizes its loss value after the 18th epoch, although it concludes with a relatively high Train_loss of 0.062. The proposed RMD-Net achieves low loss values and high accuracy within the initial 10 training epochs, demonstrating excellent learning efficiency. The loss and accuracy curves remain exceptionally smooth throughout the training process, highlighting the robustness of the model structure and the stability of the optimization process. Ultimately, the model achieves a remarkably low Train_loss value of 0.015.

Upon observing the Test_acc (test accuracy) and MAE (mean absolute error) values in Table 6, it can be found that the training curve of MD-Net stabilizes after epoch 18, showing good convergence. However, its test accuracy (90.12%) is 5.82% lower than that of RMD-Net (95.94%). The training loss (0.062) is 0.047 higher than that of RMD-Net (0.015), and the mean absolute error (0.129) is 0.025 higher than that of RMD-Net (0.104). This demonstrates that focusing on the diagnostic area and suppressing the non-diagnostic area can improve the IHC scoring results, indicating the effectiveness of the diagnostic area segmentation method FCN-DRSA proposed in this paper.

The test accuracy of RMD-Net (95.94%) is 7.03% higher than that of RM-Net (88.91%). The training loss (0.015) is 0.012 lower than that of RM-Net (0.027), and the mean absolute error (0.104) is 0.027 lower than that of RM-Net (0.131). This demonstrates that the dual-channel scoring network proposed in this paper enhances the model's ability to recognize and classify IHC images by extracting and fusing features of different modalities, thus improving scoring accuracy.

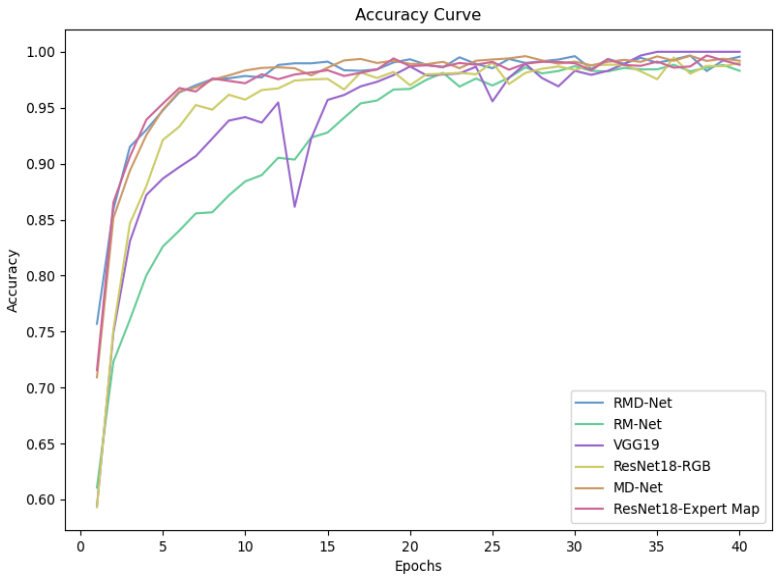


Figure 14. Accuracy curves of our model and other networks training with lung cancer IHC image datasets in 40 epochs.

Table 6. Comparison of training and testing of different models.

Model	Train_loss	Train_acc	Epochs	Test_acc	MAE
VGG19	**0.011**	0.973	40	84.48%	0.192
ResNet18-Expert Map	0.054	0.901	40	73.79%	0.367
ResNet18-RGB	0.020	0.995	40	87.30%	0.157
RM-Net	0.027	0.991	40	88.91%	0.131
MD-Net	0.062	0.989	40	90.12%	0.129
RMD-Net	0.015	**0.998**	40	**95.94%**	**0.104**

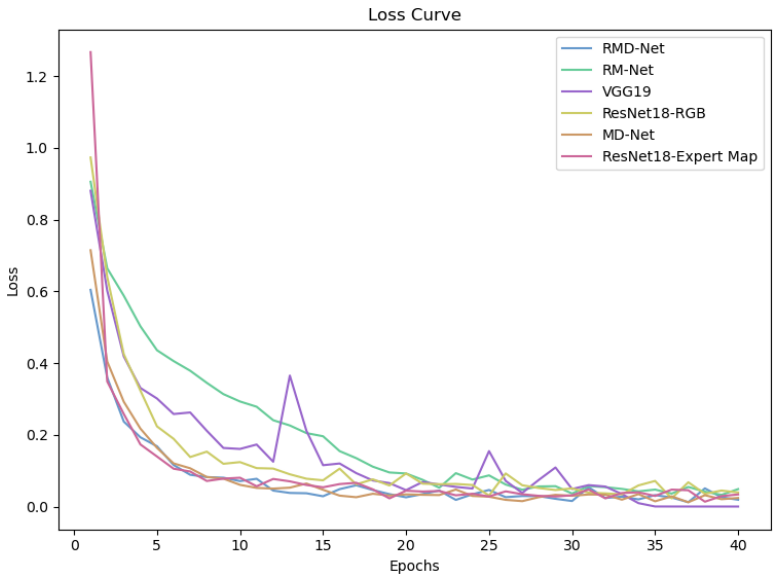


Figure 15. Loss curves of our model and other networks training with lung cancer IHC image datasets in 40 epochs.

In addition, by comprehensively comparing the Test_acc and MAE values of the RMD-Net, ResNet18-Expert Map, and ResNet18-RGB methods, it can be found that the combination of multi-modal information, such as expert robust feature description maps and original images, has a significant impact on the scoring results. This demonstrates that the expert strong feature extraction method proposed in this paper can focus on core features, objectively quantify the expression of IL-24, and improve the accuracy and robustness of IHC scoring.

3.4.1. Qualitative Analysis

To investigate the impact of various interferences in tumor tissue IHC images on the proposed scoring framework, we conducted a detailed individual analysis of selected images from the scoring results. Figure 16 illustrates the scoring outcomes of the RMD-Net and ResNet18-RGB methods proposed in this paper. Among the analyzed images, the IL-24 IHC images belonging to categories 0 and 1+ exhibit complex backgrounds, noise interference, blurred areas, and tissue adhesion. Remarkably, the scoring results generated by RMD-Net are all accurate, which contrasts with the biased scoring results produced by ResNet18-RGB. In category 0 IHC images, the weak staining intensity makes noise interference particularly prominent in the staining regions, increasing the likelihood of false positives. Additionally, blurred staining areas caused by focus offsets or image acquisition issues result in a loss of staining signal details. In category 1+ IHC images, the weak positive staining signals are often unevenly distributed, and tissue adhesion can exaggerate the positive intensity of certain regions, leading the model to mistakenly classify them as 2+ or 3+. Images of category 3+ demonstrate tissue heterogeneity, which arises due to the inherent variations in shape and structure within tumor tissue. To mitigate the influence of such variations, this paper incorporates expert knowledge to extract artificial features, thereby avoiding interference from other features. Category 2+ images display issues related to non-uniform dyeing and color variation. These discrepancies can result from non-uniform dyeing reactions or changes in lighting conditions. However, the proposed method effectively overcomes these challenges, indicating that the multi-channel scoring network exhibits superior robustness and specificity in such scenarios.

3.4.2. Quantitative Analysis

In order to analyze the scoring results of each category of IHC images, we constructed a confusion matrix, which is a commonly used visual representation in the field of pattern recognition. To enhance the clarity of the results, we integrated the confusion matrix into a heat map, as depicted in Figure 17. The results indicate that among the six comparison experiments conducted, the prediction score accuracy for IHC images in the category 0+ is relatively low. The staining of the category 0+ is typically sparse, with positive areas being very small or nearly indistinguishable from the background, making it easy for the model to confuse these regions with background signals. Furthermore, the class 0+ generally lacks prominent staining signals and is highly susceptible to noise interference. The feature attenuation in category 0+ (blurred boundaries and the absence of distinct color or shape) further diminishes the model's ability to effectively capture its features. However, the method proposed in this paper achieves the highest prediction accuracy for scores of 0+. RM-Net demonstrates the highest accuracy in predicting scores for class 1+ IHC images, while MD-Net excels in predicting scores for category 2+ IHC images. Notably, the prediction score accuracy plots for category 3+ IHC images exhibit similar trends across all six experiments. In summary, the scoring framework proposed in this article exhibits superior overall performance and robustness in predicting scores for different categories of IHC images.

