Efficient deep learning for automated corneal ulcer severity classification from fluorescein images

Rodiah[1], Indah Sinthya Permata Sari[2], Matrissya Hermita[3], Sarifuddin Madenda[2], Diana Tri Susetianingtias[4]
[1]Department of Informatics, Faculty of Industry Technology, Universitas Gunadarma, Depok, Indonesia
[2]Department of Information Technology, Graduate Program, Universitas Gunadarma, Depok, Indonesia
[3]Department of Medical Education, Faculty of Medicine, Universitas Gunadarma, Depok, Indonesia
[4]Department of Computer Systems, Faculty of Information Systems and Information Technology, Universitas Gunadarma, Depok, Indonesia

Article Info

Article history:

Received Jul 23, 2025
Revised May 13, 2026
Accepted May 24, 2026

Keywords:

Automated diagnosis
Corneal ulcer
EfficientNetV2
Fluorescein imaging
Ocular imaging

ABSTRACT

Corneal ulcers can cause permanent vision loss if not diagnosed and managed promptly, particularly in settings with limited access to ophthalmology services. This study aims to develop an automated deep learning approach for classifying corneal ulcer severity from fluorescein slit-lamp images. An EfficientNetV2-S–based model is employed, incorporating corneal area masking to suppress non-relevant regions and class distribution–based augmentation to address data imbalance. To improve evaluation reliability, a leakage-aware data splitting strategy is applied before and after augmentation. Experimental results show that the proposed approach achieves a maximum validation accuracy of 95.93% under non-leakage conditions for the category classification scenario, while maintaining high training efficiency. These results demonstrate that the proposed method provides a robust and efficient solution for automated corneal ulcer severity assessment and has the potential to support clinical decision-making in ophthalmic practice.

Corresponding Author:

Rodiah
Department of Informatics, Faculty of Industry Technology, Universitas Gunadarma
Margonda Raya 100, Pondok Cina, Depok, Indonesia
Email: rodiah@staff.gunadarma.ac.id

1. INTRODUCTION

A corneal ulcer is a pathological condition involving damage to the corneal tissue from the epithelial layer to the stroma, and may result in permanent blindness if not managed appropriately [1]. According to the World Health Organization (WHO), the annual incidence of corneal ulcers reaches approximately 710 cases per 100,000 population, with vision loss caused by ocular trauma affecting 1.5–2 million individuals worldwide, particularly in developing countries [2]. Advanced manifestations such as corneal opacity are commonly observed among populations with low educational backgrounds and high-risk occupations, including farmers, fishermen, and manual laborers, who are frequently exposed to eye trauma while having limited access to timely and high-quality ophthalmic care [3]. These conditions highlight persistent challenges related to delayed diagnosis and accurate severity assessment of corneal ulcers.

At present, corneal ulcer severity assessment predominantly relies on manual slit-lamp examination with fluorescein staining [4]. Although widely used in clinical practice, this method requires high expertise, is time-consuming, and is inherently subjective, leading to inter-observer variability and potential diagnostic delays [5]. Such limitations increase the risk of irreversible corneal damage and vision loss, underscoring the growing need for automated, fast, and reliable diagnostic tools to support clinical decision-making, particularly in resource-limited settings [6].

Recent advances in deep learning have demonstrated promising performance in automated corneal ulcer classification using fluorescein images. Vision transformer (ViT)-based approaches have achieved high accuracy across multiple classification scenarios, including category, type, and grade classification [7]. Convolutional neural network (CNN)-based methods and hybrid architectures combining deep feature extractors with classical classifiers, such as DenseNet121 with support vector machine (SVM), have also reported strong classification performance on public datasets [8]. Other studies have explored various CNN architectures for ulcer type classification [9], infectious keratitis diagnosis [10], and binary severity classification [11]. While these approaches show encouraging results, they often suffer from high model complexity, limited evaluation of inference efficiency, non-end-to-end pipelines, or restricted classification scope.

Despite their effectiveness, existing studies still exhibit several methodological gaps. Transformer-based models tend to be computationally expensive and lack analysis of deployment feasibility in clinical environments [7]. Hybrid CNN–SVM approaches rely on separated feature extraction and classification stages, which may reduce adaptability to new datasets [8]. Moreover, several CNN-based studies focus on a single classification scenario or binary severity levels and do not explicitly address visual noise from non-corneal regions, such as the sclera and eyelashes, due to the absence of corneal area masking [9], [11]. Additionally, limited attention has been paid to data leakage issues arising from improper data splitting strategies, which may compromise evaluation validity and generalization performance [12].

To address these limitations, this study proposes an automated corneal ulcer severity classification framework based on the EfficientNetV2-S architecture using fluorescein slit-lamp images. The proposed approach integrates selective corneal area masking to suppress non-relevant visual information and class distribution–based augmentation to mitigate data imbalance. Furthermore, a leakage-aware data splitting strategy is systematically designed before and after augmentation to ensure reliable model evaluation. By providing an efficient end-to-end pipeline capable of handling multiple classification scenarios: category, type, and grade, this study aims to enhance robustness, generalization, and practical applicability of automated corneal ulcer diagnosis systems.

2. METHOD

This study employs a deep learning-based approach to automatically classify the severity of corneal ulcers using fluorescein images. The overall methodology is organized into several sequential stages, as illustrated in Figure 1. The process begins with image acquisition and preprocessing to improve image quality and prepare the dataset for analysis. The next stage involves model development using a deep learning architecture for corneal ulcer severity classification. Finally, the trained model is evaluated using performance metrics to assess its classification accuracy and reliability. These stages provide a systematic and reproducible framework for the proposed research methodology.

