Multimodal Fusion Framework Using Contrastive Learning for Exposure Keratopathy

Gyutae Oh[1⋆], Yeokyoung Won[2], Donghui Lim[2⋆⋆], Jitae Shin[1⋆⋆⋆]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea
alswo740012@g.skku.edu, jtshin@skku.edu
[2] Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, South Korea
wyk900105@hanmail.net, ldhlse@gmail.com

Abstract. In this study, we propose a two-stage multimodal fusion learning framework for the automated grading of exposure keratopathy using four complementary imaging modalities: broad-beam, slit-beam, scatter, and blue-light, collected directly at a tertiary care center. In stage 1 (Grade Based Learning&Beam Based Learning), a backbone network is trained to capture and fuse the unique anatomical and pathological features inherent to each modality. In stage 2 (Dynamic Feature Fusion), we leverage the stage 1 pretrained backbone to train a modality agnostic classifier that, given only a single broad-beam image at inference time, implicitly exploits the rich multispectral information of the other three modalities. Experimental results demonstrate that our method achieves an average improvement of over 16% in both F1 score and overall accuracy (ACC) compared to single modality baselines. Ablation studies confirm the significant contribution of each component. By requiring only a single modality at inference, this framework is expected to maintain high diagnostic performance in real-world clinical settings while substantially reducing imaging requirements and patient burden. The code is publicly available at https://github.com/GYUGYUT/Multimodal-Fusion-Framework-Using-Contrastive-Learning-for-Exposure-Keratopathy.

Keywords: Exposure keratopathy, Multimodal Fusion, Contrastive Learning

1 Introduction

Exposure keratopathy is a clinical syndrome characterized by inflammation and damage to the cornea resulting from prolonged exposure of the ocular surface[18, 16]. It commonly occurs due to incomplete eye closure, as seen in conditions such

⋆ First author
⋆⋆ Corresponding author
⋆⋆⋆ Corresponding author

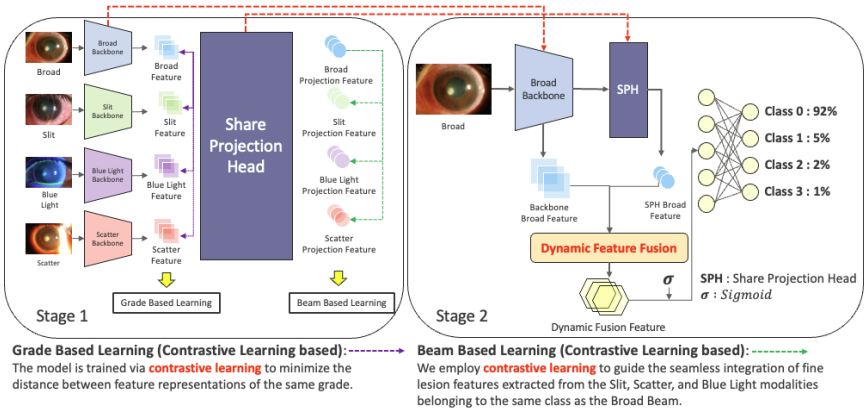

Fig. 1: This figure illustrates the overall architecture, showing the general workflow and example images for the broad-beam, slit-beam, blue-light, and scatter modalities. However, because the data used in this experiment are strictly prohibited from external distribution, we obtained prior permission from the first and corresponding authors to use similar images from [30] (CC BY 4.0) as alternatives.

as facial nerve palsy and eyelid malpositions, or reduced blink reflex[16, 18]. Elderly and critically ill patients, especially those in intensive care units (ICUs), are at significantly higher risk, due to factors such as compromised tear production, reduced blinking, lagophthalmos, and changes in vascular permeability, all of which exacerbate corneal exposure[16, 24, 26, 22]. Studies have reported that 20∼42% of ICU patients, and up to 57% of ventilated patients, experience exposure keratopathy[18, 16, 22]. Early diagnosis and prompt treatment of exposure keratopathy are essential, as the condition can often be effectively managed with simple interventions such as artificial tears or lubricating ointments[17]. However, in reality, many elderly and ICU patients face limited access to ophthalmologic care, making timely detection and management particularly challenging[19, 28]. For example, in long-term care facilities, ophthalmologists are often unavailable on-site, and patients may need to be transported to external facilities for specialized evaluation. This logistic barrier frequently delays diagnosis and treatment, leading to disease progression and vision-threatening complications such as persistent epithelial defects, corneal opacification, perforation, or even endophthalmitis[3]. In the diagnosis of corneal diseases, including exposure keratopathy, accurate assessment of lesion characteristics, including severity, location, depth, and the presence of infiltration requires dynamic examination using various illumination techniques under a slit-beam microscope. Thus, unlike fundus images, which typically involve a single, standardized imaging modality, anterior segment images inherently comprise multiple modalities that vary according to the slit-beam illumination methods: broad-beam[29], slit-beam[5], scatter[7], and blue-light[1] images. Accordingly, artificial intelligence

