

Application of geostatistical functions and deep features to kidney biopsy images to differentiate focal segmental glomerulosclerosis from minimal change disease

Justino Duarte Santos¹
Departamento de Computação
Universidade Federal do Piauí
Teresina-PI, Brasil
Email: justinoduarte@gmail.com

Romuer R. V. Silva
Curso de Sistemas de Informação
Universidade Federal do Piauí
Picos-PI, Brasil
Email: romuere@ufpi.edu.br

Rodrigo M. S. Veras
Departamento de Computação
Universidade Federal do Piauí
Teresina-PI, Brasil
Email: rveras@ufpi.edu.br

Abstract—Chronic kidney diseases arise from acute or intermittent pathologies that have not been adequately treated, such as minimal change disease (MCD) and focal segmental glomerulosclerosis (FSGS). The accurate identification of these two diseases is of paramount importance, because their treatments and prognoses are different. Thus, we propose a method that is capable of differentiating MCD from FSGS based on images from pathological examinations. In the proposed method, we use four pre-trained convolutional neural networks and geostatistical functions to extract image features. Of the 8,720 extracted features, we selected 94 based on mutual information criteria, and in the classification step, we used a random forest classifier. The proposed method obtained an accuracy of 94.3% and Kappa index of 87.9%, a level that is regarded as “almost perfect”, confirming that our method is very promising.

I. INTRODUCTION

Glomerulopathies are kidney diseases with different histopathological subtypes. Microscopic evaluation is crucial for diagnosis, since it provides prognostic data and guidance for treatment. In Brazil, glomerulopathies are among the leading causes of end-stage kidney disease (ESKD), and account for 11% of patients on dialysis [1]. In 2016, nearly 125,000 people in the USA started treatment for ESKD [2].

Nephrotic syndrome is one of the primary forms of glomerular disease, and when symptoms are persistent, is associated with a progression to chronic kidney disease (CKD). A publication by [2] from the US Department of Health & Human Services reports that 15% of US adults (37 million people) are estimated to have CKD.

Several histological abnormalities may lead to the development of nephrotic syndrome. Common causes of idiopathic nephrotic syndrome are minimal change disease (MCD) and focal segmental glomerulosclerosis (FSGS). In children, MCD is the cause of nephrotic syndrome in 90% of patients, while in adults, primary glomerular diseases such as FSGS and MCD are the cause of nephrotic syndrome in 70% of cases. When

considering only the adult population, FSGS is the leading cause of nephrotic syndrome in several countries [3].

It is essential to understand the differences between these two glomerulopathies (MCD and FSGS). From a therapeutic point of view, there are differences in the treatments, at least with regard to the duration of the attack phase in the case of corticosteroids, as well as in the rates of response to treatment and the prognoses of these two diseases [3].

Given the importance of differentiating between these two types of kidney disease, this article proposes a computational method of classification that is capable of distinguishing between them based on biopsy images. To achieve this goal, we use geostatistical texture descriptors and transfer learning (TL) techniques in convolutional neural networks (CNNs) and supervised classifiers.

This article is organized as follows. Section II presents related works; Section III describes the materials, techniques and the proposed method; Section IV presents the obtained results and a discussion; and finally, we present a conclusion and future work in Section V.

II. RELATED WORK

Transfer learning (TL) techniques that use medical images as input are widely used in computer-aided-diagnosis (CAD) systems. Other approaches perform segmentation and identification of kidney structures in images, especially those related to the glomerulus. Zhao et al. [4] proposed an automated glomerulus extraction framework based on a micrograph of the entire kidney. The results showed that 110 of 140 glomeruli from five whole-slide images were correctly extracted, with an average completeness of over 90%.

Sarder, Ginley and Tomaszewski [5] estimated the location of the glomerulus in 15 images of kidney biopsies, achieving an accuracy of 88%. Their methodology was to extract 148 smaller regions containing a single glomerulus, and to use them to segment the glomerular boundary. The accuracy of this process was 86%. In another experiment, 50 images were

¹M.Sc. dissertation.

sampled for nuclei segmentation per glomerulus, giving an accuracy of 92%.

In another work [6], they applied an integrated approach using Gabor filtering and Gaussian blurring to label glomerular textural boundaries. The best result from 400 individual glomerulus images was an accuracy of 89%.

Barros et al. [7] proposed a computer system for the detection of proliferative glomerular lesions (PGL) that could differentiate them from healthy images. They used image preprocessing, segmentation and pattern recognition methods, and the k -nearest neighbor (KNN) algorithm [8] was used as a classifier. The accuracy achieved in this work was $88.3 \pm 3.6\%$.

