Harnessing AI for Kidney Glomeruli Classification

Meghna P Ayyar*
IIIT-Delhi
Delhi, India
meghnaa@iiitd.ac.in

Puneet Mathur*
NSIT
Delhi, India
pmathur3k6@gmail.com

Rajiv Ratn Shah
IIIT-Delhi
Delhi, India
rajivratn@iiitd.ac.in

Shree G Sharma
Arkana Laboratories
Arkansas, USA
drshreegopal@gmail.com

***Abstract*—A key challenge in renal diagnosis using digital pathology has been the scarcity of reliable annotated datasets that can act as a benchmark for histological investigations. This paper uses a novel medical image dataset, titled Glomeruli Classification Database (GCDB), consisting of renal glomeruli images bifurcated into binary classes of normal and abnormal morphology. Based on this dataset, we direct our pioneering efforts to explore suitable deep neural network techniques related to kidney tissue slide imaging so as to establish a state of the art in this relatively unexplored domain. The paper focuses on classifying normal and abnormal categories of glomeruli which are the vital blood filtration units of the kidney. The results obtained using publicly available transfer learning models are held in comparison with supervised classifiers configured with image features extracted from the last layers of pre-trained image classifiers. Contrary to popular belief, transfer learning models such as ResNet50 and InceptionV3 are empirically proved to under-perform for this particular task whereas the Logistic Regression model augmented with features from the Inception-ResNetV2 show the most promising results on the GCDB dataset.**

***Index Terms*—Glomeruli Classification, Transfer Learning, Deep Neural Network, Renal Glomeruli**

I. Introduction

An important aspect of the histomorphometric (measuring the shape or form of tissue) analysis of renal biopsy images, is the spatial interrogation of the entire morphology of the biopsy tissues using multiple staining techniques. The case under consideration is of the classification of the types of glomeruli structures, the filtration units of a human kidney, from a renal biopsy. Any changes in the shape, cellularity, size or structure of the glomeruli act as one of the indicators of kidney diseases. According to Johnson et al. [1], a healthy human kidney has more than 300,000 glomeruli with an average mean size of 6.04 μm^3 (cubic micron) each.

Generally, the glomeruli are spherical in shape and may be distorted in case of swelling due to hypertension and diabetes. The size of glomeruli may also be affected by the manner in which the biopsy tissues are extracted and sectioned by a skilled pathologist. In routine clinical practice, a kidney biopsy after processing is stained with Hematoxylin and Eosin (HE), Periodic acid-Schiff (PAS), Jones Methenamine Silver (JMS) and Masson trichrome (MT) stain. We have the vision to automate the diagnosis on kidney biopsy and as a preliminary step in this direction, we have developed the requisite GCDB dataset complimented with machine-guided learning to accomplish the classification of renal glomeruli into two fundamental classes: normal and abnormal.

Scarfe et al. [2] grade a glomerulus as abnormal if the staining extends beyond the confines of the mesangium (support structure for glomerular capillary circulation located outside the capillary lumen) into the glomerular tufts. A normal glomerulus shows no proliferation in the Bowman space (sack like structure in the kidneys that performs blood filtration), no necrosis is seen and glomerular tuft is not adhered to the Bowman capsule. An abnormal glomerulus represents a departure from normal histology in terms of sclerosis (stiffening of glomerulus caused due to replacement of the original tissues by connective tissue), a proliferation of glomerular capillaries and endothelial cells.

Detecting the presence and percentage of abnormal glomeruli in the renal tissue slide is the first step a pathologist performs to decide whether the tissue is affected or healthy. Thus it becomes important to classify the glomeruli that act as primary bio-markers of the presence of a disease. To the best of our knowledge, this is the first attempt in trying to classify glomeruli based on classes like abnormal and normal.

