life

MDPI

Article

Accurate Kidney Pathological Image Classification Method Based on Deep Learning and Multi-Modal Fusion Method with Application to Membranous Nephropathy

Fang Hao [1,*], Xueyu Liu [1], Ming Li [1] and Weixia Han [2]

[1] College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China
[2] Department of Pathology, Second Hospital of Shanxi Medical University, Taiyuan 030001, China
* Correspondence: haofang0100@link.tyut.edu.cn

Abstract: Membranous nephropathy is one of the most prevalent conditions responsible for nephrotic syndrome in adults. It is clinically nonspecific and mainly diagnosed by kidney biopsy pathology, with three prevalent techniques: light microscopy, electron microscopy, and immunofluorescence microscopy. Manual observation of glomeruli one by one under the microscope is very time-consuming, and there are certain differences in the observation results between physicians. This study makes use of whole-slide images scanned by a light microscope as well as immunofluorescence images to classify patients with membranous nephropathy. The framework mainly includes a glomerular segmentation module, a confidence coefficient extraction module, and a multi-modal fusion module. This framework first identifies and segments the glomerulus from whole-slide images and immunofluorescence images, and then a glomerular classifier is trained to extract the features of each glomerulus. The results are then combined to produce the final diagnosis. The results of the experiments show that the F1-score of image classification results obtained by combining two kinds of features, which can reach 97.32%, is higher than those obtained by using only light-microscopy-observed images or immunofluorescent images, which reach 92.76% and 93.20%, respectively. Experiments demonstrate that considering both WSIs and immunofluorescence images is effective in improving the diagnosis of membranous nephropathy.

Keywords: membranous nephropathy; kidney pathology; deep learning; multi-modal fusion; whole-slide images; immunofluorescence images

check for updates

Citation: Hao, F.; Liu, X.; Li, M.; Han, W. Accurate Kidney Pathological Image Classification Method Based on Deep Learning and Multi-Modal Fusion Method with Application to Membranous Nephropathy. *Life* **2023**, *13*, 399. https://doi.org/10.3390/life13020399

Academic Editors: Binh P. Nguyen and Trang Do

Received: 28 November 2022
Revised: 25 January 2023
Accepted: 27 January 2023
Published: 31 January 2023

Copyright: © 2023 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

Membranous nephropathy (MN) is also known as membranous glomerulonephritis. Membranous nephropathy (MN) is one of the most prevalent conditions responsible for nephrotic syndrome in adults. The incidence of nephrotic syndrome in the elderly population in China is 23.4% [1]. MN is a glomerular disease with an increasing incidence. Its clinical manifestations include proteinuria, edema, and hypoalbuminemia. Hypertension and high cholesterol also frequently occur. However, these clinical manifestations are nonspecific, and membranous nephropathy must be diagnosed by renal biopsy. It is pathologically characterized by diffuse glomerular basement membrane (GBM) thickening and subepithelial immune complex deposition.

Renal pathology plays a crucial role in the clinical diagnosis, management, and prognostic evaluation of MN. Combining at least two or more of the three types of microscopy—light microscopy (LM), transmission electron microscopy (TEM), and fluorescence microscopy—is required for accurate renal pathology diagnosis. The glomeruli observed by three microscopes are shown in Figure 1. The light source system of the optical microscope is a common light source, which can magnify the tissue section being observed by 400 times. Tissue morphology at the cellular level can usually be observed under a light microscope. The relative light mirror of the transmission electron microscope has a

magnification that can reach 8000 times. It plays an important role in the early diagnosis of membranous nephropathy. Compared with LM and fluorescence microscopy, TEM is expensive and demanding in the environment. Therefore, LM and fluorescence microscopy are more common, and the images observed by them are easier to obtain. Despite these advanced techniques, observation of pathological images is performed by experienced senior pathologists. They need to manually observe glomeruli one by one, a process that is subjective and laborious. Automated and accurate methods to provide physicians with diagnostic aids are needed and promising.

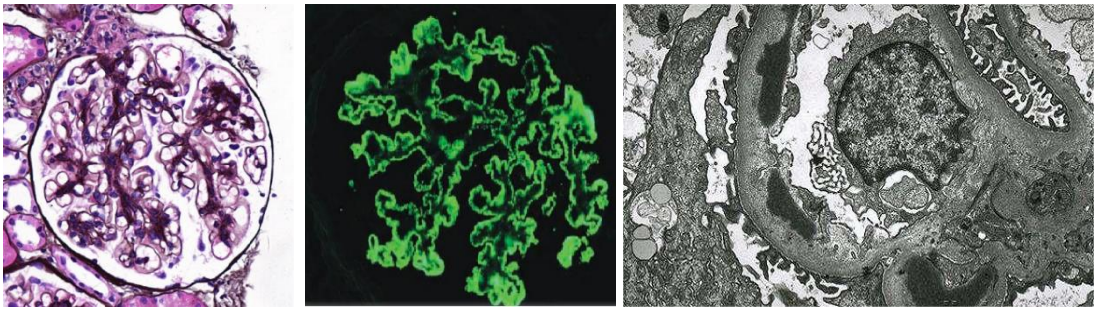


Figure 1. The glomeruli were observed by the three main techniques of Light microscopy (PASM × 400), Fluorescence microscopy (×400), and Electron microscopy (×5000) used in the pathological examination of the renal biopsy.

Because kidney pathological images are more complex and the tissues or cells to be evaluated on medical images are so similar, it is difficult for non-specialists to differentiate between them. Therefore, all of the medical image analysis and processing methods proposed in the previous ten years are non-generalized or model-based, meaning that when medical data change, researchers must devote a great deal of time and effort to designing a new method. These problems can be easily solved due to the advantages of deep learning, such as robustness and simplicity. In 2013, Ciresan first used deep learning for the detection of mitosis in pathological slices, as described in the document [2], which proved that deep learning is helpful for the research and analysis of pathological images. In 2016, researchers analyzed lung cancer samples through artificial intelligence technology [3], which can provide indicators for patient prognosis analysis. Since then, artificial intelligence techniques have been extensively explored and implemented in the field of intelligent pathological diagnosis [4]. Thus far, AI-assisted diagnosis has done a good job of finding cancer in lung, liver, breast, and pancreas pathological slice images.

