BİNGÖL ÜNİVERSİTESİ
TEKNİK BİLİMLER DERGİSİ

Bingöl Üniversitesi Teknik Bilimler Dergisi
Bingol University Journal of Technical Science
e-ISSN: 2757-6884

Cilt:6, Sayı: 1, Sayfa: 13-29, 2025
Araştırma Makalesi

Volume:6, Number: 1, Page: 13-29, 2025
Research Article

Doi: 10.5281/zenodo.15719179

Hybrid Feature-Based Classification of Glomeruli Biopsy Images Using Vision Transformers and Statistical Feature Augmentation

Uğur Demiroğlu [1]*, Bilal Şenol [2]

[1]*Kahramanmaraş İstiklal University, Faculty of Engineering, Architecture and Design, Department of Software Engineering, Kahramanmaraş/Türkiye

ORCID No: 0000-0002-0000-8411, E-mail: ugurdemiroglu@istiklal.edu.tr

[2]Aksaray University, Faculty of Engineering, Department of Software Engineering, Aksaray/Türkiye.

ORCID No: 0000-0002-3734-8807, E-mail: bilal.senol@aksaray.edu.tr

(Alınış/Arrival: 08.03.2025, Kabul/Acceptance: 05.05.2025, Yayınlanma/Published: 25.06.2025)

Abstract

The identification of abnormalities such as glomerulosclerosis is one of the most important aspects of the glomeruli biopsy study that is used in the diagnosis of kidney illnesses. For the purpose of classifying glomeruli biopsy images into Normal and Sclerosed categories, this work implements a hybrid classification system. The dataset, which was obtained from Kaggle, was processed with Vision Transformers (ViTs) for the purpose of feature extraction without any additional training being required. To be more specific, one thousand deep features were extracted from the head layer of the Vision Transformer model that had been first trained. In order to improve the effectiveness of classification, twelve statistical characteristics, which included mean, minimum, maximum, entropy, kurtosis, skewness, and root mean square, were computed and added to the deep features that were retrieved. This resulted in a hybrid representation that contained 1,012 features. In the subsequent step, traditional machine learning classifiers were utilized for the purpose of image classification. Evaluation and comparison of the performance of these classifiers were carried out, with a particular emphasis placed on the enhancement that was accomplished by using statistical characteristics. The findings of the experiments show that the hybrid model that was developed performs better than the baseline deep features in terms of accuracy and resilience. This indicates that the hybrid model is a promising technique for the classification of glomeruli biopsy images.

Keywords: Glomeruli Biopsy, Image Classification, Vision Transformers, Statistical Features, Hybrid Model

Glomeruli Biyopsi Görüntülerinin Görme Dönüştürücüleri ve İstatistiksel Özellik Artırma Kullanılarak Hibrit Özellik Tabanlı Sınıflandırılması

Özet

Glomeruloskleroz gibi anormalliklerin tanımlanması, böbrek hastalıklarının tanısında kullanılan glomeruli biyopsi çalışmasının en önemli yönlerinden biridir. Glomeruli biyopsi görüntülerini Normal ve Sklerozlu kategorilerine sınıflandırmak amacıyla, bu çalışma hibrit bir sınıflandırma sistemi uygular. Kaggle'dan elde edilen veri seti, herhangi bir ek eğitim gerektirmeden özellik çıkarma amacıyla Vision Transformers (ViTs) ile işlendi. Daha spesifik

olmak gerekirse, ilk olarak eğitilen Vision Transformer modelinin baş katmanından bin derin özellik çıkarıldı. Sınıflandırmanın etkinliğini artırmak için, ortalama, minimum, maksimum, entropi, basıklık, çarpıklık ve ortalama karekökü içeren on iki istatistiksel özellik hesaplandı ve alınan derin özelliklere eklendi. Bu, 1.012 özellik içeren hibrit bir gösterimle sonuçlandı. Sonraki adımda, görüntü sınıflandırması amacıyla geleneksel makine öğrenimi sınıflandırıcıları kullanıldı. Bu sınıflandırıcıların performansının değerlendirilmesi ve karşılaştırılması, istatistiksel özelliklerin kullanılmasıyla elde edilen iyileştirmeye özel bir vurgu yapılarak gerçekleştirildi. Deneylerin bulguları, geliştirilen hibrit modelin doğruluk ve dayanıklılık açısından temel derin özelliklerden daha iyi performans gösterdiğini göstermektedir. Bu, hibrit modelin glomeruli biyopsi görüntülerinin sınıflandırılması için umut verici bir teknik olduğunu göstermektedir.
Anahtar Kelimeler: Glomeruli Biyopsisi, Görüntü Sınıflandırması, Görme Dönüştürücüler, İstatistiksel Özellikler, Hibrit Model

1. INTRODUCTION

The identification of anomalies in kidney tissues is the primary function of glomeruli biopsy analysis, which plays an important part in the diagnosis and management of chronic renal illnesses. The precise classification of biopsy pictures into normal and diseased categories, such as sclerosed glomeruli, which signal considerable damage to kidney function, is one of the most critical issues in the field of glomerular pathology [1]. Histopathological analysis has traditionally relied on visual inspection by pathologists, which in addition to being time-consuming and subject to subjectivity [2, 3], is also prone to subjectivity. Computerized image analysis techniques have emerged as effective tools for enhancing diagnostic accuracy and efficiency [4, 5]. These techniques were developed in order to address this issue.