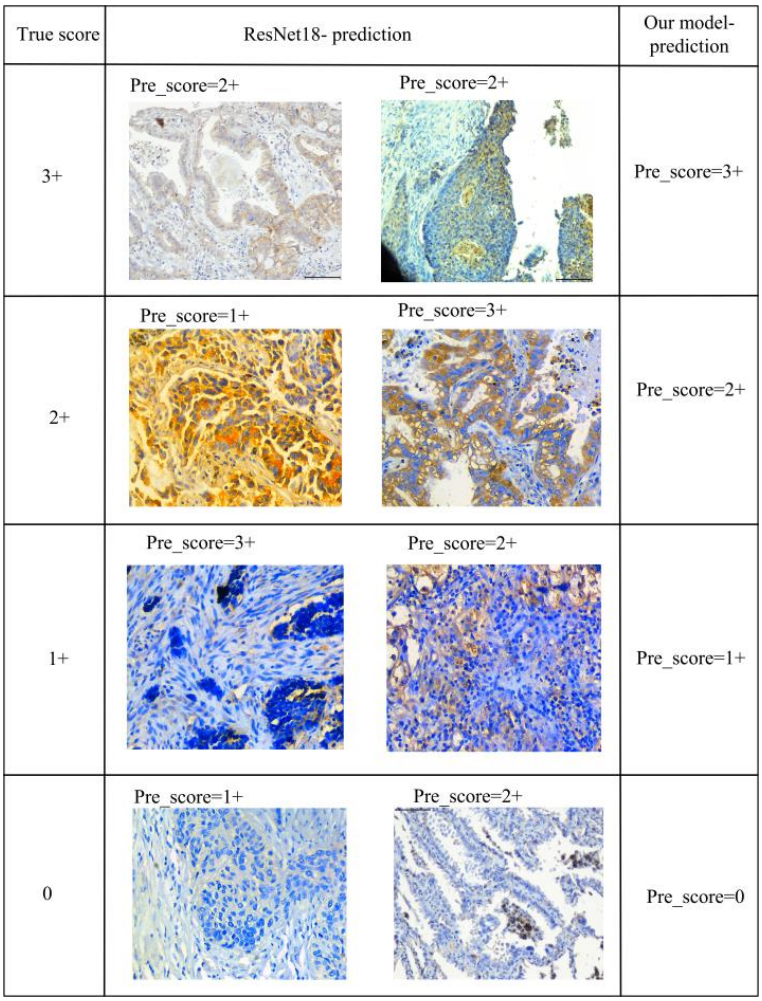


Figure 16. Examples of IL-24 images for four scoring categories. On the left are predictions by the ResNet18 model; on the right are predictions by our proposed method.

To enable a more intuitive comparison of the predictions made by different methods for each category image and emphasize the effectiveness of the method proposed in this paper, we calculated the prediction Precision, Recall, Specificity, F1 Score, and G-mean of each category based on the confusion matrix. As follows:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$Specificity = \frac{TN}{TN + FP} \tag{15}$$

$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{16}$$

$$G - mean = \sqrt{\text{Recall} \cdot Specificity} \tag{17}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively. Finally, the Precision, Recall, F1 score and G-mean of each category are shown in Figure 18a–d.

The F1 score is a metric that represents a weighted average of a model's precision and recall. It provides a comprehensive evaluation of a model's performance. The G-mean metric is particularly useful when the class distribution is imbalanced, meaning that the number of instances in one class is significantly larger than the other. From Figure 18c,d, it is evident that the F1 score and G-mean values of the RMD-Net method proposed in this article for each type of IHC image are higher compared to the results obtained from the ResNet18-Expert Map and ResNet18-RGB methods. This observation showcases the effectiveness of combining expert features with deep abstract features, as proposed in this paper. Furthermore, the F1 score and G-mean values of RMD-Net for each type of IHC image surpass those of the RM-Net method, confirming the effectiveness of the dual-channel scoring network introduced in this research. Additionally, the F1 score and G-mean values of RMD-Net for each type of IHC image outperform those of the MD-Net method, indicating that the proposed diagnostic region segmentation method effectively enhances the final IHC scoring results. From Figure 18a, it can be observed that the prediction precision for categories 2+ and 3+ is relatively high, indicating a low false alarm rate for IL-24 positive and strong positive instances. Considering that the categories 2+ and 3+ exhibit stronger staining signals, more distinct features, and are less influenced by background noise and other interferences. Additionally, regions with high staining intensity tend to have greater consistency and clearer boundaries. The DRSA module effectively enhances the model's attention to these areas, improving its feature extraction capability for the 2+ and 3+ categories and thereby achieving higher scoring accuracy. Similarly, Figure 18b reveals that the recall rates for category 1+, category 2+, and category 3+ are all higher than that of category 0+, suggesting a low false negative rate of IL-24 positivity in cancer patients from a clinical standpoint.

To assess the performance of the algorithms across the dataset, we divided the IL-24 images into four groups based on the pathologists' scores. For each group, we counted the number of IL-24

