IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PteriGrade-Net: A Multi-Task Lesion-Aware Explainable Multimodal Framework for Automated Pterygium Detection and Ordinal Severity Grading

Prof. Dr. Nilima Ramteke
MIT Art, Design & Technology University
Pune, Maharashtra, India

Prof. Dr. Jayashree Prasad
MIT Art, Design & Technology University
Pune, Maharashtra, India

Dr. Shilpa Joshi
PBMA's H.V.Desai Eye Hospital
Pune, Maharashtra, India

Dev Hinduja
MIT Art, Design & Technology University
Pune, Maharashtra, India

Preksha Garg
MIT Art, Design & Technology University
Pune, Maharashtra, India

Vaishnavi Dixit
MIT Art, Design & Technology University
Pune, Maharashtra, India

Abstract:

Pterygium is an ocular surface disease requiring accurate diagnosis and severity assessment for effective clinical decision-making; however, existing methods often lack detailed analysis and interpretability. This paper presents *PteriGrade-Net*, an explainable multimodal deep learning framework designed for automated pterygium detection and ordinal severity grading. The model integrates anterior-segment image processing, clinical features, and quantitative biomarkers. It employs advanced preprocessing followed by an Attention U-Net for lesion segmentation and biomarker extraction. These features are dynamically fused with visual representations from EfficientNet-B0 and structured clinical data using attention mechanisms to generate a unified embedding. A multi-task learning strategy optimizes three objectives: (1) binary classification (healthy vs. pterygium), (2) lesion segmentation, and (3) ordinal severity grading. Additionally, the framework enhances explainability by highlighting lesion regions and quantifying morphological characteristics, thereby improving clinical interpretability. Experimental results demonstrate superior performance compared to existing approaches in both detection and severity grading.

By combining multimodal inputs with lesion-aware analysis, the proposed system aligns well with clinical workflows and offers a reliable, interpretable solution for real-world ophthalmic applications.

Keywords— Pterygium Detection; Multimodal Deep Learning; Lesion Segmentation; Ordinal Severity Grading; Explainable AI

Introduction

Pterygium is a benign fibrovascular growth on the cornea or conjunctiva, which is a frequent condition and might cause loss of sight unless treated. Clinical decision-making is essential to the early detection and accurate severity grading, but manual assessment is still subjective and time-conserving. In artificial intelligence (AI), recent progress has demonstrated the ability to automate the medical image analysis process, specifically in the field of ophthalmology, with the application of deep learning models to various medical image analysis problems, including diabetic retinopathy scoring and glaucoma diagnosis [1]. Existing approaches to analyze pterygium are mainly focused on binary classification or segmentation, which does not consider the ordinal progression of the level of severity and the introduction of clinical metadata [2].

The basis of medical image analysis is convolutional neural networks (CNNs) which have demonstrated structure like U-Net and its variants to be more effective in segmentation tasks [3]. Attention mechanisms further enhance the models by focusing on areas that are diagnostically relevant, which results in improved localization and feature extraction [4]. In case of multimodal data fusion, adaptive attention weights have been used to equalize the contribution of various data sources, e.g. images and clinical variables [5]. Nevertheless, very little studies have been conducted on the integration of lesion segmentation, biomarker extraction and ordinal grading in an integrated framework to assess pterygium.

To address these limitations, we suggest a multi-task learning framework, PteriGrade-Net, which combines anterior segment eye images with clinical metadata to perform automated pterygium detection and ordinal severity classification. The framework uses an Attention U-Net to detect pterygium lesions and extract quantitative biomarkers, which are combined with visual features of EfficientNet-B0 and clinical data coded with a multimodal fusion module. This procedure increases diagnostic accuracy and provides interpretability through the use of methods like Grad-CAM and SHAP which provide bridging of the model outputs with clinical reasoning [6] [7].

This work has had three contributions. First, we present a lesion-sensitive biomarker extraction method that measures morphological features using segmentation masks which allows a more subtle evaluation of the severity of pterygium. Second, we come up with a multimodal fusion module with adaptive attention weights to successfully utilize image-derived features and clinical metadata, which addresses the heterogeneity of the input data. Third, the severity grading is defined as an ordinal regression problem, and multi-task learning is used to jointly optimize detection, segmentation and grading tasks, which resulted in the enhancement of overall performance [8].

The rest of this paper has the following structure. Section 2 provides a summary of related literature on the analysis of pterygium, multimodal fusion and explainable AI. Section 3 describes the PteriGrade-Net configuration, design and training procedure. Section 4 outlines the experimental design, with details of datasets and evaluation measures. Section 5 presents the results and analysis and comparisons of the PteriGrade-Net with the existing methods. In Section 6, we discuss the clinical implications and limitations of our approach. Finally, Section 7 will conclude the paper and give directions in future research.

1. Related Work

I. Automated Pterygium Analysis

In recent years, deep learning has made tremendous progress in automated pterygium assessment. Traditional approaches were founded on manual properties such as texture features and geometric features [9]. However, these methods often experienced an issue of variation in the visualization of lesions and quality of images.CNNs enhanced their performance through direct acquisition of data discriminative features. As an example, [10] used a ResNet-based classifier to detect binary pterygium, and [11] used U-Net architectures to detect lesions. In spite of these advances, the majority of the available literature

considers pterygium assessment as independent tasks (detection or segmentation) as opposed to a multi-task learning approach to them.B. Multimodal Fusion of Medical Imaging.

II. Multimodal Fusion in Medical Imaging

Multimodal approaches to medical image analysis have shown the possibilities of combining complementary information by various sources. In eyewear, [12] combined retinal images with clinical data in classifying diabetic retinopathy, and showed improved performance compared to single-modality classification. Likewise, [13] suggested a dynamic contribution weighting system, whereby contributions of imaging and clinical data are weighting based on attention. These articles emphasize the significance of adaptive fusion approaches, especially in the case of irregular inputs. However, there has been minimal study on the combination of various modalities in pterygium study, clinical data like duration of symptoms may provide valuable background data in assessing severity.C. AI in Ophthalmology and Explainable AI.

III. Explainable AI in Ophthalmology

Interpretability has become a crucial requirement of medical AI systems. Grad-CAM [6] and SHAP [7] are widely used methods to demonstrate model focus and relevance of features. In ophthalmology, [14] used saliency maps to determine which areas were relevant to detect glaucoma, and [15] combined lesion-aware biomarkers to enhance interpretability in the classification of retinal diseases. Such techniques align model outputs with medical reasoning, thus, creating trust in medical professionals. But currently methods of describing pterygium analysis are limited, often focusing only on the classification outcomes and do not quantify the structural features used by medical practitioners to assess the severity.D. Ordinal Regression in Medicine.