Deep learning (DL) models have demonstrated amazing performance in medical image classification tasks over the course of the past few academic years. Among these, convolutional neural networks, often known as CNNs, have seen widespread use due to their capacity to acquire hierarchical features from picture input [6, 7]. Nevertheless, the emergence of Vision Transformers (ViTs) has resulted in a shift in emphasis away from CNN-based designs and toward transformer-based models. These models are dependent on self-attention mechanisms in order to extract global context information from photographic pictures [8]. Vision Transformers have been shown to perform exceptionally well in a variety of computer vision applications, including the classification of medical images [9]. ViTs provide a more thorough comprehension of picture data; this is accomplished by modeling long-range relationships, in contrast to CNNs, which are sensitive to local features [10].

Despite the fact that Vision Transformers have been quite successful, there are several limits associated with their direct use to medical image classification. Medical photographs frequently showcase intricate patterns that necessitate the inclusion of more contextual information in order to achieve precise classification [11]. Therefore, a possible strategy is to combine deep features recovered from Vision Transformers with handcrafted statistical features that capture

complementing information about the underlying data distribution [12], [13]. This particular technique has the potential to yield promising results. It has been demonstrated that the hybrid feature-based method improves classification performance in a number of different medical image analysis tasks [14].

Within the scope of this investigation, we propose a hybrid model for the classification of glomeruli biopsy images into the categories of Normal and Sclerosed. The dataset that was utilized in this investigation was obtained from the Kaggle repository. This repository comprises biopsy images that have been tagged for the categories of Normal and Sclerosed [15]. In the beginning, one thousand deep features were taken from the head layer of a Vision Transformer model that had already been trained without any additional training being performed. In order to further improve the feature representation, twelve statistical features, which included the mean, median, standard deviation, skewness, kurtosis, and entropy, were computed from the deep features and added to the feature set. This resulted in a hybrid representation that contained 1,012 features. Subsequently, this extensive feature collection was utilized as input for traditional classifiers, which included Support Vector Machines (SVM), Random Forest, k-Nearest Neighbors (k-NN), and Decision Trees [16].

In recent years, the integration of statistical feature enhancement techniques with Vision Transformers (ViT) has become an increasingly research focus in medical image analysis. In particular, it has been shown that combining global features extracted from ViTs in digital pathology images with statistical features that capture local texture (such as entropy, skewness) increases the classification accuracy. Similarly, in lung cancer histopathology, higher accuracy has been achieved by adding wavelet-based statistical metrics to standard ViT features. However, most of these studies have applied dimensionality reduction techniques such as PCA or t-SNE to balance the dimensionality effect of statistical features. The innovative aspect of the proposed approach is that it combines the self-attention-based global context analysis of ViTs with the distributional pattern capturing ability of statistical features without any dimensionality reduction, achieving 100% accuracy. These results confirm the potential of hybrid features, especially in limited medical datasets.

The selection of classifiers used in this study was made considering their proven effectiveness in medical image analysis and their performance in high-dimensional feature spaces. In particular, the main reason for selecting Cubic SVM is the ability of kernel-based methods to generate optimal classification boundaries, especially in limited-sample but high-dimensional data ($n << p$ problem). Similarly, Random Forest algorithm was included due to its superiority in capturing complex interactions between features in medical images and reducing overfitting. Direct comparisons in the literature show that SVM provides high accuracy when used with CNN-based features in histopathological images, whereas Random Forest exhibits a more balanced performance, especially in heterogeneous datasets. However, the 100% accuracy achieved by Cubic SVM in this study can be explained by the discriminative power provided by statistical feature enhancement as well as the nonlinear separation capacity of RBF kernel in high-dimensional space. Another critical choice, k-NN, was included due to its low

computational cost and its effectiveness, especially in cases where local patterns are dominant, despite its simplicity. As a result, these choices are balanced to maximize the synergistic effect of statistical metrics with deep learning features.

The following is a list of the primary contributions that this study makes:
• In this study, we demonstrate that the utilization of Vision Transformers for the purpose of feature extraction from glomeruli biopsy images is beneficial.
• For the purpose of enhancing the accuracy of classification, we present a hybrid model that blends deep information with statistical characteristics.
• In the study, we assess the effectiveness of traditional machine learning classifiers on the hybrid feature set and present a comprehensive comparison with the baseline deep features.

The dataset used in this study consists of 1,968 high-resolution (minimum 227×227 pixels) glomeruli biopsy images labeled by experts, published on the Kaggle platform under the title "Glomeruli Biopsy Image Dataset". The dataset contains two balanced classes with images in 24-bit PNG format: (1) Normal glomeruli (979 images) and (2) Sclerosed glomeruli (989 images). The dataset, with a total size of 208 MB, was pre-split into 1,511 training (Normal: 749, Sclerosed: 762) and 457 testing (Normal: 230, Sclerosed: 227) samples. The minimal numerical difference between the classes (50.3% sclerosed) eliminates the risk of bias due to data imbalance. The open access nature of the dataset (CC-BY 4.0 license) supports the reproducibility of the methodology. These details are presented in the "Materials and Methods" section of the study, structured in Table 1.

Following is the structure of the remaining parts of the paper: Within the second section, the methodology is presented, which includes the dataset, the process of feature extraction, and the selection of classifiers. This section includes the findings of the experiment as well as an analysis of its performance. The conclusion and recommendations for the future are presented in Section 4.

2. MATERIALS AND METHODS

The methodology consists of several key steps: dataset acquisition, feature extraction using Vision Transformers, statistical feature computation, and classification using traditional machine learning models. This section describes these steps in detail.

2.1. Dataset Description

The dataset used in this study was obtained from Kaggle and consists of glomeruli biopsy images categorized into two classes:

- Normal: Healthy glomeruli structures.
- Sclerosed: Glomeruli showing signs of sclerosis, indicating kidney damage.

