





## Article

# An Automated Grading System Based on Topological Features for the Evaluation of Corneal Fluorescein Staining in Dry Eye Disease

Jun Feng <sup>1,†</sup>, Zi-Kai Ren <sup>2,†</sup>, Kai-Ni Wang <sup>2</sup> , Hao Guo <sup>2</sup> , Yi-Ran Hao <sup>1</sup>, Yuan-Chao Shu <sup>3</sup>, Lei Tian <sup>1</sup>, Guang-Quan Zhou <sup>2,\*</sup>  and Ying Jie <sup>1,\*</sup> 

<sup>1</sup> Beijing Ophthalmology and Visual Science Key Lab, Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China

<sup>2</sup> School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China

<sup>3</sup> College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

\* Correspondence: guangquan.zhou@seu.edu.cn (G.-Q.Z.); jie\_yingcn@aliyun.com (Y.J.); Tel.: +86-025-83790731 (G.-Q.Z.); +86-010-58268240 (Y.J.); Fax: +86-025-83792246 (G.-Q.Z.); +86-010-58268240 (Y.J.)

† These authors contributed equally to this work.

**Abstract:** Background: Corneal fluorescein staining is a key biomarker in evaluating dry eye disease. However, subjective scales of corneal fluorescein staining are lacking in consistency and increase the difficulties of an accurate diagnosis for clinicians. This study aimed to propose an automatic machine learning-based method for corneal fluorescein staining evaluation by utilizing prior information about the spatial connection and distribution of the staining region. Methods: We proposed an end-to-end automatic machine learning-based classification model that consists of staining region identification, feature signature construction, and machine learning-based classification, which fully scrutinizes the multiscale topological features together with conventional texture and morphological features. The proposed model was evaluated using retrospective data from Beijing Tongren Hospital. Two masked ophthalmologists scored images independently using the Sjögren's International Collaborative Clinical Alliance Ocular Staining Score scale. Results: A total of 382 images were enrolled in the study. A signature with six topological features, two textural features, and two morphological features was constructed after feature extraction and selection. Support vector machines showed the best classification performance (accuracy: 82.67%, area under the curve: 96.59%) with the designed signature. Meanwhile, topological features contributed more to the classification, compared with other features. According to the distribution and correlation with features and scores, topological features performed better than others. Conclusions: An automatic machine learning-based method was advanced for corneal fluorescein staining evaluation. The topological features in presenting the spatial connectivity and distribution of staining regions are essential for an efficient corneal fluorescein staining evaluation. This result implies the clinical application of topological features in dry-eye diagnosis and therapeutic effect evaluation.

**Keywords:** dry-eye disease; corneal fluorescein staining; topological features; machine learning



**Citation:** Feng, J.; Ren, Z.-K.; Wang, K.-N.; Guo, H.; Hao, Y.-R.; Shu, Y.-C.; Tian, L.; Zhou, G.-Q.; Jie, Y. An Automated Grading System Based on Topological Features for the Evaluation of Corneal Fluorescein Staining in Dry Eye Disease. *Diagnostics* **2023**, *13*, 3533. <https://doi.org/10.3390/diagnostics13233533>

Academic Editor: Stanislav Makhanov

Received: 8 October 2023

Revised: 18 November 2023

Accepted: 24 November 2023

Published: 26 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Dry eye disease (DED) is a multifactorial condition affecting the ocular surface, characterized by a loss of homeostasis of the tear film and accompanied by ocular symptoms, in which ocular surface damage plays an etiological role [1]. The prevalence of DED ranges between 5% and 50% worldwide [2]. To diagnose DED, careful assessment of ocular surface damage is essential. Recommended assessments include corneal, conjunctival, and lid-margin staining, with corneal staining being the most significant [3]. Sodium fluorescein is a common dye for corneal staining, as it can enter sick cells and be transported across altered tight junctions and micro pool at the space left by a shed epithelial cell [4].

Several subjective grading scales have been developed to evaluate corneal fluorescein staining (CFS), including the National Eye Institute (NEI)/Industry scale [5], the Oxford scale [6], and the Sjögren's International Collaborative Clinical Alliance Ocular Staining Score (OSS) scale [7]. The NEI scale divides the cornea into five zones (central, superior, inferior, nasal, and temporal) and assigns a grade from 0 to 3 to each zone. In contrast, the Oxford and OSS scales examine the entire cornea, using different criteria to evaluate staining regions. The Oxford scale employs illustrative drawings to grade corneal staining from 0 to 5, while the OSS scale counts the number of staining dots and assigns a score between 0 and 3, with an additional point for special types of dot distributions. However, there is little consensus on the features included in these scales, as they differ in their procedures and standards [8]. The variety of grading scales can be confusing for ophthalmologists and time consuming in clinical practice, and subjectivity leads to interobserver and intraobserver variability.

Therefore, computer-aided diagnosis (CAD) has made progress in evaluating CFS in recent years. Rodriguez et al. [9] developed a semi-automatic method by manually identifying staining regions using three seed points in a corneal area. Similarly, Pellegrini et al. [10] manually traced the corneal area with the ImageJ 1.51s software for CFS grading. In contrast, Daugman's operator was employed for locating the cornea, followed by an Otsu method for automatically identifying staining regions [11]. Bagbaba et al. [12] also developed a method that automatically segmented staining regions using connected component labeling. Kourukmas manually positioned the cornea and segmented the stained area with the help of ImageJ software [13]. Rather than directly predicting the CFS result, all these methods finally attempted to explore the correlation between the corresponding grade and the image-based parameters of the staining regions, such as the number of corneal staining regions and the ratio of the staining area to the total corneal area. Recently, data-driven deep neural networks (DNNs) have been used to segment staining regions or to examine the grading scale directly from the images [14–18], producing promising results in grading CFS.

The emergence of computer-aided solutions for evaluating CFS can provide objective measurements and facilitate the quantitative grading of dry eyes. However, previous studies have primarily focused on the correlation of CFS with the number or area of the staining regions [9–12,18], neglecting the distribution and texture information from these regions. This approach is inadequate for complicated corneal staining images that involve confluent staining and filaments. Although data-driven solutions such as deep learning show promising classification accuracy, their lack of interpretability restricts their usage in clinics [14–17]. Moreover, previous research has ignored spatial position information, despite the variations in the closeness and distribution between staining regions in different grades' images, indicating the potential use of the topological theory in CFS evaluation. To the best of our knowledge, the topological features have not been applied in corneal-staining evaluation.