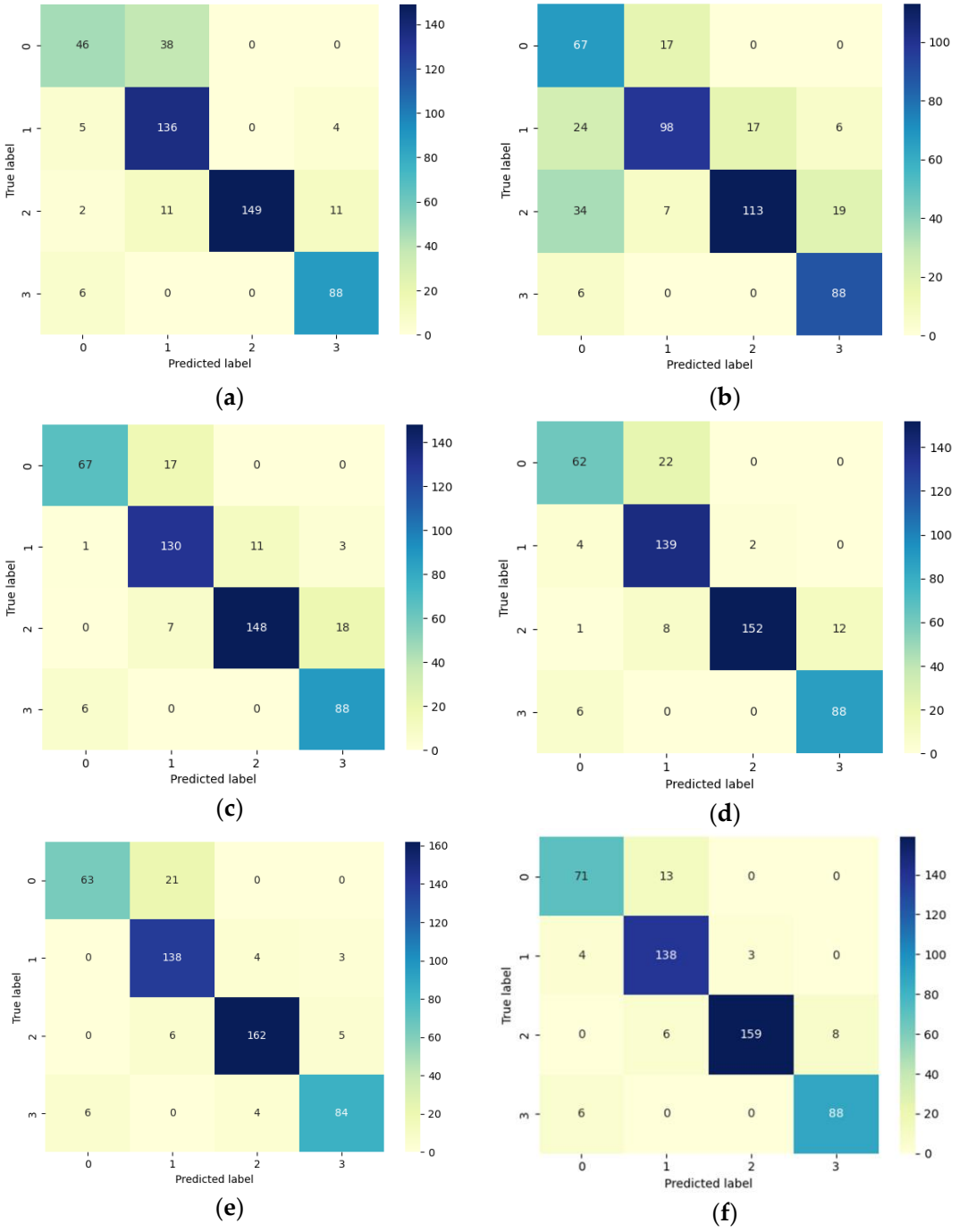


Figure 17. Confusion matrix for different models. (**a**) VGG19Net; (**b**) ResNet18-Expert Map; (**c**) ResNet18-RGB; (**d**) RM-Net; (**e**) MD-Net; (**f**) RMD-Net (Our model).

Images with absolute error (AE) of 0, 1+, 2+, and 3+, respectively. The results of different methods are shown in Figure 19. During the analysis, we identified anomaly detection in the scoring of 3+ category IHC images. Upon observing the dataset, it was found that six images exhibited poor quality, leading to incorrect predictions. Additionally, the ResNet18-Expert Map method showed the poorest prediction scores for 2+ class IHC images. On the other hand, the method proposed in this article, RMD-Net, demonstrated a reduction in abnormal prediction scores for all categories of IHC images, including 0, 1+, 2+, and 3+. This indicates that the proposed RMD-Net method effectively addresses and mitigates the issues associated with abnormal predictions in various IHC image categories.

3.5. Segmentation Accuracy Evaluation

This section aims to assess the influence of diagnostic region segmentation accuracy on IHC scoring. Table 7 illustrates the correlation between the diagnostic region segmentation accuracy of various models and their IHC score prediction. It can be observed that the performance of IHC scoring is improved with the segmentation accuracy. Among the models analyzed, the FCN-DRSA, a diagnosis-related region segmentation network developed in this study, exhibits the highest segmentation accuracy, consequently yielding the most accurate IHC score prediction. This experiment demonstrates that enhancing the specificity of biomarker expression areas through regional segmentation can improve subsequent IHC scoring accuracy. This finding contrasts with numerous studies in the literature where the entire tumor tissue image is fed into the CNN without discerning whether a region is relevant to the objective or not.

Table 7. Performance comparison of IHC scoring with different diagnostic region segmentation models.

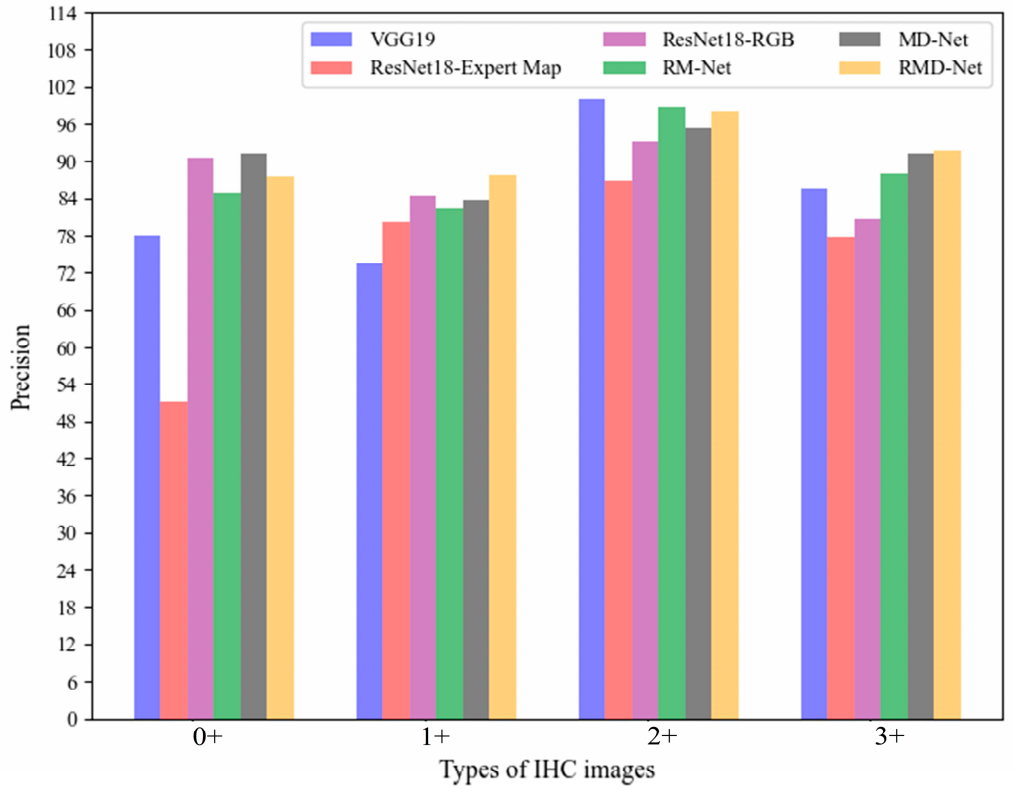
Model	Segmentation Accuracy		IHC Scoring
	DC	PA	Test_acc
SegNet	0.8134	0.8647	90.08%
U-Net	0.8254	0.8772	90.82%
FCN-8s	0.8301	0.8798	91.26%
DeepLabv3	0.8579	0.9057	91.75%
FCN-DRSA	**0.9021**	**0.9471**	**95.94%**