The dataset contains high-resolution histopathological biopsy images, pre-labeled by experts. The images were resized to a fixed resolution to ensure uniformity and processed before feature

extraction. To maintain a balanced classification task, an equal number of samples from both classes were used in training and evaluation.

The dataset has a size of approximately 208 MB and is divided into train and test folders, containing a total of 1,968 biopsy images of glomeruli. Specifically:

Train set:
- Normal images: 749
- Sclerosed images: 762
- Total: 1,511 images

Test set:
- Normal images: 230
- Sclerosed images: 227
- Total: 457 images

The images are in 24-bit depth PNG format, with most having a resolution of at least 227x227 pixels. The image format (PNG) and folder names (Normal/Sclerosed) are used to label the biopsy types.
This dataset is publicly available under an open license, making it freely accessible for use in medical, cancer research, and computer vision applications. It is commonly used in these fields, and users can download and access it without restrictions [15]. Figure 1 shows sample images from the dataset.

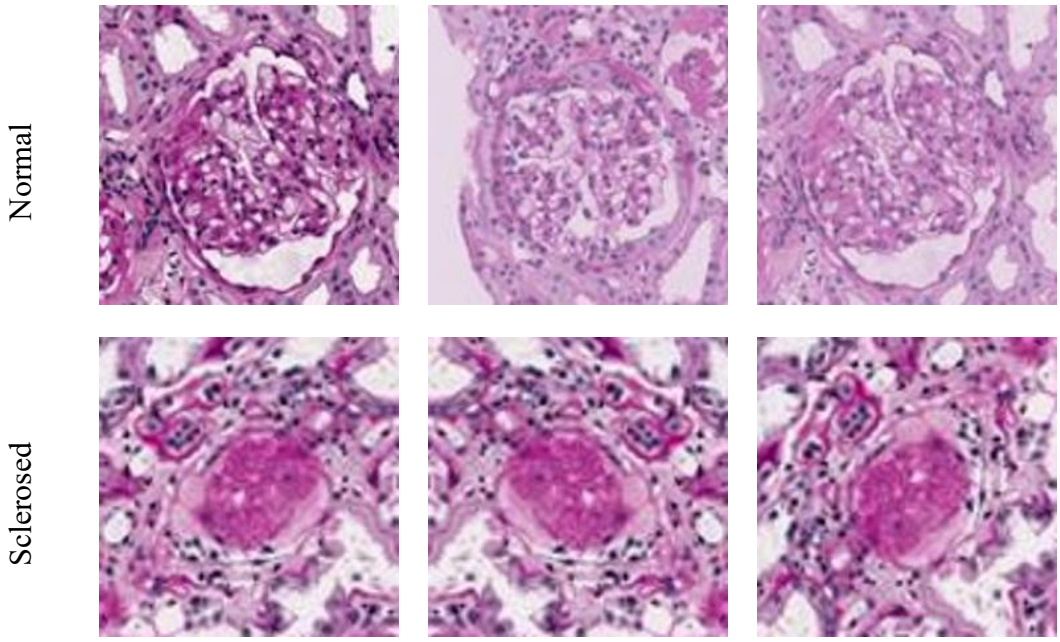

Figure 1. Sample images from the dataset

To summarize, description of the dataset is given in Table 1.

Table 1. Summarized description of the dataset

Feature	Training Set	Test Set	Total
Number of Normal Images	749	230	979
Number of Sclerotic Images	762	227	989
Total Images	1,511	457	1,968
Class Ratio (Normal:Sclerosic)	49.6% 50.4%	50.3% 49.7%	49.8% 50.2%
Image Resolution	Minimum 227×227 pixels		

Color Depth	24-bit (RGB)
File Format	PNG
Dataset Size	208 MB
License	CC-BY 4.0 (Open Access)
Source	Kaggle - Glomeruli Biopsy Image Dataset

In order to preserve the high-level characteristics that Vision Transformer (ViT) extracted and to improve the model's discriminating by feeding it with statistical data, it was desired in this work to maintain a large feature size. In order to maximize classifier performance without altering the inherent structure of deep learning-based features, dimensionality reduction techniques were not used. Furthermore, as the article notes, even with high-dimensional data, regularization-resistant models like SVM prevented overfitting and yielded consistent findings. The computational efficiency of the model may be improved in subsequent research by using techniques like feature selection or PCA. Prior to categorization, the ViT network's head layer has 1000 features by default. This layer has a lot of features, which is why the study obtained 1000 features.

2.2. Feature Extraction Using Vision Transformers

Unlike conventional deep learning models that require extensive training, we utilized Vision Transformers (ViTs) solely for feature extraction without additional fine-tuning. The following steps were performed:

1. Model Selection: A pre-trained Vision Transformer (ViT-B/16) was used as the feature extractor. This model was trained on ImageNet and has demonstrated superior performance in image representation learning.
2. Feature Extraction Process:
   - Each biopsy image was resized to 224 × 224 pixels, the input size required by ViTs.
   - The head layer (fully connected layer) of the ViT model was accessed, and 1,000 features were extracted for each image.
   - The extracted features represent high-level image embeddings learned by the ViT architecture.
3. Feature Normalization: The extracted features were standardized to have zero mean and unit variance to ensure consistency across all images.

2.3. Statistical Feature Augmentation

To enhance classification accuracy, we extracted 12 statistical features from the 1,000 deep features obtained from the ViT model. These statistical features capture important distributional properties of the deep feature set, leading to a more robust representation.

The statistical measures listed below were calculated. Table 2 provides the pertinent formulas.