Therefore, we considered whether topological attributes might uncover a more efficient underlying differentiating mechanism between different staining grades. We proposed an automated machine learning-based method for evaluating corneal fluorescein staining in dry-eye disease, which includes a set of multiscale topological features designed to capture spatial connectivity and distribution information between staining regions.

## 2. Materials and Methods

### 2.1. Data Set

The retrospective, single-center research was conducted at Beijing Tongren Hospital, Capital Medical University. Patients diagnosed with dry eye based on the Dry Eye Workshop II criteria between 2021 to 2022 were included [19]. Patients with a history of ocular injury or infection, contact lens wear within the last 6 months, surgery within the previous 12 months, or systemic autoimmune diseases such as Sjögren's syndrome and systemic

lupus erythematosus were excluded. The study adhered to the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants.

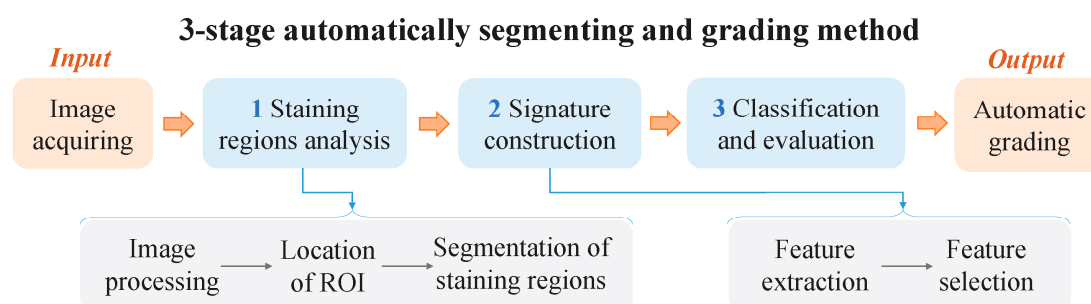
All corneal images were acquired by a single technician using a photo slit-lamp system (BX 900, Haag-Streit, Bern, Switzerland) with 10× magnification. The cornea was stained with fluorescein by introducing a wet fluorescein sodium ophthalmic strip (Tianjin Jingming New Technological Development Co., Tianjin, China) into the inferior fornix, and participants were instructed to blink [19]. The camera settings were constant for all participants, and images of the entire cornea were obtained using the blue filter and the barrier yellow filter to observe corneal staining.

## 2.2. Investigator Grading System

Two experienced ophthalmologists assessed corneal staining images independently. The results obtained by the two ophthalmologists were compared and discrepancies were resolved with another ophthalmologist. In dry-eye disease, the inferior zone typically presents the most severe grade of corneal staining. Compared to other areas, the inferior zone has a higher density of staining and provides more features [20]. We assumed that the center of the cornea was located at the middle of pupil, and the distance from the center to the bottom edge of the cornea was regarded as the radius. Therefore, the region of interest (ROI) was located at the lower sector areas, with an included angle of 90°. The ROI in the inferior cornea was finally graded according to the OSS scale. The OSS score was calculated as follows: 0, no staining; 1, one to five staining dots; 2, six to 30 staining dots; 3, more than 30 staining dots [7]. An additional point was added if: (1) there were one or more patches of confluent staining, (2) there were one or more filaments, or (3) staining occurred in the central cornea. As the ROI was the inferior zone of the cornea, the maximum score for each ROI was 5 in our study. The final OSS score was regarded as the true label.

## 2.3. Staining Region Analysis

As shown in Figure 1, the stage of staining region analysis was developed to extract the regions of staining. The extraction process consisted of three primary steps: image preprocessing, identification of the region of interest (ROI) in the cornea, and segmentation of staining regions within the ROI. The precise determination of both the ROI and staining regions was critical for achieving optimal performance of the automated grading system.

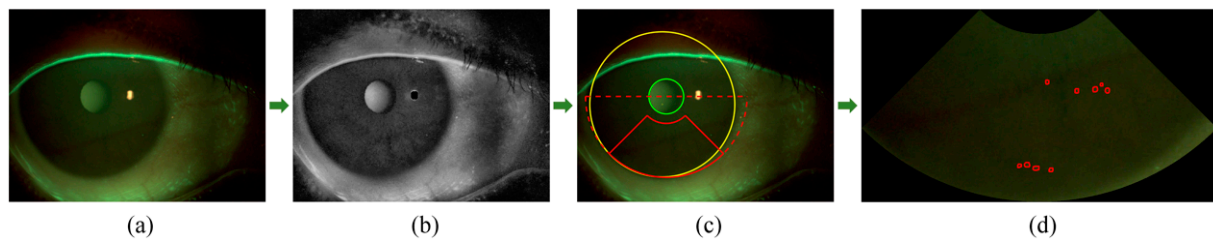


**Figure 1.** Overview of our proposed automatic segmenting and grading method for corneal fluorescein staining evaluation.

### 2.3.1. Image Preprocessing

To improve the clarity of corneal edges and eliminate reflective areas, we applied a series of image preprocessing operations before determining the ROI. In cases where low-contrast images were obtained, likely due to variations in the light environment during image capture, we employed image enhancement methods to increase the visibility of corneal edges. First, we extracted the red channel from the input color image to enhance the edges and minimize the possible impact of staining (Figure 2a). We then utilized median filtering and contrast-limited adaptive histogram equalization (CLAHE) to decrease noise and improve image contrast [11,21], respectively. Additionally, to address potential

reflective regions in the images, Otsu thresholding was applied to detect and eliminate the brightest areas (Figure 2b) [22].



**Figure 2.** Representative result of image digital analysis: (a) original image; (b) image after preprocessing; (c) location of the region of interest (ROI) (red lines), with the cornea in yellow lines, the pupil in green lines and the semicircle applied to determine ROI in red dashed lines; (d) segmentation of staining regions in the ROI (red lines).