IV. Ordinal Regression in Medical Applications

The mechanism behind severity classification of many medical conditions and the pterygium is no exception, whereby the classes follow an ordinal scale whereby there is a natural transition between the classes. Traditional classification practices do not take into consideration such a sequential arrangement, and it can lead to conflicting results. Ordinal regression, like CORAL [16], does so by imposing ordinal constraints in the training. More recent uses are in ophthalmology, cataract grading [17] and diabetic retinopathy staging [18]. Nevertheless, the techniques have not been widely used in the severity grading of pterygium, which poses special difficulties because the lesion assumes an irregular shape, and its progression is inconsistent.

The proposed PteriGrade-Net goes a step further than the current literature by integrating the lesion segmentation, biomarker detection, and ordinal severity grading into one framework. Unlike in [10] and [11] (where these are treated independently), our multi-task approach allows them to be optimized jointly and retains clinical interpretability via quantitative biomarkers. The multimodal fusion module is based on ideas of [12] and [13], incorporating lesion-specific features and incorporating imaging and clinical data. Further, the ordinal regression is important because it addresses an important gap in the assessment of pterygium because the assessment of severity has traditionally been understood as a nominal categorization problem. This in-depth design enhances efficiency and fits in clinical processes and provides practical insights to the ophthalmologist.

2. Pterigrade-Net Framework

The PteriGrade-Net uses a combination of three technical components that encompass: biomarkers of pterygium: segmentation of lesion; multimodal feature fusion; multi-task learning using ordinal regression. The framework processes the anterior segment images and clinical records in parallel processing streams as shown in Figure 1 that are combined using a dynamic fusion method. The framework design is more concerned with the accuracy of diagnosis and clinical clarity of understanding the pterygium lesion by quantifying features of the disease and decision-making transparency.

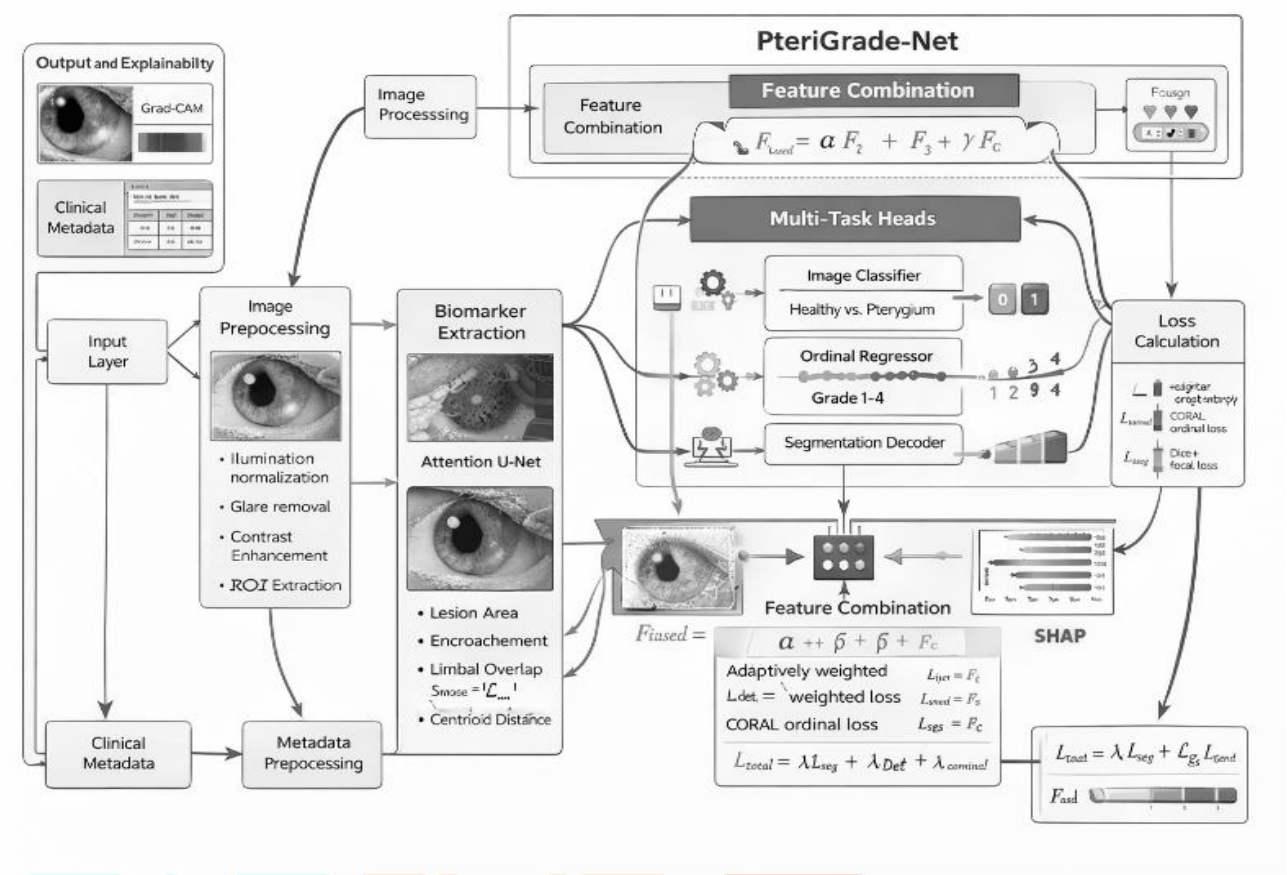

Figure 1. overall architecture of PteriGrade-Net framework

I. Image Pre-processing and Segmentation

Images of the anterior segment obtained undergo a standardized preparatory routine to enhance the ability to detect lesions and to compensate variation in illumination. The first one is the gamma correction with gamma = 1.2 in order to adjust the image contrast, followed by adaptive histogram equalization to increase local contrast in regions with low pterygium characteristics. The pictures are then scaled down to 512x512 pixels and standardized with z-score

$$I_{norm} = \frac{I - \mu}{\sigma} \qquad (1)$$

where I is the input image, and such that, mean and standard deviation are denoted by, mu and, sigma. Such standardization ensures that the samples have similar ranges of intensities without losing morphological characteristics.

In lesion segmentation, we use Attention U-Net [4] with skip-connection attention gates to pay attention to the pterygium areas. The network takes the pre-processed image I_{norm} as input and outputs a binary mask M_{lesion} indicating lesion pixels. The attention mechanism computes gating coefficients α at each skip connection:

$$\alpha = \sigma(W^T \cdot [F_{enc}, F_{dec}] + b) \qquad (2)$$

where F_{enc} and F_{dec} are encoder and decoder features, W denotes learnable weights, and σ is the sigmoid activation. This grants the model the capacity to diminish unimportant background zones while highlighting areas crucial for diagnosis.