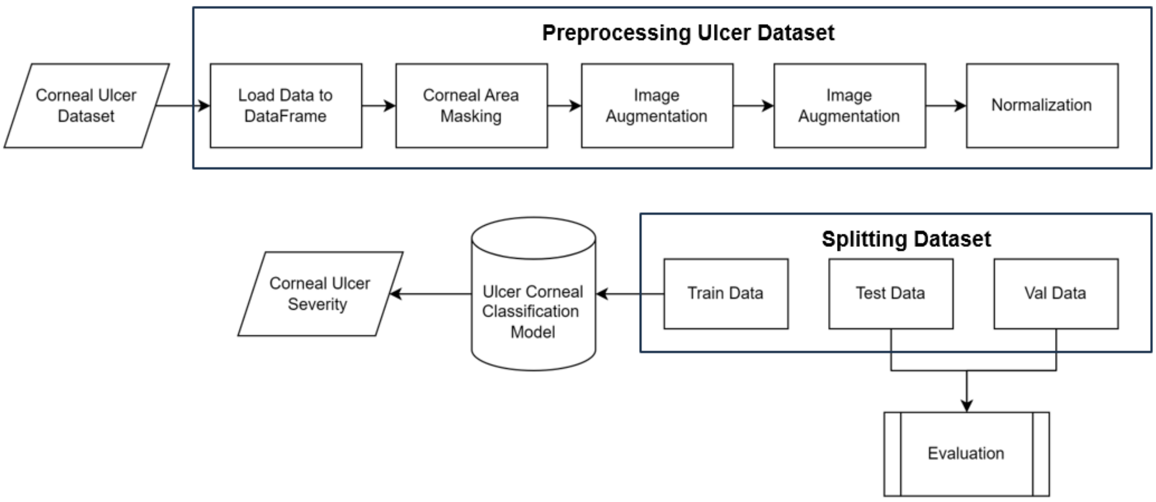


Figure 1. Research methodology for corneal ulcer severity classification using deep learning-based approach

2.1. Corneal ulcer dataset

The fluorescein eye image dataset used in this study is a secondary dataset. This dataset was obtained from the research [13]. This dataset consisted of 712 fluorescein-stained eye images collected from patients with varying degrees of corneal ulcers at the Zhongshan Ophthalmology Center at Sun Yat-sen University. The fluorescein eye image dataset is divided into three categories, namely category, type, and grade [13]. The following is a detailed explanation of the three scenarios.

i) The category scenario represents the general type of corneal ulcer and is divided into three subclasses: category 0, category 1, and category 2. Category 0 (point-like corneal ulcer) refers to small, dot-shaped ulcers that are generally mild and commonly appear during the early stage of inflammation or healing. Category 1 (point-flaky mixed ulcer) represents a combination of point and flaky ulcer patterns irregularly distributed across the cornea, indicating a more severe condition. Category 2 (flaky corneal ulcer) refers to larger flaky lesions with clear boundaries that may lead to corneal scarring and vision loss [14].
ii) The type scenario provides a more detailed description of ulcer morphology based on lesion shape and size. This scenario consists of five subclasses: type 0, type 1, type 2, type 3, and type 4. Type 0 indicates a normal corneal epithelium without ulceration. Type 1 (micro punctate) represents small scattered lesions on the corneal surface. Type 2 (macro punctate) indicates larger punctate lesions, while type 3 (coalescent macro punctate) describes larger lesions that merge or appear closely grouped. Type 4 refers to ulcer patches measuring ≥1 mm, indicating more severe corneal damage [13].
iii) The grade scenario describes the severity and spatial distribution of corneal ulcers based on the affected corneal quadrants [15]. This scenario is divided into five subclasses: grade 0, grade 1, grade 2, grade 3, and grade 4. Grade 0 indicates no ulceration, whereas grades 1 to 3 represent ulcer involvement in one, two, and three or four corneal quadrants, respectively. Grade 4 indicates ulcers located in the central optical zone of the cornea, which is the most critical region affecting vision [16]. The representation of the corneal quadrants used in the grade scenario is illustrated in Figure 2.

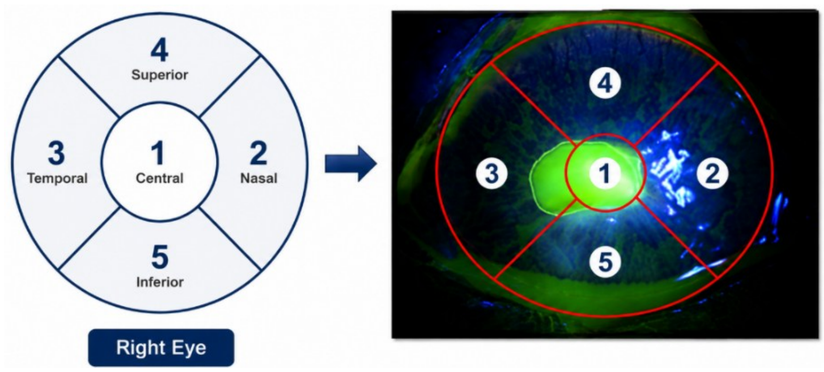


Figure 2. Quadrant representation in corneal ulcers [10]

2.2. Preprocessing of corneal ulcers

Preprocessing is the initial stage of data preparation applied to fluorescein eye images before they are used for training machine learning models. This stage aims to improve data consistency, reduce irrelevant visual information, and ensure that the images are suitable for effective feature learning. In this study, preprocessing consists of data organization, corneal area masking, data augmentation, image resizing, and normalization.