(AI) diagnostic models for retinal diseases, such as those by Ebrahimi et al. [8] and Pao et al. [25], are based on uniform fundus images. In contrast, the diagnosis of corneal diseases demand a more flexible imaging approach that integrates diverse modalities to capture the full spectrum of pathological features. Nevertheless, most existing AI studies on corneal diseases have utilized only a single imaging modality. For instance, Li et al. [20] and Hung et al. [12] proposed infectious keratitis classification models based solely on broad-beam images. In parallel, there is growing interest in multimodal approaches in medial AI, where diagnostic models integrate several types of data to enhance diagnostic performance[9, 13, 31]. For example, Ma et al. [21] significantly improved the accuracy of early biliary atresia diagnosis by combining ultrasound images with clinical data, and Kaczmarczyk et al. [14] demonstrated that a multimodal AI model fusing images and clinical data outperforms single-modal models. Hou et al. [11] reported similar findings. However, prior studies have predominantly focused on combining text-based clinical information (e.g., electronic medical records) with a single imaging modality or with large language models, offering limited examples of leveraging the unique characteristics of the diverse imaging actually used in clinical settings[2, 10, 32]. Therefore, we aimed to propose a two-stage AI framework for the diagnosis and severity prediction of exposure keratopathy based on four types of anterior segment images. In the first stage, we employ a contrastive learning-based approach to train the backbone network to effectively learn each modality specific features, enabling the detection of early lesion signs that may not be visible with a single modality alone[6, 27]. However, in medically underserved settings, acquiring all four modalities is often impractical due to limitations in time, cost, and equipment availability. To address this, the second stage of our framework introduces a classifier trained to use only a single imaging modality broad-beam image, which represents the most common format captured by standard cameras or smartphones at inference time, while implicitly leveraging the information learned from the other modalities during training. To our knowledge, this is the first work in exposure keratopathy to combine multimodal image fusion with single modality based classification.

Table 1: Train/Val/Test Data Distribution for Four Imaging Modalities

	Photo 1 (Broad-beam)					Photo 2 (Slit-beam)					Photo 3 (Blue-light)					Photo 4 (Scatter)				
Split	Total	0	1	2	3	Total	0	1	2	3	Total	0	1	2	3	Total	0	1	2	3
train	346	83	93	88	82	1035	149	398	238	250	603	76	256	184	87	552	121	212	142	77
val	43	10	12	11	10	129	19	50	29	31	75	9	32	23	11	69	15	26	18	10
test	44	11	12	11	10	130	18	50	30	32	76	10	32	23	11	69	15	27	18	9

2 Method

The method proposed in this paper comprises two sequential training stages. The overall training process is illustrated in Fig. 1. In the first stage, we employ two contrastive learning techniques (Grade Based Learning and Beam Based Learning) to ensure that the broad-beam features corresponding to the same grade naturally incorporate the fine lesion characteristics extracted from the slit-beam, scatter, and blue-light modalities of the same grade. To maximize the efficacy of contrastive learning, the grade composition within each modality's mini-batch is configured to match the broad-beam mini-batch in both quantity and positional arrangement. In the second stage, we introduce a Dynamic Feature Fusion strategy that adaptively integrates the features extracted by the backbone with those produced by a shared projection head. This fusion mechanism is designed to allow complementary interactions between heterogeneous feature representations, thereby enhancing the overall performance of the downstream classifier.