### 2.3.2. Location of the ROI

Following the preprocessing procedures, we identified the ROI by first detecting the cornea and the pupil. We used Otsu thresholding to binarize and segment the sclera areas, which were outside and brighter than the cornea. We then determined the center and radius of the cornea by identifying the largest diameter in the inner edges of the binarized sclera, corresponding to the circular edges of the cornea. Daugman's integrodifferential operator was subsequently used to detect the pupil within the cornea [23]. Specifically, we searched for the largest outward radial gradient of the grey intensity of the circular ring under any proposed combination of the center  $(x_0, y_0)$  and radius  $r$ , consisting of the circle of the pupil. Daugman's equation, as presented in Equation (1), was used to compute the pupil's position. Once the cornea and the pupil were detected, the ROI's location was determined as the lower sector area with an included angle of  $90^\circ$ . Figure 2c shows the detected cornea, pupil, and ROI.

$$\max_{(r, x_0, y_0)} \left| G_\sigma(r) \times \frac{\partial}{\partial r} \int \frac{I(x, y)}{2\pi r} ds \right| \quad (1)$$

### 2.3.3. Segmentation of Staining Regions in the ROI

To segment the staining regions of the dots or patches in the located ROI image, we employed the top-hat by reconstruction operation in morphology [24]. Specifically, we first extracted the green channel to highlight the green staining regions and then performed an open operation to eliminate small bright details. Next, the ROI image was reconstructed in grayscale using the result from the open operation as a template, which effectively removed the highlights while preserving the background. Finally, the staining regions were segmented by taking the difference between the ROI and the reconstructed result (Figure 2d). Compared with the commonly used threshold segmentation methods, such as Otsu, the top-hat by reconstruction operation showed superior detection performance to staining regions under different contrasts.

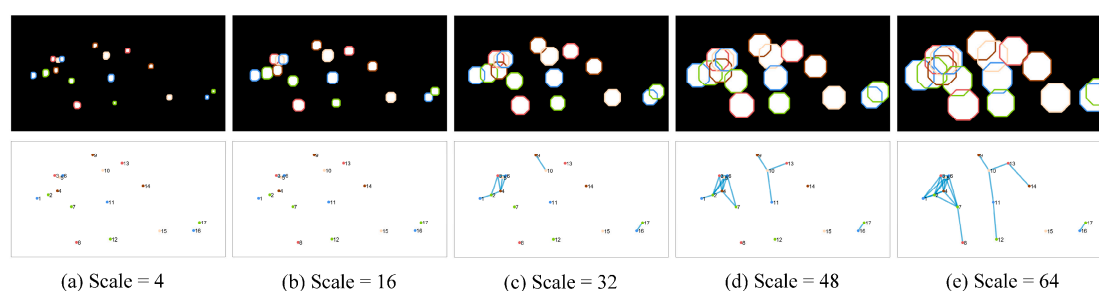
### 2.4. Signature Construction

After segmenting the desired stained regions, we proceeded with signature construction, which involved feature extraction and selection to meet the requirements of the classification model. In this stage, we extracted conventional radiomic features, including textural and morphological features. Additionally, we developed a set of multiscale topological features to uncover spatial connectivity and distribution information between staining regions [25]. These additional features aimed to capture the intricate relationships and patterns among the stained regions.

### 2.4.1. Feature Extraction

Radiomic features, encompassing both textural and morphological characteristics, were calculated, as the previous study mentioned [26,27]. A total of 837 texture features were extracted to analyze the background grayscale textural and pattern information of ROI images. Specifically, this included 18 first-order histogram statistical features and 75 second-order gray matrix features, such as autocorrelation and cluster prominence in the gray-level co-occurrence matrix (GLCM), the gray-level non-uniformity and gray-level non-uniformity normalized in gray-level size zone matrix (GLSZM), the gray-level variance and high gray-level run emphasis in gray-level run length matrix (GLRLM), and so on. These features were calculated in nine forms of images, respectively [26]. In addition to the original image, three image transformation operations were performed to enrich the textural information of the ROI. These operations included Laplacian of Gaussian filtering (LoG) ( $\sigma = 1, 2, 3$ ), four components of wavelet decomposition, and local binary patterns (LBPs). Furthermore, to avoid ignoring the information from staining regions, nine morphological features were extracted by calculating the shape characteristics, such as circularity and the area of each segmented staining region [10].

To capture the spatial connectivity and distribution information of staining regions, a set of multiscale graph-theoretic topological features were developed. These features aimed to address characteristics like regional clustered distribution resulting from confluent staining, which cannot be accurately described by textural and morphological features alone [25]. The approach involved considering the centroid of each staining region as the vertex in a graph and estimating the connectivity between staining regions using dilation operations at multiple scales. When two staining regions overlapped after dilation, their nodes were connected by an edge, thus generating a topological graph that reflected the connectivity. As the scale of dilation increased from small to large, the number of edges and the connectivity of the graph exhibited a positive correlation. The construction process of the topological graph is shown in Figure 3. Eight topological features, including vertex degree and eccentricity, were calculated at each scale. These calculations were performed using different disk dilation operators, ranging in size from 4 to 64. A total of 128 multiscale topological features were extracted in 16 scales. The texture features we extracted were the same as those in the literature, and the morphological and topological features extracted are shown in Table 1 [25].



**Figure 3.** Representative process of constructing topological graphs. The scales shown from column (a) to column (e) are 4, 16, 32, 48, and 64. Top row: the results of dilation on the staining regions at different scales; bottom row: the results of the constructed topological graph at corresponding scales.

### 2.4.2. Feature Selection

In order to handle the large number of features (974 normalized features) obtained after calculating all three kinds of features, a four-step feature selection method was employed for signature construction. This approach aimed to reduce the dimensionality of the features set and prevent overfitting of the model without compromising its performance. One-way analysis of variance (ANOVA) and a Pearson redundancy-based filter (PRBF) were applied to eliminate redundant terms in the feature vector [28]. Subsequently, a backward feature selection method based on the linear regression model was used to remove non-significant



features [29]. The final selection for important features was determined using a decision-tree method [30], where the importance of each feature was measured by the gini impurity at each node. Higher importance indicated stronger separability of the feature across different classes of the model. Multiple alternative signatures were constructed, based on different combinations of the importance ranking and feature categories.