Deep convolutional neural networks (CNNs) are commonly utilized for the supplementary diagnosis of WSIs as a result of their successful application in the kidney pathology area. The tissue structure of kidney disease images is complex, and the identification of lesion features is easily affected by factors such as staining and background tissue, which make the identification results inaccurate. Therefore, studying the detection and recognition methods of pathological images and assisting the relevant hospital staff in performing qualitative and quantitative analysis have become current and future research hotspots. Gurcan et al. [5] trained a CNN on kidney pathological WSIs to identify glomeruli with an accuracy of 87.8 percent. Pedraza et al. [6] implemented the classification of glomerular and non-glomerular areas in kidney disease images using deep neural networks in 2017. Barros et al. [7] created a pathological computing system to detect lesions in digital kidney pathological images to detect proliferative glomerular lesions with an 88.3% accuracy rate. Gallego et al. [8] implemented glomerular classification and WSI detection using overlapping sliding windows based on the AlexNet framework. They confirmed the need for color normalization to identify glomeruli with varied degrees of staining. Marsh et al. [9] developed a CNN model to recognize and categorize sclerotic and non-sclerotic glomeruli in WSIs of frozen section kidney cases. By expanding a pre-trained network model, a diagnostic accuracy similar to that of experienced renal pathologists was achieved with

only 48 abnormal sections. Kannan et al. [10] developed a deep learning framework for glomerular identification and segmentation in three kinds of stained images of kidney pathology specimens from 171 chronic kidney disease patients with an accuracy of up to 82.02%. Ginley et al. [11] built a CNN model to detect glomerular boundaries and nuclei from WSIs and defined a set of features to quantify and classify diabetic nephropathy. Uchino et al. [12] used 283 pathological images of kidney biopsies and a fine-tuned InceptionV3 convolutional neural network to classify 7 types of glomerular lesions, and the accuracy rate of glomerular global sclerosis classification was as high as 98.6%. Chagas et al. [13] proposed an architecture that utilizes CNN to extract features and support vector machines for classification, achieving near-perfect average results on a binary dataset of proliferating and normal glomeruli. The algorithm proposed by Salvi et al. [14] was used to assess glomerulosclerosis by employing a fully automated approach to detect key tissues, such as glomeruli.

The primary method for applying deep learning to immunofluorescence images is to stack networks hierarchically. These works are able to recognize texture features. Only one literary work achieves this objective. Through the use of a convolutional neural network, Ligabue et al. [15] were able to identify the distribution and intensity of the glomerular deposit.

Here, the most pressing challenges in renal pathology's image-based MN classification are highlighted. First, when pathologists make a clinical diagnosis, they often need to confirm the results of a variety of examination methods, including immunofluorescence, electron microscopy, and light microscopy. Moreover, only observing one of the microscope images from a light microscope or fluorescence microscope may not necessarily show the typical features of membranous nephropathy. Using only one modality, such as light microscopy, different glomeruli may still show features of different stages of the lesion, which has a confounding effect on the classification of MN. It is also a concern to improve the classification effect of membranous nephropathy under the premise of reducing the workload of labeling. Third, a whole-slide image (WSI) is a high-resolution image, and a kidney pathological WSI is generally gigabyte-level, which is a difficult problem for the training and testing of deep learning networks. To make sure the data are accurate and available, the lesion labeling work can only be performed by an experienced pathologist. This makes the labeling job more difficult, so it is difficult to obtain a lot of professionally labeled pathological images.

To address these issues, it is proposed to classify MN using a network framework that combines periodic acid-silver methenamine (PASM)-stained light-microscope images and IgG antibody-positive immunofluorescence images. This network framework is mainly divided into three stages. The first stage is the segmentation network, which separates the glomeruli from the original image. The second stage separately classifies glomeruli on images from the two modalities and obtains confidence coefficients. The third stage fuses the two confidence coefficients. This classification network framework segments glomeruli first, then fuses the classification results of two types of images: optical microscope images and immunofluorescence images, which can effectively improve the accuracy of MN classification.

The following is the outline for this paper: classification networks and multi-modal fusion methods are discussed in Section 2. The proposed classification algorithm is introduced in three stages in Section 3. Section 4 contains a description of the datasets and experiments used in this paper. Finally, Section 5 provides some conclusions.

2. Related Work

2.1. Classification Networks

As an important method of deep learning, deep convolutional neural networks are widely used in computer vision tasks, especially in image classification tasks, which use convolution to automatically extract features and determine image categories. AlexNet, as the earliest deep convolutional neural network, has attracted a lot of attention since it won

the ImageNet competition classification task champion in 2012. Since then, classification algorithms based on deep convolutional neural networks have emerged one after another. At present, the most classic classification algorithms are VGGNet, GoogLeNet, Inception, ResNet [16], etc. CNN is quickly becoming a common technique for image classification screening, and its medical applications are vast. Using CNN to automatically learn discriminative features to categorize mammography lesions, Arevalo et al. [17] established a paradigm for feature learning for breast cancer diagnosis using CNN to automatically learn discriminative features. CNN is also commonly used to classify objects or lesions. Kawahara et al. classified skin lesions using a CNN with multiple procedures for processing images at varying resolutions. Jiao et al. used CNN to extract deep features at many levels, which improved the classification accuracy of breast cancer [18]. CNN is also used in bone age assessment and sex determination [19], tuberculosis detection [20], medical image alignment [21], medical image retrieval [22], and diabetic retinopathy detection [23]. However, these classification CNN algorithms also have certain limitations. They pay more attention to the overall information of the image, and the image details are eliminated in the process of convolution. The structure of renal pathological images is more complex, and the observation process pays more attention to the details of some specific tissues. The similarity of tissues makes the classification results of the CNN network easily disturbed. Therefore, it does not show high accuracy on a single type of renal pathology image.

2.2. Multi-Modal Fusion Methods

Multi-modal fusion is also called multi-source information fusion, and multi-sensor fusion. Multi-modal fusion refers to the process of combining the information of two or more modalities for prediction. During prediction, the effective information contained in a single modality is limited, so the results cannot be accurately predicted. The multi-modal fusion process excavates potential knowledge and realizes information supplementation by synthesizing the information of multiple modalities. Multi-modal fusion broadens the coverage of the information contained in the input data, improves the accuracy of the prediction results, and obtains a robust prediction model.

Multi-modal fusion is used in applications such as audiovisual speech recognition (AVSR) [24], multi-modal emotion recognition [25,26], and medical-aided diagnosis [27].

Generally, fusion methods can be put into four groups based on the stage of fusion: signal-level fusion, data-level fusion, feature-level fusion [28], and decision-level fusion [29].