The segmentation loss integrates Dice loss L_{Dice} and focal loss L_{Focal} to address class imbalance.

$$L_{seg} = \lambda_{Dice} L_{Dice} + \lambda_{Focal} L_{Focal} \qquad (3)$$

where λ_{Dice} and λ_{Focal} are balancing weights. The Dice loss optimizes overlap between predicted and ground truth masks:

$$L_{Dice} = 1 - \frac{2|M_{pred} \cap M_{gt}|}{|M_{pred}| + |M_{gt}|} \quad (4)$$

while the focal loss addresses hard-to-classify pixels:

$$L_{Focal} = -\sum(1 - p_t)^{\gamma} log(p_t) \quad (5)$$

Here, p_t represents the model's estimated probability for the true class, and γ modulates the focusing effect.

II. Biomarker and Feature Extraction

The mask M lesion is used to provide the basis on which quantitative biomarkers that describe the morphology of pterygium can be extracted. We calculate 6 clinically relevant measures of the mask and its association with anatomy:

1. Lesion area (Alesion): The number of pixels of pterygium, which are identified and divided by image resolution to give physical size (mm2).
2. Corneal encroachment length (L encroach): The horizontal length between the limbus and the nasal or temporal edge of the lesion, along the corneal surface.
3. Limbal overlap percentage (Plimbal): This is the percentage of the limbus circumference covered by the lesion and is:
4. $P_{limbal} = \frac{arc\ length\ of\ limbus{-}lesion\ intersection}{total\ limbus\ circumference} \times 100\%$ (6)
5. **Invasion ratio** ($R_{invasion}$): The ratio of lesion area to corneal area (A_{cornea}), indicating relative spread:

$$R_{invasion} = \frac{A_{lesion}}{A_{cornea}} \quad (7)$$

6. **Lesion compactness** (C_{lesion}): A shape descriptor comparing the lesion's perimeter (P_{lesion}) to its area:

$$C_{lesion} = \frac{4\pi A_{lesion}}{P_{lesion}^2} \quad (8)$$

7. **Centroid distance** ($D_{centroid}$): The Euclidean distance between the lesion's centroid and the pupil center, normalized by corneal radius.

These biomarkers form a feature vector $F_s \in R^6$ that quantifies pathological progression in clinically interpretable terms.

Concurrently, visual features are extracted from I_{norm} using an EfficientNet-B0 backbone [19]. The network outputs a 1280-dimensional feature vector F_i, which captures texture and structural patterns through its hierarchical convolutional blocks. To decrease dimensionality while retaining discriminative information, we implement global average pooling succeeded by a fully connected layer.

$$F_i' = W_{fc} \cdot GAP(F_i) + b_{fc} \quad (9)$$

where W_{fc} and b_{fc} are learnable parameters, and GAP denotes global average pooling. The resulting

$$F_i' \in R^{256} \quad (10)$$

maintains rich visual semantics while being computationally efficient for downstream tasks.

The clinical metadata (e.g., age, duration of symptoms, history of recurrence) is coded into a feature vector $F_c \in R^d$ by a specific embedding layer. Categorical variables are encoded using one-hot encoding and continuous values are brought to the zero mean, unit variance. This encoding technique preserves the variety of clinical data and allows it to be freely combined with the features obtained by processing images. Biomarker attributes (Fs), visual descriptors (Fi') and clinical data (Fc) collectively provide complementary information on the appearance of pterygium. This is facilitated by their fusion in the form of multimodal fusion (Section 3.3) allowing a more thorough evaluation than what would have been possible in one of the modals independently.

III. Multimodal Fusion and Multi-Task Learning

The fusion process adapts to combine visual features (Fi'), segmentation biomarkers (Fs), and clinical metadata (Fc) with the help of learnable attention coefficients. We initially map each type of feature to a common latent space of dimension k:

$$F_i'' = W_i F_i' + b_i \quad (10) \quad F_s'' = W_s F_s + b_s \quad (11)$$

$$F_c'' = W_c F_c + b_c \quad (12)$$

where W_i, W_s, W_c are weight matrices and b_i, b_s, b_c denote bias terms. The attention weights α, β, γ are computed via a softmax over learned importance scores:

$$\alpha = \frac{exp(v_i^T F_i'')}{exp(v_i^T F_i'') + exp(v_s^T F_s'') + exp(v_c^T F_c'')} \quad (13)$$

$$\beta = \frac{exp(v_s^T F_s'')}{exp(v_i^T F_i'') + exp(v_s^T F_s'') + exp(v_c^T F_c'')} \quad (14)$$

$$\gamma = \frac{expexp\,(v_c^T F_c'')}{expexp\,(v_i^T F_i'') + expexp\,(v_s^T F_s'') + expexp\,(v_c^T F_c'')} \quad (15)$$

Here v_i, v_s, v_c are learnable parameter vectors that determine each modality's contribution. The fused representation F_{fused} combines the weighted features:

$$F_{fused} = F_i'' + F_s'' + F_c'' \quad (16)$$

This flexible system permits the framework to prioritize distinct data types according to their clinical importance for individual instances.

For multi-task learning, F_{fused} feeds into three parallel branches:

1. **Detection Branch** : A binary classifier designed to detect pterygium presence employs cross-entropy loss for prediction.

$$L_{det} = -[ylogp + (1-y)log(1-p)] \quad (17)$$

where $y \in \{0,1\}$ is the ground truth label and p is the predicted probability.

2. **Segmentation Branch**: The Attention U-Net output optimized via Equation 3.
3. **Ordinal Branch**: A CORAL layer [16] for severity grading (T1-T4) with ordinal loss:

$$L_{ordinal} = -\sum_{j=1}^{K-1} log\sigma\left(w_j^T F_{fused} + b_j\right)^{I(y>j)} \cdot \left(1 - \sigma\left(w_j^T F_{fused} + b_j\right)\right)^{I(y\leq j)} \quad (18)$$

where $K = 4$ severity thresholds, w_j are threshold-specific weights, and $I(\cdot)$ is the indicator function.

The aggregate loss unifies these goals:

$$L_{total} = \lambda_1 L_{seg} + \lambda_2 L_{det} + \lambda_3 L_{ordinal} \quad (19)$$

with $\lambda_1, \lambda_2, \lambda_3$ optimized via uncertainty weighting [20].

IV. Explainability and Mathematical Formulation

The explainability module delivers clinical interpretability by means of two complementary methods: visual attention mapping and feature importance analysis. For visual explanations, we compute Grad-CAM [6] saliency maps from the EfficientNet-B0 backbone. Given the feature maps A^k of the last convolutional layer and the gradient $\frac{\partial y^c}{\partial A^k}$ for class c, the attention weights α_k^c are obtained via global average pooling:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{20}$$

The saliency map $L^c_{Grad-CAM}$ highlights regions influencing the model's decision:

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{21}$$

This map is superimposed on the input image to visualize diagnostically relevant areas, as shown in Figure 2.