**Table 1.** Extracted topological features.

Feature Categories	Feature Names
Topological features	Number of subgraphs
	Average vertex degree
	Maximum vertex degree
	Average vertex eccentricity
	Diameter
	Average clustering coefficient
	Giant connected component ratio
	Percentage of isolated points
Morphological features	Mean area
	Total area
	Mean perimeter to area ratio
	Total perimeter to area ratio
	Mean circularity
	Mean perimeter
	Total perimeter
	Minimum external rectangle aspect ratio
	Number

## 2.5. Classification and Evaluation

The developed machine learning-based model aimed to classify corneal staining images into different grades, based on selected features. To determine the optimal approach for our auto-grading method and evaluate the effect of different categories of features, we extensively explored the performance of different combinations of signatures and models. Specifically, four groups of features were applied in the same support vector machine (SVM) model in order to determine the one that was associated with best performance and accept it as the final signature. Additionally, different combinations of topological and non-topological features (textural and morphological) were investigated to evaluate the contributions of topological features in the classification process. This included the overall top-10 important features, the top-10 important non-topological features, and the top-5 important topological features and non-topological features, respectively. These resulting signatures were then accessed using five kinds of machine learning-based classification models, namely, SVM, Naive Bayes (NB), Decision Tree (DT), Boosting Tree (BT), K-nearest neighbors (KNN), and Random Forest (RF). In our experiments, the original dataset was divided into a training dataset (307 images) and a testing dataset (77 images), in an 8:2 ratio. A ten-fold cross-validation method was used to train and validate each model on the training dataset, while the classification performance of the model was evaluated with the testing dataset.

## 2.6. Statistical Analysis

An intraclass correlation coefficient (ICC) was calculated to assess the interobserver repeatability between two resident ophthalmologists;  $ICC \geq 0.8$  indicated good reliability. Accuracy (ACC), sensitivity (SEN), precision (PRE), specificity (SPE), F-measure of the confusion matrix, and the area under the receiver operating characteristic (ROC) curve (AUC) were calculated to evaluate the performance of the proposed method. The confusion matrix showed the predicted label and true label of each class, and the ACC was the ratio of the total number of correctly classified samples to the total number of samples. Due to the imbalance between different classes in the dataset, the F-measure was also used to evaluate

the harmonic mean of precision and sensitivity. The ROC is a function of sensitivity and specificity, which was established by changing the threshold of the classifier. Generally, AUC was used to evaluate the accuracy of the classification model. The larger the AUC, the better the performance. Kendall's tau-b correlation coefficient ( $r_K$ ) was used to evaluate the correlation between selected features and true labels. All experiments were two-sided tests;  $p < 0.05$  was considered to be statistically significant and  $p < 0.01$  was considered to be statistically extremely significant.

### 2.7. Implementation Details

The detection and classification experiments in this study were conducted on a personal computer with an Intel Core i9-9900K, 3.60 GHz, and 16 cores. The programming language used for the implemented algorithms was MATLAB software (version R2022a, RRID:SCR\_001622). All input images were standardized to a size of  $1944 \times 2592$ , and the extracted ROI size was normalized to  $596 \times 1104$ . In the preprocessing phase, median filtering was applied with a neighborhood size of  $21 \times 21$ . For pupil detection using Daugman's integrodifferential operator, the preselected range for the pupil center ( $x_0, y_0$ ) was a rectangle with the corneal center as the center and 0.35 times the corneal radius as the length. The preselected range for the pupil radius ( $r$ ) was from 0.15 times to 0.35 times the length of the corneal radius. The calculation angle range for the integrodifferential operator was from 0 degrees to 360 degrees. In the reconstruction of the top-hat operation, a circular structural element with a radius of 10 was used. For machine learning model training, the dataset was divided into a training set and a test set in an 8:2 ratio. Model parameters were optimized using the training set, and the model's performance was evaluated using the test set. For the SVM model, the kernel function was a Gaussian kernel. The DT model had a maximum split criterion of 4 at each node. The boosting tree model underwent 30 training cycles. The KNN model had a neighborhood size of 10.

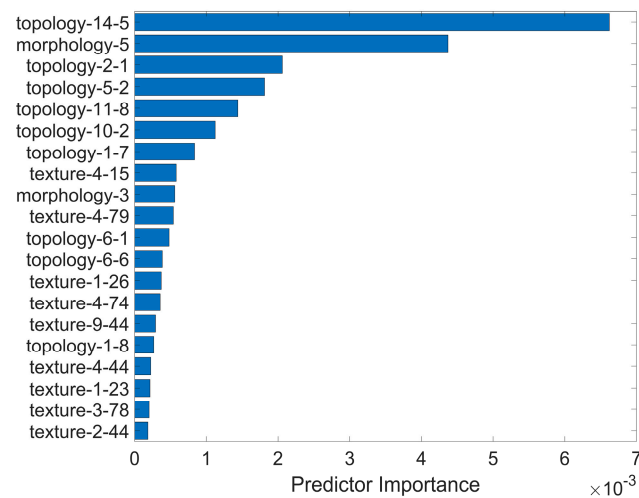
## 3. Results

### 3.1. Analysis of Dataset and Subjective Grades

A total of 421 images were acquired and 382 images were enrolled in the study, excluding blurred or poor-exposure images. No image was observed with filament, allowing for the implementation of a five-category model ranging from 0 to 4. There were 41 images (10.7%) with grade 0; 57 images (14.9%) with grade 1; 112 images (29.3%) with grade 2; 59 images (15.5%) with grade 3; and 113 images (29.6%) with grade 4. The median score was 2 (range, 0–4). The ICC values between the two ophthalmologists were 0.789 (95% confidence interval: 0.562–0.881,  $p < 0.001$ ).

### 3.2. Selection and Evaluate of Features

In this study, 992 initial features were extracted from the ROI and the staining regions of each image. These included 837 textural features, 9 morphological features, and 128 topological features. Specifically, texture features were directly computed within the localized ROI image, while morphological and topological features involved operations on the detected stained regions. Morphological features calculated information for each connected component of the stained regions, while topological features were evaluated from the graph structure constructed from the overall results of the stained region detection. During the process of feature selection, ANOVA retained 556 features, PRBF preserved 199 features, and, finally, backward feature selection based on a linear regression model yielded 44 features. Additionally, the decision-tree feature selection method predicted the importance of features in classification. The top-20 important features are shown in Figure 4, demonstrating that six out of the top seven ranked features were topological features.



**Figure 4.** The top-20 importance ranking list for the selected features.

The top-20 important features are shown in Figure 4. We evaluated the performance of four groups of features with an SVM model on the training (10 FCV) and testing datasets, respectively. The classification results are summarized in Table 2. The overall top-10 important features provided the best performance, with an ACC of 82.67% and an AUC of 96.59%. These results were significantly higher than those of the other feature groups. Similarly, the topological top-5 important features exhibited superior performance (ACC: 74.67%, AUC: 95.36%) compared to the non-topological top-5 (ACC: 64.00%, AUC: 84.14%). This finding suggests that topological features made a more significant contribution than other types of features.