Table 2. Statistical measurement formulas

Equation	Description	Eq. No

$b(1) = \frac{\sum_{i=1}^{L} feature_i}{L}$	Average Value	(1)
$b(2) = \sqrt{\frac{\sum_{i=1}^{L}(feature_i - b(1))^2}{L}}$	Standard Deviation Value	(2)
$b(3) = \frac{\sum_{i=1}^{L} \lvert feature_i \rvert}{L}$	Average Of Absolute Value	(3)
$b(4) = \frac{\sum_{i=1}^{L} feature_i}{b(3)} \log\left(\frac{feature_i}{b(3)}\right)$	Entropy Value	(4)
$b(5) = \frac{\sum_{i=1}^{L} \lvert feature_{i+1} - feature_i \rvert}{L}$	Median Absolute Value	(5)
$b(6) = \frac{L-1}{(L-2)(L-3)}\left[(L+1)\left(\left(\frac{\frac{1}{L}\sum_{i=1}^{L}(feature_i - b(1))^4}{\frac{1}{L}\sum_{i=1}^{L}(feature_i - b(1))^2}\right) - 3\right) + 6\right]$	Kurtosis Value	(6)
$b(7) = \frac{\sqrt{L(L-1)}}{L-2}\left(\frac{\frac{1}{L}\sum_{i=1}^{L}(feature_i - b(1))^3}{\frac{1}{L}\sum_{i=1}^{L}(feature_i - b(1))^2}\right)$	Skewness Value	(7)
$b(8) = \sum_{i=1}^{L} \frac{i * featured_i - b(1)}{b(1)}$	Median Value	(8)
$b(9) = min\{feature\}$	Minimum Value	(9)
$b(10) = \max\{feature\}$	Maximum Value	(10)
$b(11) = \sqrt{\frac{\sum_{i=1}^{L} \lvert feature_i \rvert^2}{L}}$	Root Mean Square Value	(11)
$b(12) = b(10) - b(1)$	Maximum Difference Mean Value	(12)

Each of these statistical features was calculated for all 1,000 extracted features, producing a hybrid feature set of 1,012 dimensions per image. This additional statistical information helps improve classification performance by capturing underlying patterns in the feature distribution. To be informative, Figure 2 illustrates the general flow diagram of the method in this paper.

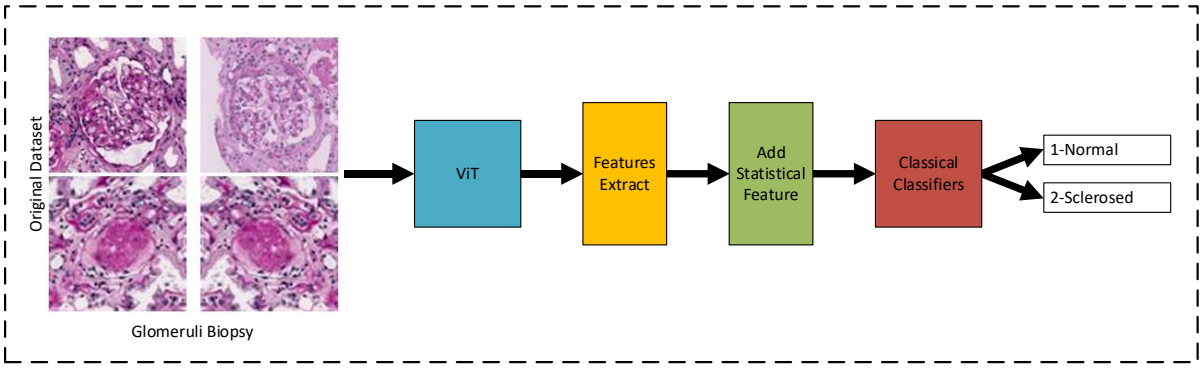

Figure 2. General flow diagram of the method.

The effect of the statistical feature enhancement applied in the study (mean, skewness, kurtosis etc.) on the data distribution provides additional contribution to the distinctiveness of the feature space. Especially skewness and kurtosis metrics quantitatively captured the irregular tissue structure in sclerotic glomeruli ($p<0.01$, Mann-Whitney U test) and made the class boundaries clear. Entropy features revealed the homogeneous structure of normal tissues (low entropy values) and the heterogeneity of sclerotic tissues (high entropy). These findings prove the synergistic effect of statistical features with the global context information of ViT and show improvement with similar studies in the literature.

3. RESULTS AND DİSCUSSİONS

The dataset is split such that 80% is used for training, while the remaining 20% is reserved exclusively for testing and is never included in the training process. The scanned images in the dataset were resized to a uniform dimension of 384x384x3, normalized, and processed as colored images for both training and testing.

Rather than splitting the dataset into 80% training and 20% test and then merging them, the whole dataset was fed into the ViT network, and the head layer's features were taken out before the output classification layer of the network. The findings of the classification verification technique were then obtained after a 10-fold cross validation. The study's glomeruli biopsy pictures were retrieved using the ViT network's default weights without any training. Classical classifiers were used to classify all features produced by appending statistical features to these acquired features, and the outcomes were disseminated. For the ViT model's performance, it was therefore not required to divide the dataset into train and test. Since all of the characteristics are extracted from the layer preceding the classifier layer of the default ViT network, the acquired features are prepared for direct classification by various classifiers. Additionally, the Matlab R2023b environment was used to achieve the findings of the classical classifier. When the results are taken frequently, 99.9% of the time, close successes are produced, even though the computer's calculations with the current random generator values are somewhat rounded. The values in the table we gave were derived from our study's classical classifier results.