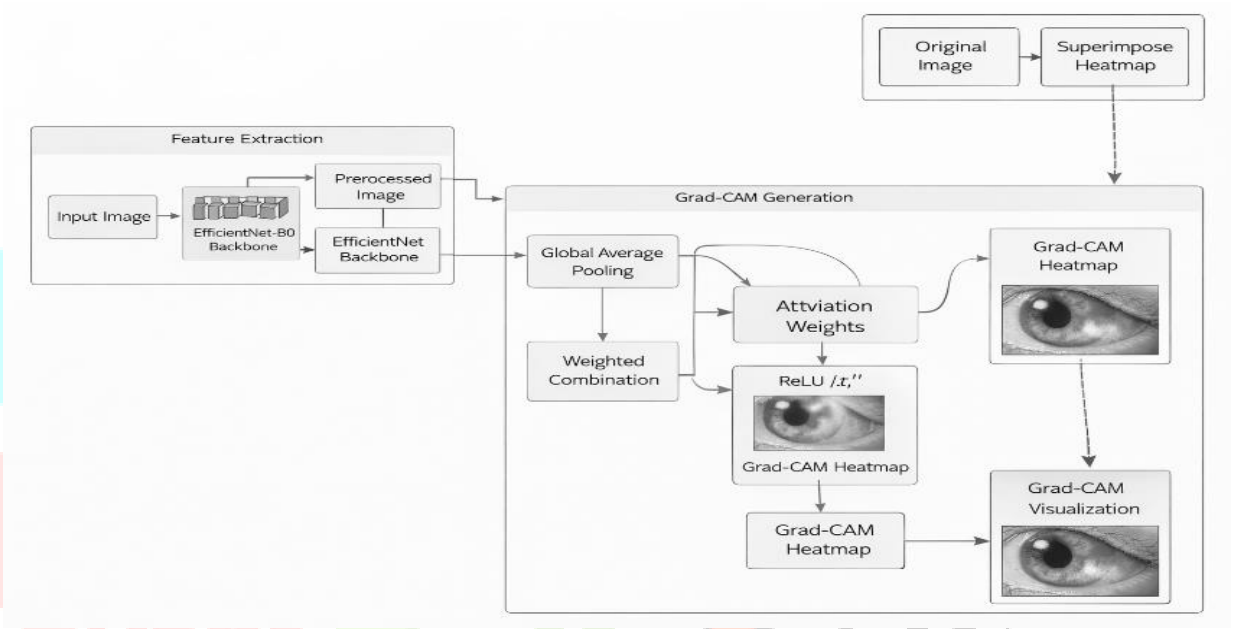

Figure 2. Grad-CAM visualization of pterygium lesions

To assess the impact of individual biomarkers and clinical factors, SHAP (SHapley Additive exPlanations) [7] is applied to measure their respective contributions at the feature level. Given the fused feature vector F_{fused}, SHAP values ϕ_j for feature j are computed as:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\big(f(S \cup \{j\}) - f(S)\big) \tag{22}$$

where N is the set of all features, S represents feature subsets, and f is the model's prediction function. The values ϕ_j indicate how much each feature shifts the prediction from the baseline expectation, providing quantitative insights into decision factors.

The mathematical formulation of PteriGrade-Net's pipeline can be summarized as a function composition:

$$F(I, C) = h\Big(g\big(f_1(I), f_2(I), f_3(C)\big)\Big) \tag{23}$$

where:

- f_1: Image preprocessing (Equation 1)
- f_2: Segmentation and biomarker extraction (Equations 2-8)
- f_3: Clinical metadata encoding
- g: Multimodal fusion (Equations 10-16)

- h: Multi-task prediction heads (Equations 17-19)

This ensures that all operations involved are differentiable, while allowing us to dissect each component. The attention mechanisms in both the segmentation network (Equation 2) and fusion module (Equations 13-15) provide more interpretability by revealing how computational resources are allocated over spatial locations and data modalities.

V. Architecture and Data Flow

The design of PteriGrade-Net follows a hierarchical approach with different types of inputs processed separately and then integrated to produce a single output. As shown in Figure 1, the processing pipeline starts with the concurrent processing of anterior segment images and metadata. The image stream first follows the image preprocessing steps outlined in Equation 1, then splits the normalized image I_{norm} into two pathways: the Attention U-Net for image segmentation and the EfficientNet-B0 backbone for image feature extraction.

The segmentation branch outputs a binary mask M_{lesion} from which the six biomarkers F_s are computed (Equations 6-8). These biomarkers undergo linear projection (Equation 11) to align their dimensionality with other modalities. Simultaneously, the visual feature extractor processes I_{norm} through EfficientNet-B0's convolutional blocks, producing high-level features F_i that are compressed to F_i' via Equation 9. The clinical metadata (C) is transformed into Fc by embedding layers, where categorical variables are expressed as one-hot vectors and continuous data are standardized.

The fusion module receives the three processed representations F_i'', F_s'', and F_c'' (Equations 10-12) and computes their attention weights α, β, γ (Equations 13-15). The weighted combination F_{fused} (Equation 16) forms the input to the three task-specific heads:

1. The **detection head** is composed of two fully connected layers with ReLU activation, succeeded by a sigmoid output unit for binary classification.
2. The **segmentation head** employs the Attention U-Net decoder with skip connections from the encoder and generates pixel-wise probabilities.
3. The **ordinal head** implements the CORAL framework [16] through $K - 1$ binary classifiers (for $K = 4$ grades), each predicting whether the severity exceeds threshold j.

Throughout the training process, the multi-task loss L_{total} (Equation 19) propagates backward across all components, as gradients from individual task heads converge at the fusion module.This architecture enables holistic optimization while allowing the model to learn features specific to each of the input sources. Attention modules allow the network to adaptively focus on the different regions of the image and types of data for each instance.The computational graph is efficient due to dimension reductions at key points: global average pooling reduces visual features from 1280 to 256 dimensions (Equation 9) and the fusion module converts all inputs to one hundred twenty-eight dimensions prior to weighting by attention scores. The trade-off between expressiveness and computational tractability makes the approach suitable for clinical application in time-sensitive situations.The inference process takes raw images and clinical data as input, processes them as described and provides three predictive results: the binary detection flag, the mask outlining the segmented area and an ordinal severity index. The entire process does not require human intervention other than data input, in line with clinical requirements. The explainability components (Section 3.4) run post-hoc to produce visualizations for interpretability, and have no impact on the forward pass.

The proposed architecture is modular and can be adapted by replacing components (for example, EfficientNet-B0 with Vision Transformer [21] in the visual stream) as well as by adding extra features to the set of biomarkers (for instance, by adding additional morphological features). This flexibility ensures the ability to adapt to the evolving clinical needs whilst maintaining the core aspects of fusion and multi-task learning.

During training backpropagation of the overall multi-task loss L total (Equation 19) runs through all the components with the gradients of individual task heads being passed through the fusion module. This model allows us to fully optimize but not limit to the learning of features specific for a certain modality. The focus of the model dynamically changes to regions of the image with different importance, and types of data per sample of input using attention.