**Table 2.** Classification results of different groups of features in SVM.

Feature Groups	ACC	SEN	PRE	SPE	F-M	AUC
The overall top-10 important features	82.67	83.71	80.91	80.91	81.80	96.59
The top-10 important non-topological features	64.00	53.53	57.88	57.88	54.22	90.64
The top-5 important topological features	74.67	65.82	70.91	70.91	67.60	95.36
The top-5 important non-topological features	64.00	46.48	56.36	56.36	49.90	86.14

ACC, accuracy; SEN, sensitivity; PRE, precision; SPE, specificity; F-M, F-measure; AUC, area under the receiver operating characteristic curve.

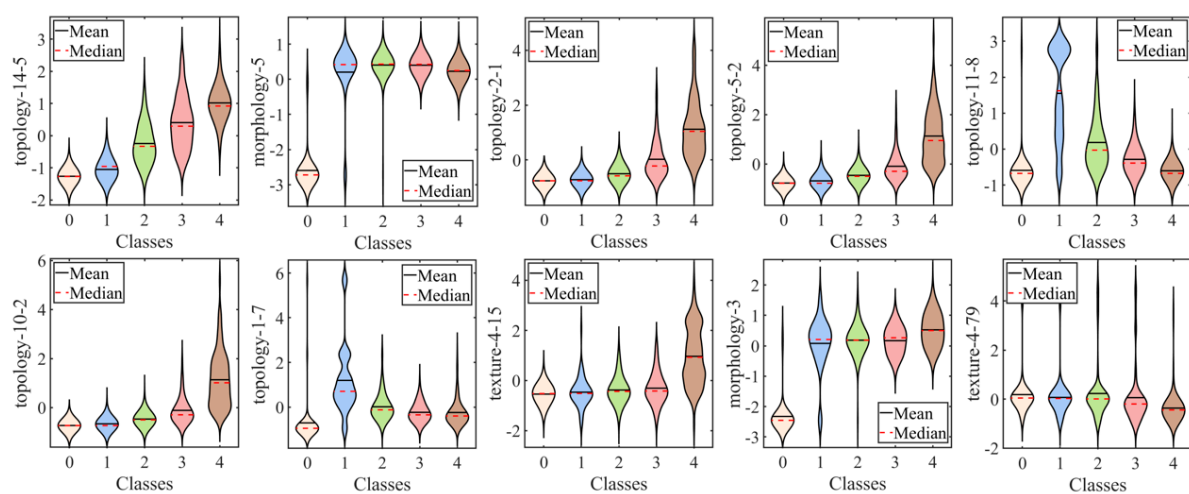
The overall top-10 important features consisted of two textural, two morphological, and six topological features. Their explanations are shown in Table 3. For more detail, the distribution of each feature in the five classes is presented in Figure 5. The correlation between features and true labels is shown in Table 4. All topological features except “topology 11-8” and “topology 1-7” were strongly correlated to true labels ( $r_K > 0.7$ ,  $p < 0.001$ ).

Since some features in score 0 images had zero values, which may interfere with the potential correlation, the result of removing the score 0 images is also shown in Table 4. Please note that “topology-11-8” showed a negative relationship ( $r_K = -0.6$ ,  $p < 0.001$ ).



**Table 3.** Explanation of the top-10 important features.

Feature Names	Feature Explanations
Topology-14-5	The diameter in topological features of scale 56
Morphology-5	The mean circularity in morphological features
Topology-2-1	The number of subgraphs in topological features of scale 8
Topology-5-2	The average vertex degree in topological features of scale 20
Topology-11-8	The percentage of isolated points in topological features of scale 44
Topology-10-2	The average vertex degree in topological features of scale 40
Topology-1-7	The giant connected component ratio in topological features of scale 4
Texture-4-15	The skewness in textural features of LOG transform of sigma 3
Morphology-3	The mean area perimeter ratio in morphological features
Texture-4-79	The large area low gray-level emphasis in textural features of GLSZM of LOG transform of $\sigma = 3$

**Figure 5.** The violin plots of the top-10 important features.**Table 4.** The correlation between the top-10 important features and true labels.

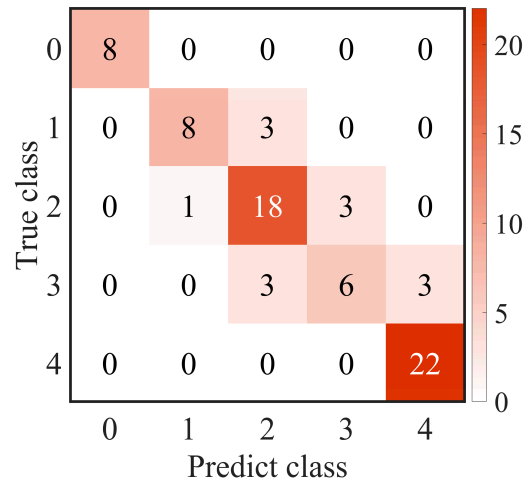
Feature Names	OSS 0 to 4		OSS 1 to 4	
	$r_K$	$p$ Value	$r_K$	$p$ Value
Topology-14-5	0.74	<0.001 *	0.68	<0.001 *
Morphology-5	−0.02	0.694	−0.30	<0.001 *
Topology-2-1	0.74	<0.001 *	0.71	<0.001 *
Topology-5-2	0.72	<0.001 *	0.68	<0.001 *
Topology-11-8	−0.31	<0.001 *	−0.60	<0.001 *
Topology-10-2	0.75	<0.001 *	0.71	<0.001 *
Topology-1-7	−0.10	0.006	−0.39	<0.001 *
Texture-4-15	0.39	<0.001 *	0.40	<0.001 *
Morphology-3	0.35	<0.001 *	0.18	<0.001 *
Texture-4-79	−0.31	<0.001 *	−0.31	<0.001 *

OSS, the Sjögren's International Collaborative Clinical Alliance Ocular Staining score;  $r_K$ , Kendall's tau-b correlation coefficient; \*,  $p$  value < 0.05.