Araújo et al. [9] used images of single glomeruli to detect segmental glomerulosclerosis. Their architecture had the typical structure, consisting of a digital image processing and pattern recognition system. Three feature vectors were extracted and supplied to four classifiers: KNN, support vector machine (SVM) [10], a neural network, and naive Bayes. These authors achieved an accuracy of 84.8% for hematoxylin-eosin (H&E) stained samples and 81.3% for periodic acid-Schiff (PAS) stained samples.

Ginley et al. [11] proposed an approach to defining the structural progression of human glomeruli in diabetic nephropathy. The authors segmented glomerular compartment boundaries and quantified 47 features from each glomerulus, using texture analysis, morphological, distance, and handcrafted features. They used a naive Bayesian classifier on the feature set, and reported that this method was able to distinguish pathological stage IIa from stage III with sensitivity/specificity 0.89/0.93, and stage IIb from stage III with sensitivity/specificity 0.7/0.8, for 514 glomeruli taken from 13 human biopsies with diagnosed diabetic nephropathy, and five human renal tissues with no histological abnormalities.

Sheehan and Korstanje [12] developed a method for identifying and collecting quantitative data from glomeruli. This approach was semi-automatic, since it required intervention from a specialist. The authors used contrast enhancement and Gaussian blurring, followed by a size filter to identify regions of interest corresponding to glomeruli tufts. Three features were extracted: mesangial matrix expansion (MME) [13], the number of nuclei, and capillary openness, which were classified using a random forest (RF) approach [14]. A strong correlation was reported between MME and the analyzed phenotypes.

Marsh et al. [15] described the development of a deep learning model that identifies and classifies non-sclerosed and sclerosed glomeruli in whole-slide images of frozen biopsy sections of donor kidneys. This differentiation is meaningful because the criterion for accepting or rejecting the donor kidneys relies heavily on the pathologist's determination of the percentages of glomeruli that are normal and sclerotic. The proposed approach fine-tuned the VGG-16 [16] convolution neural network using 48 whole-slide images. According to the authors, the model achieved a precision of 81.28% in the identification of non-sclerosed glomeruli. They concluded that the method outperformed another model trained on image

patches of isolated glomeruli, in terms of accuracy and speed.

Chagas et al. [17], like in [7], also worked on PGL detection, but their proposal did more than binary classification (lesioned or healthy), also performing classification in specific PGL subcategories: endocapillary, mesangial, and, both. These authors built a CNN-based architecture to extract features from glomerulus images, and these features were then supplied to an SVM classifier. In the classification task, their proposed method achieved an accuracy of 82%.

Other works use renal biopsy image databases, but with different purposes such as segmentation [18], or classification between global or local sclerosis [19]. But to the best of our knowledge, there are no databases or works that use computational methods to differentiate between FSGS and MCD. This is, therefore, the main scientific contribution of the current article.

III. MATERIALS AND METHODS

In this section, we present a method that is capable of differentiating between kidney biopsy images with FSGS vs. MCD. We performed experiments using texture as the Haralick features [20] and geostatistical, pre-trained CNNs such as VGG-16, VGG-19 [16], Xception [21], and ResNet50 [22].

We evaluated the potential individual description of each of these approaches, and concluded that a combination of different characteristics would lead to better classification rates. We therefore formed a new image representation using a set of features.

A. Proposed Method

Figure 1 presents an overview of the proposed method. It can be seen that it contains four steps: pre-processing, feature extraction, feature selection and classification.

From performing a set of experiments, we observed that a concatenation of the geostatistical functions with the last fully connected layer in the VGG-16, Xception, and ResNet50 CNNs produced the best results. In the CNN feature extraction process, only resizing was done at the pre-processing step, whereas in the extraction by geostatistical functions, we computed the local binary patterns (LBPs) [23] of the image channels. We applied the mutual information (MI) algorithm to rank the features from the most to the least significant, and performed feature selection. Finally, the RF classifier was used to determine whether the image contained MCD or FSGS.