Grade Based Learning(GBL,Stage 1): As shown in Fig. 1, we define four imaging modalities broad-beam (b), slit-beam (s), blue-light (bl), and scatter (sc) as $\mathcal{M} = \{b, s, bl, sc\}$, $\mathcal{M} \subset \mathbb{R}^{H\times W}$. For any modality $m \in \mathcal{M}$, let I_m denote the input image of modality m, and B_m denote the modality specific backbone. The feature extracted through the backbone is expressed as $F_m = B_m(I_m) \in \mathbb{R}^{C\times H\times W}$. In this study, we partially modify the SupCon loss [15] by defining positive pairs as samples from different modalities sharing the same grade label as the broad-beam grade and training them to be close in the embedding space. The embeddings $\mathbf{F}_{b,i}, \mathbf{F}_{M,i} \in \mathbb{R}^{H\times W}$ obtained from two modalities are flattened and unit normalized to form $\mathbf{f}_{b,i}, \mathbf{f}_{M,i} \in \mathbb{R}^{D}$, and a loss is defined to reduce the distance between samples with the same grade label as follows: Here, N denotes the total number of normalized embedding vectors, i indexes the anchor embedding vector, j indexes positive candidates sharing the same grade as anchor i, and k indexes all embedding vectors compared against anchor i, including both positive and negative samples. Additionally, the embedding vector f_b is fixed as the anchor. $\tau > 0$ is the temperature parameter (set to 0.5), and the indicator function $\mathbb{I}[y_i = y_j] = \begin{cases} 1, & \text{if } y_i = y_j \\ 0, & \text{otherwise} \end{cases}$ is used. y is labels. By maximizing the exponentiated inner product between embeddings of the same grade, the distances between embeddings within each grade are effectively reduced. We refer to this as Grade Based Learning.

$$\mathcal{L}_{GBL(f_b,f_m)} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbb{I}[\,y_{b,i}=y_j\,]\,\exp\!\left(\mathbf{f}_{b,i}^{\top}\mathbf{f}_j\,/\,\tau\right)}{\sum_{k=1}^{N}\exp\!\left(\mathbf{f}_{b,i}^{\top}\mathbf{f}_k\,/\,\tau\right)}, \tag{1}$$

Beam Based Learning(BBL,Stage 1): This section explains the method by which broad-beam features of the same grade obtained in Stage 1 implicitly

reflect the fine lesion characteristics extracted from the slit-beam, scatter, and blue-light modalities of the same grade. The Shared Projection Head (SPH) used for Feature Embedding Projection follows [4], where a nonlinear projection layer is added after the Backbone output. This separates the Backbone representation from the contrastive loss space, encouraging the Backbone to learn more generalizable features. The SPH is shared across all B_m, and after passing through the SPH, the projection is expressed as $P_m = \mathrm{SPH}(F_m) \in \mathbb{R}^{B\times d}$, where d is the projection dimension and B is the batch size, both of which are configurable. Additionally, we partially modify the NT-Xent loss from [4] to enable P_b to implicitly learn the fine lesion features of other P_m. Applying L_2 normalization to both P_b and P_m, we generate unit vectors $z_b, z_m \in \mathbb{R}^{B\times d}$, which are stacked to form a matrix $R \in \mathbb{R}^{2B\times d}$. The cosine similarity matrix is then computed by $S = RR^\top \in \mathbb{R}^{2B\times 2B}$. A temperature parameter τ is applied to S. Since we need to compare positives and negatives, self similarity entries are masked to prevent interference during training. For indices, if $i \le B$, the positive sample index is defined as $i^+ = i+B$, and if $i > B$, $i^+ = i-B$. The positive similarity is $s_i^+ = S_{i,i^+}$, and the negative similarity is $s_{ij}^- = S_{i,j}, \quad j = 1,\ldots,2B, \quad j \neq i, \quad j \neq i^+$. The general form of the loss function is as follows:

$$\mathcal{L}_{BBL}(P_b, P_m) \;=\; \frac{1}{2B}\sum_{i=1}^{2B}\Bigg[-\log\frac{\exp\big(s_i^+\big)}{\displaystyle\sum_{\substack{j=1\\ j\neq i}}^{2B}\exp\big(S_{i,\,j}\big)}\Bigg]. \tag{2}$$