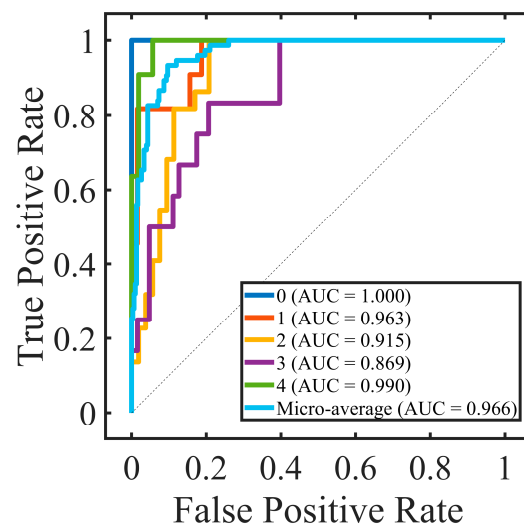
### 3.3. Effectiveness of Selected Features among Different Classification Models

The effectiveness of the selected three groups of features was evaluated using five different machine learning-based classification models (SVM, DT, BT, NB, KNN, and RF) to assess their generalization performance. As shown in Table 5, among the three feature groups, the overall top-10 important features consistently achieved the highest performance across all classification models. The topological top-5 features outperformed the non-topological features within the same range, indicating a robust effectiveness of the selected topological features. Moreover, SVM showed the best performance compared

with other models (ACC: 82.67%, AUC: 96.59%). To further analyze the grading process, the confusion matrix and ROC curve are presented in Figures 6 and 7, respectively. The confusion matrix highlights excellent performance in distinguishing class 0 and class 4, while some errors were observed between class 3 and class 2, as well as class 3 and class 4.



**Figure 6.** The confusion matrix with the top-10 important features in support vector machines (SVMs).



**Figure 7.** The results of receiver operating characteristic (ROC) and area under the curve (AUC) with the top-10 important features in support vector machines (SVMs) for each class. The performance metrics for the average ROC curve were computed using the micro-averaging method.

**Table 5.** Classification results for different groups of features in SVM, DT, BT, NB, and KNN.

Models	Feature Groups	ACC	SEN	PRE	SPE	F-M	AUC
SVM	A-10	82.67	83.71	80.91	80.91	81.80	96.59
	T-5	74.67	65.82	70.91	70.91	67.60	95.36
	NT-5	64.00	46.48	56.36	56.36	49.90	86.14
DT	A-10	73.33	78.40	71.97	71.97	73.66	90.67
	T-5	69.33	62.55	65.45	65.45	62.60	87.79
	NT-5	57.33	41.68	50.23	50.23	44.58	84.97

Table 5. Cont.

Models	Feature Groups	ACC	SEN	PRE	SPE	F-M	AUC
BT	A-10	74.67	75.77	74.55	74.55	74.88	93.25
	T-5	69.33	68.08	68.03	68.03	67.57	91.48
	NT-5	68.00	60.02	59.17	59.17	56.39	87.62
NB	A-10	76.00	78.50	76.21	76.21	76.78	88.35
	T-5	68.00	71.24	67.58	67.58	68.53	85.41
	NT-5	62.67	57.72	56.14	56.14	55.10	88.19
KNN	A-10	80.00	79.66	78.18	78.18	78.26	95.44
	T-5	74.67	74.88	73.18	73.18	73.59	91.43
	NT-5	60.00	55.72	55.23	55.23	55.05	86.40
RF	A-10	77.33	78.52	78.03	78.03	78.13	94.86
	T-5	72.00	72.32	70.61	70.61	70.37	90.96
	NT-5	68.00	65.94	64.17	64.17	64.73	90.28

A-10, the overall top-10 important features; T-5, the top-5 important topological features; NT-5, the top-5 important non-topological features; ACC, accuracy; SEN, sensitivity; PRE, precision; SPE, specificity; F-M, F-measure; AUC, area under the receiver operating characteristic curve; SVM, support vector machine; DT, Decision Tree; BT, Boosting Tree; NB, Naïve Bayes; KNN, K-Nearest Neighbor; RF, Random Forest.

#### 4. Discussion

In this study, we developed an automatic machine learning-based method for corneal fluorescein staining evaluation. Our findings highlighted the significance of topological features as clinical characteristics in the evaluation of corneal fluorescein staining in dry-eye disease. This end-to-end model consisted of segmenting staining regions, extracting their radiomic characteristics, and differentiating them into the corresponding grades. The resulting radiomic signature was constructed from a combination of multiscale topological features, as well as textural and morphological features. The experimental results of our clinical dataset demonstrated the good performance of the proposed model with ACC 82.67% and AUC 95.36%. Moreover, we identified the 10 most important interpretable image-based features, with topological features playing a crucial role. These topological features captured the spatial connectivity and distribution information among staining regions, thereby holding promising in guiding clinical decision-making for the diagnosis and treatment of dry-eye diseases.

The key component in our proposed method was the usage of multiscale topological features that scrutinized the spatial connectivity and distribution among staining regions. These features significantly contributed to the discriminatory ability of the model. This is evident in Figure 4, where the top-ranked radiomic features after selection include six topological attributes among the top seven characteristics. Additionally, the superior performance of topological features compared to non-topological features was consistently observed across different machine learning-based classification models, as shown in Table 5. This indicated the effectiveness of topological features in distinguishing between different grades of corneal staining. Moreover, Table 4 reveals a significant positive or negative correlation between topological characteristics and subjective gradings, which was not observed in other textural and morphological features. These findings underscore the unique contribution of topological features in capturing relevant spatial information related to corneal staining.

The observation of an increasing diameter of scale 56 (“topology-14-5”) with higher true labels suggested a linear distribution of fluorescein staining. In the context of clinical practice, we interpreted this linear distribution as being horizontal. The horizontal distribution is associated with incomplete blinking, which refers to a blink cycle where the upper lid turns to the upstroke before making contact with the lower lid [31,32]. Incomplete blinking is commonly observed in dry-eye patients [33]. Incomplete blinks extend the period of horizontal strip area exposure to desiccating stress, leading to more corneal epithelium defects in that region [34]. Therefore, the presence of a linear distribution of staining may serve as an indicator of incomplete blinking and the associated increased risk of corneal

epithelial damage. Other topological features revealed the aggregation of corneal staining in the inferior zone, which has rarely been reported in previous studies. The number of subgraphs of scale 8 ("topology-2-1") and the average vertex degree of scale 20 ("topology-5-2") and 40 ("topology-10-2") positively correlated with the true label, which indicated a tendency for staining to cluster. The aggregation phenomenon may be attributed to the main mechanism of corneal epithelium cell damage in DED, which is tear hyperosmolarity. It has been proven that hyperosmolarity can be localized [3]. Additionally, tear film is more likely to break up in areas in damaged epithelium, leading to an increase in osmolarity in those regions. This creates a vicious cycle, resulting in more corneal epithelium defects in the same area. Furthermore, the more staining in the fixed area (the inferior zone of the cornea) inevitably affects the value of such topological features. These topological features highlight the crucial role they play in distinguishing the severity of corneal staining and provide valuable insights into the spatial connectivity and distribution of staining regions in a clinical setting.