1) *Pre-processing*: In the pre-processing step, the images were resized to the default input dimensions for the CNNs: 224×224 for VGG-16, VGG-19 and ResNet50, and 299×299 for Xception. This was performed without border cutout or addition of padding, although this occasionally changed the original aspect ratio of the input image. In order to extract texture information with the geostatistical functions, the images were also resized to 299×299 , due to hardware and time issues, and then represented in the form of LBPs for each color channel. To extract Haralick features, the images were used in their original sizes and smoothed via the application of a median filter with a 9×9 -sized window looking for a

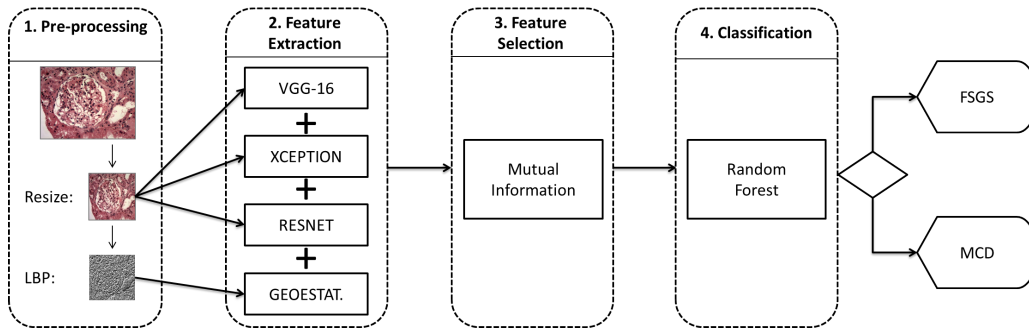


Fig. 1. Flowchart for the proposed method of automatic differentiation between MCD and FSGS in biopsy images.

manual inspected trade-off between noise reduction and details retaining.

2) *Feature Extraction*: Geostatistics are statistics about a population with a known address (i.e., coordinates). The fundamental theory of geostatistics is based on the assumption that on average, samples that are near to each other in time and space are more similar than those that are distant [24].

In this work, we use four geostatistical functions: a semivariogram, a semimadogram, a covariogram and a correlogram. These functions summarize the strengths of the associations between responses as a function of distance and possibly direction [25], and can describe the texture of a given image through the degree of spatial association between spatially referenced pixels of the input image as follows [26]:

We use vectors (h) obtained from a combination of four directions (0° , 45° , 90° or 135°) and 11 distances, of which six are fixed (1, 2, 5, 10, 15, 20) and five are proportional to the image width (1.25%, 2%, 5%, 10%, and 20%). Thus, by combining the 44 distance vectors (h) with the four geostatistical functions and applying these to the R, G, and B image channels, the resulting vector contains 528 characteristics.

The Haralick features are calculated based on the gray-level co-occurrence matrix (GLCM), a texture descriptor that analyzes the co-occurrences between pairs of pixels and stores their relative intensities in a square matrix, with dimensions equal to the number of *levels* of gray (i.e. 256 in the case of 8-bit image levels). The probabilities of co-occurrences ($P_{i,j}$) are calculated between two levels of gray, i and j , using an angle θ (0° , 45° , 90° or 135°) and a distance called the pixel pair spacing. For this purpose, we use 11 distance values: six of which are fixed (1, 2, 5, 10, 15, 20) and five that are proportional to the image width (1.25%, 2%, 5%, 10%, and 20%).

Although we can define several characteristics based on the GLCM, the most important of these are contrast, dissimilarity, homogeneity, angular second moment (ASM), and correlation. We extract the texture feature using the three channels from the RGB image. Thus, for the Haralick vector, we have 660 attributes for each image:

CNNs are commonly applied in the field of machine learning for image classification. A significant advantage of these techniques is their ability to automatically detect essential

features, since their deep architectures allow us to extract a set of characteristics at multiple levels of abstraction [27]. CNNs have been proposed to aid in diagnosis, and have outperformed conventional methods of extracting features, obtaining better accuracy rates [28].

The architecture of a CNN typically includes two sections. The first is formed of a sequence of convolution operations followed by pooling operations, and the second is composed of fully connected layers.

After the sequence of convolutions and pooling, the generated feature map forms the input for the fully connected layers. In this step, the architecture and mechanism of operation are similar to those of a traditional neural network, and the last layer generates the output data (classification).

The training of a CNN is a task with high computational cost, and requires a large amount of data to achieve satisfactory results in terms of the power of generalization. To avoid the training stage, we use TL, which allows the domains, tasks, and distributions used in training and testing to be different [29]. That is, the goal is to reuse the knowledge learned in one field and to apply it to another correlate [27].

We apply TL in feature extraction by taking the output vectors of the last fully connected layer (before the classification layer) of four CNNs: VGG-16, VGG-19, Xception and ResNet50. All CNNs were pre-trained on the ImageNet database [30], which contains more than 1.2 million images and 1,000 classes.