However, when performing Beam Based Learning, unlike the conventional NT-Xent method, the model is encouraged solely to generate feature vectors where the broad-beam modality is similar to other modalities without considering grade labels. This increases the likelihood that the model will fail to classify grades correctly. Therefore, to mitigate this issue, Grad Based Learning, which clusters samples of the same grade, is conducted beforehand. Consequently, the overall loss used in stage 1 can be defined as follows, where λ controls the influence of Beam Based Learning:

$$\mathcal{L}_{Stage1} = \sum_{m\in\{\,s,\,bl,\,sc\}} \mathcal{L}_{GBL}(f_b, f_m) + \lambda \sum_{m\in\{\,s,\,bl,\,sc\}} \mathcal{L}_{BBL}(P_b, P_m). \tag{3}$$

Dynamic Feature Fusion(Stage 2): Stage 2 introduces the method for training the classifier. We use only the B_b and SPH models learned in stage 1 and discard the other components. Accordingly, the training utilizes only data from the broad-beam modality. Through Grade Based Learning, we trained F_b to generate similar embedding vectors when it shares the same grade with F_s, F_{bl}, F_{sc}. Also, via Beam Based Learning, P_b was encouraged to have a distribution similar to P_s, P_{bl}, P_{sc}. This enables B_b to have robust characteristics for grade classification, and P_b, even when using only broad-beam as input, to generate features implicitly reflecting fine lesion characteristics from the slit-beam, scatter, and blue-light modalities. Therefore, to effectively utilize these two characteristics,

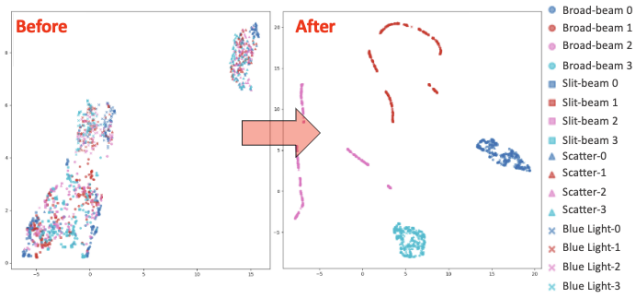

Fig. 2: This figure uses UMAP[23] to visualize the features before and after training with our proposed method for each independently trained backbone, grouped by modality and grade. Detailed information for each modality and grade is shown on the right side of the figure. Identical colors represent the same grade, and identical shapes represent the same modality.

we propose Dynamic Feature Fusion. First, since $f_b \in \mathbb{R}^D$ and $P_b \in \mathbb{R}^d$, we transform both embeddings to a common dimension D' via a simple MLP. The transformed embeddings are expressed as $h_1, h_2 \in \mathbb{R}^{D'}$ where D' is a tunable dimension. Next, the concatenated vector of h_1 and h_2 is passed through a sigmoid function to compute the weight α, and the features are finally fused as follows: where α is a scalar weight applied to each embedding vector.

$$h_{\text{fused}} = \alpha \cdot h_1 + (1 - \alpha) \cdot h_2 \tag{4}$$

Dynamic Feature Fusion is not simply an averaging of h_1 and h_2, but a method that dynamically adjusts the contribution of each feature through a learnable parameter α. This allows the model to autonomously assess the importance of the input features and combine them effectively. The extracted h_{fused} is first passed through a sigmoid activation(σ), and the resulting value is then used as the input to the final classifier.

3 EXPERIMENTS

3.1 Experiment Data collection & Setting :

The study was approved by the Institutional Review Board (IRB) of the Samsung Medical Center (IRB no. 2024-12-004). This study retrospectively collected data from 61 patients diagnosed with and treated for exposure keratopathy at the department of ophthalmology, Samsung Medical Center, between April 2008 and February 2025. All patients had multiple follow-up visits over the course of their illness, during which various types of anterior segment images were obtained. Images were captured using the SL-D7 (Topcon Healthcare Solutions, Inc.) and D850 (Nikon, Inc.) cameras. Each image was annotated by two ophthalmology specialists according to the severity of exposure keratopathy, categorized into

Table 2: Comparison of Individual and Combined Training Results for broad-beam, slit-beam, scatter, and blue-light Using Supervised Learning and Our Method.