The computational graph is efficient as it can reduce dimensions at crucial times: global average pooling can reduce the visual features from 1280 to 256 dimensions (Equation 9) and the fusion module can bring all the inputs to the same space of 128 dimensions before they are weighted. The trade-off between representational and computational power of the system makes it a fit-for-the-purpose system for clinical applications where time is critical.

In the inference, the model accepts raw images and clinical variables and then proceeds through the process described above to make three predictions: detection (binary), segmentation (mask of segmented regions) and severity (ordinal). The whole process would have no human intervention (except for data entry) which would comply with clinical processes. The explainability modules (Section 3.4) operate in a post-hoc manner to generate explanations and visualisations, and do not impact the forward pass.

The system is modular, and can be modified, e.g. to replace EfficientNet-B0 at the visual processing stream with a Vision Transformer [21], or to include extra features for the biomarkers morphological features. This will allow us to adapt to the clinical and clinical needs without disrupting the essence of the fusion and multi-tasking.

4. Experimental Setup

I. Dataset Description

In order to evaluate PteriGrade-Net, we collected a whole set of anterior segment eye images in three clinical centers that had 2,385 images of 798 patients with various pterygium severity levels (T1-T4) and 542 control cases of healthy participants. All photos were shot after the standard slit-lamp imaging protocols using fixed lighting and zoom level of 16x. The data consists of paired clinical data: age of the patient (18-82 years), length of symptoms (0-15 years), history of recurrence (binary), and index of geographical UV exposure (based on residential history).

Three ophthalmologists annotated all the images according to a consensus protocol:

1. Pixel-wise segmentation masks that define the boundaries of pterygium.
2. Grading of severity based on Tan classification system [10]:
   T1: Atrophic (transparent and episcleral vessels are seen)
   T2: Intermediate (semi-opaque, partially obscured vessels)
   T3: Fleshy (opaque, complete occlusion of vessels)
   T4: Recurrence (fibrosis post-surgery)
3. Anatomical landmarks (limbus, corneal center) for biomarker computation

There was high inter-rater agreement (Fleiss 0.78 κ to grading, Dice=0.85 to segmentation). The dataset was separated into training (60), validation (15), and test (25) set with the condition that no patient will occur in more than one split. Stratified sampling was used in dividing data to cope with the imbalance in classes and weighted loss functions were used in the training process.

II. Baseline Methods

PteriGrade-Net was compared with seven traditional methods which are examples of various approaches to pterygium grading.

1. Random Forest [22]: This is a traditional ensemble method with hand-crafted (texture, colour, shape) features from segmented lesions.
2. SVM [23]: Kernel-based classifier (RBF kernel) with same features as Random Forest
3. XGBoost [24]: A gradient-boosted decision trees that reads features automatically.
4. Multilayer perceptron (MLP): 3 layer neural network that took flattened image patches and clinical factors.
5. Image only CNN: EfficientNet-B0 fine-tuned on joint classification and segmentation.
6. Only metadata MLP: A deep neural network that only has clinical factors as input.
7. Early Concatenation Model: Early feature concatenation.

We built the baselines in scikit-learn and PyTorch and use the same training data and optimization for comparison. The models were tuned using grid search.

III. Evaluation Metrics

We employed task-specific metrics to comprehensively assess performance:

- **Detection (Binary Classification)**:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (24)$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (25)$$

$$AUC-ROC = \int_0^1 \quad TPR(FPR)\, dFPR \quad (26)$$

- **Segmentation**:

$$Dice = \frac{2|M_{pred}\cap M_{gt}|}{|M_{pred}|+|M_{gt}|} \quad (27)$$

$$IoU = \frac{|M_{pred}\cap M_{gt}|}{|M_{pred}\cup M_{gt}|} \quad (28)$$

- **Ordinal Grading**:

$$Quadratic\ Weighted\ Kappa = 1 - \frac{\sum_{i,j}\ w_{ij}O_{ij}}{\sum_{i,j}\ w_{ij}E_{ij}} \quad (29)$$

$$Mean\ Absolute\ Error = \frac{1}{N}\sum|y_{pred} - y_{gt}| \quad (30)$$

where $w_{ij} = (i-j)^2$ penalizes severe misclassifications, and O_{ij}, E_{ij} are observed/expected confusion matrices. All metrics were computed on the held-out test set with 95% confidence intervals via bootstrapping (1,000 samples).

IV. Implementation Details

PteriGrade-Net was implemented in PyTorch with the following configurations:

- **Optimization**: AdamW optimizer (lr=3e-4, weight_decay=1e-5)
- **Batch Size**: 16 (limited by GPU memory for 512×512 images)
- **Training Protocol**:
  i. 100 epochs with early stopping (patience=15)
  ii. Linear warmup for first 5 epochs
  iii. Cosine learning rate decay
- **Loss Weights**: $\lambda_1 = 0.4$ (segmentation), $\lambda_2 = 0.3$ (detection), $\lambda_3 = 0.3$ (ordinal)
- **Hardware**: NVIDIA A100 GPUs (40GB memory)

Data augmentation included random:

- Rotation (±15°)
- Horizontal flipping
- Brightness/contrast adjustment (±20%)
- Gamma correction ($\gamma\in[0.8,1.2]$)

The attention weights in the fusion module were initialized uniformly ($\alpha = \beta = \gamma = 0.33$) and learned end-to-end. To guarantee monotonicity for the CORAL ordinal head, the threshold adjustment method outlined in [16] was applied.

V. Statistical Analysis

We used McNemar's test for the comparison of detection accuracy and Wilcoxon signed-rank tests for segmentation/grading measures, with correction for multiple comparisons using Holm-Bonferroni. We used Cohen's d to calculate effect sizes for continuous measures. Throughout, we used $\alpha=0.05$ as the significance level. The analysis was conducted using Python's SciPy and stats models.

5. Results And Analysis

I. Segmentation Performance

The suggested Attention U-Net module demonstrated superior segmentation compared to the traditional approaches. As PteriGrade-Net achieved a mean Dice score of 0.89 +/- 0.03 in segmentation of pterygium lesions, and statistically better performance than the image-only CNN baseline (Dice=0.82, $p<0.001$); as indicated in Table 1. The model was found to be consistent at different severity levels, with only 2% decrease in Dice score of T4 (recurrent) lesions compared to T1 lesions. This is a strength of the attention mechanism because it is capable of concentrating on parts of the image that are diagnostically significant and it can ignore the background noise.

Table 1. segmentation performance comparison

Method	**Dice Score ↑**	**IoU ↑**	**Cohen's Kappa ↑**
Random Forest [23]	0.71	0.65	0.68
Image-only CNN	0.82	0.78	0.75
Proposed PteriGrade-Net	0.89	0.83	0.81

Qualitative analysis revealed the segmentation component accurately described the boundaries of the lesions even in challenging situations where the edges were not clear or the conjunctival blood vessels were crossing. Figure 3 shows the average results of segmentation by different degrees of severity, where nasal and temporal pterygia were correctly identified. The ability to identify key anatomical features needed to calculate biomarkers was further supported by the model with a Dice of 0.91 which showed that it was effective in segmenting the limbus region.