Table I shows a summary of all the individual feature vectors extracted. In addition to these six individual vectors, we also evaluated all of their combinations, giving 63 vectors in total.

In some of the sections below, we use the following abbreviations to refer the feature vectors to improve comprehension and layout: v16=VGG-16, v19=VGG-19, xce=Xception, rsnet=ResNet50, hrlk:Haralick and geo=geostatistics.

3) *Feature Selection*: We performed a feature selection process for each of the 63 feature vectors obtained in the previous step, sorting the features in the vector in descending order of relevance. We used the F statistic of variance analysis (ANOVA-F) and mutual information (MI) to calculate the relevance of each feature [31].

TABLE I
SUMMARY OF THE FEATURES EVALUATED IN THIS WORK

Type	Method	Features	Number of features
Texture	Geostatistic	Semivariogram, semimadogram, covariogram and correlogram	528
	Haralick	Contrast, dissimilarity, homogeneity, angular second moment and correlation	660
Deep features	VGG-16	Features extracted from the last fully connected layer	4,096
	VGG-19		4,096
	Xception		2,048
	ResNet50		2,048
TOTAL			13,476

4) *Classification:* We performed empirical tests to define the dimensionality of the final feature vector, always searching for a trade-off between better results and vector size. We used the ranked features (from ANOVA-F and MI), starting from the one with the highest relevance and adding the others incrementally to complete all features. We used a random forest approach in our proposed method, but also evaluated other two classifiers, SVM and KNN.

The initial parameters used in the SVM were a penalty of 1.0, and a radial basis function (RBF) kernel with a gamma coefficient of 1/number of features. For KNN, we chose $K = 5$ and the Euclidian distance. Finally, for RF, we used 100 trees, and no limit in-depth grow. After analyzing the classification results, we chose the five best results and performed a search for the best set of classifier hyperparameters.

In order to evaluate the selected classifiers, the input images were grouped into training and test sets using the stratified cross-validation technique ($k = 5$). We computed a confusion matrix for each fold, and from this, the accuracy and kappa metrics were calculated. We computed the arithmetic mean of the five values achieved from each classifier studied.

B. Evaluation Metrics

The confusion matrix confronts the classifier predicted results and the actual results for the same set of tests. In our case, we have two classes, FSGS and MCD, and our approach is therefore a binary classifier. Thus, there are four values in this matrix: the true positive (TP), which indicates the quantity of images correctly classified as MCD; the true negative (TN), corresponding to the number of correct FSGS classifications; the false positive (FP), representing the number of images classified as MCD, but which are actually FSGS; and finally, the false negative (FN), which refers to the number of images erroneously classified as FSGS.

The results in this work were analysed using four metrics from the literature: accuracy, precision, recall and kappa (κ) [32].

We used Cohen’s kappa [32] as the primary evaluation metric. This gives a value that represents the degree of agreement between nominal classifications performed by two evaluators

TABLE II
CHARACTERISTICS OF THE IMAGES USED.

Width	Height	Ratio H/W	Number of images	
			FSGS	MCD
1024	768	0.75	17	40
2048	1644	0.80	6	0
3264	2448	0.75	2	0
4456	2976	0.67	1	0
4096	3288	0.80	0	12
5312	2988	0.56	7	2
Total			33	54

– in the present case, those predicted by the classifier and those annotated by the pathologist. It can be calculated based on Equation 1.

$$\kappa = \frac{(\text{observed} - \text{expected})}{1 - \text{expected}} \times 100 \quad (1)$$

The maximum value of kappa is 100%, which indicates perfect agreement among the evaluators. The result is qualified according to the value of κ as follows: $\kappa \leq 20\%$: Slight; $20\% < \kappa \leq 40\%$: Fair; $40\% < \kappa \leq 60\%$: Moderate; $60\% < \kappa \leq 80\%$: Substantial e $\kappa > 80\%$: Almost Perfect.

C. Image Database

The image database used in this work was composed of 87 colored images, of which a specialist had classified 33 as FSGS and 54 as MCD.

We obtained the images using microscopes Nikon e220, and a Nikon e200 adapted with immunofluorescence, using different objective lenses and consequently different magnification rates. Pigmentation was applied to the slides using the following dyes: H&E, Masson’s trichrome, PAS and silver methenamine. The resolutions and aspect ratios of the acquired images were heterogeneous, as indicated in Table II, but all were larger than the input image size of the CNNs used.