i) Data organization: corneal ulcer images were organized based on their directory structure and file naming conventions to ensure systematic data handling. The images and corresponding labels were integrated into a dataframe using the Pandas library [17], including the creation of image path references and standardized scenario labels. This step ensures consistency between image data and labels for subsequent preprocessing and modeling stages [18].
ii) Corneal area masking: corneal masking was performed using the OpenCV library [19] to remove non-corneal regions, such as the sclera and surrounding background structures. A binary mask image in PNG format was used to define the corneal region. The masking process involved inverting the mask using the bitwise_not function, followed by applying the bitwise_and operation to preserve the corneal area while suppressing irrelevant regions. As a result, the output image retained only the

corneal region as the primary area of interest for model training. An example of the masking result is presented in Figure 3.

iii) Data augmentation: data augmentation was applied after dataset splitting to increase data variability and balance the number of samples across classes. The augmentation process utilized the ImageDataGenerator module from TensorFlow [4] with several transformations, including horizontal and vertical flipping, rotation, and zooming. Augmentation was selectively applied to minority classes to maintain class balance [20]. The training and validation data ratio remained consistent at 80:20 after augmentation. A summary of the augmented dataset distribution is presented in Table 1.

iv) Image resizing: the EfficientNetV2-S model has an input image size of 224×224 pixels [21]. In the original dataset, the fluorescein eye image was 2592×1728 pixels. In this study, the image size was adjusted to 224×224 pixels. This step is important to ensure that the image matches the specifications required by the model without omitting important details in the image.

v) Data normalization: it is the last stage of the sub-stage of data preprocessing carried out in this study. The purpose of normalization is to help the model training process run faster and more effectively. Each pixel on the input image has a value from the range of 0–255. Therefore, normalization is done on each grayscale image by dividing the pixel value by 255, resulting in a new value with a range of 0–1 [22].

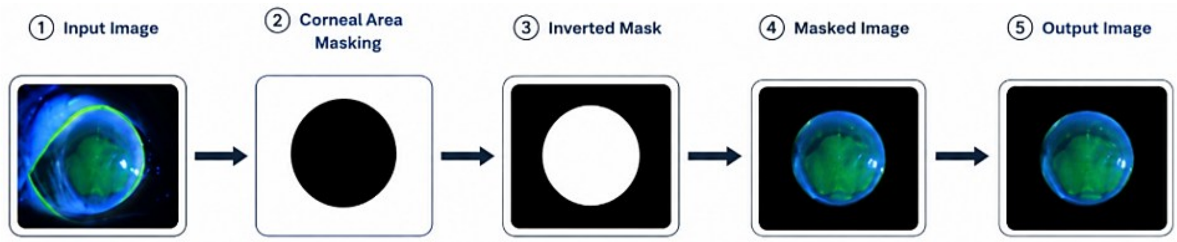


Figure 3. Example of corneal area masking applied to fluorescein images

Table 1. Summary of dataset distribution after augmentation

Scenario	Original train/val	Added train/val	Final train/val
Category	569/143	289/73	858/216
Type	569/143	1,621/407	2,190/550
Grade	569/143	521/132	1,090/275

2.3. Data splitting

The data is divided into two, namely training data (train_df) and validation (val_df) with a ratio of 8:2. This means that as much as 80% is for training data and 20% is for validation data. The division of data is divided into two classes because of the small amount of data and there is an imbalance in the amount of data between the classes. Each scenario implements .copy() to duplicate the initial dataframe. Each dataframe will contain only the file name, image directory, and label name attributes that are relevant to that scenario. The dataframe of each scenario is defined as df_cat for categories, df_type for the type, and df_grade for grade. Thus, the data for each scenario can be managed separately, making it easier to analyze and train specific models on each data characteristic. Next, each scenario dataframe will be divided into training and validation data. The data sharing ratio is 80% for training data and 20% for validation data. Data sharing is carried out using the functions in the ScikitLearn Module, namely train_test_split() [23]. Variable Dataframe differentiated by data type and scenario type.

2.3.1. Data splitting to prevent data leakage

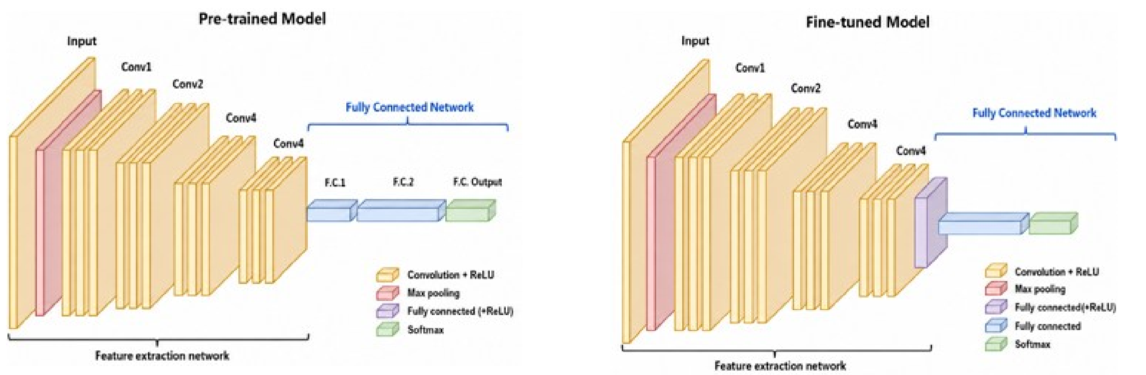
In processing medical images, especially eye images with fluorescein, data leakage is very important to repair. This is because medical images are obtained from different patients, so that later it will affect follow-up research, namely diagnosis [24]. Data leakage occurs when image augmentation is done before data sharing. Data leakage can be avoided by sharing data first before the augmentation stage to ensure that the validation data does not contain images of the same patients as the training data. Thus, validation data will provide more accurate results because it consists of data that the model has never seen and studied. Therefore, data sharing is important to do after image masking and before augmentation in data preprocessing [25].