3.6. Comparative Analysis

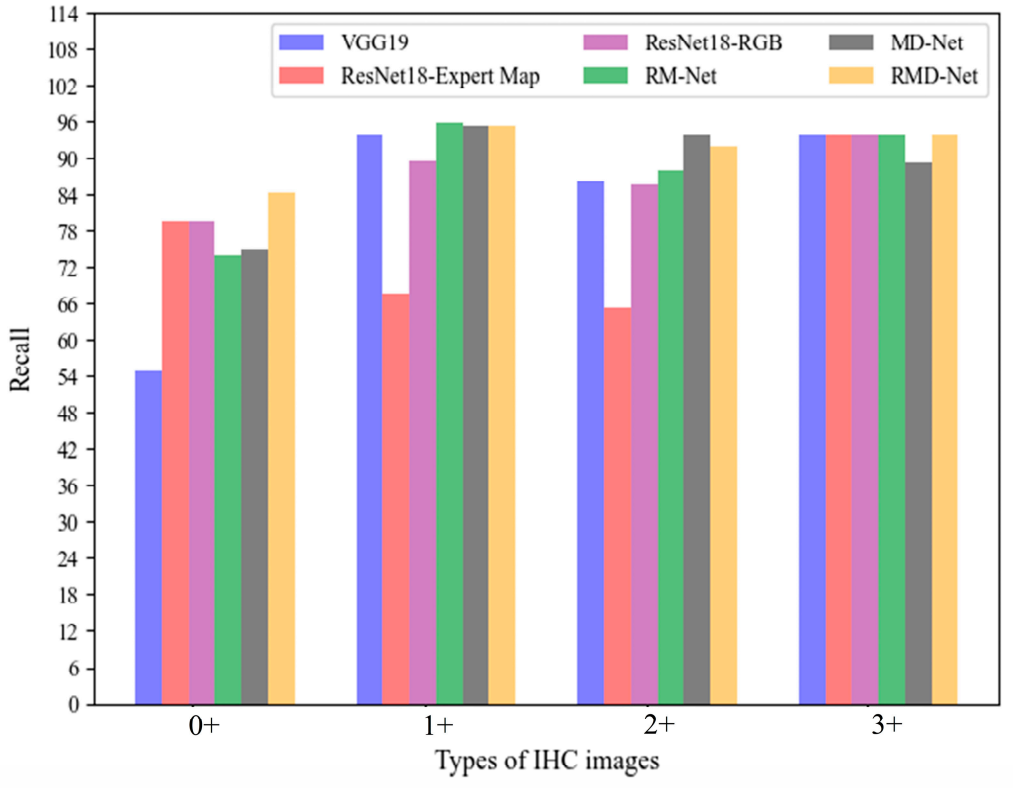
(a) Comparison with popular methods: To evaluate the performance of RMD-Net in terms of accuracy, robustness, and computational efficiency, we conducted a systematic comparison with several popular methods, including DenseNet, ResNet, and MobileNet, using the lung cancer IHC image dataset described in this study. Additionally, to assess the robustness of RMD-Net, we introduced interference factors, such as Gaussian noise and brightness adjustments, into the test set. As shown in Table 8, DenseNet-121 and ResNet-50 demonstrate strong IHC scoring performance, achieving accuracies of 85.26% and 83.74%, respectively. However, both models incorporate numerous dense and residual connections, resulting in high parameter counts and computational complexity. In contrast, GoogleNet and MobileNet (V1) are more suitable for tasks with limited computational resources and lower precision requirements. Notably, RMD-Net surpasses other mainstream models in both accuracy and F1 score while maintaining computational complexity within an acceptable range.

Table 8. Performance and efficiency comparison of RMD-Net and popular models.

Model	Scoring Performance		Computational Efficiency	
	Accuracy	F1 Score	Parameters	GFLOPS
EfficientNet(B7)	81.38%	79.60	61M	20
DenseNet121	85.26%	83.29	15.6M	7.8
GoogleNet	80.82%	77.08	10.5M	6.5
ResNet50	83.74%	81.87	30.7M	4.1
MobileNet(V1)	77.23%	75.61	7.9M	3.2
RMD-Net	**94.21%**	**87.55**	25.4M	10.3

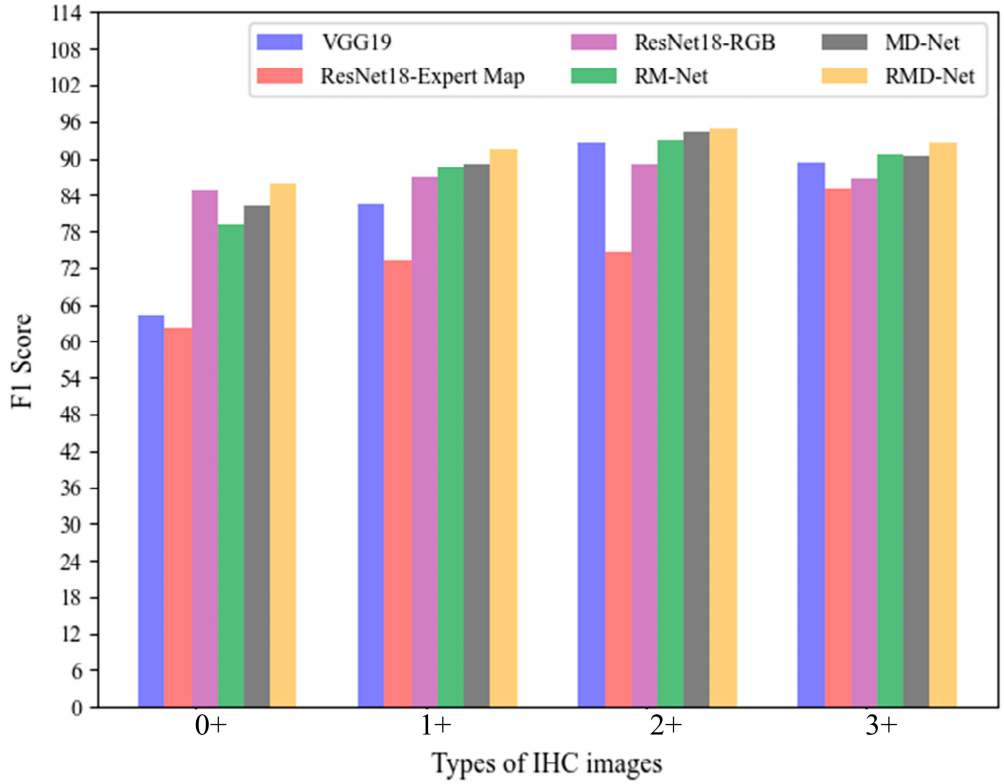


(**a**)

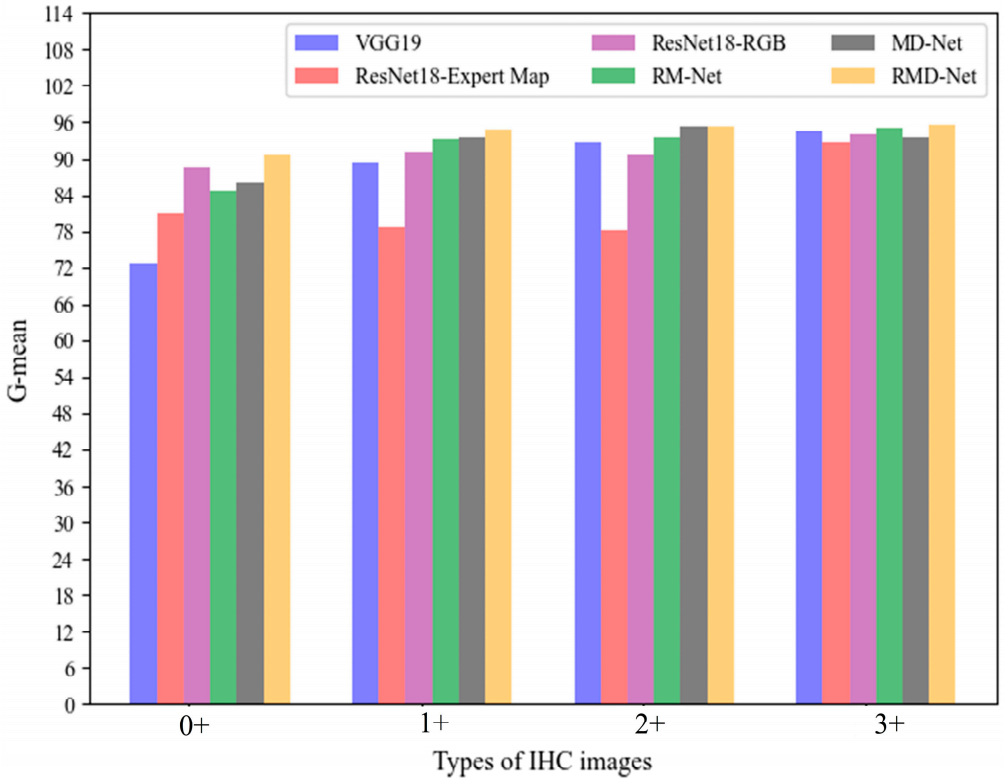


(**b**)