Figure 2 shows examples of these images. In many cases, visual heterogeneity is observed between images belonging to the same class, as well as similarity between images in distinct classes, which is always very challenging for classification.

IV. RESULTS AND DISCUSSION

In this study, we evaluated 378 combinations of tests (63 feature vectors, two selectors and three classifiers). We applied a incremental feature selection methodology by computing the importance of each feature using the MI and ANOVA-F algorithms.

To define the best scenario, we calculated the highest mean kappa reached for each of the 378 tested models using the incremental attribute selection approach. As an example, Figure 3 shows the mean values of accuracy and kappa for the five runs of the stratified k-fold achieved with the RF classifier, using the attributes of the concatenated v16+xc+rsnet+geo vector, selected based on mutual information.

Tables III, IV and V show the 5 best classification results from the SVM, KNN and RF classifiers, respectively. These

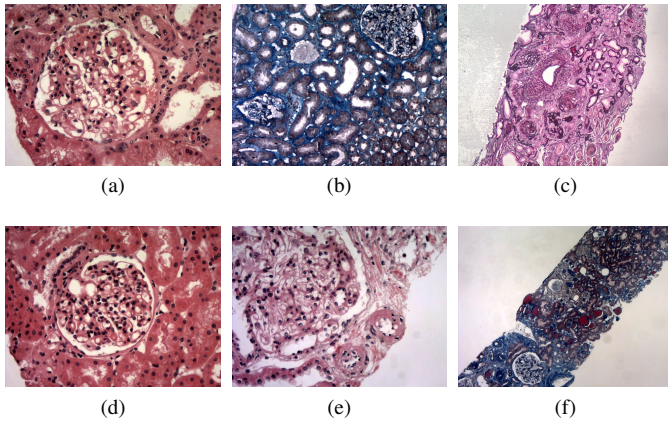


Fig. 2. Samples of images from the database. a-c: focal segmental glomerulosclerosis images; d-f: minimal change disease images.

TABLE III
FIVE BEST RESULTS ACHIEVED FOR KAPPA WITH SVM CLASSIFIER
(RESULTS OBTAINED WITH MI).

Vector	#F	accuracy	kappa	precision	recall
v19+rsnet	101	91.0±7.5	80.6±15.8	90.0±6.3	96.4±7.3
v16+v19+rsnet	136	91.0±7.5	80.6±15.8	90.0±6.3	96.4±7.3
rsnet	52	91.0±4.5	80.6±9.8	91.5±4.9	94.5±4.5
hrlk+rsnet	59	91.0±4.5	80.6±9.8	91.5±4.9	94.5±4.5
v19+xce+rsnet	257	91.0±7.5	80.6±15.7	89.8±6.3	96.4±7.3

-Best results are in bold.

results were obtained using MI to select the features (in these tables, column #F indicates the number of selected features).

From analyzing the results in Tables III, IV and V, it can be seen that the 5 highest kappa values were achieved using the features selected by mutual information and classified using the RF classifier.

To improve the classification results, we calculated the five

TABLE IV
FIVE BEST RESULTS ACHIEVED FOR KAPPA USING THE KNN CLASSIFIER
(RESULTS OBTAINED WITH MI).

Vector	#F	accuracy	kappa	precision	recall
v16+v19+xce	68	92.1±7.5	83.3±15.8	93.0±6.7	94.4±7.4
v19+xce+rsnet	181	91.0±7.5	81.1±15.8	92.9±6.8	92.7±6.8
v16+xce+rsnet	165	91.0±5.6	81.1±11.5	92.5±3.7	92.7±6.8
v16+v19+rsnet	406	91.0±7.5	80.6±15.7	89.8±6.3	96.4±7.3
v16+rsnet	169	89.9±5.5	78.4±11.4	91.3±4.9	92.7±6.8

-Best results are in bold.

TABLE V
FIVE BEST RESULTS ACHIEVED FOR KAPPA USING THE RF CLASSIFIER
(RESULTS OBTAINED WITH MI).

Vector	#F	accuracy	kappa	precision	recall
v16+v19+xce	62	93.3±6.5	85.5±14.1	91.9±7.2	98.2±3.6
v16+v19+hrlk+geo	216	93.3±6.5	85.5±14.1	91.9±7.2	98.2±3.6
v16+v19+geo	434	93.3±6.5	85.5±14.1	91.9±7.2	98.2±3.6
v16+xce+geo	61	93.2±4.1	85.4±9.0	93.3±5.9	96.4±4.5
v16+xce+rsnet+geo	94	93.2±4.1	85.2±8.9	91.6±4.9	98.2±3.6

-Best results are in bold.