2.3.2. Data splitting under data leakage conditions

Data leakage needs to be tested by creating conditions where data leakage occurs to compare the results of data leakage and not. Using the image data generator (IDG) in the data preprocessing in each augmentation process will result in different image variations, thus avoiding duplication [26]. In this study, data leakage can be carried out by re-arranging training and validation data that have been augmented and separated previously. Data augmentation is done once, and the results can be compared. If the data has been combined, then the data will be redistributed using the stratified sampling method [27] with the same ratio of 80:20. This can cause the image data of eye ulcers with fluorescein from the same patient to be in both training and validation data. The results of the model performance evaluation using data leakage can be used to compare with the results of model performance evaluation without data leakage conditions.

2.4. Proposed model architecture for automated classification of corneal ulcer severity

Figure 4 illustrates the proposed EfficientNetV2-S architecture for corneal ulcer severity classification, consisting of the pre-trained model and the fine-tuned model. Figure 4(a) presents pre-trained EfficientNetV2-S architecture used for feature extraction, while Figure 4(b) shows the fine-tuned architecture proposed for corneal ulcer severity classification. The pre-trained EfficientNetV2-S model used in this study was obtained from the tf.keras.applications module and had been previously trained on the ImageNet dataset containing 1,000 classes. Fluorescein images with a resolution of 224×224 pixels were used as input, while the extracted features were processed using a GlobalAveragePooling2D layer [28], followed by a dense layer with ReLU activation and a final softmax classification layer [29]. The proposed architecture was designed as an efficient end-to-end model with relatively low computational complexity and potential integration into clinical decision-support systems for early corneal ulcer diagnosis [30]. EfficientNetV2-S was selected because it is a lightweight and modern architecture that remains relatively underexplored for fluorescein image classification in ophthalmology, while the selective fine-tuning strategy was intended to improve model generalization on limited datasets.



(a) (b)

Figure 4. Proposed EfficientNetV2-S architecture for corneal ulcer severity classification (a) pre-trained model and (b) fine-tuned model

2.5. Classification model training

In this study, the model was trained for 50 epochs using early stopping and adaptive learning rate strategies to improve convergence efficiency. The early stopping mechanism was configured with a patience value of 10 epochs, meaning that training was automatically terminated if no performance improvement was observed for 10 consecutive epochs. In addition, the learning rate was adaptively adjusted using the ReduceLROnPlateau strategy based on the validation loss metric, with a patience value of 5 epochs and a reduction factor of 0.2. The minimum learning rate was limited to 0.000001 to prevent excessively slow training [31]. For the final performance evaluation, the validation dataframe (val_df) was also used as the test dataframe (test_df). These strategies were implemented to maintain training stability and improve the model's generalization capability for corneal ulcer classification.

2.6. Evaluation metrics

To comprehensively evaluate the performance of the proposed corneal ulcer classification model, multiple evaluation metrics are employed in addition to overall accuracy. In medical image analysis tasks, relying solely on accuracy is widely recognized as insufficient due to class imbalance and the high clinical cost of misclassification, particularly in disease detection scenarios [32], [33]. Therefore, this study evaluates model performance using accuracy, sensitivity (recall), specificity, and the area under the curve (AUC), which are commonly recommended metrics in medical artificial intelligence evaluation frameworks [34], [35].

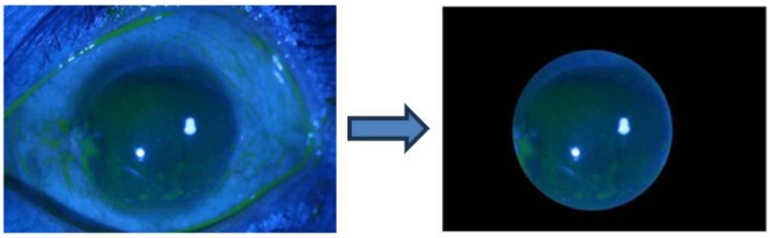
Accuracy measures the overall correctness of the classification, while sensitivity reflects the model's ability to correctly identify positive ulcer cases, which is critical for minimizing missed diagnoses and delayed treatment [36]. Specificity evaluates the ability of the model to correctly identify negative cases, thereby reducing false-positive predictions that may lead to unnecessary clinical interventions [37]. In addition, the AUC metric provides a threshold-independent assessment of classification performance by quantifying the trade-off between sensitivity and specificity, making it particularly suitable for evaluating diagnostic reliability in ophthalmic imaging applications [38], [39]. For multi-class classification scenarios (category, type, and grade), sensitivity and specificity are computed using a macro-averaging strategy to ensure balanced performance evaluation across all classes, regardless of class distribution [40]. Collectively, these metrics offer a more clinically meaningful assessment of the proposed model and facilitate fair comparison with previous studies in automated corneal ulcer classification.

3. RESULT AND DISCUSSION

The data preprocessing stage is carried out to increase the variety of training data and prepare the data so that the training process can run effectively. This stage starts from loading data into the dataframe, masking the corneal area, and adding data with augmentation. It continues with adjusting the image size and normalizing.

3.1. Results of the image masking stage

The masking process was carried out to eliminate noise outside the corneal area using the OpenCV module. Figure 5 illustrates the corneal image preprocessing step, showing the comparison between the original fluorescein image and the masked corneal region. Figure 5(a) shows the original fluorescein image, while Figure 5(b) presents the masking result in which regions outside the cornea are suppressed using black pixels, allowing the corneal area to become the primary focus for model training. After masking, data augmentation was applied to generate additional image variations and balance the number of samples in each class. The augmentation process was performed using ImageDataGenerator with several transformations, including horizontal and vertical flipping, rotation, and zooming. The dataset was divided into training and validation sets using an 80:20 ratio across the category, type, and grade classification scenarios. To avoid data leakage, augmented images derived from the same original image were carefully separated between the training and validation sets. This strategy ensured that model evaluation was conducted using unseen data and provided a more reliable assessment of model performance.