We extensively explored different combinations of feature signatures and models to identify the approach that yielded the best performance for our computer-aided diagnosis method. The selection of the support vector machine (SVM) model as the best-performing classification model in our study was based on its high accuracy (ACC) of 82.67% and area under the receiver operating characteristic curve (AUC) of 96.59%. The high ACC and AUC values obtained with the SVM model indicated its effectiveness in accurately classifying corneal staining images into different grades.

There were some previous studies that reported similar results in staining-regions evaluation. While some studies focused on semi-automatic methods for locating the region of interest (ROI) in the cornea [9,10], our method was fully automated, providing accurate localization and segmentation of staining regions. Additionally, previous studies mainly relied on image processing or deep-learning techniques for corneal region segmentation, without specifically analyzing the spatial connectivity and distribution of staining regions [11,12,18]. In contrast to approaches that only consider parameters such as the number and area of staining regions, we introduced topological features to capture the spatial characteristics and the confluence of staining regions. This provided a more comprehensive analysis of corneal staining patterns. By incorporating these topological features along with textural and morphological traits, our machine learning-based classification model demonstrated improved performance in grading corneal staining.

The effectiveness of the topological features identified in our study holds promise for clinicians in diagnosing dry-eye disease. The integration of topological features with other radiomic characteristics enhances the classification performance and provides valuable insights into the spatial characteristics of corneal staining. This information can, potentially, aid clinicians in diagnosing and managing dry-eye disease more effectively.

There were several limitations in our study that should be acknowledged. First, the segmentation accuracy of large staining patches could be improved. Due to the low contrast and ambiguous edges of the patches, the segmentation accuracy was compromised. Future studies could explore more efficient methods, such as deep learning, for accurately locating the ROI and segmenting staining regions. Second, the performance of the proposed method should be further validated on a larger dataset. Although our study included a substantial number of images, a larger dataset would provide more robust evidence of the effectiveness of the proposed method. Lastly, the classification performance for score 3 was relatively insufficient. It is possible that score 3 included two distinct conditions: true score 3 and score 2 with an additional score for confluent staining. This may have resulted in a misclassification as score 4 or score 2. Future research could focus on designing a more refined classification model to improve the differentiation of score 3 from other grades. Addressing these limitations in future studies will help to further enhance the accuracy and applicability of the proposed method in corneal fluorescein staining evaluation.

## 5. Conclusions

Our study demonstrated the effectiveness of a fully automatic machine learning-based method for corneal fluorescein staining analysis. The integration of topological features with other radiomic characteristics enhanced the classification performance and provided valuable insights into the spatial characteristics of corneal staining. This information can, potentially, aid clinicians in diagnosing and managing dry-eye disease more effectively.

**Author Contributions:** Conceptualization, L.T., G.-Q.Z. and Y.J.; Methodology, J.F., Z.-K.R., K.-N.W., H.G., L.T., G.-Q.Z. and Y.J.; Software, Z.-K.R., K.-N.W. and H.G.; Formal analysis, J.F. and Z.-K.R.; Investigation, J.F. and Y.-R.H.; Resources, L.T., G.-Q.Z. and Y.J.; Data curation, J.F., Z.-K.R. and Y.-R.H.; Writing—original draft, J.F., Z.-K.R., K.-N.W. and H.G.; Writing—review & editing, J.F., Z.-K.R., Y.-C.S., L.T., G.-Q.Z. and Y.J.; Visualization, J.F. and Z.-K.R.; Supervision, L.T., G.-Q.Z. and Y.J.; Project administration, L.T., G.-Q.Z. and Y.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the Research Development Fund of Beijing Municipal Health Commission (2019-4), the National Natural Science Foundation of China (62371121), and the Jiangsu Provincial Key R & D Program, China (BE2022827).

**Institutional Review Board Statement:** The study was approved by the Ethics Committee of Beijing Tongren Hospital (TRECK2021-238).

**Informed Consent Statement:** Written informed consent was obtained from all the participants.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author, upon reasonable request.

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

## References

1. Craig, J.P.; Nichols, K.K.; Akpek, E.K.; Caffery, B.; Dua, H.S.; Joo, C.K.; Liu, Z.; Nelson, J.D.; Nichols, J.J.; Tsubota, K.; et al. TFOS DEWS II definition and classification report. *Ocul. Surf.* **2017**, *15*, 276–283. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Stapleton, F.; Alves, M.; Bunya, V.Y.; Jalbert, I.; Lekhanont, K.; Malet, F.; Na, K.-S.; Schaumberg, D.; Uchino, M.; Vehof, J.; et al. TFOS DEWS II epidemiology report. *Ocul. Surf.* **2017**, *15*, 334–365. [\[CrossRef\]](#) [\[PubMed\]](#)
3. McMonnies, C.W. An examination of the relationship between ocular surface tear osmolarity compartments and epitheliopathy. *Ocul. Surf.* **2015**, *13*, 110–117. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Bron, A.J.; Argüeso, P.; Irkeç, M.; Bright, F.V. Clinical staining of the ocular surface: Mechanisms and interpretations. *Prog. Retin. Eye Res.* **2015**, *44*, 36–61. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Lemp, A. Report of the National Eye Institute/Industry workshop on clinical trials in dry eyes. *Eye Contact Lens* **1995**, *21*, 221–232.
6. Bron, A.J.; Evans, V.E.; Smith, J.A. Grading of corneal and conjunctival staining in the context of other dry eye tests. *Cornea* **2003**, *22*, 640–650. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Whitcher, J.P.; Shiboski, C.H.; Shiboski, S.C.; Heidenreich, A.M.; Kitagawa, K.; Zhang, S.; Hamann, S.; Larkin, G.; McNamara, N.A.; Greenspan, J.S.; et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjögren's Syndrome International Registry. *Am. J. Ophthalmol.* **2010**, *149*, 405–415. [\[CrossRef\]](#)
8. Begley, C.; Caffery, B.; Chalmers, R.; Situ, P.; Simpson, T.; Nelson, J.D. Review and analysis of grading scales for ocular surface staining. *Ocul. Surf.* **2019**, *17*, 208–220. [\[CrossRef\]](#)
9. Rodriguez, J.D.; Lane, K.J.; Ousler, G.W.; Angjeli, E.; Smith, L.M.; Abelson, M.B. Automated grading system for evaluation of superficial punctate keratitis associated with dry eye. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 2340–2347. [\[CrossRef\]](#)
10. Pellegrini, M.; Bernabei, F.; Moscardelli, F.; Vagge, A.; Scotto, R.; Bovone, C.; Scoria, V.; Giannaccare, G. Assessment of corneal fluorescein staining in different dry eye subtypes using digital image analysis. *Transl. Vis. Sci. Technol.* **2019**, *8*, 34. [\[CrossRef\]](#)
11. Chun, Y.S.; Yoon, W.B.; Kim, K.G.; Park, I.K. Objective assessment of corneal staining using digital image analysis. *Investig. Ophthalmol. Vis. Sci.* **2014**, *55*, 7896–7903. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Bağbaba, A.; Şen, B.; Delen, D.; Uysal, B.S. An automated grading and diagnosis system for evaluation of dry eye syndrome. *J. Med. Syst.* **2018**, *42*, 1–14. [\[CrossRef\]](#)
13. Kourukmas, R.; Roth, M.; Geerling, G. Automated vs. human evaluation of corneal staining. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2022**, *260*, 2605–2612. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Wang, S.; He, J.; He, X.; Liu, Y.; Lin, X.; Xu, C.; Zhu, L.; Kang, J.; Wang, Y.; Li, Y.; et al. AES-CSFS: An automatic evaluation system for corneal sodium fluorescein staining based on deep learning. *Ther. Adv. Chronic Dis.* **2023**, *14*, 20406223221148266. [\[CrossRef\]](#) [\[PubMed\]](#)