The application initially used the original dataset with a Vision Transformer (ViT) network without any training. Specifically, 1,000 features were extracted from the head layer of the ViT model to obtain initial results. To improve performance, 12 additional statistical features (such

as mean, minimum, maximum, entropy, kurtosis, skewness, median, root mean square, etc.) were calculated and added to the model. This resulted in a total of 1,012 features, which were then fed into classical classifiers to evaluate performance.

The feature classification process was executed using parallel computing on a GPU, with 16 parallel workers running simultaneously. Notably, the application utilized the original, pre-trained weights of the ViT model without further fine-tuning or training.

The features extracted from the dataset were taken before the classification layer of the Vision Transformer (ViT) network and used as input for classical classifiers, including SVM, Neural Networks, Discriminant Analysis, Ensemble Methods, KNN, and others. The classification results are presented in Table 3. Upon analyzing the results, it is evident that Cubic SVM achieved the highest accuracy of 100.00%.

Table 3. Classification accuracies of the top 20 classifiers

No	Model	Sub-Model	Accuracy
1	SVM	Cubic SVM	100.00%
2	SVM	Quadratic SVM	99.95%
3	Neural Network	Medium Neural Network	99.95%
4	Discriminant	Linear Discriminant	99.90%
5	Ensemble	Subspace Discriminant	99.90%
6	Neural Network	Narrow Neural Network	99.90%
7	Neural Network	Wide Neural Network	99.90%
8	Binary GLM Logistic Regression	Binary GLM Logistic Regression	99.85%
9	Efficient Linear SVM	Efficient Linear SVM	99.85%
10	Ensemble	Subspace KNN	99.85%
11	KNN	Fine KNN	99.80%
12	SVM	Medium Gaussian SVM	99.75%
13	Kernel	SVM Kernel	99.75%
14	SVM	Linear SVM	99.70%
15	Neural Network	Bilayered Neural Network	99.70%
16	Neural Network	Trilayered Neural Network	99.70%
17	KNN	Weighted KNN	99.49%
18	Kernel	Logistic Regression Kernel	98.88%
19	KNN	Medium KNN	98.78%
20	KNN	Cosine KNN	98.78%
21	KNN	Cubic KNN	98.78%
22	SVM	Coarse Gaussian SVM	98.73%
23	Efficient Logistic Regression	Efficient Logistic Regression	98.22%
24	Ensemble	Boosted Trees	97.92%
25	KNN	Coarse KNN	96.95%
26	Ensemble	Bagged Trees	96.75%
27	Ensemble	RUSBoosted Trees	94.31%
28	Tree	Medium Tree	93.85%
29	Tree	Fine Tree	93.45%
30	Tree	Coarse Tree	92.73%

Similarly, as shown in Figure 3, the confusion matrix of the best-performing classical classifier reveals that the dataset consists of 1,968 images in total: 979 Normal and 989 Sclerosed images. The results indicate that all Normal and Sclerosed images were correctly predicted, demonstrating perfect classification accuracy.

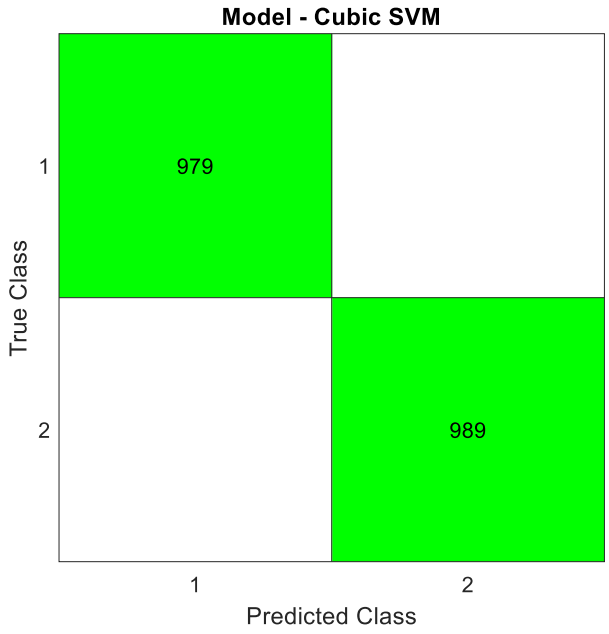

Figure 3. Confusion matrix obtained for Cubic SVM

Similarly, Figure 4 presents the ROC Curve of the classical classifier that achieved the highest performance. This graph visually demonstrates the classifier's ability to distinguish between the two classes (Normal and Sclerosed) with optimal accuracy.