(a) (b)

Figure 5. Corneal images before and after masking (a) the original fluorescein image and (b) the masked corneal region

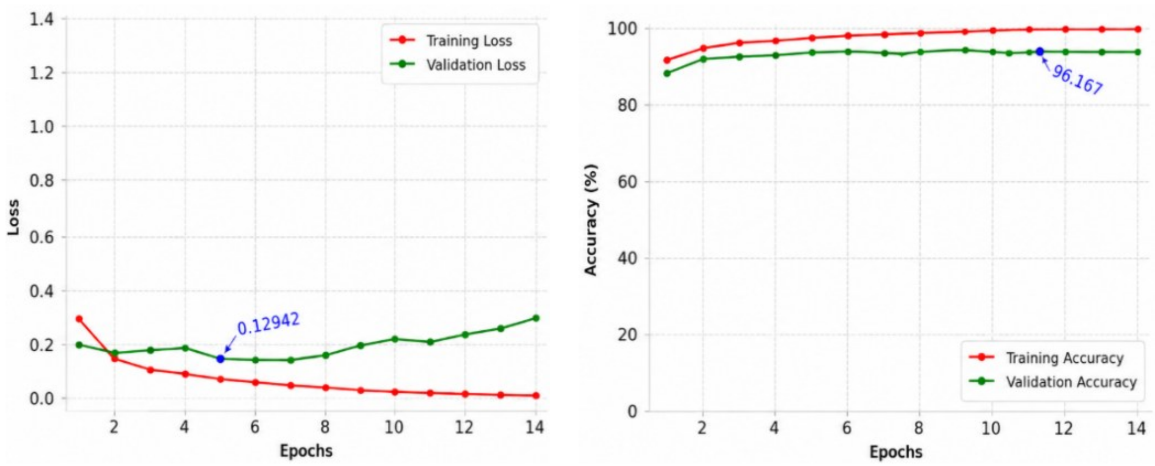
3.2. Classification model training results

The results of the machine learning modeling stages are divided into three stages, namely the results of model initiation, the results of model training, and the results of model training evaluation. This stage produced three pre-trained CNN models. These models were able to classify corneal ulcers based on three scenarios, namely category, type, and grade.

3.2.1. Model initiation results

The model initialization in this study was performed by combining the pre-trained EfficientNetV2-S architecture with fully connected layers for category, type, and grade classification of corneal ulcers. The EfficientNetV2-S model was obtained from the tf.keras.applications module and had been previously trained on the ImageNet dataset. The category classification model consisted of 21,646,179 total parameters, while the type and grade classification models each contained 21,648,229 parameters.

Model training evaluation was conducted to analyze the learning performance during the training process. Figure 6 illustrates the training evaluation results of the category classification model using the EfficientNetV2-S architecture over 50 epochs. Figure 6(a) presents the training and validation loss curves, while Figure 6(b) shows the training and validation accuracy curves for the category classification scenario using the EfficientNetV2-S architecture. The model achieved the highest validation accuracy of 95.93% at epoch 24, while the lowest validation loss of 0.1292 was observed during the early training stage. These results indicate that the proposed model was able to learn relevant image features effectively while maintaining stable validation performance.



(a) (b)

Figure 6. Evaluation results of the category classification model training (a) training and validation loss curves and (b) training and validation accuracy curves

Overall, the training process for the non-leakage category scenario required approximately 33 minutes. In this scenario, the model achieved the highest validation accuracy of 95.93% at epoch 24. Under leakage conditions, the category classification model obtained a maximum validation accuracy of 87.91% at epoch 30, while the lowest validation loss of 0.5292 was recorded at epoch 10. Meanwhile, the type classification scenario with leakage data achieved the highest validation accuracy of 97.99% at epoch 36, with the lowest validation loss of 0.1070 obtained at epoch 22. These results indicate that the proposed preprocessing and data-splitting strategy contributed to stable model performance and reduced the impact of data leakage during evaluation. The summarized training results are presented in Table 2.

Table 2. Summary of fine-tuned EfficientNetV2-S training results

Scenario	Data condition	Best validation accuracy (%)	Best validation loss	Epoch
Category	Non-leakage	95.93	0.9193	24
Category	Leakage	87.91	0.5292	30
Type	Leakage	97.99	0.1070	36

3.3. Comparison with previous studies and performance discussion

Table 3 summarizes a comparison between this study and relevant previous works on corneal ulcer classification using fluorescein images, highlighting differences in methodology, classification scope, and performance. The comparison focuses on the learning architectures, classification scenarios, and reported evaluation results to provide a clear contextual understanding of existing approaches. This analysis helps position the contributions of this study in relation to prior research and supports a more comprehensive interpretation of the experimental findings.

As summarized in Table 3, previous studies have shown that deep learning–based approaches are effective for corneal ulcer classification using fluorescein images. Transformer-based methods [7] achieved high classification accuracy but generally require higher computational complexity, while hybrid CNN-classifier approaches [8] may have limited adaptability due to their non end-to-end structure. Other CNN-based studies [9], [11] mainly focused on specific classification scenarios, limiting their applicability in broader clinical settings.