Figure 18. *Cont.*

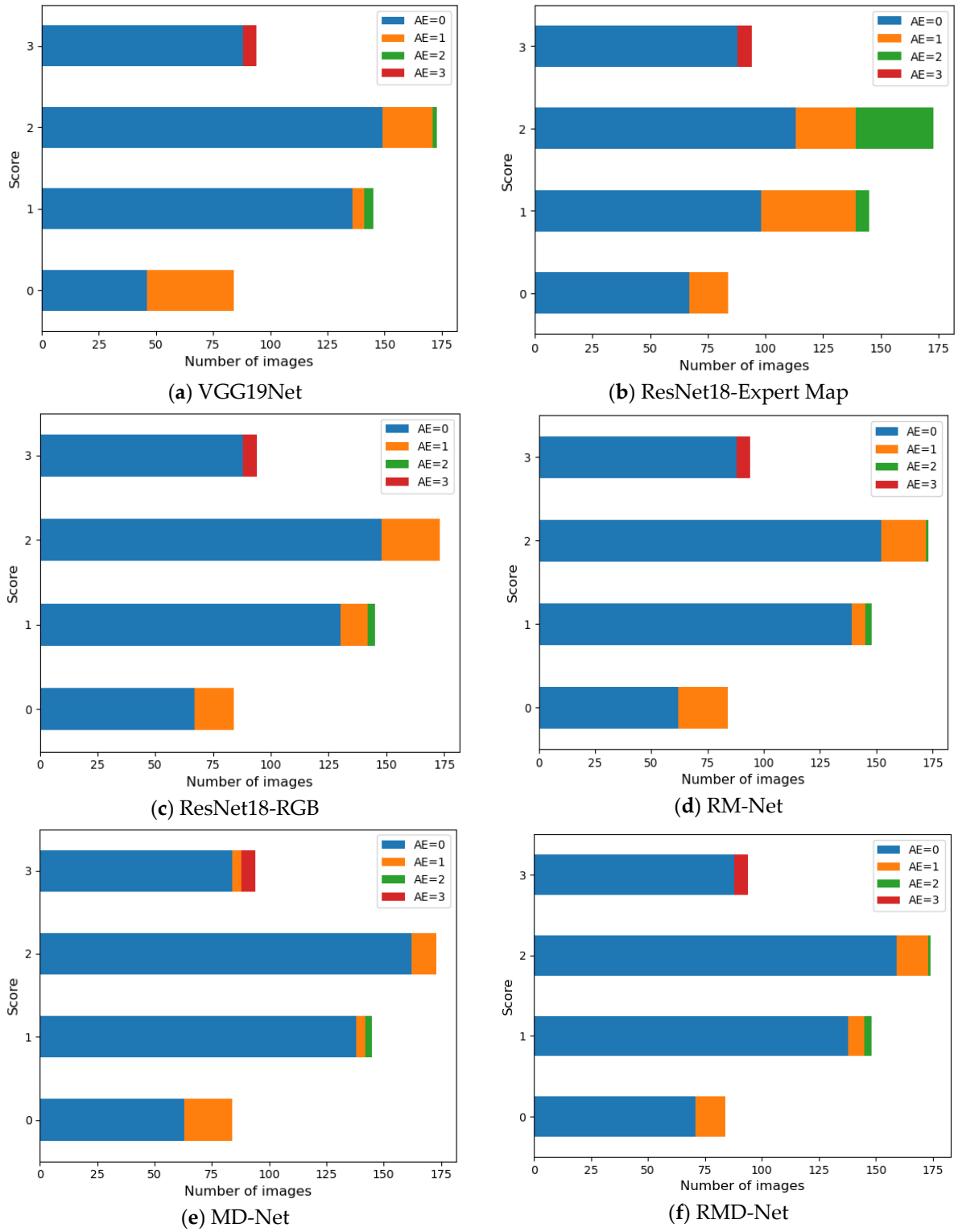


(**c**)



(**d**)

Figure 18. Comparison of Precision, Recall, and F1 indicators of different model structures. (**a**) Precision (%) per class, (**b**) Recall (%) per class, (**c**) F1 score per class, (**d**) G-mean per class.



(**a**) VGG19Net (**b**) ResNet18-Expert Map

(**c**) ResNet18-RGB (**d**) RM-Net

(**e**) MD-Net (**f**) RMD-Net

Figure 19. Comparison of performances AE of different methods in different score groups.

In addition, RMD-Net demonstrates exceptional robustness in IHC scoring tasks, primarily due to its DRSA mechanism. This mechanism enhances the model's ability to focus on diagnosis-relevant areas by generating attention maps, effectively suppressing interference from non-diagnostic regions and artifacts. When integrated with the DAB-LAMT algorithm, RMD-Net accurately extracts expert-level features under varying staining conditions and maintains stability even in samples with significant brightness variations or complex noise. Furthermore, the model integrates these expert-level features with deep

learning's abstract features, enhancing its adaptability and interpretability. This design ensures consistent and accurate performance in complex clinical scenarios.

(b) Comparison with state-of-the-art methods: To evaluate the efficacy of our proposed method in quantifying IL-24 immune molecule expression, we conducted a comparative analysis using a unified lung cancer IHC image dataset from the Institute of Life Sciences and Bioengineering, Beijing Jiaotong University. As shown in Table 9, this approach ensures an equitable and statistically robust comparison across all analytical techniques examined. Niyas et al. [61] employed an improved LadderNet architecture for biomarker expression segmentation and quantification, achieving 83.20% classification accuracy. However, LadderNet's multi-scale feature processing and long skip connections result in extended training times and high memory consumption, limiting its efficiency. Feng et al. [16] utilized a DenseNet-based scoring model, attaining 87.32% accuracy and a 76.65 F1 score. While effective, this method does not incorporate regional attention or expert knowledge, potentially limiting its performance on complex IL-24 expression patterns. Ghoshal et al. [40] proposed HBNet, a hybrid Bayesian neural network combining CNN and manual features. Despite its simpler structure and lower memory requirements, the manual feature extraction process extends preprocessing time and yields lower scoring accuracy. Lai et al. [62] proposed an integrated multi-model data processing system. However, when applied to IL-24 IHC image analysis in lung cancer, this approach achieved only 66.78% accuracy and a 60.46 F1 score. These results indicate its limitations in addressing the unique challenges of color overlap and irregular histomorphology in IL-24 expression quantification for lung cancer tissues. Ghahremani et al. [63] introduced DeepLIIF, a multi-task deep learning framework for IHC image analysis, achieving 90.01% accuracy and 87.72 F1 score. However, this approach lacks integration of expert knowledge in feature extraction and fails to adequately focus on key diagnostic regions. These limitations result in reduced processing efficiency, interpretability, and robustness, particularly for complex lung cancer IHC images. He et al. [64] proposed a "block" attention mechanism and a DarkNet-based scoring network, achieving a scoring accuracy of 91.71%. It can process images of different resolutions through clustering preprocessing operations. The feature extraction block is primarily based on residual blocks, featuring simple structural parameters and low memory usage.

In contrast, our RMD-Net framework specifically addresses IL-24 IHC analysis in lung cancer. By combining FCN-DRSA, DAB-LAMT, and a multi-channel scoring network, we integrate deep learning with expert knowledge. This approach enhances accuracy and interpretability, closely aligning with pathologist assessments and advancing automated IL-24 IHC scoring in lung cancer.