TABLE VI
COMPARISON OF CLASSIFICATION RESULTS BEFORE AND AFTER
HYPERPARAMETER TUNING USING V16+XCE+GEO AND
V16+XCE+RSNET+GEO DESCRIPTORS.

Vector	Metric	Before	After
v16+xce+geo	accuracy	93.2±4.1%	94.2±3.5%
	kappa	85.4±9.0%	87.4±7.8%
	precision	93.3±5.9%	93.4±5.9%
	recall	96.4±4.5%	98.2±3.6%
v16+xce+rsnet+geo	accuracy	93.2±4.1%	94.3±3.5%
	kappa	85.2±8.9%	87.9±7.4%
	precision	91.6±4.9%	94.7±4.3%
	recall	98.2±3.6%	96.4±4.5%

-Best results are in bold.

best results and performed a search for the best hyperparameters of the RF classifier. For this, we used the same validation criterion (stratified 5-fold cross-validation) and used the set of selected features as input. We evaluated a total of 211,200 hyperparameter configurations using the RF classifier.

The hyperparameter tuning, when was performed on the top three kappa results (first three lines of Table V) did not lead to better results, but for the fourth and fifth positions, the results were improved. Table VI allows for a comparison between the results reached before and after the RF tuning, using v16+xce+geo and v16+xce+rsnet+geo descriptors. It can be seen that almost all metrics were improved, and the standard deviation decreased in most cases.

Finally, after tuning, we defined the best feature vector: v16+xce+rsnet+geo. This vector was formed from the 94 features selected based on MI and classified using the RF classifier. Figure 4 shows the number of features chosen from each descriptor and their distributions along the positions of the vector, sorted by the mutual information algorithm applied to the v16+xce+rsnet+geo feature vector. It is notable that the geostatistical features, which had only 528 attributes (compared with at least 2048 for the CNNs), obtained a significant rate on the essential characteristics subset.

We evaluated 211,200 parameters combination and the best parameters found from the tuning process of the RF classifier were number of trees: 30; minimum number of samples required to split an internal node: 5; Minimum number of samples required to be at a leaf node: 1; grow limit in the depth way: unlimited; split quality measure function: gini impurity; amount of features to consider when looking for the best split: $\log_2(N)$, where N is number of features; value which grow trees in best-first fashion: unlimited.

V. CONCLUSION AND FUTURE WORK

Glomerulopathies can lead to end-stage kidney disease. Microscopic evaluation is crucial in diagnosis, as this can provide prognostic data and guidance for treatment. Several histological abnormalities may lead to the development of nephrotic syndrome, and of these, MCD and FSGS are highlighted as causes of idiopathic nephrotic syndrome. Differentiating MCD from FSGS is essential to determine drug regimens, as these are strongly dependent on the diagnosis.

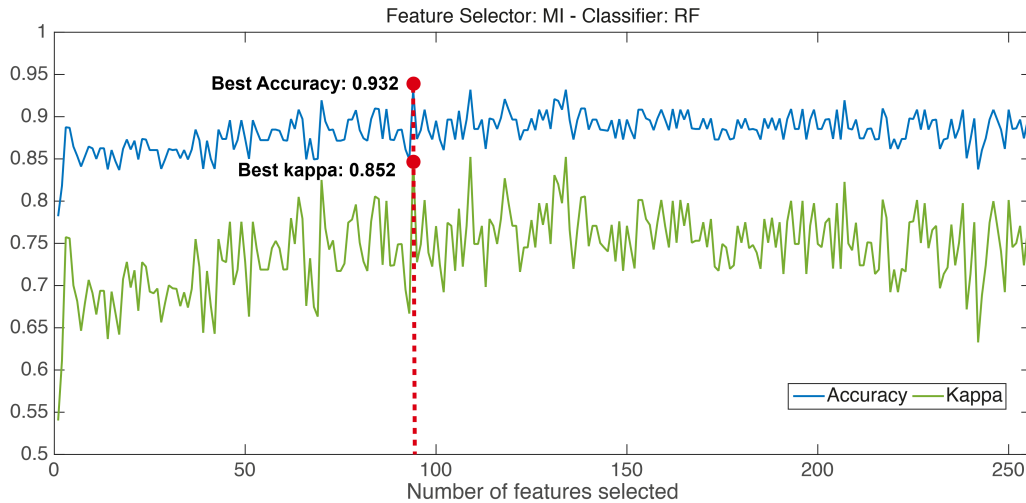


Fig. 3. Values of accuracy and kappa achieved using the random forest classifier with a different number of features from the v16+xce+rsnet+geo vector, sorted by mutual information (for easy visualization, only up to 256 features are represented).