In contrast, this study addresses category, type, and grade classification using a unified EfficientNetV2-S framework. In addition, the proposed approach explicitly evaluates model performance under leakage and non-leakage conditions to improve evaluation reliability. The results indicate that corneal area masking and class distribution–based augmentation contribute to improved model generalization, particularly under non-leakage settings. Despite having lower model complexity than transformer-based approaches, EfficientNetV2-S achieved competitive validation performance, suggesting a balanced trade-off between accuracy, efficiency, and robustness for automated corneal ulcer diagnosis applications. Although external dataset validation was not conducted, the leakage-aware evaluation strategy and comparison with previous studies provide indirect evidence of the model's generalization capability.

Table 3. Category leakage training results

Study	Dataset	Method	Classification task	Key limitations
Alakuş and Baykara [7]	SUSTech-SYSU (712 images)	ViT	Category, type, grade	High model complexity; inference efficiency not evaluated
Susila [8]	SUSTech-SYSU (712 images)	DenseNet121+SVM	Category, type, grade	Non end-to-end pipeline; limited adaptability
Sevli [9]	SUSTech-SYSU (712 images)	CNN-based models (AlexNet, VGG16)	Ulcer type classification	Focused on a single classification scenario
Hu *et al.*[10]	Clinical dataset (2757 images)	Multiple DL models incl. EfficientNetV2M	Keratitis diagnosis	Lower accuracy; not focused on severity grading
Gross *et al.* [11]	SUSTech-SYSU (712 images)	CNN (VGG, ResNet)	Binary severity classification	Limited to two severity levels
This study	Fluorescein slit-lamp images	EfficientNetV2-S+corneal masking+leakage-aware splitting	Category, type, grade	Focused on image-based features only

3.4. Discussion on generalization and external validity

Although this study primarily used internal validation on the SUSTech-SYSU fluorescein image dataset, several strategies were implemented to improve model generalization. A leakage-aware data splitting strategy was applied to prevent overly optimistic performance estimation and ensure strict separation between training and validation data. In addition, corneal area masking and class distribution–based augmentation were used to reduce dataset-specific bias and improve robustness to variations in illumination, lesion appearance, and imaging conditions commonly found in clinical practice. Due to the limited availability of publicly accessible fluorescein datasets with consistent annotations, external validation could not be performed. Nevertheless, the experimental findings are consistent with previous studies [7]–[11], as summarized in Table 3. Future work will focus on multi-center and cross-dataset evaluations to further assess model generalization in diverse clinical settings.

4. CONCLUSION

This study presents an end-to-end deep learning framework for automated corneal ulcer severity classification based on the EfficientNetV2-S architecture using fluorescein slit-lamp images. The primary technical contribution lies in the integration of corneal area masking and leakage-aware data splitting into the training pipeline, which effectively enhances feature focus and mitigates evaluation bias commonly present in medical image classification tasks. Experimental results demonstrate that the proposed strategy improves model stability and generalization, with non-leakage training achieving a peak validation accuracy of 95.93% in the category classification scenario, while leakage conditions consistently lead to performance degradation due to overfitting. From an artificial intelligence perspective, the proposed framework offers a robust and computationally efficient solution that balances model complexity and predictive performance, making it suitable for real-world deployment in intelligent clinical decision support systems. The lightweight nature of EfficientNetV2-S enables fast inference while maintaining high classification accuracy across multiple severity scenarios. Future studies will extend this work by evaluating the proposed model on external and multi-center datasets to further validate its generalizability and support scalable clinical adoption.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kemdiktisaintek) for funding this research under the fundamental research–regular scheme. The authors also acknowledge the Research Institutions of Universitas Gunadarma for their administrative support, research facilities, and continuous assistance throughout the study.

FUNDING INFORMATION

This research was funded by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kemdiktisaintek) under the fundamental research – regular scheme, as stipulated in the director of research decree no. 0419/C3/DT.05.00/2025 dated May 22, 2025, main contract no. 124/C3/DT.05.00/PL/2025 dated May 28, 2025, subcontract no. 10974/LL3/AL.04/2025 dated June 4, 2025, and subcontract no. 16.28/LP/UG/VI/2025 dated June 5, 2025.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rodiah	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Indah Sinthya Permata Sari	✓	✓		✓	✓		✓			✓	✓	✓		
Matrissya Hermita		✓	✓			✓		✓		✓				
Sarifuddin Madenda	✓			✓	✓					✓	✓		✓	
Diana Tri Susetianingtias		✓			✓	✓	✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

This study did not involve human participants; therefore, informed consent was not required.

ETHICAL APPROVAL

This study utilized a secondary dataset of fluorescein slit-lamp images and did not involve direct interaction with human participants or vertebrate animal subjects. Therefore, ethical approval was not required.

DATA AVAILABILITY

The dataset used in this study is the publicly available SUSTech-SYSU corneal ulcer dataset, which consists of fluorescein-stained slit-lamp images collected at the Zhongshan Ophthalmology Center, Sun Yat-sen University. The dataset is accessible through the original publication by Deng *et al.* and is commonly used for benchmarking automated corneal ulcer classification and grading tasks. All data utilized in this research are publicly available and can be accessed for research purposes, allowing reproducibility of the experimental setup. Detailed information regarding data preprocessing steps, corneal area masking procedures, data augmentation strategies, and data splitting configurations implemented in this study is available from the corresponding author upon reasonable request.