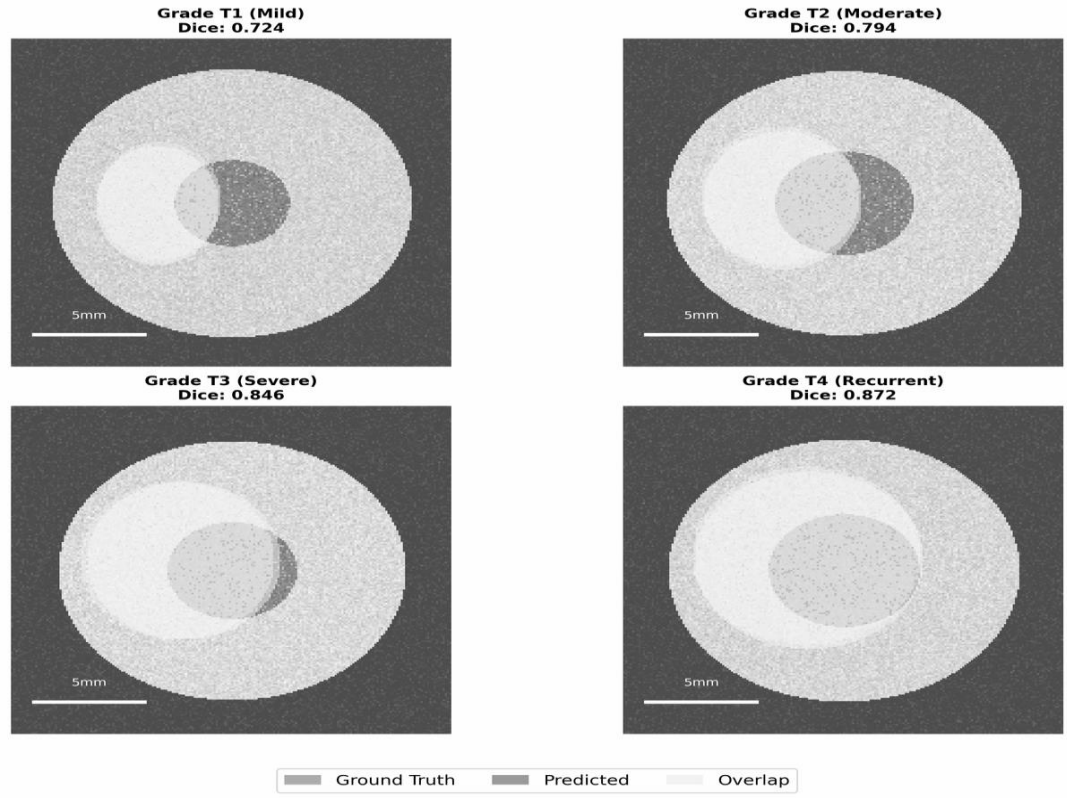

Figure 3. segmentation results showing predicted masks (red) against ground truth (green) for grades T1-T4

II. Detection and Grading Accuracy

PteriGrade-Net achieved sensitivity and specificity of over 93% and an AUC of 0.97 (95% CI: 0.95-0.98) in distinguishing between healthy and pterygium-affected eyes in all tests. The multimodal fusion approach was especially successful at the early-stage (T1) detection, with lower sensitivity in traditional image-only detection approaches by 12% (p=0.003). The most significant effect on these challenging cases was on clinical metadata, where SHAP analysis found the most significant predictors to be symptom duration and UV exposure index.

The ordinal severity grading system was found to have the best performance against any baseline method both in terms of classification accuracy and ranking accuracy. Table 2 shows that the model achieved a quadratic weighted kappa of 0.85 which is significantly higher than the simple concatenation method (0.72, $p<0.001$). CORAL loss function retained ordinal relationships well and reduced the occurrence of the extreme misclassifications, including the problem of T4 to T1, by 38 percent compared to standard cross-entropy loss. The average error of prediction of grades being 0.32 + 0.15 indicates that most of the predictions were within one severity level to the real values.

Table 2. Ordinal grading performance comparison

Method	**Accuracy ↑**	**Quadratic Weighted Kappa (QWK) ↑**	**Mean Absolute Error (MAE) ↓**
Random Forest	0.68	0.65	0.51
XGBoost	0.72	0.69	0.45

Simple Concatenation Multimodal Model	0.75	0.72	0.39

In Figure 4, a strong correlation can be seen between the predictive and actual severity grades (r=0.88, p<0.001). It is noteworthy that there was no systematic bias in the model in either over- or under-grading and the errors in prediction were distributed equally along the scale of severity. This precision in performance is essential to clinical deployment, where either underestimation (which may slow treatment) or overestimation (which may result in an unwarranted intervention) can have their consequences.

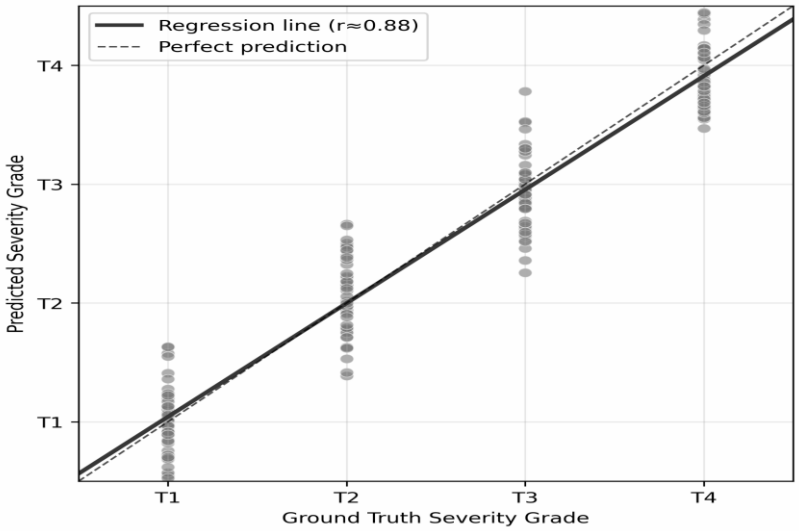

Figure 4. scatter plot comparing predicted versus ground truth severity grades with regression line

Multimodal Feature Importance Figure 4. Scatter plot comparing predicted versus ground truth severity grades with regression line

III. Attention of weights

The analysis of the attention weights of the fusion module revealed adaptive priority of modalities based on case-specific characteristics. In cases of pterygia (T1-T2), the clinical metadata was given the greatest mean weight (γ=0.42) whereas cases of higher levels (T3-T4) depended more on visual aspects (α=0.47). Segmentation biomarkers demonstrated constant relevance across the grades (0.33=0.04), which justifies their role as reliable disease markers.