Transfer Learning models, ResNet50 and InceptionV3, used for large-scale image classification have been used as a baseline. Secondary supervised classifiers used in this study are Logistic Regression (LOGREG), Random Forest (RF) and Naive Bayes (NB). Each of these classifiers takes the feature vectors of the image data as the input and outputs the predicted label after classification. The input features are extracted from the last layers of the pre-trained image classification models-InceptionResNetV2, InceptionV3, ResNet50 and VGG19 by convolving the image RGB descriptors with the weights of the last layers. These transformed architectures used as feature extractors, are referred to as IRFE (InceptionResNetV2 Feature Extractor), IFE (InceptionV3 Feature Extractor), RFE (ResNet50 Feature Extractor) and VFE (VGG19 Feature Extractor) throughout the paper.

The main contributions can be summarized as follows:

- The dataset images has been collected from annotations made to anonymous patient records by multiple in-house experts and then verified by our expert nephropathologist.The database thus curated has undergone multiple rounds of scrutiny to eliminate any incorrect annotation.
- Experimentation to ascertain the applicability of transfer learning for classification of glomeruli images.

*Authors contributed equally

- Experimentation to analyze the performance of supervised secondary classifiers that derive weighted image feature vectors from pre-trained transfer learning architectures such as ResNet50, InceptionV3, etc. as inputs.

II. Related Work

Analogous classification tasks [3] have earlier demonstrated the use of one-shot classifiers to classify normal, abnormal, and confounder classes in prostate histopathology that support our argument to account textural as well as edge features as an important contributor for minimizing intra-class homogeneity while maximizing inter-class heterogeneity. Earlier efforts also emphasize the application of converging squares algorithm and the wedge filter technique for location and segmentation [4]. Locality sensitive deep learning frameworks have also been known to automatically detect and classify individual nuclei in colon histology images [5]. Kidney biopsies form an integral part of the diagnosis of renal ailments and there have been studies in detail to observe various idiopathic nephrotic syndromes in high-risk patients [6]. Interstitial fibrosis serves as a hallmark to correlate chronic kidney diseases with the extent of scarring using appropriate histological stains [7]. Keeping in mind the need for standardization of renal biopsy interpretations, Banff schema [8] was put forward that distinctly explained glomerulitis as an accumulation of mononuclear inflammatory cells in glomerular capillaries. Recently, transfer learning has been employed for detection of glomerulus in biopsiesIn [9], the authors have used a deep neural networks for the detection and segmentation of glomeruli from kidney tissue. Earlier efforts also emphasize the application of converging squares algorithm and the wedge filter technique for location and segmentation [4].

III. Methodology

This section we describe histogram equalization and data augmentation (Sections III-A and III-B) followed by transfer learning (Section III-C) and supervised classification using pre-trained feature extraction (Section III-D) for classification.

A. Histogram Equalization

Histogram equalization has been practiced as a standard data preprocessing technique for enhancement of image contrast. Adaptive Histogram Equalizers (AHE) have been observed to be advantageous for contrast-enhancement. However, a worthwhile adversity of AHE application is the intensification of noise in the image. Hence, we have particularly used Contrast-Limited Adaptive Histogram Equalizer (CLAHE) [10] for preprocessing GCDB dataset images in order to reduce the noise magnification to a large extent.

B. Data Augmentation

Data augmentation has demonstrated a significant contribution in suitably handling the obstacles in training like data inadequacy, data imbalance and lack of uniform modalities across the dataset. Beneficial approaches that have been applied in our work include elastic deformations that generate plausible transformations of existing samples without distorting the original label information. In order to aggrandize the robustness of CNN models, we have introduced affine transformations that have been listed out in Table I.

Parameter	Value
Horizontal Flip	True
Vertical Flip	True
Fill Mode	Nearest
Zoom Range	0.1
Width Shift Range	0.2
Height Shift Range	0.2
Rotational Range	180

TABLE I: Data augmentation parameters

C. Transfer Learning

Fig.1 showcases the complete flow of glomerulus detection using transfer learning models. The GCDB dataset is initially converted into a format acceptable to the CNN models through histogram equalization and data augmentation as discussed above. The proposed architectures used for transfer learning consist of basic implementation of ResNet50 and InceptionV3 models with their pre-evaluated weights initialized beforehand respectively. The models are re-trained keeping only the last three layers as trainable but augmented with three dense fully-connected layers having sizes 1024, 256 and 2 units, respectively. The activation function applied is Rectified Linear Unit (ReLU) for the second and third last dense layers followed by *'Softmax '* in the last dense layer. Each customized layer is provided with a dropout of 0.5 units. ResNet50 and InceptionV3 are trained with an 80:20 train-test split with data points segregated in the ratio of the class frequencies so as to achieve unbiased class-weighted outcomes. The models are trained with epochs and batch size empirically selected as 25 and 10 respectively for optimal performance. The loss function used was binary cross-entropy, with Adam Optimizer and L2 normalization. Batch size was set to 10 and epochs set to 25.