Pixel-level fusion is the most basic form of multi-modal image fusion. The image obtained after pixel-level image fusion has more detailed information, such as edges and textures. However, pixel-level fusion has certain limitations. Pixel-level fusion means higher computation and more computation time, and the resolution of the fused images should not be too different. Feature-level image fusion needs to first extract the feature information in the image and then process, analyze, and integrate this feature information to finally obtain the fused image features. The step of extracting key features results in improved computational speed and accuracy. Decision-level image fusion is to fuse the decision results of each modal image according to the actual problem. This method requires the least amount of computation, but its accuracy is highly dependent on the credibility of the decision-making process. At present, the area undergoing the most research and application is pixel-level image fusion. In addition, there is no general fusion algorithm, and a reasonable fusion model needs to be established according to the actual application scenario.

3. Methods

This section demonstrates how to use PASM-stained whole-slide images and IgG immunofluorescence images to determine whether or not a patient has MN. Our suggested model's whole design consists of three components. The process of our investigation is depicted in Figure 2. The framework consists of three modules: the module for the region-of-interest extraction, the module for confidence coefficient creation, and the module for

cross-domain aggregation. First, the glomeruli are identified from the immunofluorescence pictures, or WSIs, by using the author's previously trained networks, which exactly match the region-of-interest extraction module. The confidence-coefficient-generating module then combines many attention mechanisms to classify MN and derive the confidence coefficient for two multi-modal images. After the training phase of classification is complete, the fusion module combines the results of classification to produce a single label.

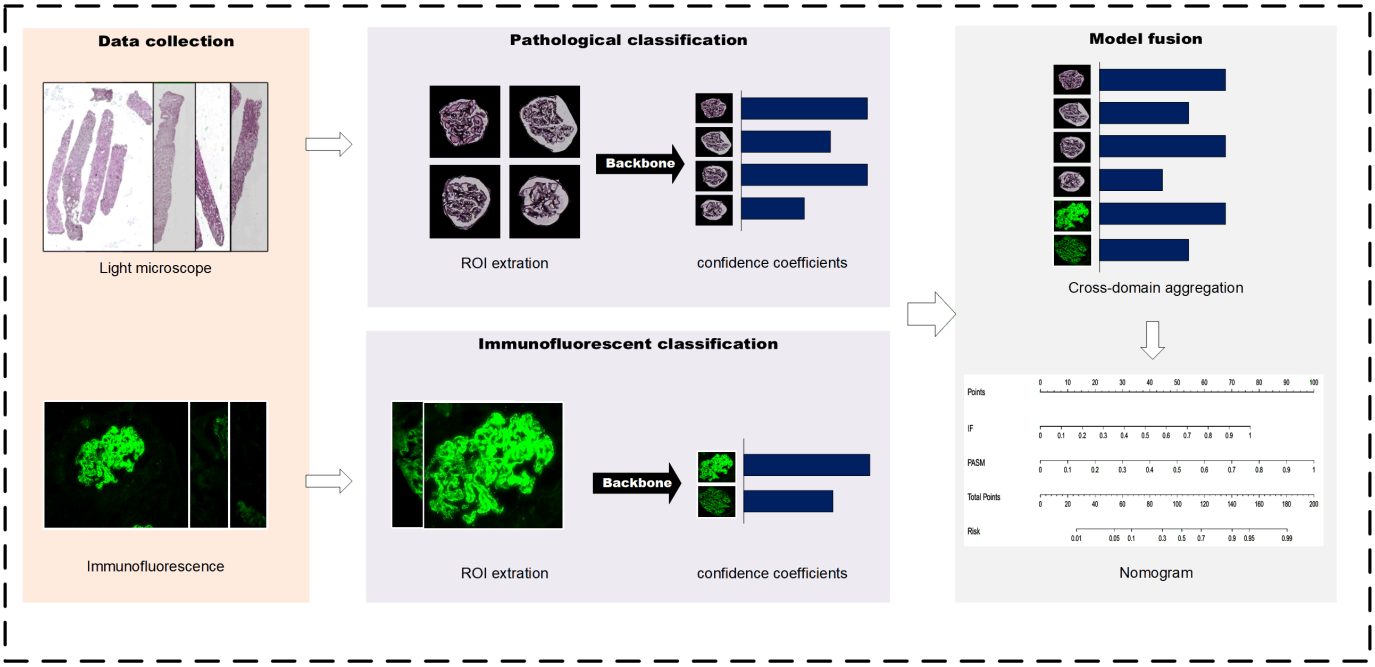


Figure 2. The workflow of this study.

3.1. Region-of-Interest Extraction Module

According to the pathogenesis of membranous nephropathy (MN), MN is formed due to the deposition of immune compounds on the GBM of the glomerulus. Therefore, the diagnosis process for MN mainly focuses on the glomerulus. There are other tissues and noises, such as proximal and distal tubules, around the glomerulus that can alter the classification accuracy of the CNN, regardless of whether you are seeing WSIs or immunofluorescence images. It is necessary to preprocess the WSIs or immunofluorescence images, crop out the glomeruli, and suppress noise interference to accurately classify the glomeruli.

Thus, the first step is to segment the glomeruli in WSIs and immunofluorescence images. This is performed with an automatic segmentation architecture based on deep learning algorithms. Segmentation results for glomeruli are shown in Figure 3.

The glomerular object locator (GOL) proposed in our previous work was used to efficiently localize and segment glomeruli in WSI or immunofluorescence images. Specifically, for the high-resolution and fine-grained characteristics of WSIs of pathological slides, our recently proposed GCPANet32 structure can segment each glomerulus from a kidney pathological WSI.

Glomeruli on immunofluorescence images are a salient object. UNet++ is utilized to pre-segment glomeruli on immunofluorescence images due to its superior ability to segment salient objects.