Train_Method	Train_Data	Accuracy	Specificity	Sensitivity	F1 Score
Supervised	Broad-beam	0.7045	0.9008	0.7015	0.7049
Supervised	Slit-beam	0.6692	0.8787	0.6686	0.6510
Supervised	Blue-light	0.6447	0.8758	0.6407	0.6368
Supervised	Scatter	0.6957	0.8900	0.6685	0.6729
Supervised	Total	0.6865	0.8880	0.6783	0.6789
Ours	Stage 1:Total / Stage2:Broad	**0.8409**	**0.9462**	**0.8382**	**0.8379**

four grades: Normal, Mild, Moderate, and Severe. For normal data, the fellow eyes of patients diagnosed with exposure keratopathy, that were confirmed to have no abnormal findings upon examination, were included in the dataset. The entire collected dataset was split into training, validation, and test sets in an 8:1:1 ratio. The number of samples in each split is summarized in Table 1. For experimental consistency, all experiments were conducted under the following settings: image size of 224×224, GPU V100, batch size of 16, learning rate of 0.0001, early stopping patience of 10, embedding dimensions D, d, $D' = 256$, weighting parameter $\lambda = 0.1$, temperature parameter $\tau = 0.5$, DenseNet-121 as the backbone network, Adam optimizer, ReduceLROnPlateau learning rate scheduler, and PyTorch version 2.4.1+cu118. Additionally, evaluations were performed every 100 iterations.

3.2 Experiment Results

This section demonstrates that the proposed method more effectively integrates various image modalities compared to other approaches, leading to superior results. All experimental settings, including backbone architectures and hyperparameters, strictly followed the configurations described in Section 3.1 to ensure a fair comparison. The results of training on each single modality individually and the combined training on all four modalities are presented in Table 2. These results show that training with individual modalities generally yields suboptimal performance, and surprisingly, the combined training using all modalities (Total experiment), despite the substantially increased data volume, fails to achieve meaningful performance improvements. This suggests that conventional supervised learning alone is insufficient to effectively capture the complex lesion characteristics present across multiple image modalities. In contrast, the proposed method achieved an average improvement of over 16% in accuracy and an average improvement of 13% across other evaluation metrics. Notably, while most modalities exhibit F1-scores in the range of 60 to 70, the proposed method demonstrates superior performance, exceeding 83, strongly validating the necessity and effectiveness of integrated learning across multiple image modali-

ties. Furthermore, despite the data imbalance evident in Table 1, the proposed method achieved outstanding training results. As shown in Fig. 2, the proposed approach not only effectively learns the complex lesion characteristics across multiple image modalities but also enables clear separation between severity grades.

Table 3: Ablation Study: Performance Evaluation by Removing Grade Based Learning (GBL), Beam Based Learning (BBL), and Dynamic Feature Fusion (DFF).

Method			**Metrics**			
GBL	BBL	DFF	Accuracy	Specificity	Sensitivity	F1 Score
✓			0.7500	0.9154	0.7492	0.7563
	✓		0.7727	0.9235	0.7700	0.7710
✓		✓	0.7727	0.9232	0.7678	0.7685
	✓	✓	0.7954	0.9315	0.7969	0.7958
✓	✓		0.6590	0.8858	0.6640	0.6634
✓	✓	✓	**0.8409**	**0.9462**	**0.8382**	**0.8167**

3.3 Ablation study

We evaluated the impact of our proposed techniques on performance. All experiments were conducted under the same hyperparameter settings specified in Section 3.1. In experiments without dynamic feature fusion, the features extracted by the SPH module were fed into the final layer; for a fair comparison, we also evaluated the results of applying the GBL and BBL modules individually after SPH. In Table 3, checkmarks (✓) in the "Method" column indicate which modules were used in each experiment. The ablation studies showed that performance was maximally enhanced when all modules were used together, whereas individual application prevented effective feature fusion, yielding only marginal gains or even degraded performance. These results confirm that the combined use of our proposed techniques produces the optimal outcome.