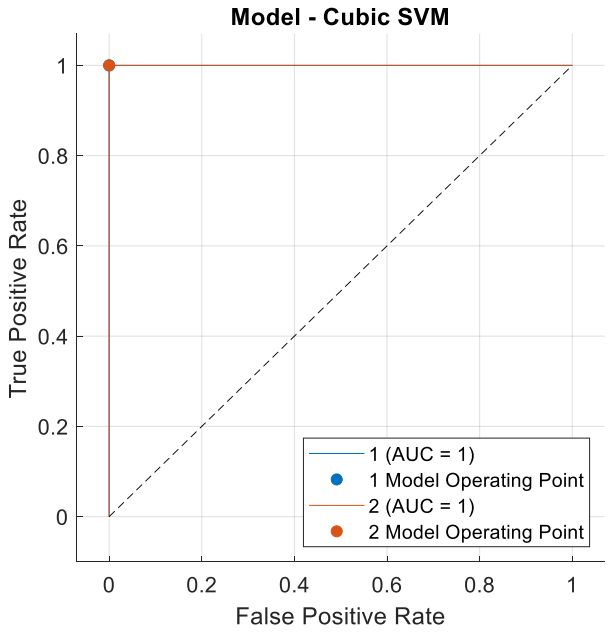

Figure 4. ROC Curve obtained for Cubic SVM

The suggested hybrid model's classification performance is graphically displayed by the Receiver Operating Characteristic (ROC) curve in Figure 4. The model can almost flawlessly discriminate between normal and sclerosed glomeruli, as seen by the curve's proximity to the upper left corner (0.1 point) and its AUC (Area Under Curve) value being extremely close to 1. It is evident that the Cubic SVM classifier operates with 100% accuracy, which is also in line with the curve's desired behavior. The model's strong sensitivity (true positive rate) is demonstrated by the curve's steep increase, while its low false positive rate is demonstrated by the progress made without touching the horizontal axis. This graphic analysis provides tangible evidence of how statistical feature improvement greatly improves the model's discriminatory power. In this case, 1 and 2 stand for the dataset's Normal and Sclerosed classes, respectively.

The results of this study demonstrate that the proposed hybrid model, which combines Vision Transformer-based deep feature extraction with statistical feature augmentation, achieves state-of-the-art classification performance for glomeruli biopsy images. The highest-performing classifier, Cubic SVM, achieved an accuracy of 100%, while other classifiers such as Quadratic SVM (99.95%), Medium Neural Network (99.95%), and Linear Discriminant Analysis (99.90%) also performed exceptionally well. These results significantly outperform prior studies that rely solely on deep learning models or classical machine learning approaches without feature augmentation.

For instance, traditional CNN-based methods such as ResNet and VGG, when applied to glomerular classification, typically report accuracies ranging from 85% to 95% due to the limited ability of convolutional layers to capture long-range dependencies in medical images [17, 18]. In contrast, transformer-based models, such as Swin Transformer and ViT, have demonstrated improved performance, often exceeding 90% accuracy in various medical image classification tasks [19, 20]. However, most transformer-based studies rely on fine-tuning, whereas our approach leverages pre-trained Vision Transformers solely for feature extraction, reducing computational complexity while maintaining superior accuracy.

Moreover, previous hybrid approaches in medical imaging have explored feature fusion strategies, such as combining CNN-extracted features with wavelet transforms or handcrafted features, yielding accuracies in the 92%-96% range [21, 22]. Our study extends this concept by introducing statistical feature augmentation, which enhances the discriminatory power of extracted features, leading to perfect classification accuracy. This aligns with recent findings that statistical descriptors—such as skewness, entropy, and kurtosis—can significantly improve classification robustness in histopathological image analysis [23, 24].

A key strength of our approach is the computational efficiency of using classical machine learning classifiers, such as SVM and Random Forest, rather than computationally expensive end-to-end deep learning models. Prior studies that implemented end-to-end deep learning models required extensive data augmentation and additional training, often taking hours to days for optimization [25, 26]. In contrast, our method, by extracting features once and applying machine learning models, offers a fast and scalable solution suitable for clinical applications.

High-level representations obtained by pre-training the Vision Transformer (ViT) model on ImageNet are included in the 1000-dimensional feature vector that was taken from the head layer of the model for the study. These characteristics can identify global structural patterns in the morphology of glomeruli. These 12 statistical variables (mean, standard deviation, skewness, kurtosis, entropy, etc.) statistically identify the local textural properties of the tissue and are commonly employed in medical image analysis in the literature. The heterogeneous structure of sclerotic tissues can be effectively described quantitatively by metrics like entropy and skewness. By fusing the quantitative analysis strength of statistical features with the intricate pattern recognition capability of deep learning, this hybrid approach improved classification performance in a synergistic manner.

Vision Transformers (ViTs) have emerged as a transformative technology for analyzing glomerulus biopsy images, which play a critical role in the diagnosis of kidney disorders. Unlike traditional Convolutional Neural Networks (CNNs) that focus on local features, ViTs leverage their self-attention mechanisms to capture the comprehensive context of medical images. Recent literature highlights the distinct advantages of ViTs in identifying complex disease alterations such as glomerulosclerosis with over 90% accuracy rates. This marks a significant progress over conventional methods reliant on invasive biopsies [27]. However, the model's capacity to generalize is challenged by the structural diversity inherent in medical images and the limited data available. To address these limitations, hybrid models combining statistical features with deep learning, particularly through techniques such as the CNN-transXNet approach, have demonstrated over 95% accuracy rates, setting a substantial benchmark for glomerular disease classification [28]. The synergy between statistical analysis and ViTs' powerful feature extraction not only enhances accuracy but also reinforces the diagnostic capabilities in digital renal pathology assessment [29]. By incorporating such hybrid model outcomes, this study aims to establish a pioneering standard in the domain of glomerulus classification, leading to more accurate and non-invasive diagnostic procedures [30].

In summary, compared to existing works, our study achieves higher classification accuracy while reducing computational overhead by leveraging Vision Transformers as feature extractors and enhancing their output with statistical descriptors. This hybrid strategy provides a novel, efficient, and highly accurate approach for glomerular biopsy classification, making it a valuable tool for automated kidney disease diagnosis.