REFERENCES

[1] G. R. V.-Duarte *et al.*, "Neurotrophic keratopathy: general features and new therapies," *Survey of Ophthalmology*, vol. 69, no. 5, pp. 789–804, 2024, doi: 10.1016/j.survophthal.2024.04.004.
[2] F. Stapleton, "The epidemiology of infectious keratitis," *The Ocular Surface*, vol. 28, no. 9, pp. 351–363, 2023, doi: 10.1016/j.jtos.2021.08.007.
[3] K. Waslan and R. Oktaliani, "A case report: the strategy of corneal ulcers," *Jambi Medical Journal: Jurnal Kedokteran dan Kesehatan*, vol. 11, no. 1, pp. 68–73, 2023.
[4] D. Manawongsakul and K. Patanukhom, "A segmentation-based automated corneal ulcer grading system using deep learning and Hough circle transform," *Algorithms*, vol. 17, no. 9, 2024, doi: 10.3390/a17090405.
[5] S. Ghafarian *et al.*, "Clinical evaluation of corneal ulcer with a portable and smartphone-attachable slit lamp device: smart eye camera," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-025-87820-z.
[6] D. D. Pai, I. M. R. R, and H. Sunil, "Artificial intelligence in clinical decision support systems: architectures, applications, and ethical challenges," *International Journal of Engineering Research and Reviews*, vol. 14, no. 1, pp. 26–31, 2026.

[7] T. B. Alakuş and M. Baykara, "Classification and determination of severity of corneal ulcer using vision transformer based on the analysis of public image dataset of fluorescein-stained corneas," *Diagnostics*, vol. 14, no. 8, 2024, doi: 10.3390/diagnostics14080786.

[8] S. J. G. Susila, "Advanced detection and classification of corneal ulcer using DenseNet121 and SVM classifier," *Journal of Electrical Systems*, vol. 20, no. 7, pp. 1867–1883, May 2024, doi: 10.52783/jes.3892.

[9] O. Sevli, "Deep learning-based classification of corneal ulcers on ocular staining images," *Acta Infologica*, vol. 7, no. 2, pp. 281–292, Aug. 2023, doi: 10.26650/acin.1173465.

[10] S. Hu *et al.*, "Automatic diagnosis of infectious keratitis based on slit lamp images analysis," *Journal of Personalized Medicine*, vol. 13, no. 3, 2023, doi: 10.3390/jpm13030519.

[11] J. Gross, J. Breitenbach, H. Baumgartl, and R. Buettner, "High-performance detection of corneal ulceration using image classification with convolutional neural networks," in *54th Hawaii International Conference on System Sciences*, 2021, pp. 3416–3425, doi: 10.24251/HICSS.2021.415.

[12] D. Passos and P. Mishra, "A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks," *Chemometrics and Intelligent Laboratory Systems*, vol. 223, Apr. 2022, doi: 10.1016/j.chemolab.2022.104520.

[13] L. Deng, J. Lyu, H. Huang, Y. Deng, J. Yuan, and X. Tang, "The SUSTech-SYSU dataset for automatically segmenting and classifying corneal ulcers," *Scientific Data*, vol. 7, no. 1, Jan. 2020, doi: 10.1038/s41597-020-0360-7.

[14] H. Alquran, Y. Al-Issa, M. Alsalatie, W. A. Mustafa, I. A. Qasmieh, and A. Zyout, "Intelligent diagnosis and classification of keratitis," *Diagnostics*, vol. 12, no. 6, May 2022, doi: 10.3390/diagnostics12061344.

[15] K. Czarnacka *et al.*, "Diagnosis and treatment of corneal ulcers in the course of systemic autoimmune diseases," *Ophtha Therapy. Conservative Treatment*, vol. 10, no. 1, pp. 43–47, Mar. 2023, doi: 10.24292/01.OT.210323.4.

[16] A. A. Mahmoud, M. S. A.-Rahman, T. A. Mohamed, and D. M. El-Sebaity, "Is corneal collagen cross-linking beneficial as an adjunct to the conventional treatment of bacterial keratitis?" *The Open Ophthalmology Journal*, vol. 16, no. 1, pp. 1–8, Apr. 2022, doi: 10.2174/18743641-v16-e2203150.

[17] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.

[18] I. A. Qasmieh, H. Alquran, A. Zyout, Y. Al-Issa, W. A. Mustafa, and M. Alsalatie, "Automated detection of corneal ulcer using combination image processing and deep learning," *Diagnostics*, vol. 12, no. 12, Dec. 2022, doi: 10.3390/diagnostics12123204.

[19] T. Inneci and H. Badem, "Detection of corneal ulcer using a genetic algorithm-based image selection and residual neural network," *Bioengineering*, vol. 10, no. 6, May 2023, doi: 10.3390/bioengineering10060639.

[20] L. Lv, M. Peng, X. Wang, and Y. Wu, "Multi-scale information fusion network with label smoothing strategy for corneal ulcer classification in slit lamp images," *Frontiers in Neuroscience*, vol. 16, Nov. 2022, doi: 10.3389/fnins.2022.993234.

[21] M. Tan and Q. V. Le, "EfficientNetV2: smaller models and faster training," in *38 th International Conference on Machine Learning*, 2021, vol. 139, pp. 10096–10106.

[22] L. Alzubaidi *et al.*, "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, Apr. 2023, doi: 10.1186/s40537-023-00727-2.

[23] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[24] Y. Deng *et al.*, "Advanced and interpretable corneal staining assessment through fine grained knowledge distillation," *npj Digital Medicine*, vol. 8, no. 1, May 2025, doi: 10.1038/s41746-025-01706-y.

[25] F.-F. Li *et al.*, "Integrating prior knowledge with deep learning for optimized quality control in corneal images: a multicenter study," *Computer Methods and Programs in Biomedicine*, vol. 267, Jul. 2025, doi: 10.1016/j.cmpb.2025.108814.

[26] S. Mehrtabar *et al.*, "Ethical considerations in patient privacy and data handling for AI in cardiovascular imaging and radiology," *Journal of Digital Imaging*, vol. 39, pp. 2004–2025, 2026, doi: 10.1007/s10278-025-01656-7.