4 Conclusion

This study addresses not only multimodal research but also the integration of multiple imaging modalities in a manner that closely reflects real-world diagnostic workflows. The proposed framework jointly learns complex features from multiple imaging modalities in stage 1, followed by the application of Dynamic Feature Fusion in stage 2 to effectively integrate these diverse representations. Remarkably, our method achieves an accuracy exceeding 84.09%, significantly outperforming existing approaches. Furthermore, it demonstrates that even in

the absence of all imaging modalities, accurate diagnosis can still be achieved using only a single broad-beam image. Comprehensive ablation studies further validate that the strategic combination of the proposed components leads to substantial performance gains. In conclusion, this study presents a practical and efficient framework for the rapid and accurate diagnosis of exposure keratopathy in clinical settings. The proposed method requires only a single-modal input at inference yet achieves superior classification performance compared to existing approaches, enabling early diagnosis that minimizes corneal damage and facilitates timely treatment.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgment

This work was supported by the BK21 FOUR Project. and This study was supported by SMC-SKKU Future Convergence Research Program Grant.

References

1. Bandamwar, K.L., Papas, E.B., Garrett, Q.: Fluorescein staining and physiological state of corneal epithelial cells. Contact Lens and Anterior Eye **37**(3), 213–223 (2014)
2. Bhattacharya, S., Prusty, S., Pande, S.P., Gulhane, M., Lavate, S.H., Rakesh, N., Veerasamy, S.: Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging. Frontiers in Human Neuroscience **19**, 1552178 (2025)
3. Bird, B., Dingley, S., Stawicki, S.P., Wojda, T.R.: Exposure keratopathy in the intensive care unit: Do not neglect the unseen. In: Vignettes in Patient Safety-Volume 2. IntechOpen (2018)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
5. Clover, J.: Slit-lamp biomicroscopy. Cornea **37**, S5–S6 (2018)
6. Curiel, L., Chopra, R., Hynynen, K.: Progress in multimodality imaging: truly simultaneous ultrasound and magnetic resonance imaging. IEEE transactions on medical imaging **26**(12), 1740–1746 (2007)
7. Denion, E., Béraud, G., Marshall, M.L., Denion, G., Lux, A.L.: Sclerotic scatter. Journal Français d'Ophtalmologie **41**(1), 62–77 (2018)
8. Ebrahimi, B., Le, D., Abtahi, M., Dadzie, A.K., Rossi, A., Rahimi, M., Son, T., Ostmo, S., Campbell, J.P., Paul Chan, R., et al.: Assessing spectral effectiveness in color fundus photography for deep learning classification of retinopathy of prematurity. Journal of biomedical optics **29**(7), 076001–076001 (2024)
9. Fukuzawa, F., Yanagita, Y., Yokokawa, D., Uchida, S., Yamashita, S., Li, Y., Shikino, K., Tsukamoto, T., Noda, K., Uehara, T., et al.: Importance of patient history in artificial intelligence–assisted medical diagnosis: comparison study. JMIR Medical Education **10**, e52674 (2024)