3.7. Limitations and Future Directions

Considering the intricacies of the clinical setting, we delved deeper into the influence of the quality of lung cancer tissue microarray slides on our proposed methodology. Our findings indicate that issues such as defocusing, heterogeneity, and inconsistent staining in the test slides could skew the predictions of our approach. Specifically, inaccurate focusing during microscope imaging can result in blurred regions, adversely impacting the segmentation of diagnostic areas and allowing non-diagnostic regions to influence scoring results, particularly in 0+ and 1+ weakly stained images. Additionally, the heterogeneity of tumor tissue leads to uneven distribution of staining signals, which can blur the boundaries of stained regions and further complicate the scoring process. Variations in staining intensity within IHC images from the same tissue slice, such as weak staining or overstaining, also interfere with the model's ability to accurately determine specific grades.

Table 9. Comparative analysis of existing methods for assessing lung cancer IHC image status.

Reference	Method	Remarks
Niyas et al. [61] (2023)	LadderNet: A network combining multiple encoders and decoders via long skip connections	83.20% Accuracy 72.38 F1 score
Feng et al. [16] (2020)	The small field-of-view image centered on the cell nucleus is input into the DenseNet backbone network for classification and scoring	87.32% Accuracy 76.65 F1 score
Ghoshal et al. [40] (2021)	HBNet, for extracting deep image features	89.25% Accuracy 83.31 F1 score
Lai et al. [62] (2020)	An integrative DNN framework	66.78 Accuracy 60.46 F1 score
Ghahremani et al. [63] (2022)	A multi-task deep learning framework referred to as DeepLIIF	90.01% Accuracy 87.72 F1 score
He et al. [64] (2024)	Block attention mechanism; Improved lung cancer IHC scoring network based on Darknet	91.71% Accuracy 88.79 F1 score
Our method	RMD-Net(FCN-DRSA; multi-modal feature extraction; the dual-channel scoring model)	**95.94% Accuracy 89.00 F1 score**

In response to these identified constraints, a straightforward solution is to incorporate low-quality images into the training set, thereby enhancing the model's resilience. In addition, incorporating image processing modules, such as focus optimization, background denoising, or staining normalization, before feeding images into the scoring network could help mitigate the impact of low-quality images on model predictions. Additionally, reinforcement learning techniques could be employed to adaptively adjust low-quality images, enhancing the network's sensitivity and specificity when handling suboptimal images.

Evaluation and detection of IHC biomarkers in tumor tissues play a pivotal role in determining the suitable treatment for cancer patients. This research will delve into the nuances and commonalities of automatic detection and scoring methods for various biomarkers across different cancers. Our insights aim to pave the way for future advancements in the automated detection of clinical tumor immunohistochemistry.

4. Conclusions

In this paper, we introduce a scoring network framework named RMD-Net, specifically tailored for the evaluation of lung cancer IHC IL-24. The framework comprises three stages: initial segmentation of the diagnostic area, subsequent extraction of robust expert feature maps, and the ultimate integration of multiple modal features into the multi-channel scoring network, culminating in the output of IHC scoring results. In our study, we meticulously crafted six groups of comparative experiments, performing both qualitative and quantitative analyses of the outcomes. Furthermore, we conducted a thorough comparison with contemporary state-of-the-art methods. Our approach, as presented in this paper, outperforms other experimental results, boasting a precision of 89.25%, an F1 score of 89.00, and an accuracy of 95.94%. In summary, the regional attention mechanism and expert strong feature extraction method proposed in this framework meet the requirements of pathology experts and ensure the high accuracy and interpretability of the scoring results. Additionally, the multi-modal features are fused through the multi-channel scoring network to improve the model's scoring performance, demonstrating its effectiveness and superiority in processing complex image tasks. The final scoring results are statistically

significantly consistent with those of multiple expert pathologists. This fully demonstrates that the scoring framework proposed in this paper has practical potential and value and can be used as an innovative and effective tool for the clinical diagnosis and treatment of lung cancer.

Author Contributions: Z.H.: Methodology, Software development, Data processing, Writing—original draft; D.J.: Conceptualization, Supervision, Project administration, Writing—review and editing; Y.S.: Clinical data analysis, Validation, Investigation; Z.L.: Data curation, Software testing; N.W. and F.Z.: Writing—review and editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundamental Research Funds for the Central Universities, China [grant numbers 2024YJS007].

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author (dongyaojia1974@163.com) upon reasonable request. The immunohistochemical images were obtained from the Institute of Life Sciences and Bioengineering at Beijing Jiaotong University.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Abbreviations

IHC	Immunohistochemical
IL-24	Interleukin-24
CAD	Computer-Aided Diagnosis
ROI	regions of interest
HER2	Human Epidermal Growth Factor Receptor 2
DAB	Diaminobenzidine
PR	Progesterone Receptor
ER	Estrogen Receptor
Ki-67	Ki-67 antigen
PD-L1	Programmed Death Ligand 1
EGFR	Epidermal Growth Factor Receptor
PCAN	Proliferating Cell Nuclear Antigen
H-Score	Histological Score
TC	Tumor Cells
HEP-2	Human Epithelial type 2
WSI	Whole Slide Image
CNN	Convolutional Neural Network
R-CNN	Region-CNN
SSD	Single Shot MultiBox Detector
FCN	Fully Convolutional Network
HSL	Hue, Saturation, and Lightness
AE	Absolute Error

References

1. Leiter, A.; Veluswamy, R.R.; Wisnivesky, J.P. The global burden of lung cancer: Current status and future trends. *Nat. Rev. Clin. Oncol.* **2023**, *20*, 624–639. [CrossRef] [PubMed]
2. Viswanathan, V.S.; Toro, P.; Corredor, G.; Mukhopadhyay, S.; Madabhushi, A. The state of the art for artificial intelligence in lung digital pathology. *J. Pathol.* **2022**, *257*, 413–429. [CrossRef] [PubMed]
3. Conte, P.; Ascierto, P.; Patelli, G.; Danesi, R.; Vanzulli, A.; Sandomenico, F.; Tarsia, P.; Cattelan, A.; Comes, A.; De Laurentiis, M.; et al. Drug-induced interstitial lung disease during cancer therapies: Expert opinion on diagnosis and treatment. *ESMO Open* **2022**, *7*, 100404. [CrossRef]