D. Supervised Classifier with pre-trained Feature Extraction

Extracting generic image descriptors through a pre-trained convolutional neural network has been known to show superior performance when compared to standard instance retrieval algorithms [11]. Features obtained through last layers of pre-trained deep learning convolutional networks outperform SIFT and HOG descriptors [12] and thus there exists a possibility to investigate similar methodologies with Logistic Regression, Naive Bayes and Random Forest (the scikit-learn [13] implementation of the machine learning classifiers have been used) respectively. The pre-trained networks adopted for the feature extraction in our work include InceptionV3, InceptionResNet, ResNet50 and VGG19. One could reason that the deeper layers of the CNN could possibly be skewed in extracting the higher level of features of the image training set, but contrary to this opinion, the features extracted at the last levels accommodate the intricacies specific to the medical image dataset that are difficult to detect through regular image descriptors as these images show minute variations across the breadth of the dataset. The activations of the fully connected (FC) layers

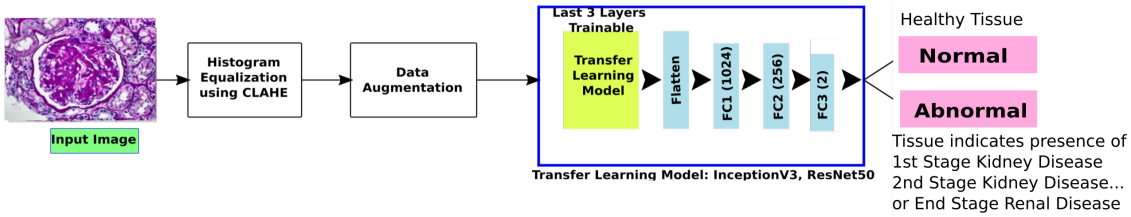

Fig. 1: Framework of pre-trained models used for transfer learning

capture the overall shape of the glomeruli structures while the last output layer preserves the spatial information. The local spatial information such as the several substructures of the biopsy tissue image may be lost when propagated through the fully connected layers which explains our hypothesis that transfer learning models would not be as effective on GCDB.

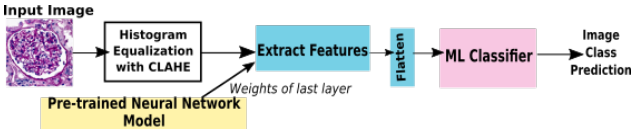

Fig. 2: Framework of supervised classifier with pre-trained feature extraction

Fig. 2 illustrates the execution of glomerulus classification via the proposed methodology. The dataset is preprocessed through histogram equalization before being vectorized. The proposed strategy is divided into three steps:

1) Extraction of last level feature vectors from a pre-trained CNN model.
2) Matrix multiplying the obtained weights with image vectors to form image specific feature vectors. The output of the feature extractor is generally of the form $w*h*d$ (w is the width, h is the height, d is the number of channels in the convolutional layers). These 2D arrays of d dimension are flattened.
3) Running a supervised classifier with the image feature vectors as an input to Logistic Regression, Naive Bayes and Random Forest models each.

The final layers from which the feature weights were extracted are for IRFE is CONV_7B, IFE is MIXED10, RFE is AVG_POOL and VFE is FC1. The hyper-parameters for random forest classifier were fine-tuned using 10-fold cross-validation and the results were found to be optimal when n_estimators, max_depth and max_features were fixed at 1000, 15 and *log2* respectively. Other parameters for the three classifiers were initialized to default values as set in [13].