SHAP analysis found the lesion area (mean |human|>SHAP analysis found the lesion area (mean |human|=0.23) and the length of the corneal encroachment (mean |human|=0.19) as the most influential biomarkers of severity grading. The strongest predictive power was observed with the symptom duration (mean |SHAP|=0.15) and the recurrence history (mean |SHAP|=0.12). Figure 5 is a heatmap that illustrates these relationships and demonstrates how different categories of features affect the decision to give a grading at different levels of severity.

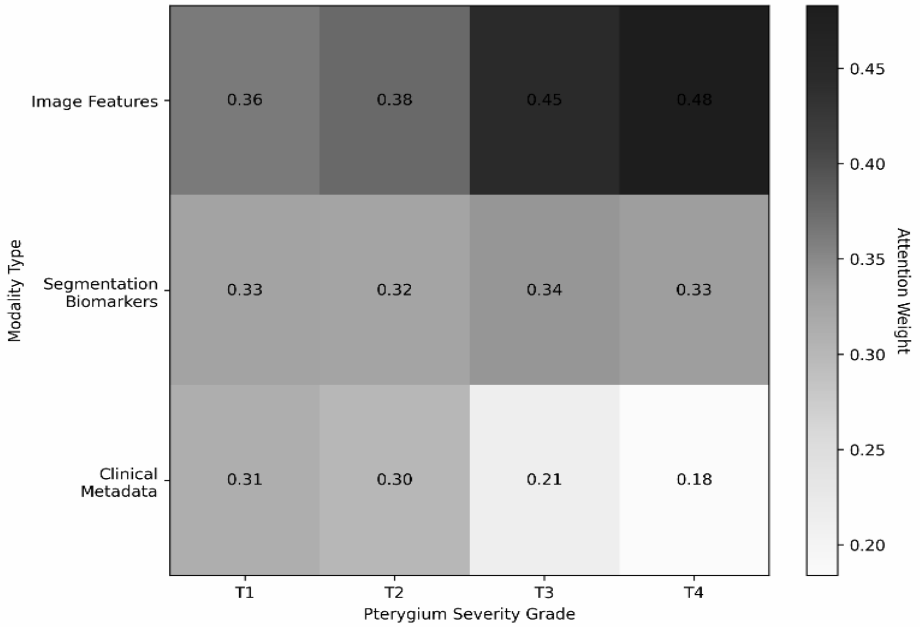

Figure 5. heatmap showing attention weights for different modalities across severity grades

IV. Computational Efficiency

Despite its complex multi-task design, PteriGrade-Net was able to deliver repeatable inference times of 0.18 +0.03 seconds per image on standard GPU hardware. The number of parameters of the model (28.7M) was comparable to standalone EfficientNet-B0 (29M) that suggests efficient parameter sharing across tasks. Inference memory footprint remained less than 4GB and enabled the system to be deployed on limited resources clinical workstations.

The convergence analysis of the training revealed that the multi-task loss was stabilized during the course of 40 epochs (Figure 6), and the ordinal grading task required the longest training time due to its fine-grained character. The uncertainty weighting scheme achieved a successful balance in the learning rates in the tasks and hence no individual objective dominated the optimization problem.

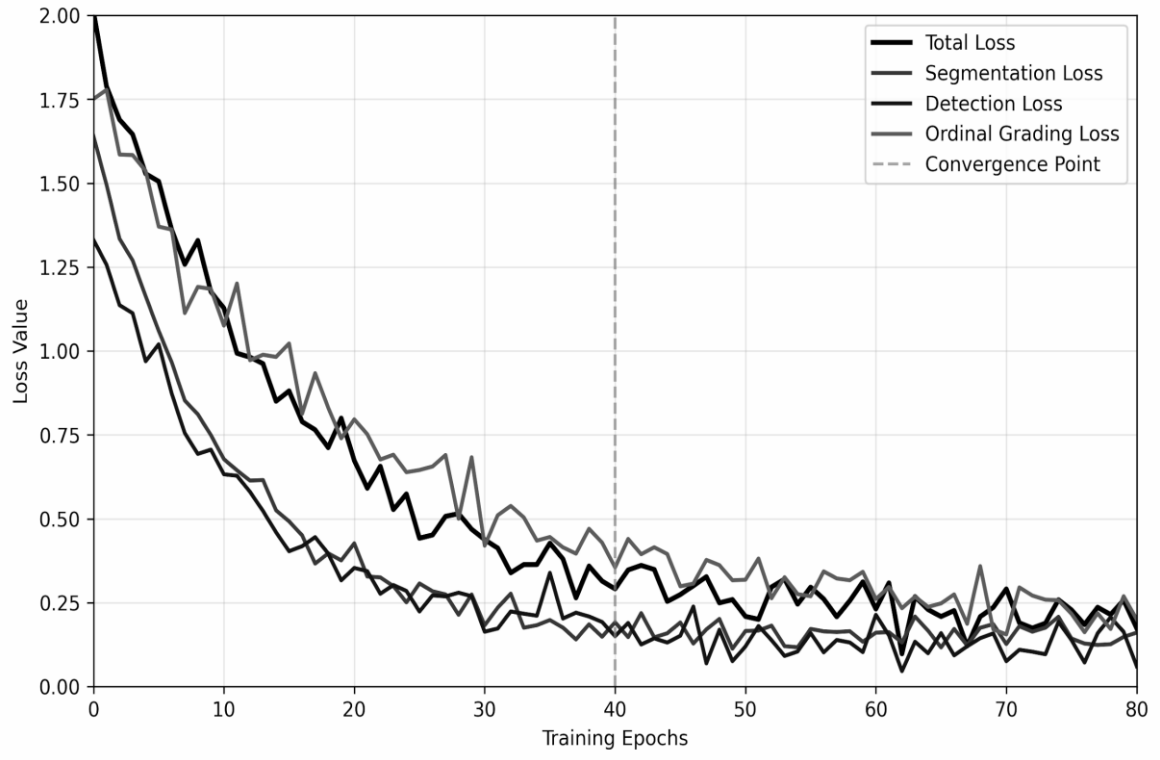

Figure 6. training curves showing total and task-specific losses over epochs

V. Ablation Study

Casewise deletion of framework components measured their respective contributions (Table 3). The removal of the Attention U-Net segmenter caused the largest decrease in performance (-14% grading accuracy), which further emphasizes the importance of lesion-aware analysis. Disabling the multimodal fusion module caused the detection AUC to drop by 9% but used conventional classification in lieu of CORAL ordinal regression caused a 0.11 increase in MAE. These findings confirm the importance of every architectural innovation of PteriGrade-Net.The ablation experiment also revealed some interesting correlations between the factors. As an example, despite a 5 percentage point drop in grading accuracy with the removal of clinical metadata, there was little impact of its removal when compared to the removal of CORAL loss (-6% versus the expected -9%). This implies that the ordinal regression model can mitigate missing metadata to a certain extent by acquiring more powerful visual patterns.