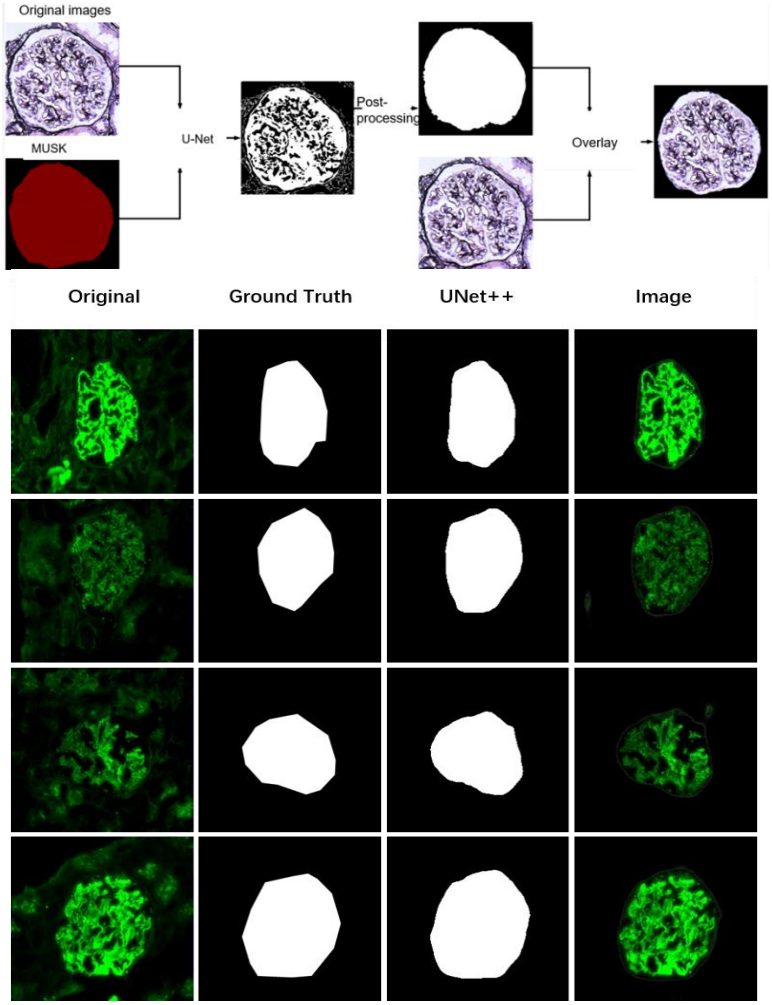


Figure 3. Glomerular segmentation on PASM staining ($\times 400$) (**above**) WSIs and immunofluorescence images ($\times 400$) (**below**).

3.2. Confidence Coefficient Generation Module

In the clinical diagnosis of MN, the three inspection techniques of electron microscopy, immunofluorescence microscopy, and light microscopy are required to confirm each other, and generally, at least two inspection techniques are required. In our study, annotations for each image are derived from patients' electronic pathology records, which are given by senior pathologists. Due to the diffuse character of MN, almost all glomeruli in the case of MN show immune complex deposition. Thus, all of the segmented images of glomeruli from the same recording are marked as "positive", while those without MN are marked as "negative".

Then, using PASM-stained WSIs and immunofluorescence images, a deep convolutional-neural-network-based classification model is made that could classify whether each glomerulus image is positive or negative on PASM-stained or immunofluorescence images. In the traditional CNN binary classification model, each image outputs a confidence coefficient in the last layer and then passes a threshold (usually 0.5) to determine whether it is in a positive or negative category. However, different thresholds calculate different recall and precision. A fixed threshold cannot fit every model. As a result, the image category, which is the output of this phase, is replaced with the positive confidence coefficients of C_{P_i} and C_{F_i} of the PASM-stained WSIs and immunofluorescence images to eliminate the influence of the confidence coefficient on the final classification result. The generated confidence coefficients are then fed to the multi-modal module to synthesize the final diagnosis for each patient.

In this way, a positive degree can better reflect the probability of MN than having a fixed threshold for a single class of each image.

3.3. Multi-Modal Fusion Module

Given a case, all segmented glomerular images on PASM-stained WSIs and immunofluorescence images, as well as MN confidence coefficients for glomerular images under the classification branch, could be obtained. Because WSIs are high-resolution images, detailed information about tissue morphology can be observed, while immunofluorescence images can only roughly observe the stained tissue morphology and fluorescence intensity, so WSIs and immunofluorescence images belong to two different domains. In this paper, in the diagnosis of MN, it is necessary to combine the two image domains to give a comprehensive judgment. Therefore, the multi-modal module is used to take MN probabilities on PASM-stained WSIs and immunofluorescence images into account in a complete way. Given that each patient has multiple glomeruli images in both kinds of images, the overall MN confidences C_P and C_F for each patient's PASM-stained WSIs and immunofluorescence images can be calculated as follows:

$$C_P = \frac{1}{n} \sum_{i=1}^{n} C_{P_i} \tag{1}$$

$$C_F = \frac{1}{m} \sum_{i=1}^{m} C_{F_i} \tag{2}$$

where n and m represent the number of glomeruli in each patient's PASM-stained WSIs and immunofluorescence images, respectively. Then, a classification model is built, such as a LASSO logistic regression model, to predict MN using a ten-fold cross-validation strategy. The probability of MN in the PASM-stained images and immunofluorescence images is aggregated into one predicted class per patient.

3.4. Nomogram Construction of Membranous Nephropathy

Above, the confidence coefficients C_P and C_F of glomeruli on PASM-stained WSIs and immunofluorescence images are added together to predict MN classification. As the classification of MN is based on two factors, a nomogram is needed to give each value of the confidence coefficient for the two types of glomeruli images. This is because each factor affects the outcome variable (the size of the regression coefficient) in a different way. The final score is calculated by adding the points from each individual's evaluation. Through a functional transformation between the total score and the probability of MN occurrence, a predicted value for each image may be determined.

4. Experiments Results

This section focuses mostly on the results calculated by the proposed membranous nephropathy classification method on the kidney pathology images dataset. The recommended algorithms consist primarily of two sections, segmentation and classification networks, which are trained and tested in the same environment. As a reminder, the environment is set-up as follows: The central processing unit and graphics processing unit are comprised of an Intel(R) Xeon(R) Gold 5118 CPU at 2.30 GHz and an NVIDIA Tesla V100 PCIe GPU with 32 GB of RAM, respectively. CUDA version 10.2 is installed. The code for the entire implementation is penned in Python 3.6 using the Pytorch 1.6.0 platform and is run on a 64-bit installation of Ubuntu 18.2 LTS.

4.1. Datasets Preparation

A total of 1206 patients with MN and 733 patients without MN who had a renal biopsy at the Second Hospital of Shanxi Medical University were screened between 2014 and 2019. Two types of pathological image data were collected from the above cases. In total, 1206 PASM-stained WSIs and 1262 immunofluorescence images were obtained. On the

one hand, renal biopsy tissues with periodic acid silver methenamine (PASM) stains were scanned to WSIs using a KF-PRO-005-EX digital slide scanner (KFbio, Ningbo, China) with a 40× objective and a 10× eyepiece at a resolution of 0.25 µm/pixel. On the other hand, fluorescence images with IgG immunoglobulins from the above patients were obtained by fluorescence microscopy with a size of 1360 × 1024. The kidney biopsy samples were all utilized for regular clinical diagnosis, and the diagnoses of all cases used in this paper were given by senior pathologists. The patient's personal information was removed throughout the collaboration, because it would be detrimental to the patient's privacy.