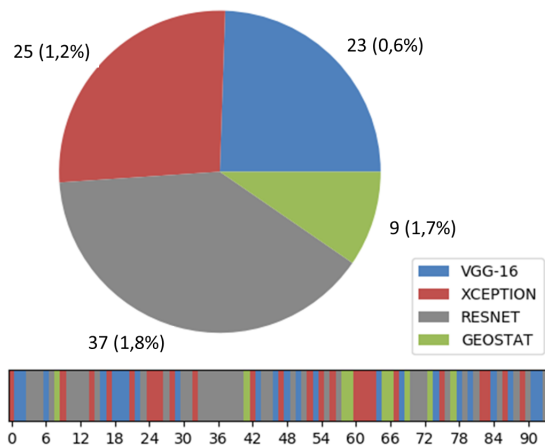


Fig. 4. Origin of the 94 attributes with best mutual information for the v16+xce+rsnet+geo vector. The bar at the bottom depicts the source of the characteristics at each position.

This article has proposed a computer vision method to differentiate between MCD and FSGS in microscopic images. The tests performed to define the final approach included feature extraction based on traditional texture features, i.e. Haralick features, geostatistical functions, and deep features from pretrained CNNs. These features were then selected using two criteria: mutual information and ANOVA-F. We then used these as input for three supervised classifiers: SVM, KNN, and RF.

The results indicate that the use of a concatenated feature vector of VGG-16, Xception, ResNet50, and geostatistics gave the best scores, followed by feature selection based on mutual information and classification using a random forest classifier. This model gave results that indicate near-perfect agreement with the pathologist's diagnosis.

The characterization of the image using a single type of

descriptor shows little power of representation. On the other hand, creating a mixed descriptor, such as deep features and texture, can extract attributes from the image in a more productive manner.

The feature selection is significant, not only because it reduces the dimensionality of the descriptors, and consequently relieves the hardware, but also because it removes attributes that have little relevance to the classification that may hinder the classifier's task. In this sense, the selection using the Mutual Information criterion is better than ANOVA-F in the task proposed here.

The Random Forest classifier proves to be superior to the others tested, due to its nature of being a committee of decision trees internally, and even more resistant to overfitting due to the formed trees considering different subsets of attributes.

As future work, we intend to evaluate methods that use learning transfer, but in the form of fine-tuning in convolutional neural networks. This technique takes a pre-trained CNN on a large database and re-trains some network layers over a specific dataset with a low learning rate doing an adjust in their weights.

VI. PUBLICATIONS

- Classificação de Imagens de Biópsias Renais com Glomerulosclerose Segmentar e Focal ou com Lesões Mínimas Utilizando Transfer Learning em CNN. Online version: <https://sol.sbc.org.br/index.php/sbcas/article/view/6244>.
- A Hybrid deep feature space to differentiate focal segmental glomerulosclerosis from minimal change disease in kidney biopsy images (submitted paper to Computers in Biology and Medicine).

REFERENCES

- [1] D. M. d. N. Costa, L. M. Valente, P. A. d. C. Gouveia, F. W. Sarinho, G. V. Fernandes, M. A. G. d. M. Cavalcante, C. B. L. d. Oliveira, C. d. A. J. d. Vasconcelos, and E. S. C. Sarinho, "Comparative analysis