4. Hsu, M.T.; Wang, Y.K.; Tseng, Y.J. Exosomal proteins and lipids as potential biomarkers for lung cancer diagnosis, prognosis, and treatment. *Cancers* **2022**, *14*, 732. [CrossRef]
5. Rasoolian, M.; Kheirollahi, M.; Hosseini, S.Y. MDA-7/interleukin 24 (IL-24) in tumor gene therapy: Application of tumor penetrating/homing peptides for improvement of the effects. *Expert Opin. Biol. Ther.* **2019**, *19*, 211–223. [CrossRef]
6. Jin, K.-T.; Du, W.-L.; Liu, Y.-Y.; Lan, H.-R.; Si, J.-X.; Mou, X.-Z. Oncolytic Virotherapy in Solid Tumors: The Challenges and Achievements. *Cancers* **2021**, *13*, 588. [CrossRef]
7. Kumar, A.; Das, S.K.; Emdad, L.; Fisher, P.B. Applications of tissue-specific and cancer-selective gene promoters for cancer diagnosis and therapy. *Adv. Cancer Res.* **2023**, *160*, 253–315.
8. Taube, J.M.; Akturk, G.; Angelo, M.; Engle, E.L.; Gnjatic, S.; Greenbaum, S.; Greenwald, N.F.; Hedvat, C.V.; Hollmann, T.J.; Juco, J.; et al. The Society for Immunotherapy in Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *J. Immunother. Cancer* **2020**, *8*, e000155. [CrossRef]
9. Camarillo, G.F.; Furuzawa-Carballeda, J.; Martínez-Benítez, B.; Barreto-Zúñiga, R.; Yamamoto-Furusho, J.K. Role of the interleukin 24 in patients with ulcerative colitis. *Inflamm. Bowel Dis.* **2011**, *17*, 2209–2210. [CrossRef]
10. Glenn McCluggage, W. Immunohistochemistry as a diagnostic aid in cervical pathology. *Pathology* **2007**, *39*, 97–111. [CrossRef]
11. Vrtis, M.C. Oncology Care in Non-Small Cell Lung Cancer—Personalizing the Treatment Plan. *Home Healthc. Now* **2023**, *41*, 68–77. [CrossRef] [PubMed]
12. Yatabe, Y.; Dacic, S.; Borczuk, A.C.; Warth, A.; Russell, P.A.; Lantuejoul, S.; Beasley, M.B.; Thunnissen, E.; Pelosi, G.; Rekhtman, N.; et al. Best practices recommendations for diagnostic immunohistochemistry in lung cancer. *J. Thorac. Oncol.* **2019**, *14*, 377–407. [CrossRef] [PubMed]
13. Boyd, J.; Villa, I.; Mathieu, M.C.; Deutsch, E.; Paragios, N.; Vakalopoulou, M.; Christodoulidis, S. Region-guided cyclegans for stain transfer in whole slide images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 356–365.
14. Harms, P.W.; Frankel, T.L.; Moutafi, M.; Rao, A.; Rimm, D.L.; Taube, J.M.; Thomas, D.; Chan, M.P.; Pantanowitz, L. Multiplex immunohistochemistry and immunofluorescence: A practical update for pathologists. *Mod. Pathol.* **2023**, *36*, 100197. [CrossRef] [PubMed]
15. Jin, Y.; Shi, K.; Gao, X.; Langenbach, S.Y.; Li, M.; Harris, T.; Stewart, A.G. An automated quantitative method to analyze immunohistochemistry and immunocytochemistry images. *J. Eng. Sci. Med. Diagn. Ther.* **2020**, *3*, 044503. [CrossRef]
16. Feng, M.; Chen, J.; Xiang, X.; Deng, Y.; Zhou, Y.; Zhang, Z.; Zheng, Z.; Bao, J.; Bu, H. An advanced automated image analysis model for scoring of ER, PR, HER-2 and Ki-67 in breast carcinoma. *IEEE Access* **2020**, *9*, 108441–108451. [CrossRef]
17. Tan, W.C.; Nerurkar, S.N.; Cai, H.Y.; Ng, H.H.; Wu, D.; Wee, Y.T.; Lim, J.C.; Yeong, J.; Lim, T.K. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun.* **2020**, *40*, 135–153. [CrossRef]
18. Bencze, J.; Szarka, M.; Kóti, B.; Seo, W.; Hortobágyi, T.G.; Bencs, V.; Módis, L.V.; Hortobágyi, T. Comparison of semi-quantitative scoring and artificial intelligence aided digital image analysis of chromogenic immunohistochemistry. *Biomolecules* **2021**, *12*, 19. [CrossRef]
19. Cornish, T.C. Clinical application of image analysis in pathology. *Adv. Anat. Pathol.* **2020**, *27*, 227–235. [CrossRef]
20. Ram, S.; Vizcarra, P.; Whalen, P.; Deng, S.; Painter, C.L.; Jackson-Fisher, A.; Pirie-Shepherd, S.; Xia, X.; Powell, E.L. Pixelwise H-score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. *PLoS ONE* **2021**, *16*, e0245638. [CrossRef]
21. Zhong, Y.; Zhang, X.; Chong, W. Interleukin-24 immunobiology and its roles in inflammatory diseases. *Int. J. Mol. Sci.* **2022**, *23*, 627. [CrossRef]
22. Kabakçı, K.A.; Çakır, A.; Türkmen, İ.; Töreyin, B.U.; Çapar, A. Automated scoring of CerbB2/HER2 receptors using histogram based analysis of immunohistochemistry breast cancer tissue images. *Biomed. Signal Process. Control* **2021**, *69*, 102924. [CrossRef]
23. Liang, Q.; Ma, D.; Gao, R.-F.; Yu, K.-D. Effect of Ki-67 expression levels and histological grade on breast cancer early relapse in patients with different immunohistochemical-based subtypes. *Sci. Rep.* **2020**, *10*, 7648. [CrossRef] [PubMed]
24. Alataki, A.; Zabaglo, L.; Tovey, H.; Dodson, A.; Dowsett, M. A simple digital image analysis system for automated Ki67 assessment in primary breast cancer. *Histopathology* **2021**, *79*, 200–209. [CrossRef] [PubMed]
25. Lu, W.; Toss, M.; Dawood, M.; Rakha, E.; Rajpoot, N.; Minhas, F. SlideGraph+: Whole slide image level graphs to predict HER2 status in breast cancer. *Med. Image Anal.* **2022**, *80*, 102486. [CrossRef]
26. Wu, J.; Mayer, A.T.; Li, R. Integrated imaging and molecular analysis to decipher tumor microenvironment in the era of immunotherapy. In *Seminars in Cancer Biology*; Academic Press: Cambridge, MA, USA, 2022; Volume 84, pp. 310–328.
27. Tewary, S.; Mukhopadhyay, S. HER2 molecular marker scoring using transfer learning and decision level fusion. *J. Digit. Imaging* **2021**, *34*, 667–677. [CrossRef]