To ensure the reliability of the model, the dataset was screened. For the positive samples, the PASM-stained WSIs images with the highest-quality scanning were kept. In order to ensure that the negative samples did not interfere with the classification of MN, images of cases that were diagnosed as focal segmental glomerulosclerosis (FSGS) and minimal change disease (MCD) were selected as the normal control group. For immunofluorescence images, IgG immunofluorescence images, which are of great guideline significance for the diagnosis of MN, were selected.

Because the pathological features of membranous nephropathy are mainly manifested in the glomeruli, the glomerulus is our ROI (region of interest). The morphologically distinct glomeruli are then segmented from PASM-stained WSIs and IgG-positive immunofluorescence images of selected cases. The glomerular segmentation part is implemented by the network proposed in our previous work [30,31]. In the training phase of glomeruli segmentation, the Adam optimization algorithm is used to iterate the hyperparameter on both types of images. The batch size is set to 36, and the basic learning rate is set to 10^{-3}. After 200 epoch iterations of training, testing is performed, and the weight of the glomerular segmentation network is set to the parameter with the highest Dice coefficient in testing. Finally, 8297 images of glomeruli with PASM staining from 611 patients and 1896 IgG immunofluorescence images from 1665 patients are used for confidence coefficient extraction network training, validation, and testing, as shown in Table 1. These glomeruli are all selected from sections with high staining quality and uniform thickness, and the tissues in these glomeruli are clearly visible. The confidence coefficient of the glomerulus obtained by the classification network is used to train the subsequent model fusion module. If the dataset is only divided into a training set and a test set, it is equivalent to all the data participating in the tuning of network hyperparameters, which reduces the credibility of the training data used for the model fusion module. To prevent overfitting and improve the credibility of the model, a validation set is added. During classification model training, different data augmentation techniques, such as cropping (both randomly and centered), flipping (both vertically and horizontally), and rotating, are used to make the model more reliable. It is important to note that while data augmentation techniques are used, there is no data repetition in either the training or testing datasets. In particular, glomeruli from the cases that have both the PASM-stained glomeruli and the immunofluorescence image should be considered. The glomeruli used to train and test the cross-domain aggregation model come from the test dataset of the confidence coefficient extraction network. This is performed to avoid the effects of overfitting.

Table 1. The detailed information of image-level and patient-level datasets.

		PASM			**Immunofluorescence**			**Fusion**
		train	val	test	train	val	test	10-fold cross-validation
Patient-level	positive	194	47	66	885	209	100	66
	negatives	120	81	103	378	51	42	42
Image-level	positive	3603	914	1353	992	248	81	1353
	negatives	1534	171	722	463	42	70	70

4.2. Performance Metrics

To quantitatively evaluate the performance of the MN cross-domain classification model proposed in this paper, the following evaluation metrics are chosen: precision, recall

(also known as sensitivity), and F_1-score. These evaluation metrics play an important role in the optimization of the trained model, as they can reveal potential parts of the model that cannot be predicted correctly. The formulas of these evaluation indicators are expressed as follows:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F_1 - Score = \frac{2TP}{2TP+FP+FN} \tag{5}$$

where TP is True Positive, or the number of positive samples predicted to be positive; FP is false positive, or the number of positive samples that are predicted to be negative; FN is false positive, i.e., the number of negative samples predicted to be positive.

4.3. Performance Comparison of Different Classification Models under Different Thresholds

In order to validate the performance of classification models, Resnet [6] and Densenet [32] architectures with different layers are chosen. The classification loss function uses the cross-entropy function as follows:

$$Loss = \frac{1}{N}\sum_i -(y_i \log(P_i) + (1-y_i)\log(1-P_i)) \tag{6}$$

where y_i is the sample's label: 1 indicates a positive classification and 0 indicates a negative one; P_i is the probability that the sample i belongs to the positive classification.

After the sample data go through the classification network, the confidence level of whether the sample belongs to the category will be determined. This is a continuous number from 0 to 1 that needs to be mapped to the binary value space using a threshold.

To demonstrate the impact of the threshold, evaluation indicators such as recall, precision, and F_1-score are listed, with thresholds of 0.3, 0.5, and 0.7, respectively, in each architecture. The results are shown in Table 2. It could be observed that the best performance of a model is different under different thresholds. Appropriate thresholds need to be found to output classification results for different application scenarios.

Table 2. Performance comparison of different images.

Backbone	PASM Stained Images				Immunofluorescence Images			
	Threshold	**Recall**	**Precision**	**F_1-Score**	**Threshold**	**Recall**	**Precision**	**F_1-Score**
Resnet34	0.3	0.8752	0.8900	0.8825	0.3	0.6475	0.7804	0.7077
	0.5	0.8993	0.9070	0.9031	**0.5**	**0.7770**	**0.7951**	**0.7859**
	0.7	**0.9046**	**0.9081**	**0.9063**	0.7	0.7670	0.7898	0.7833
Resnet50	0.3	0.9200	0.9250	0.9225	0.3	0.8345	0.8711	0.8524
	0.5	0.9234	0.9265	0.9250	**0.5**	**0.9281**	**0.9360**	**0.9320**
	0.7	**0.9267**	**0.9285**	**0.9276**	0.7	0.8849	0.8956	0.8902
Resnet101	0.3	0.9128	0.9173	0.9150	0.3	0.6115	0.7669	0.6804
	0.5	0.9176	0.9208	0.9192	0.5	0.6835	0.7949	0.7350
	0.7	**0.9200**	**0.9220**	**0.9210**	**0.7**	**0.8993**	**0.8991**	**0.8992**
Densenet121	0.3	0.8564	0.8823	0.8692	0.3	0.7410	0.7926	0.7659
	0.5	0.8872	0.9034	0.8953	**0.5**	**0.8201**	**0.8204**	**0.8203**
	0.7	**0.9075**	**0.9142**	**0.9108**	0.7	0.7914	0.8174	0.8042
Densenet161	**0.3**	**0.9234**	**0.9278**	**0.9256**	0.3	0.8633	0.8774	0.8703
	0.5	0.9210	0.9214	0.9212	0.5	0.8705	0.8702	0.8703
	0.7	0.9051	0.9050	0.9050	**0.7**	**0.8705**	**0.8852**	**0.8778**
Densenet169	0.3	0.9012	0.9121	0.9066	0.3	0.8273	0.8511	0.8391
	0.5	**0.9128**	**0.9164**	**0.9146**	**0.5**	**0.8561**	**0.8561**	**0.8561**
	0.7	0.9027	0.9021	0.9024	0.7	0.8058	0.8441	0.8245

Figure 4 shows the ROC curves and AUC indicators for six models in PASM staining images and immunofluorescence images from the test set shown in Table 1. It is

observed that Resnet101 achieves the best performance in classification models in both image domains.