IV. Evaluation

This section describes the dataset details (Section IV-A) followed by the results in Section IV-B.

A. Dataset

Glomerulus Classification Database (GCDB) is a dataset of 935 images of renal glomeruli obtained from human renal biopsies. The dataset has been constructed to have 619 abnormal and 316 normal images for the respective class labels. The imbalance of dataset is encouraged so as to represent a realistic picture that is usually found in a majority of pathologies. The identified biopsy images were sourced from Arkana Laboratories[1] and were procured between October 2017 to February 2018. The kidney tissues have been extracted through incisional biopsies and were processed and stained according to published standards. The tissue samples were digitized using MoticEasyScan[2] at 20X (0.5 micron/pixel) and 40X (0.25 micron/pixel) resolution. The images are obtained in TIFF-based SVS format that was converted into JPEG format before preprocessing. The static images from the glomeruli were captured using the Olympus camera. The digitized biopsy images consist of several renal substructures such as interstitial tissue, tubules, blood vessels and glomeruli. Out of these, the coordinates of the interested glomerular sections were marked and cropped out into patches for classification. The database was subjected to further cleaning where images with inappropriate length-width ratio and borderless glomeruli structures were excluded. Images having insufficient staining, poor light intensity and fragmented glomeruli portions are not included in the dataset. All the kidney biopsy images and

(a) Abnormal glomerulus (b) Normal glomerulus

Fig. 3: Sample glomeruli from each class

their annotations meet the medical standards. The glomeruli were categorized into normal and abnormal categories by multiple pathologists during the real-time patient-testing phase to maintain credibility in reported test outcomes used for clinical diagnosis. The GCDB dataset was extracted from the pool of archived image data available with the laboratory and was further subjected to verification by an independent expert pathologist. Every step was performed with due medical diligence [14] and in consensus with the kidney pathologist. Sample instances of GCDB have been illustrated in Fig. 3 with one sample each of abnormal and normal glomeruli.

B. Results

Table II portrays the experimental results obtained through retraining on the pre-trained models on GCDB dataset. Weighted metrics were calculated in each case as the class

[1]https://www.arkanalabs.com
[2]https://www.motic.com

Metric	ResNet50	InceptionV3
Accuracy	72.35	68.56
Precision	0.7456	0.7135
Recall	0.7256	0.6892
F1 Score	0.7297	0.6901

TABLE II: Results for transfer learning

Classifiers		Accuracy	Precision	Recall	F1
IRFE	*LogReg*	**0.8823**	0.8813	**0.8823**	**0.8808**
	Random Forest	0.8074	0.8508	0.8074	0.7811
	Naive Bayes	0.7219	0.7480	0.7219	0.7284
IFE	*LogReg*	0.8663	0.8653	0.8663	0.8635
	Random Forest	0.8021	0.8476	0.8021	0.7738
	Naive Bayes	0.7542	0.7540	0.7540	0.7540
RFE	*LogReg*	0.8342	0.8316	0.8342	0.8307
	Random Forest	0.8449	0.8574	0.8449	0.8341
	Naive Bayes	0.6791	0.7139	0.6791	0.6873
VFE	*LogReg*	0.8716	0.8702	0.8716	0.8699
	Random Forest	0.8770	**0.8962**	0.8770	0.8684
	Naive Bayes	0.6042	0.7439	0.6042	0.6049

TABLE III: Metrics of supervised classification with pre-trained feature extractor (LogReg: Logistic Regression).

imbalance may skew the model performance to unilaterally favor the dominant class and obscure the results. The preliminary results depicted herein were treated as baselines for later comparisons. The findings suggest to be supporting our hypothesis that transfer learning models unexpectedly suffer from misclassification due to high congruity and intricate variations in biopsy image. Amongst ResNet50 and InceptionV3, ResNet50 clearly outperforms InceptionV3 primarily due to the presence of shortcut connections known as residual networks in its architecture.