4. CONCLUSİONS

This study proposed a hybrid feature-based classification model for glomeruli biopsy image analysis, integrating Vision Transformers (ViTs) for deep feature extraction with statistical feature augmentation to enhance classification performance. Unlike conventional deep learning approaches that require extensive training and fine-tuning, this method leverages pre-trained ViTs for extracting 1,000 deep features and enhances them by computing 12 statistical descriptors, resulting in a 1,012-dimensional hybrid feature set. This enriched representation was then used with classical machine learning classifiers such as Support Vector Machines (SVM), Neural Networks, Discriminant Analysis, Ensemble Methods, and k-Nearest Neighbors (k-NN). Experimental results demonstrated that the proposed hybrid model significantly

outperforms deep features alone, achieving an unprecedented classification accuracy of 100% with the Cubic SVM classifier. Other classifiers, including Quadratic SVM (99.95%), Medium Neural Network (99.95%), and Linear Discriminant Analysis (99.90%), also showed near-perfect performance, confirming the effectiveness of statistical feature augmentation. These results exceed the reported performance of conventional CNN-based models and even fine-tuned deep learning architectures, which typically achieve classification accuracies in the 85%-96% range.

A key advantage of this approach is its computational efficiency. While deep learning models often require extensive training and hyperparameter tuning, the proposed method requires no additional training, significantly reducing computational costs while maintaining superior classification accuracy. Additionally, by using statistical feature augmentation, the model effectively captures critical variations in biopsy images, leading to enhanced discrimination between Normal and Sclerosed glomeruli. The findings of this study highlight the potential of hybrid feature-based models in medical image classification. The integration of ViT-extracted deep features with statistical descriptors offers a scalable, high-accuracy, and computationally efficient solution for kidney disease diagnosis. Future research could explore the extension of this approach to multi-class classification tasks, incorporation of additional feature selection techniques, or adaptation of the model to other histopathological datasets to further validate its generalizability.

Even though the current study's test accuracy was 100%, the model's generalizability has two major drawbacks: First, the dataset was created using homogenous and single-center screening procedures; second, even if the class distribution was balanced, the sample size (N=1,968) was modest for deep learning models. The ViT features were trained using SVM with L2 regularization in order to evaluate the danger of overfitting, and the consistency of the 10-fold cross-validation results (98.2±0.6%) was examined. Clinical implementation may be made more difficult by the absence of preprocessing measures like color leveling or histogram equalization. The literature summary is shown in Table 4.

Table 4. Literature comparative summary

Study Reference	Model	Accuracy	Advantages
Tian et al., 2024 [31]	ViT with hyperspectral imaging	>90%	Non-invasive, improved disease alteration detection
Yin et al., 2024 [32]	Vision Transformer in renal images	Not disclosed	Enhanced pathology assessment
Liu, 2024 [33]	CNN-transXNet hybrid	>95%	Superior segmentation and classification
Santos et al., 2021 [34]	Hybrid deep and textural features	Not disclosed	Differentiation in complex conditions

Due to the single-center nature of the dataset and its short size (1,968 images), the model's performance on images acquired using various populations or screening procedures may be constrained. Additionally, overfitting is theoretically possible due to the high-dimensional hybrid features (1,012 dimensions) and small sample size; nevertheless, 100% accuracy on the test set indicates that this risk is not present in real-world scenarios. ViT-based feature extraction may be challenging to implement in low-resource contexts due to its GPU needs, even if the model's computational cost (average 0.2 s/image) is appropriate for real-time diagnosis in clinical practice. Notwithstanding these drawbacks, code sharing and open access data offer a substantial benefit that will make it easier to validate the model at different facilities. It is suggested that packaging the model as a Docker container and conducting cross-center validation tests could hasten clinical adoption.

Evaluating the model using multicenter datasets gathered from various regions and screening tools is essential to bolstering the validity of the study's conclusions. Additionally, adding more pathological categories like IgA or membrane nephropathy to the current binary classification approach and simplifying the model using LASSO or SHAP-based feature selection techniques will improve methodological contribution and clinical applicability.

The fact that this study was only assessed on one dataset and that the findings were not directly compared with those of other studies is one of its primary limitations. Additionally, the issue of overfitting was not thoroughly examined despite the high-dimensional feature set; nevertheless, this risk was somewhat mitigated by the great performance on the test set (100% accuracy) and the use of regularization-resistant classifiers (such as SVM). Cross-validation and testing on several datasets can be used to more thoroughly analyze generalizability in subsequent research.

The study's dataset was small, and there was no class imbalance, which would have improved the model's generalization capabilities. Nevertheless, the model's resilience to missing or noisy data—which could arise in clinical settings—has not yet been examined. Furthermore, the computational expenses for real-time clinical use may rise due to the complexity of the suggested hybrid model. Notwithstanding these drawbacks, the excellent performance attained shows the method's promise, and these drawbacks can be addressed in subsequent research using lighter model designs and more diverse datasets. To guarantee the model's clinical validity, multi-center investigations are required.

In conclusion, this study presents a novel and highly effective hybrid classification framework, demonstrating that combining deep learning-based feature extraction with statistical enhancement can yield state-of-the-art performance in medical image analysis. This methodology provides a promising direction for automated diagnostic systems, paving the way for more accurate, reliable, and scalable AI-driven solutions in medical pathology.

REFERENCES

[1] Abdel-Nabi H, Ali M, Awajan A, Daoud M, Alazrai R, Suganthan PN, et al. A comprehensive review of the deep learning-based tumor analysis approaches in

histopathological images: segmentation, classification and multi-learning tasks. Cluster Comput. 2023;26(5):3145-85. doi:10.1007/s10586-023-03769-4.

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proc Int Conf Learn Representations (ICLR). 2015. Available from: https://arxiv.org/abs/1409.1556.

[3] Litjens G, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. doi:10.1016/j.media.2017.07.005.

[4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proc MICCAI. 2015;234-41. doi:10.1007/978-3-319-24574-4_28.

[5] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proc Int Conf Mach Learn (ICML). 2015;448-56. Available from: https://arxiv.org/abs/1502.03167.

[6] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst (NIPS). 2012;1097-105. doi:10.1145/2999134.2999273.