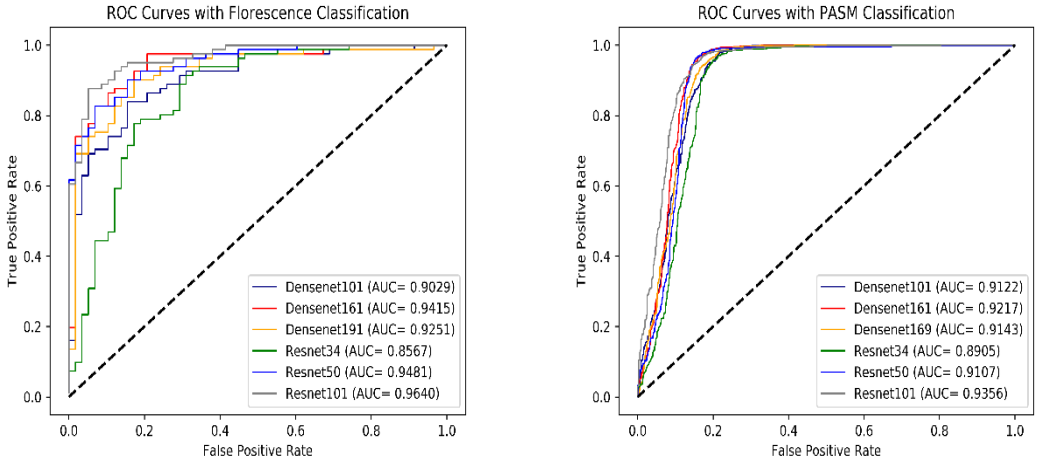


Figure 4. ROC curves of different models in PASM-stained WSIs (**left column**) and Immunofluorescence image (**right column**).

4.4. Performance Comparison of Different Fusion Models

In the previous step, it is possible to obtain the glomerular classification model on PASM-stained WSIs and immunofluorescence images, but these two classification models can only obtain the confidence coefficient of whether a single glomerulus on a certain image belongs to MN or not. This paper introduces a multi-modal module that not only integrates the classification results of the two image domains but also the classification results of all glomerular images of the same patient. That is to say, after the fusion module, the glomerular-level classification information is converted into patient-level classification information.

Here, the aggregation module adopts five classification models based on machine learning: logistic regression, random forest, gradient boosting classifier, Xgboost [33], and lightGBM [34].

Finally, 66 patients with MN and 42 patients without MN are used to evaluate the performance of the model that aggregates the two image domains. For these cases, immunofluorescence images as well as PASM-stained WSIs images are obtained.

Then, 10-fold cross-validation is performed to find the best model and make sure it does not fit too well. The average results are shown in Table 3. The best performance in diagnosing MN is achieved (the F_1-score is 0.9732) when the backbone is Densenet 161 and the aggregation model is logistic regression. The fusion model is compared with the best classification results obtained when only PASM-stained WSIs or immunofluorescence images are used on the same backbone. The classification results improve to a certain extent, and the improvement is shown in the sixth column of Table 3. It can also be seen in the table that, in the case of the same backbone, compared with the highest F_1-score on a single modality without the model fusion method, the model fusion method does show a certain improvement in the MN classification performance. The aggregation models improve F_1-scores by an average of 4.5 percent (the maximum is 5 percent).

The results of classification using images of different modalities are compared in Table 4. When only using the CNN network for MN classification of PASM-stained WSIs or immunofluorescence images, the best performing network is Resnet50, and the best F_1-scores are 0.9276 and 0.9320, respectively. When the method proposed in this paper is used for MN classification with images of two modalities, the F_1-score reaches 0.9732. This proves that the method proposed in this paper is effective.

Table 3. Performance comparison of different fusion models.

Backbone	Fusion Model	Precision	Recall	F_1-Score	Improvement
Resnet34	Logistic regression	0.9471	0.9390	0.9430	0.0498
	Random forest	0.9542	0.9467	0.9505	
	Gradient boosting classifier	0.9480	0.9400	0.9440	
	Xgboost	0.9582	0.9515	0.9549	
	lightGBM	**0.9597**	**0.9526**	**0.9561**	
Resnet50	Logistic regression	0.9670	0.9621	0.9646	0.0404
	Random forest	**0.9742**	**0.9708**	**0.9724**	
	Gradient boosting classifier	0.9532	0.9460	0.9496	
	Xgboost	0.9664	0.9602	0.9633	
	lightGBM	0.9640	0.9575	0.9607	
Resnet101	Logistic regression	0.9595	**0.9519**	0.9557	0.0353
	Random forest	0.9590	0.9479	0.9534	
	Gradient boosting classifier	0.9560	0.9482	0.9521	
	Xgboost	0.9537	0.9420	0.9478	
	lightGBM	**0.9608**	0.9518	**0.9563**	
Densenet121	Logistic regression	0.9560	0.9458	0.9509	0.0472
	Random forest	0.9593	0.9524	0.9559	
	Gradient boosting classifier	0.9390	0.9245	0.9317	
	Xgboost	0.9611	0.9517	0.9564	
	lightGBM	**0.9619**	**0.9542**	**0.9580**	
Densenet161	Logistic regression	**0.9748**	**0.9716**	**0.9732**	0.0476
	Random forest	0.9531	0.9469	0.9732	
	Gradient boosting classifier	0.9359	0.9260	0.9301	
	Xgboost	0.9504	0.9414	0.9458	
	lightGBM	0.9571	0.9545	0.9596	
Densenet169	Logistic regression	**0.9670**	**0.9621**	**0.9646**	0.0500
	Random forest	0.9603	0.9525	0.9563	
	Gradient boosting classifier	0.9408	0.9328	0.9368	
	Xgboost	0.9542	0.9444	0.9492	
	lightGBM	0.9658	0.9603	0.9630	

Table 4. Performance comparison of different image models.

	Image Modal	Precision	Recall	F_1-Score
Resnet 50	PASM stained WSIs	0.9267	0.9285	0.9276
Resnet 50	Immunofluorescence images	0.9281	0.9360	0.9320
Densenet161 + Logistic regression	PASM stained WSIs + Immunofluorescence images	0.9748	0.9716	0.9732

Furthermore, Resnet101's AUC indicators in PASM staining images and immunofluorescence images are both highest, whereas performance is not achieved after aggregation models. This is probably because, even if the AUC indicator weakens the impact of threshold selection, the confidence coefficient, compared with the classification result, can better reflect the cognitive level of the samples.