28. Aeffner, F.; Zarella, M.D.; Buchbinder, N.; Bui, M.M.; Goodman, M.R.; Hartman, D.J.; Lujan, G.M.; Molani, M.A.; Parwani, A.V.; Lillard, K.; et al. Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. *J. Pathol. Inform.* **2019**, *10*, 9. [CrossRef]
29. Zidane, M.; Makky, A.; Bruhns, M.; Rochwarger, A.; Babaei, S.; Claassen, M.; Schürch, C.M. A review on deep learning applications in highly multiplexed tissue imaging data analysis. *Front. Bioinform.* **2023**, *3*, 1159381. [CrossRef]
30. Feng, M.; Deng, Y.; Yang, L.; Jing, Q.; Zhang, Z.; Xu, L.; Wei, X.; Zhou, Y.; Wu, D.; Xiang, F.; et al. Automated quantitative analysis of Ki-67 staining and HE images recognition and registration based on whole tissue sections in breast carcinoma. *Diagn. Pathol.* **2020**, *15*, 1–12. [CrossRef]
31. Xu, B.; Liu, J.; Hou, X.; Liu, B.; Garibaldi, J.; Ellis, I.O.; Green, A.; Shen, L.; Qiu, G. Look, investigate, and classify: A deep hybrid attention method for breast cancer classification. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 914–918.
32. Khameneh, F.D.; Razavi, S.; Kamasak, M. Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network. *Comput. Biol. Med.* **2019**, *110*, 164–174. [CrossRef]
33. Zhang, J.; Guo, X.; Cai, W.; Cao, Y.; Ge, W.; Huang, P.; Hou, D.; Zheng, S.; Zhang, G. Automatic Scoring Method for Tumor IHC Images Based on Deep Learning and Its Application on P53 Protein. *IEEE Access* **2023**, *11*, 64005–64015. [CrossRef]
34. Jin, Q.; Cui, H.; Sun, C.; Song, Y.; Zheng, J.; Cao, L.; Wei, L.; Su, R. Inter-and intra-uncertainty based feature aggregation model for semi-supervised histopathology image segmentation. *Expert Syst. Appl.* **2024**, *238*, 122093. [CrossRef]
35. Guo, R.; Ma, L.; Bai, X.; Miao, L.; Li, Z.; Yang, J. A scoring method for Immunohistochemical staining on Ki67. *Appl. Immunohistochem. Mol. Morphol.* **2021**, *29*, e20–e28. [CrossRef] [PubMed]
36. Liu, J.; Zheng, Q.; Mu, X.; Zuo, Y.; Xu, B.; Jin, Y.; Wang, Y.; Tian, H.; Yang, Y.; Xue, Q.; et al. Automated tumor proportion score analysis for PD-L1 (22C3) expression in lung squamous cell carcinoma. *Sci. Rep.* **2021**, *11*, 15907. [CrossRef] [PubMed]
37. Baxi, V.; Edwards, R.; Montalto, M.; Saha, S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod. Pathol.* **2022**, *35*, 23–32. [CrossRef]
38. Durán-Díaz, I.; Sarmiento, A.; Fondón, I.; Bodineau, C.; Tomé, M.; Durán, R.V. A Robust Method for the Unsupervised Scoring of Immunohistochemical Staining. *Entropy* **2024**, *26*, 165. [CrossRef]
39. Xue, Z.-Z.; Wu, Y.; Gao, Q.-Z.; Zhao, L.; Xu, Y.-Y. Automated classification of protein subcellular localization in immunohistochemistry images to reveal biomarkers in colon cancer. *BMC Bioinform.* **2020**, *21*, 398. [CrossRef]
40. Ghoshal, B.; Hikmet, F.; Pineau, C.; Tucker, A.; Lindskog, C. DeepHistoClass: A novel strategy for confident classification of immunohistochemistry images using deep learning. *Mol. Cell. Proteom.* **2021**, *20*, 100140. [CrossRef]
41. Fassler, D.J.; Abousamra, S.; Gupta, R.; Chen, C.; Zhao, M.; Paredes, D.; Batool, S.A.; Knudsen, B.S.; Escobar-Hoyos, L.; Shroyer, K.R.; et al. Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images. *Diagn. Pathol.* **2020**, *15*, 100. [CrossRef]
42. Hameed, K.S.; Abubacker, K.S.; Banumathi, A.; Ulaganathan, G. Immunohistochemical analysis of oral cancer tissue images using support vector machine. *Measurement* **2021**, *173*, 108476. [CrossRef]
43. Rmili, H.; Mouelhi, A.; Solaiman, B.; Doghri, R.; Labidi, S. A novel pre-processing approach based on colour space assessment for digestive neuroendocrine tumour grading in immunohistochemical tissue images. *Pol. J. Pathol.* **2022**, *73*, 134–158. [CrossRef]
44. Mahanta, L.B.; Hussain, E.; Das, N.; Kakoti, L.; Chowdhury, M. IHC-Net: A fully convolutional neural network for automated nuclear segmentation and ensemble classification for Allred scoring in breast pathology. *Appl. Soft Comput.* **2021**, *103*, 107136. [CrossRef]
45. Anaam, A.; Al-Antari, M.A.; Hussain, J.; Samee, N.A.; Alabdulhafith, M.; Gofuku, A. Deep active learning for automatic mitotic cell detection on HEp-2 specimen medical images. *Diagnostics* **2023**, *13*, 1416. [CrossRef] [PubMed]
46. Molnar, C.; Tavolara, T.E.; Garcia, C.A.; McClintock, D.S.; Zarella, M.D.; Han, W. IHCScoreGAN: An unsupervised generative adversarial network for end-to-end ki67 scoring for clinical breast cancer diagnosis. In Proceedings of the Medical Imaging with Deep Learning, Paris, France, 3–5 July 2024.
47. Wen, Z.; Luo, D.; Wang, S.; Rong, R.; Evers, B.M.; Jia, L.; Fang, Y.; Daoud, E.V.; Yang, S.; Gu, Z.; et al. Deep Learning–Based H-Score Quantification of Immunohistochemistry-Stained Images. *Mod. Pathol.* **2024**, *37*, 100398. [CrossRef] [PubMed]
48. Springenberg, M.; Frommholz, A.; Wenzel, M.; Weicken, E.; Ma, J.; Strodthoff, N. From modern CNNs to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology. *Med. Image Anal.* **2023**, *87*, 102809. [CrossRef]
49. Taylor, N.J.; Nikolaishvili-Feinberg, N.; Midkiff, B.R.B.; Conway, K.; Millikan, R.C.; Geradts, J. Rational manual and automated scoring thresholds for the immunohistochemical detection of TP53 missense mutations in human breast carcinomas. *Appl. Immunohistochem. Mol. Morphol.* **2016**, *24*, 398–404. [CrossRef]
50. Hameed KA, S.; Banumathi, A.; Ulaganathan, G. Performance evaluation of maximal separation techniques in immunohistochemical scoring of tissue images. *Micron* **2015**, *79*, 29–35. [CrossRef]

51. Irshad, H.; Oh, E.-Y.; Schmolze, D.; Quintana, L.M.; Collins, L.; Tamimi, R.M.; Beck, A.H. Crowdsourcing scoring of immunohistochemistry images: Evaluating performance of the crowd and an automated computational method. *Sci. Rep.* **2017**, *7*, 43286. [CrossRef]
52. Sode, M.; Thagaard, J.; Eriksen, J.O.; Laenkholm, A. Digital image analysis and assisted reading of the HER2 score display reduced concordance: Pitfalls in the categorisation of HER2-low breast cancer. *Histopathology* **2023**, *82*, 912–924. [CrossRef]
53. Zilenaite, D.; Rasmusson, A.; Augulis, R.; Besusparis, J.; Laurinaviciene, A.; Plancoulaine, B.; Ostapenko, V.; Laurinavicius, A. Independent prognostic value of intratumoral heterogeneity and immune response features by automated digital immunohistochemistry analysis in early hormone receptor-positive breast carcinoma. *Front. Oncol.* **2020**, *10*, 950. [CrossRef]
54. Vahadane, A.; Peng, T.; Sethi, A.; Albarqouni, S.; Wang, L.; Baust, M.; Steiger, K.; Schlitter, A.M.; Esposito, I.; Navab, N. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1962–1971. [CrossRef]
55. Onder, D.; Zengin, S.; Sarioglu, S. A review on color normalization and color deconvolution methods in histopathology. *Appl. Immunohistochem. Mol. Morphol.* **2014**, *22*, 713–719. [CrossRef] [PubMed]
56. Kumari, N.; Anwar, S.; Bhattacharjee, V. Automated visual stimuli evoked multi-channel EEG signal classification using EEGCapsNet. *Pattern Recognit. Lett.* **2022**, *153*, 29–35. [CrossRef]
57. Zhou, Y.; Dou, Q.; Chen, H.; Qin, J.; Heng, P.-A. Sfcn-opi: Detection and fine-grained classification of nuclei using sibling fcn with objectness prior interaction. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
58. Xing, F.; Cornish, T.C.; Bennett, T.; Ghosh, D.; Yang, L. Pixel-to-pixel learning with weak supervision for single-stage nucleus recognition in ki67 images. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3088–3097. [CrossRef] [PubMed]
59. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
60. Yurtkulu, S.C.; Şahin, Y.H.; Unal, G. Semantic segmentation with extended DeepLabv3 architecture. In Proceedings of the 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 24–26 April 2019; pp. 1–4.
61. Niyas, S.; Bygari, R.; Naik, R.; Viswanath, B.; Ugwekar, D.; Mathew, T.; Kavya, J.; Kini, J.R.; Rajan, J. Automated molecular subtyping of breast carcinoma using deep learning techniques. *IEEE J. Transl. Eng. Health Med.* **2023**, *11*, 161–169. [CrossRef]
62. Lai, Y.-H.; Chen, W.-N.; Hsu, T.-C.; Lin, C.; Tsao, Y.; Wu, S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Sci. Rep.* **2020**, *10*, 4679. [CrossRef]
63. Ghahremani, P.; Li, Y.; Kaufman, A.; Vanguri, R.; Greenwald, N.; Angelo, M.; Hollmann, T.J.; Nadeem, S. Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. *Nat. Mach. Intell.* **2022**, *4*, 401–412. [CrossRef]
64. He, Z.; Jia, D.; Zhang, C.; Li, Z.; Wu, N. An automatic Darknet-based immunohistochemical scoring system for IL-24 in lung cancer. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107485. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.