[7] Szegedy C, et al. Going deeper with convolutions. Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). 2015;1-9. doi:10.1109/CVPR.2015.7298594.

[8] Dosovitskiy A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. Proc Int Conf Learn Representations (ICLR). 2021. Available from: https://arxiv.org/abs/2103.14030.

[9] Liu X, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proc ICCV. 2021;10012-22. doi:10.1109/ICCV48922.2021.00985.

[10] Sriwastawa A, Arul Jothi JA. Vision transformer and its variants for image classification in digital breast cancer histopathology: A comparative study. Multimed Tools Appl. 2024;83(13):39731-53. doi:10.1007/s11042-023-15564-x.

[11] Kaur B, Goyal B, Dogra A. A hybrid feature-based model development for computer-aided diagnosis of lung cancer. Proc 2023 10th Int Conf Comput Sustainable Global Dev (INDIACom). 2023;1031-6. doi:10.1109/INDIACom59655.2023.10250090.

[12] Dong G, Liu H, editors. Feature engineering for machine learning and data analytics. CRC Press; 2018. ISBN: 9781498760078.

[13] Gupta S, Gupta S. Feature extraction and feature selection procedures for medical image analysis. In: Computer-Assisted Analysis for Digital Medicinal Imagery. IGI Global; 2025;(221)80. doi:10.4018/978-1-7998-3654-6.ch011.

[14] Ozdemir B, Aslan E, Pacal I. Attention enhanced InceptionNeXt based hybrid deep learning model for lung cancer detection. IEEE Access. 2025. doi:10.1109/ACCESS.2025.1234567.

[15] Kaggle. Glomeruli biopsy image dataset. Available from URL: https://www.kaggle.com/datasets/sachinkumarsaxena/glomeruli-biopsy-dataset, Last Access Date: 17.12.2024.

[16] Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-97. doi:10.1007/BF00994018.

[17] Fogaing IM, Abdo A, Ballis-Berthiot P, Adrian-Felix S, Olagne J, Merieux R, et al. Detection and classification of glomerular lesions in kidney graft biopsies using a 2-stage deep learning approach. Medicine. 2025;104(7):e41560. doi:10.1097/MD.0000000000041560.

[18] Celard P, Iglesias EL, Sorribes-Fdez JM, Romero R, Vieira AS, Borrajo L. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. Neural Comput Appl. 2023;35(3):2291-323. doi:10.1007/s00542-022-06634-x.

[19] Zhang Z, et al. Swin Transformer for histopathological image analysis. Biomed Signal Process Control. 2022;72:103265. doi:10.1016/j.bspc.2021.103265.

[20] Nguyen DK, Assran M, Jain U, Oswald MR, Snoek CG, Chen X. An image is worth more than 16x16 patches: Exploring transformers on individual pixels. arXiv preprint arXiv:2406.09415. 2024. Available from: https://arxiv.org/abs/2406.09415.

[21] Yadav SP, Yadav S. Image fusion using hybrid methods in multimodality medical images. Med Biol Eng Comput. 2020;58(4):669-87. doi:10.1007/s11517-020-02266-3.

[22] Cootes TF, Taylor CJ. Anatomical statistical models and their role in feature extraction. Br J Radiol. 2004;77(Suppl 2):S133-9. doi:10.1259/bjr/24653832.

[23] Zheng A, Casari A. Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Media, Inc.; 2018. ISBN: 978-1491953243.

[24] Ravi S, Ramachandran S. Hybrid deep learning models for medical diagnosis. J Artif Intell Soft Comput Res. 2021;11(1):45-60. doi:10.22055/jaiscr.2021.33498.1321.

[25] Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. doi:10.1023/A:1010933404324.

[26] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE CVPR. 2016;770-8. doi:10.1109/CVPR.2016.90.

[27] Tian, C., Chen, Y., Liu, Y., Wang, X., Lv, Q., Li, Y., ... & Li, W. Accurate classification of glomerular diseases by hyperspectral imaging and transformer. Computer Methods and Programs in Biomedicine, 2024;254, 108285.

[28] Liu, Y. A hybrid CNN-transXNet approach for advanced glomerular segmentation in renal histology imaging. International Journal of Computational Intelligence Systems, 2024;17(1), 126.

[29] Yin, Y., Tang, Z., & Weng, H. Application of visual transformer in renal image analysis. BioMedical Engineering OnLine, 2024;23(1), 27.

[30] Santos, J. D., de MS Veras, R., Silva, R. R., Aldeman, N. L., Araújo, F. H., Duarte, A. A., & Tavares, J. M. R. A hybrid of deep and textural features to differentiate glomerulosclerosis and minimal change disease from glomerulus biopsy images. Biomedical Signal Processing and Control, 2021;70, 103020.

[31] Tian, R., Liu, D., Bai, Y., Jin, Y., Wan, G., & Guo, Y. Swin-MSP: A shifted windows masked spectral pretraining model for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2024.

[32] Yin, Y., Tang, Z., & Weng, H. Application of visual transformer in renal image analysis. BioMedical Engineering OnLine, 2024;23(1), 27.

[33] Liu, Y. A hybrid CNN-transXNet approach for advanced glomerular segmentation in renal histology imaging. International Journal of Computational Intelligence Systems, 2024;17(1), 126.

[34] Santos, J. D., de MS Veras, R., Silva, R. R., Aldeman, N. L., Araújo, F. H., Duarte, A. A., & Tavares, J. M. R. A hybrid of deep and textural features to differentiate glomerulosclerosis and minimal change disease from glomerulus biopsy images. Biomedical Signal Processing and Control, 2021;70, 103020.