The second phase of feature-based supervised classification was trained on the same dataset with stratified sampling while doing K-fold cross-validation for parameter tuning in order to account for class imbalance. The results compiled in Table III gives plausible credibility to our proposition to use CNN-based feature extraction as a preliminary step for the glomerulus classification task. Out of multiple pre-trained models used on numerous classifiers, the Logistic Regression supplemented with features imparted by InceptionResNetV2 is most successful in terms of accuracy, recall and F1 score. The precision value, in this case, is secondary only to the Random Forest classifier supported by features through VGG19 architecture. Feature extraction based classifiers showed a drastic improvement in contrast to transfer learning based methods. Amongst contemporary secondary classifiers, Logistic Regression shows greater ability to adapt to the fine-grained features of the glomeruli images with high capacity to fit onto a small-sized, immensely correlated dataset. On the other hand, the InceptionResNetV2 framework gives the most favorable results compared to its corollaries followed by VGG19 as evident by slight degradation in the metric values.

V. Conclusion and Future Work

The experiments prove that despite cumbersome computations, transfer learning models ResNet50 and InceptionV3 did not surpass feature-enriched Logistic Regression, Naive Bayes and Random Forest models owing to high interclass similarities as highlighted in the discussion of results.

The current findings aim to establish a state of the art in the area of renal histopathology. The dataset that has been treated as ground truth for this paper can be extended to include unreported categories of glomeruli such as Sclerotic and Crescentic. The work is a progressive step in a positive direction to determine the severity of the kidney diseases such as Focal Segmental Glomerulosclerosis (FSGS) as they can be predicted from the extent and count of the glomeruli having abnormal morphology [15].

References

[1] K. J. Johnson, N. G. Wreford, W. E. Hoy, and J. F. Bertram, "Estimating total glomerular number in human kidneys with a physical disector/fractionator combination," *Image Analysis & Stereology*, vol. 19, no. 2, pp. 105–108, 2011.

[2] L. Scarfe, A. Rak-Raszewska, S. Geraci, D. Darssan, J. Sharkey, J. Huang, N. C. Burton, D. Mason, P. Ranjzad, S. Kenny *et al.*, "Measures of kidney function by minimally invasive techniques correlate with histological glomerular damage in scid mice with adriamycin-induced nephropathy," *Scientific reports*, vol. 5, p. 13601, 2015.

[3] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC bioinformatics*, vol. 13, no. 1, p. 282, 2012.

[4] L. O'Gorman, A. C. Sanderson, and K. Preston, "A system for automated liver tissue image analysis: methods and results," *IEEE Transactions on biomedical engineering*, no. 9, pp. 696–706, 1985.

[5] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[6] D. Ledbetter, L. Ho, and K. V. Lemley, "Prediction of kidney function from biopsy images using convolutional neural networks," *arXiv preprint arXiv:1702.01816*, 2017.

[7] A. B. Farris and C. E. Alpers, "What is the best way to measure renal fibrosis?: A pathologist's perspective," *Kidney international supplements*, vol. 4, no. 1, pp. 9–15, 2014.

[8] K. Solez, R. A. Axelsen, H. Benediktsson, J. F. Burdick, A. H. Cohen, R. B. Colvin, B. P. Croker, D. Droz, M. S. Dunnill, P. F. Halloran *et al.*, "International standardization of criteria for the histologic diagnosis of renal allograft rejection: the banff working classification of kidney transplant pathology," *Kidney international*, vol. 44, no. 2, pp. 411–422, 1993.

[9] J. Gallego, A. Pedraza, S. Lopez, G. Steiner, L. Gonzalez, A. Laurinavicius, and G. Bueno, "Glomerulus classification and detection based on convolutional neural networks," *Journal of Imaging*, vol. 4, no. 1, p. 20, 2018.

[10] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[11] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[12] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2014, pp. 512–519.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[14] P. D. Walker, T. Cavallo, and S. M. Bonsib, "Practice guidelines for the renal biopsy," *Modern Pathology*, vol. 17, no. 12, p. 1555, 2004.

[15] R. Hogg, "Focal segmental glomerulosclerosis in children with idiopathic nephrotic syndrome. a report of the southwest pediatric nephrology study group 1," *Kidney International*, vol. 27, no. 2, pp. 442–449, 1985.