4.5. Construction of the Aggregation Nomogram

The above experiments demonstrate the effectiveness of the aggregation model constructed based on the confidence coefficients obtained from PASM images and immunofluorescence image classification models for classifying MN. MN's classification performance improves after the aggregation model. According to the state-of-the-art models in this study (logistic regression with Densenet 161), an aggregation nomogram incorporating these two coefficients is built in Figure 5. By connecting its corresponding value to the Points line in an upward-sloping line, each confidence coefficient is assigned a point value. Multiple sclerosis risk is proportional to the number of points on the Total Points line.

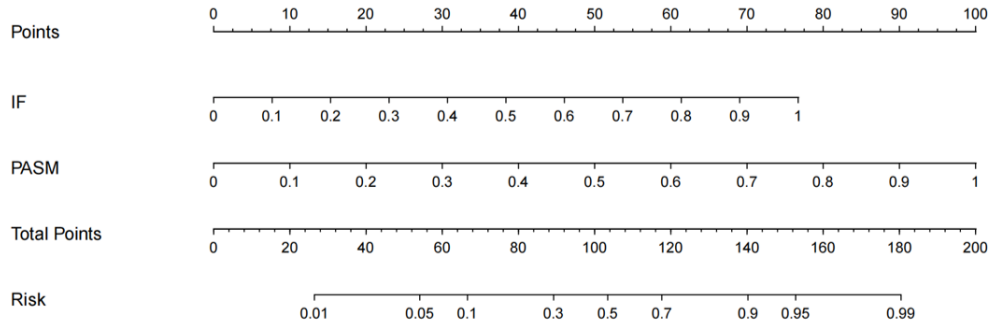


Figure 5. MN risk nomogram. IF and PASM denote two confidence coefficients.

Through the nomogram, the scores of the confidence coefficients of the two images can be added to obtain the total score, and finally, the predicted value of whether the individual has MN is calculated through the functional conversion relationship between the total score and the probability of whether it is MN.

5. Conclusions

This paper proposes a classification algorithm for membranous nephropathy using a deep learning algorithm and a multi-modal fusion algorithm. PASM-stained WSIs and immunofluorescent images are combined for the classification of MN. The framework used in this paper is mainly divided into three modules: region-of-interest extraction module, confidence coefficient generation module, and multi-modal fusion. Among them, the region-of-interest extraction module identifies and segments the glomeruli on the two types of images to reduce the influence of other surrounding tissues on the subsequent confidence coefficient generation module. The confidence coefficient generation module obtains the confidence coefficients of glomeruli on the two modalities of images by training an MN classifier. Multi-modal fusion fuses the confidence coefficients of glomeruli across the two modalities through machine learning algorithms to obtain the final classification result. The results show that via logistic regression with Densenet 161, the F_1-score of MN classification using images of two modalities achieves 0.9732 and is about 5 percentage points higher than that of images using only one modality, which proves that the research idea is feasible.

Under experimental conditions, this deep learning method can provide adjuvant assessment comparable to that of advanced pathologists. Compliance bias can occur because there are not many images of kidney pathology, light microscopy and immunofluorescence microscopy only show a small number of features, images are not spread out evenly in some categories, among others.

Although deep convolutional neural networks have shown promising results in automatically identifying kidney tissue, lesions, and related markers, the accuracy of these networks mainly depends on the size, complexity, algorithmic structure, and noise of the dataset used. Learning algorithms for detecting digital image heterogeneity such as lesions require further study.

Author Contributions: Conceptualization, M.L.; methodology, F.H.; software, X.L.; data curation, W.H.; writing—original draft preparation, F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China: Grant No. 61901292, Grant No. 11771321 and Grant No. 6210137.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Second Hospital of Shanxi Medical University ((2021)YX(026) and 15 March 2021).

Informed Consent Statement: Patient consent was waived in the approval document of the Ethics Committee.

Data Availability Statement: Not applicable.

Acknowledgments: We are very grateful to the pathologists of the Second Hospital of Shanxi Medical University for providing sufficient data, labeling, explaining the data, and providing guidance for our research based on their rich experience. Grants from the National Natural Science Foundation of China (Grant No. 61901292, Grant No. 11771321, and Grant No. 62101376) supported this study.