[27] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions," *Healthcare Analytics*, vol. 5, Jun. 2024, doi: 10.1016/j.health.2024.100340.

[28] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, 2024, doi: 10.1038/s41467-024-44824-z.

[29] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[30] D. S. W. Ting *et al.*, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167–175, 2019, doi: 10.1136/bjophthalmol-2018-313173.

[31] O. A. Ramwala *et al.*, "Establishing a validation infrastructure for imaging-based artificial intelligence algorithms before clinical implementation," *Journal of the American College of Radiology*, vol. 21, no. 10, pp. 1569–1574, Oct. 2024, doi: 10.1016/j.jacr.2024.04.027.

[32] D. C. Lepcha *et al.*, "Deep learning in medical image analysis: a comprehensive review of algorithms, trends, applications, and challenges," *Computer Modeling in Engineering & Sciences*, vol. 145, no. 2, pp. 1487–1573, Oct. 2025, doi: 10.32604/cmes.2025.070964.

[33] S. U. R. Khan, S. Asif, M. Zhao, W. Zou, Y. Li, and X. Li, "Optimized deep learning model for comprehensive medical image analysis across multiple modalities," *Neurocomputing*, vol. 619, no. 9, Feb. 2025, doi: 10.1016/j.neucom.2024.129182.

[34] G. Uzoechina and T. Osajiuba, "Deep learning in clinical diagnostics: a scoping review of innovations shaping future healthcare delivery," *Artificial Intelligence in Medicine*, vol. 146. Nov. 18, 2025, doi: 10.21203/rs.3.rs-7876598/v1.

[35] B. Kocak *et al.*, "Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations," *European Journal of Radiology: Artificial Intelligence,* vol. 3, Sep. 2025, doi: 10.1016/j.ejrai.2025.100030.

[36] H. Samma, A. S. B. Sama, and Q. S. Hamad, "Optimized deep learning model for medical image diagnosis," *Journal of Engineering Research*, vol. 13, no. 4, pp. 3147–3155, Dec. 2025, doi: 10.1016/j.jer.2024.11.003.

[37] H. Lu, Y. Lin, Z. Li, M. L. Yiu, Y. Gao, and S. Uddin, "Toward fair medical advice: addressing and mitigating bias in large language model-based healthcare applications," *Artificial Intelligence in Medicine*, vol. 168, no. 4, Oct. 2025, doi: 10.1016/j.artmed.2025.103216.

[38] A. M. Carrington *et al.*, "Deep ROC analysis and AUC as balanced average accuracy for improved classifier selection, audit and explanation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329–341, Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.

[39] Z. Lan, Y. Hu, S. Yang, M. Wu, Y. Huang, and H. Zhang, "Deep learning in ophthalmic image analysis: a task-driven review of segmentation, diagnosis, and progress prediction," *Neurocomputing*, vol. 667, Feb. 2026, doi: 10.1016/j.neucom.2025.132282.

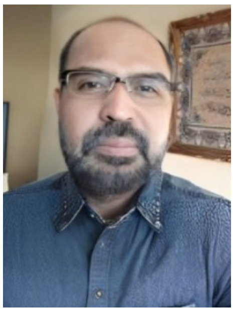
[40] S. Hisaria, P. Sharma, R. Gupta, and K. Sumalatha, "An analysis of multi-criteria performance in deep learning-based medical image classification: a comprehensive review," *npj Digital Medicine*, vol. 7, no. 1. pp. 1–12, Mar. 21, 2024, doi: 10.21203/rs.3.rs-4125301/v1.

BIOGRAPHIES OF AUTHORS

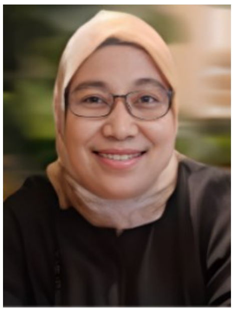
Rodiah is an associate professor, researcher, and lecturer, and currently serves as the vice head of Postgraduate Academic System Development at Universitas Gunadarma. Since 2012, she has received nine competitive research grants from the Indonesian Directorate General of Higher Education (DIKTI, Ministry of Research and Technology). Her research interests primarily focus on medical image processing. She can be contacted at email: rodiah@staff.gunadarma.ac.id.

Indah Sinthya Permata Sari is a lecturer at Universitas Gunadarma. Her research interests include artificial intelligence, deep learning, image analysis, and data-driven health informatics. She can be contacted at email: indahsinthya@gmail.com.

Matrissya Hermita is an associate professor at the Faculty of Psychology and a lecturer at the Faculty of Medicine, Universitas Gunadarma, Indonesia. She holds a bachelor's degree in Medicine, as well as master's and doctoral degrees in Communication and Psychology. Her research focuses on the biopsychosocial aspects of humans in the context of information and communication technology (ICT), particularly cognitive processes, user experience (UX), and the social dimensions of ICT. She can be contacted at email: matrissya@staff.gunadarma.ac.id.



Sarifuddin Madenda is a professor who holds a Doctorate in Electronics and Image Processing from Université de Bourgogne, France. He is currently active as the head of the Ph.D. program in Information Technology and serves as a lecturer in the doctoral program at Universitas Gunadarma, Indonesia. His research interests include image and video processing, multimedia data compression, content-based image, and video retrieval. He can be contacted at email: sarif@staff.gunadarma.ac.id.



Diana Tri Susetianingtias is an associate professor, researcher, and lecturer at Universitas Gunadarma. Since 2017, she has received four competitive research grants from the Indonesian Directorate General of Higher Education (DIKTI, Ministry of Research and Technology). Her research interests focus on medical image processing, particularly retinal fundus imaging, and retinal biometrics. She can be contacted at email: diants@staff.gunadarma.ac.id.