10. Hegselmann, S., von Arnim, G., Rheude, T., Kronenberg, N., Sontag, D., Hindricks, G., Eils, R., Wild, B.: Large language models are powerful electronic health record encoders. arXiv preprint arXiv:2502.17403 (2025)
11. Hou, C., Huang, T., Hu, K., Ye, Z., Guo, J., Zhou, H.: Artificial intelligence-assisted multimodal imaging for the clinical applications of breast cancer: a bibliometric analysis. Discover Oncology **16**(1), 537 (2025)
12. Hung, N., Shih, A.K.Y., Lin, C., Kuo, M.T., Hwang, Y.S., Wu, W.C., Kuo, C.F., Kang, E.Y.C., Hsiao, C.H.: Using slit-lamp images for deep learning-based identification of bacterial and fungal keratitis: model development and validation with different convolutional neural networks. Diagnostics **11**(7), 1246 (2021)
13. Ji, H., Kim, S., Sunwoo, L., Jang, S., Lee, H.Y., Yoo, S., et al.: Integrating clinical data and medical imaging in lung cancer: Feasibility study using the observational medical outcomes partnership common data model extension. JMIR Medical Informatics **12**(1), e59187 (2024)
14. Kaczmarczyk, R., Wilhelm, T.I., Martin, R., Roos, J.: Evaluating multimodal ai in medical diagnostics. npj Digital Medicine **7**(1), 205 (2024)
15. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020)
16. Kousha, O., Kousha, Z., Paddle, J.: Exposure keratopathy: Incidence, risk factors and impact of protocolised care on exposure keratopathy in critically ill adults. Journal of Critical Care **44**, 413–418 (2018)
17. Kousha, O., Kousha, Z., Paddle, J.: Incidence, risk factors and impact of protocolised care on exposure keratopathy in critically ill adults: a two-phase prospective cohort study. Critical Care **22**, 1–8 (2018)
18. Kuruvilla, S., Peter, J., David, S., Premkumar, P.S., Ramakrishna, K., Thomas, L., Vedakumar, M., Peter, J.V.: Incidence and risk factor evaluation of exposure keratopathy in critically ill patients: A cohort study. Journal of critical care **30**(2), 400–404 (2015)
19. Labreche, T., Stolee, P., McLeod, J.: An optometrist-led eye care program for older residents of retirement homes and long-term care facilities. Canadian Geriatrics Journal: CGJ **14**(1), 8 (2011)
20. Li, Z., Xie, H., Wang, Z., Li, D., Chen, K., Zong, X., Qiang, W., Wen, F., Deng, Z., Chen, L., et al.: Deep learning for multi-type infectious keratitis diagnosis: a nationwide, cross-sectional, multicenter study. NPJ Digital Medicine **7**(1), 181 (2024)
21. Ma, Y., Yang, Y., Du, Y., Jin, L., Liang, B., Zhang, Y., Wang, Y., Liu, L., Zhang, Z., Jin, Z., et al.: Development of an artificial intelligence-based multimodal diagnostic system for early detection of biliary atresia. BMC medicine **23**(1), 127 (2025)
22. McHugh, J., Alexander, P., Kalhoro, A., Ionides, A.: Screening for ocular surface disease in the intensive care unit. Eye **22**(12), 1465–1468 (2008)
23. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
24. Ousler III, G.W., Hagberg, K.W., Schindelar, M., Welch, D., Abelson, M.B.: The ocular protection index. Cornea **27**(5), 509–513 (2008)
25. Pao, S.I., Lin, H.Z., Chien, K.H., Tai, M.C., Chen, J.T., Lin, G.M.: Detection of diabetic retinopathy using bichannel convolutional neural network. Journal of Ophthalmology **2020**(1), 9139713 (2020)

26. Patel, V., Daya, S.M., Lake, D., Malhotra, R.: Blink lagophthalmos and dry eye keratopathy in patients with non-facial palsy: clinical features and management with upper eyelid loading. Ophthalmology **118**(1), 197–202 (2011)
27. Rodríguez-Palomares, J.F., Fernández, G., et al.: Integrating multimodal imaging in clinical practice: The importance of a multidisciplinary approach. Revista Espanola de Cardiologia (English ed.) **69**(5), 477–479 (2016)
28. Sommer, A., Taylor, H.R., Ravilla, T.D., West, S., Lietman, T.M., Keenan, J.D., Chiang, M.F., Robin, A.L., Mills, R.P., of the American Ophthalmological Society, C., et al.: Challenges of ophthalmic care in the developing world. JAMA ophthalmology **132**(5), 640–644 (2014)
29. Witmer, M.T., Kiss, S.: Wide-field imaging of the retina. Survey of ophthalmology **58**(2), 143–154 (2013)
30. Won, Y.K., Lee, H., Kim, Y., Han, G., Chung, T.Y., Ro, Y.M., Lim, D.H.: Deep learning-based classification system of bacterial keratitis and fungal keratitis using anterior segment images. Frontiers in Medicine **10**, 1162124 (2023)
31. Yang, L., Wan, Y., Pan, F.: Enhancing chest x-ray diagnosis with a multimodal deep learning network by integrating clinical history to refine attention. Journal of Imaging Informatics in Medicine pp. 1–16 (2025)
32. Zhu, L., Mou, W., Lai, Y., Chen, J., Lin, S., Xu, L., Lin, J., Guo, Z., Yang, T., Lin, A., et al.: Step into the era of large multimodal models: a pilot study on chatgpt-4v (ision)'s ability to interpret radiological images. International Journal of Surgery **110**(7), 4096–4102 (2024)