15. Susila, S.J.G.; Kavitha, D. Corneal ulcer feature extraction and image classification using a deep convolutional network and the VGG 16 mode. In Proceedings of the 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 13–15 December 2022.
16. Portela, H.M.; Veras RD, M.; Vogado, L.H.; Leite, D.; Ambrósio, P.E.; Cardoso de Paiva, A.; Tavares, J.M.R. A corneal ulcer segmentation approach using U-Net and stepwise fine-tuning. In *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*; Taylor & Francis: Abingdon, UK, 2023; pp. 1–13. [\[CrossRef\]](#)
17. Qu, J.H.; Qin, X.R.; Li, C.D.; Peng, R.M.; Xiao, G.G.; Cheng, J.; Gu, S.F.; Wang, H.-K.; Hong, J. Fully automated grading system for the evaluation of punctate epithelial erosions using deep neural networks. *Br. J. Ophthalmol.* **2021**, *107*, 453–460. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Su, T.Y.; Ting, P.J.; Chang, S.W.; Chen, D.Y. Superficial punctate keratitis grading for dry eye screening using deep convolutional neural networks. *IEEE Sens. J.* **2019**, *20*, 1672–1678. [\[CrossRef\]](#)
19. Wolffsohn, J.S.; Arita, R.; Chalmers, R.; Djalilian, A.; Dogru, M.; Dumbleton, K.; Gupta, P.K.; Karpecki, P.; Lazreg, S.; Pult, H.; et al. TFOS DEWS II diagnostic methodology report. *Ocul. Surf.* **2017**, *15*, 539–574. [\[CrossRef\]](#)
20. Woods, J.; Hutchings, N.; Srinivasan, S. Geographic distribution of corneal staining in symptomatic dry eye. *Ocul. Surf.* **2020**, *18*, 258–266. [\[CrossRef\]](#)
21. Woods, J.; Hutchings, N.; Srinivasan, S.; Jones, L. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368.
22. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
23. Daugman, J. How Iris Recognition Works. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 21–30. [\[CrossRef\]](#)
24. Vincent, L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans. Image Process.* **1993**, *2*, 176–201. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Chen, Z.; Strange, H.; Oliver, A.; Denton, E.R.; Boggis, C.; Zwiggelaar, R. Topological modeling and classification of mammographic microcalcification clusters. *IEEE Trans. Biomed. Eng.* **2014**, *62*, 1203–1214. [\[CrossRef\]](#)
26. Van Griethuysen, J.J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [\[CrossRef\]](#)
27. Li, M.; Zhu, L.; Zhou, G.; He, J.; Jiang, Y.; Chen, Y. Predicting the pathological status of mammographic microcalcifications through a radiomics approach. *Intell. Med.* **2021**, *1*, 95–103. [\[CrossRef\]](#)
28. Biesiada, J.; Duch, W. *Feature Selection for High-Dimensional Data: A Kolmogorov-Smirnov Correlation-Based Filter*. *Computer Recognition Systems*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 95–103.
29. Erişti, H.; Yildirim, Ö.; Erişti, B.; Demir, Y. Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. *Int. J. Electr. Power Energy Syst.* **2013**, *49*, 95–103. [\[CrossRef\]](#)
30. Gordon, A.D. Classification and Regression Trees. *Biometrics* **1984**, *40*, 874. [\[CrossRef\]](#)
31. Cruz, A.A.; Garcia, D.M.; Pinto, C.T.; Cechetti, S.P. Spontaneous eyeblink activity. *Ocul. Surf.* **2011**, *9*, 29–41. [\[CrossRef\]](#)
32. Jie, Y.; Sella, R.; Feng, J.; Gomez, M.L.; Afshari, N.A. Evaluation of incomplete blinking as a measurement of dry eye disease. *Ocul. Surf.* **2019**, *17*, 440–446. [\[CrossRef\]](#)
33. Wang, M.T.; Tien, L.; Han, A.; Lee, J.M.; Kim, D.; Markoulli, M.; Craig, J.P. Impact of blinking on ocular surface and tear film parameters. *Ocul. Surf.* **2018**, *16*, 424–429. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Bron, A.J.; de Paiva, C.S.; Chauhan, S.K.; Bonini, S.; Gabison, E.E.; Jain, S.; Knop, E.; Markoulli, M.; Ogawa, Y.; Perez, V.; et al. Tfos dews ii pathophysiology report. *Ocul. Surf.* **2017**, *15*, 438–510. [\[CrossRef\]](#) [\[PubMed\]](#)

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