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

References

1. Xu, X.; Wang, G.; Chen, N.; Lu, T.; Nie, S.; Xu, G.; Zhang, P.; Luo, Y.; Wang, Y.; Wang, X.; et al. Long-Term Exposure to Air Pollution and Increased Risk of Membranous Nephropathy in China. *J. Am. Soc. Nephrol.* **2016**, *27*, 3739–3746. [CrossRef] [PubMed]
2. Ciresan, D.; Giusti, A.; Gambardella, L.; Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 411–418.
3. Yu, K.-H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Ré, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. [CrossRef] [PubMed]
4. Liu, Z.; Su, W.; Ao, J.; Wang, M.; Jiang, Q.; He, J.; Gao, H.; Lei, S.; Nie, J.; Yan, X.; et al. Instant diagnosis of gastroscopic biopsy via deep-learned single-shot femtosecond stimulated Raman histology. *Nat. Commun.* **2022**, *13*, 4050. [CrossRef] [PubMed]
5. Sarder, P.; Ginley, B.; Tomaszewski, J.E. Automated renal histopathology: Digital extraction and quantification of renal pathology. *Int. Soc. Opt. Photonics* **2016**, *9791*, 1–12. [CrossRef]
6. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
7. Barros, G.; Navarro, B.; Duarte, A.; Dos-Santos, W.L. Pathospotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys. *Sci. Rep.* **2017**, *7*, 46769. [CrossRef]
8. Gallego, J.; Pedraza, A.; Lopez, S.; Steiner, G.; Gonzalez, L.; Laurinavicius, A.; Bueno, G. Glomerulus classification and de-tection based on convolutional neural networks. *J. Imaging* **2018**, *4*, 20. [CrossRef]
9. Marsh, J.N.; Matlock, M.K.; Kudose, S.; Liu, T.-C.; Stappenbeck, T.S.; Gaut, J.P.; Swamidass, S.J. Deep Learning Global Glomerulosclerosis in Transplant Kidney Frozen Sections. *IEEE Trans. Med. Imaging* **2018**, *37*, 2718–2728. [CrossRef]
10. Kannan, S.; Morgan, L.A.; Liang, B.; Cheung, M.G.; Lin, C.Q.; Mun, D.; Nader, R.G.; Belghasem, M.E.; Henderson, J.M.; Francis, J.M.; et al. Segmentation of Glomeruli Within Trichrome Images Using Deep Learning. *Kidney Int. Rep.* **2019**, *4*, 955–962. [CrossRef]
11. Ginley, B.; Lutnick, B.; Jen, K.-Y.; Fogo, A.B.; Jain, S.; Rosenberg, A.; Walavalkar, V.; Wilding, G.; Tomaszewski, J.E.; Yacoub, R.; et al. Computational Segmentation and Classification of Diabetic Glomerulosclerosis. *J. Am. Soc. Nephrol.* **2019**, *30*, 1953–1967. [CrossRef]
12. Uchino, E.; Suzuki, K.; Sato, N.; Kojima, R.; Tamada, Y.; Hiragi, S.; Yokoi, H.; Yugami, N.; Minamiguchi, S.; Haga, H.; et al. Classification of glomerular pathological findings using deep learning and nephrologist–AI collective intelligence approach. *Int. J. Med. Inform.* **2020**, *141*, 104231. [CrossRef]
13. Chagas, P.; Souza, L.; Araújo, I.; Aldeman, N.; Duarte, A.; Angelo, M.; Dos-Santos, W.L.; Oliveira, L. Classification of glomerular hypercellularity using convolutional features and support vector machine. *Artif. Intell. Med.* **2020**, *103*, 101808. [CrossRef]
14. Salvi, M.; Mogetta, A.; Gambella, A.; Molinaro, L.; Barreca, A.; Papotti, M.; Molinari, F. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. *Comput. Med. Imaging Graph.* **2021**, *90*, 101930. [CrossRef]
15. Ligabue, G.; Pollastri, F.; Fontana, F.; Leonelli, M.; Furci, L.; Giovanella, S.; Alfano, G.; Cappelli, G.; Testa, F.; Bolelli, F.; et al. Evaluation of the classification accuracy of the kidney biopsy direct immunofluorescence through convolutional neural net-works. *Clin. J. Am. Soc. Nephrol.* **2020**, *15*, 1445–1454. [CrossRef]
16. Wu, Z.; Shen, C.; Van Den Hengel, A. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recog.* **2019**, *90*, 119–133. [CrossRef]
17. Arevalo, J.; González, F.A.; Ramos-Pollán, R.; Oliveira, J.L.; Lopez, M.A.G. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput. Methods Programs Biomed.* **2016**, *127*, 248–257. [CrossRef]
18. Kawahara, J.; Hamarneh, G. Multi-resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers. In Proceedings of the PRICAI 2019: Trends in Artificial Intelligence, Athens, Greece, 17 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 164–171.
19. Nguyen, Q.H.; Nguyen, B.P.; Nguyen, M.T.; Chua, M.C.; Do, T.T.; Nghiem, N. Bone age assessment and sex determination using transfer learning. *Expert Syst. Appl.* **2022**, *200*, 116926. [CrossRef]
20. Nguyen, Q.H.; Nguyen, B.P.; Dao, S.D.; Unnikrishnan, B.; Dhingra, R.; Ravichandran, S.R.; Chua, M.C. Deep learning models for tuberculosis detection from chest X-ray images. In Proceedings of the 2019 26th International Conference on Telecommunications (ICT), Hanoi, Vietnam, 8–10 April 2019; pp. 381–385.
21. Ban, Y.; Wang, Y.; Liu, S.; Yang, B.; Liu, M.; Yin, L.; Zheng, W. 2D/3D Multimode Medical Image Alignment Based on Spatial Histograms. *Appl. Sci.* **2022**, *12*, 8261. [CrossRef]

22. Zhuang, Y.; Chen, S.; Jiang, N.; Hu, H. An Effective WSSENet-Based Similarity Retrieval Method of Large Lung CT Image Databases. *KSII Trans. Internet Inf. Syst.* **2022**, *16*, 2359–2376.
23. Nguyen, Q.H.; Muthuraman, R.; Singh, L.; Sen, G.; Tran, A.C.; Nguyen, B.P.; Chua, M. Diabetic retinopathy detection using deep learning. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing, Haiphong, Vietnam, 17–19 January 2020; pp. 103–107.
24. Potamianos, G.; Neti, C.; Gravier, G.; Garg, A.; Senior, A. Recent advances in the automatic recognition of audiovisual speech. *Proc. IEEE* **2003**, *91*, 1306–1326. [CrossRef]
25. Soleymani, M.; Pantic, M.; Pun, T. Multimodal Emotion Recognition in Response to Videos. *IEEE Trans. Affect. Comput.* **2011**, *3*, 211–223. [CrossRef]
26. Myers, G.K.; Snoek, C.G.M.; Nevatia, R.; Nallapati, R.; van Hout, J.; Pancoast, S.; Sun, C.; Habibian, A.; Koelma, D.C.; van de Sande, K.E.A.; et al. Evaluating Multimedia Features and Fusion for Example-Based Event Detection. In *Fusion in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 109–133.
27. Fu, J.; Li, W.; Du, J.; Xiao, B. Multimodal medical image fusion via laplacian pyramid and convolutional neural network re-construction with local gradient energy strategy. *Comput. Biol. Med.* **2020**, *126*, 104048. [CrossRef] [PubMed]
28. Jin, K.; Yan, Y.; Chen, M.; Wang, J.; Pan, X.; Liu, X.; Liu, M.; Lou, L.; Wang, Y.; Ye, J. Multimodal deep learning with feature level fusion for identification of choroidal neovascularization activity in age-related macular degeneration. *Acta Ophthalmol.* **2021**, *100*, e512–e520. [CrossRef] [PubMed]
29. Du, J.; Li, W.; Lu, K.; Xiao, B. An overview of multi-modal medical image fusion. *Neurocomputing* **2016**, *215*, 3–20. [CrossRef]
30. Zhang, L.; Li, M.; Wu, Y.; Hao, F.; Wang, C.; Han, W.; Niu, D.; Zheng, W. Classification of renal biopsy direct immunofluo-rescence image using multiple attention convolutional neural network. *Comput. Methods Programs Biomed.* **2022**, *214*, 106532. [CrossRef]
31. Liu, X.; Li, M.; Wu, Y.; Chen, Y.; Hao, F.; Zhou, D.; Wang, C.; Ma, C.; Shi, G.; Zhou, X. An efficient glomerular object locator for renal whole slide images using proposal-free network and dynamic scale evaluation method. *AI Commun.* **2022**, *34*, 245–258. [CrossRef]
32. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawai, 21–26 July 2017; pp. 4700–4708.
33. Xgboost: Extreme Gradient Boosting. R Package Version 2015. Available online: https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf (accessed on 27 November 2022).
34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.

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.