- of primary and secondary glomerulopathies in the northeast of Brazil: data from the Pernambuco registry of glomerulopathies-repeg," *Brazilian Journal of Nephrology*, vol. 39, no. 1, pp. 29–35, 2017.
- [2] Centers for Disease Control and Prevention, "Chronic kidney disease in the united states, 2019," 2019, gA: US Department of Health and Human Services, Centers for Disease Control and Prevention.
 - [3] L. R. Moura, M. F. Franco, and G. M. Kirsztajn, "Minimal change disease and focal segmental glomerulosclerosis in adults: response to steroids and risk of renal failure," *Brazilian Journal of Nephrology*, vol. 37, no. 4, pp. 475–480, 2015.
 - [4] Y. Zhao, E. F. Black, L. Marini, K. McHenry, N. Kenyon, R. Patil, A. Balla, and A. Bartholomew, "Automatic glomerulus extraction in whole slide images towards computer aided diagnosis," in *12th International Conference on e-Science (e-Science)*. IEEE, 2016, pp. 165–174.
 - [5] P. Sarder, B. Ginley, and J. E. Tomaszewski, "Automated renal histopathology: digital extraction and quantification of renal pathology," in *Medical Imaging 2016: Digital Pathology*, vol. 9791. International Society for Optics and Photonics, 2016, p. 97910F.
 - [6] B. Ginley, J. E. Tomaszewski, and P. Sarder, "Automatic computational labeling of glomerular textural boundaries," in *Medical Imaging 2017: Digital Pathology*, vol. 10140. International Society for Optics and Photonics, 2017, p. 101400G.
 - [7] G. O. Barros, B. Navarro, A. Duarte, and W. L. Dos-Santos, "Pathspotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Scientific reports*, vol. 7, p. 46769, 2017.
 - [8] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, Jan 1991.
 - [9] I. C. d. Araújo, L. Schnitman, A. A. Duarte, and W. L. dos Santos, "Automated detection of segmental glomerulosclerosis in kidney histopathology," in *XIII Brazilian Congress on Computational Intelligence*, 2017.
 - [10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995.
 - [11] B. G. Ginley, J. E. Tomaszewski, K.-Y. Jen, A. Fogo, S. Jain, and P. Sarder, "Computational analysis of the structural progression of human glomeruli in diabetic nephropathy," in *Proceedings of SPIE Medical Imaging*, vol. 10581, 2018, pp. 105 810A–1–105 810A–6.
 - [12] S. M. Sheehan and R. Korstanje, "Automatic glomerular identification and quantification of histological phenotypes using image analysis and machine learning," *American Journal of Physiology-Renal Physiology*, vol. 315, no. 6, pp. F1644–F1651, 2018.
 - [13] G. A. Noordmans, C. R. Caputo, Y. Huang, S. M. Sheehan, M. Bulthuis, P. Heeringa, J.-L. Hillebrands, H. van Goor, and R. Korstanje, "Genetic analysis of mesangial matrix expansion in aging mice and identification of far2 as a candidate gene," *Journal of the American Society of Nephrology*, vol. 24, no. 12, pp. 1995–2001, 2013.
 - [14] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, Aug 1995, pp. 278–282.
 - [15] J. N. Marsh, M. K. Matlock, S. Kudose, T. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass, "Deep learning global glomerulosclerosis in transplant kidney frozen sections," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2718–2728, Dec 2018.
 - [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
 - [17] P. Chagas, L. Souza, I. Araújo, N. Aldeman, A. Duarte, M. Angelo, W. L. dos Santos, and L. Oliveira, "Classification of glomerular hypercellularity using convolutional features and support vector machine," *arXiv preprint arXiv:1907.00028*, 2019.
 - [18] S. Kannan, L. A. Morgan, B. Liang, M. G. Cheung, C. Q. Lin, D. Mun, R. G. Nader, M. E. Belghasem, J. M. Henderson, J. M. Francis *et al.*, "Segmentation of glomeruli within trichrome images using deep learning," *Kidney international reports*, vol. 4, no. 7, pp. 955–962, 2019.
 - [19] N. P. Pavinkurve, K. Natarajan, and A. J. Perotte, "Deep vision: learning to identify renal disease with neural networks," *Kidney International Reports*, vol. 4, no. 7, p. 914, 2019.
 - [20] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
 - [21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint*, pp. 1610–02 357, 2017.
 - [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 770–778.
 - [23] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.
 - [24] E. Isaaks, R. Srivastava, and K. (Firm), *An introduction to Applied Geostatistics*. Oxford University Press, 1989.
 - [25] A. C. Silva, P. C. P. Carvalho, and M. Gattass, "Analysis of spatial variability using geostatistical functions for diagnosis of lung nodule in computerized tomography images," *Pattern Analysis and Applications*, vol. 7, no. 3, pp. 227–234, 2004.
 - [26] J. A. de Sousa, A. C. de Paiva, J. D. S. de Almeida, A. C. Silva, G. B. Junior, and M. Gattass, "Texture based on geostatistic for glaucoma diagnosis from fundus eye image," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 19 173–19 190, 2017.
 - [27] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
 - [28] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.
 - [29] S. J. Pan, Q. Yang *et al.*, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
 - [30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec 2015.
 - [31] T. M. Cover and J. A. Thomas, "Entropy, relative entropy and mutual information," *Elements of information theory*, vol. 2, pp. 1–55, 1991.
 - [32] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.