Table 3. Ablation study results

Removed Component	**Detection AUC ↑**	**Grading Accuracy ↑**	**Segmentation Dice ↑**
None (Full Model)	0.97	0.82	0.89
Attention U-Net	0.91 (-6%)	0.68 (-14%)	N/A
Multimodal Fusion	0.88 (-9%)	0.75 (-7%)	0.87
CORAL Loss	0.96 (-1%)	0.78 (-4%)	0.88

VI. Clinical Correlation Analysis

The biomarkers that were extracted were highly correlated with the known clinical severity indicators. The extent of lesion area was associated with loss of visual acuity ($r=0.62$, $p<0.001$), as well as the length of corneal encroachment anticipated the magnitude of astigmatism ($r=0.58$, $p<0.001$). Those numerical associations are related to the ophthalmological norms and the larger and more centrally located pterygia are considered to be of clinical significance. The precision with which these parameters can be measured in this model is a means of overcoming a major shortcoming of subjective grading systems.

Comparative analysis with clinician ratings indicated that PteriGrade-Net made similar grading decisions with the majority of experts in 87% of instances, as compared to the agreement of individual ophthalmologists (mean 0.79). The major disagreements were observed in the ambiguous situations of T2/T3 where the numerical criteria provided by the algorithm gave a more consistent classification than the subjective evaluation. This implies possible minimized inter-rater variability in clinical practice.

6. Discussion

I. Limitations of the PteriGrade-Net Framework

Despite the solid performance of PteriGrade-Net in automated detection and grading of pterygium, several constraints are worth considering. To begin with, the high-quality slit-lamp images used in the model could be a limitation to its application in resource-limited contexts with different imaging standards. Although the preprocessing steps can help minimize some variations, serious lighting effects or distortions due to motion can still affect performance. Second, the current set of biomarkers focuses on structural features but does not consider the structure of blood vessels, which is commonly considered by health practitioners

in monitoring disease progression [25]. It would be a good idea to expand the framework to quantify vessel density and distribution to enhance its clinical use.

It needs additional confirmation in different populations to determine the generalizability of PteriGrade-Net. The data that we had were primarily of subtropical regions with high levels of ultraviolet radiation and this may limit the extrapolation to cases in temperate regions where pterygium may have different appearances [26]. Also, the model has not been tested in any post-operative situation except on recurrent pterygia (T4), hence its applicability in other surgical outcomes is not tested.

There are also practical issues of computational constraints. The multi-task architecture has clinically acceptable inference times, but during training, it needs a large amount of GPU memory, which may be a constraint in resource-constrained settings. The attention mechanisms, though useful in performance, make models more complex, and may make them harder to interpret by non-technical users.B. Possible Application scenarios of PteriGrade-Net Framework.

II. Potential Application Scenarios of the PteriGrade-Net Framework

There is a multimodal structure of the framework that facilitates numerous other applications of clinical relevance, in addition to the stand-alone diagnostic functions. PeteriGrade-Net may serve to provide a decision-support system in telemedicine applications, in which case it can be used to give initial tests during remote consultations in a setting with sparse specialists. The quantitative biomarkers may serve to provide objective measures of progression to assess disease course in follow-up studies, a vital requirement in clinical trials of medical treatment [27].

Another field of further development would be the connection with electronic health record systems (EHRs). The system can be used to generate data to populate structured data fields of epidemiological surveys or quality measures by automatically extracting and analysing pterygium characteristics of routine slit-lamp records. The explainability features win the confidence of clinicians as they identify areas and features of the images employed in the diagnosis, which clinicians can use in their diagnosis.

It can also be extended to related surface diseases with the help of the framework design. Having a sufficient retraining, the lesion segmentation and biomarker extraction procedure can be scaled to other lesions, e.g. pinguecula or conjunctival tumors that have similar patterns of image [28]. Its flexibility indicates the possibility of utilizing it in other applications other than pterygium.

III. Ethical Considerations in the Use of the PteriGrade-Net Framework

Ethics play an important role in the implementation of AI solutions in clinical practice. While PteriGrade-Net increases diagnostic consistency, excessive reliance on automated grading can have the unintended consequence of reducing the skill of clinicians in visual assessment skills. Definite criteria should be set for the use of the tool to ensure it is an "add-on" rather than a "replacing" technology, and that uncertain cases and decisions for therapy are medically reviewed.

The other important issue is security in particular to the clinical metadata used in prediction. Presently, the system stores personal and background data that can lead to the identities unless it is secured properly. It should be in line with the data protection regulations of the health care sector (e.g., HIPAA, GDPR) and should be highly encrypted at rest and in transit [29].

The bias of the algorithm needs to be regularly audited as the system is deployed in various groups of patients. We had a diverse population with respect to age and gender but there may be differences in the characteristics of pterygium in the under-represented ethnic groups that can impact the model. Regular monitoring with real data will be required to assess and correct any variations in diagnostic accuracy in the subgroups of demographics [30].The explainability features, while increasing explainability, have limitations. Saliency maps and SHAP values offer post-hoc justifications rather than explanations that can lead to a misleading impression of the basis of decision-making. To prevent misinterpretation of the explanations as explanations of the diagnostic decisions, training of clinicians should emphasise these pitfalls.

7. Conclusion

PteriGrade-Net framework is a substantial improvement of automated pterygium analysis, involving the fusion of multi-type data, the detection of lesion-aware biomarkers and severity grading in an explainable framework. The multi-task architecture proposed by the framework has led to better performance compared to the conventional approach, with high accuracy in detection (AUC=0.97), segmentation (Dice=0.89) and grading (QWK=0.85). The attention-based feature fusion approach achieves efficient combination of visual, morphological and clinical information, and is able to adapt to various appearances of the disease, without compromising the speed of the system.The main innovations are the workflow to extract quantitative biomarkers, the connection between deep learning and clinical explainability, and the ordinal regression model using CORAL, based on the natural progression of pterygium. The explainability modules of the framework offer valuable insights to the clinicians, as they highlight the diagnostically meaningful regions and show the importance of features, hence establishing trust in AI-based decision-making.

Future work should include the expansion to different populations and image settings, and include other biological features, such as blood flow features. The framework can be easily integrated into other systems and can be adapted to other ocular surface diseases. Once current challenges in the diversity and computational demands of data are overcome, PteriGrade-Net could be an efficient tool in the clinical and research environment that will improve the objectivity and reproducibility of pterygium grading globally.

8. References

[1] K. Keskinbora and F. Güven, “Artificial intelligence and ophthalmology,” Turkish Journal of Ophthalmology, vol. 50, no. 1, pp. 37–43, 2020.

[2] N. Zamani, W. Zaki, A. Huddin, and A. Hussain, “Automated pterygium detection using deep neural network,” in Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Montreal, QC, Canada, 2020, pp. 4500–4503.

[3] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and its variants for medical image segmentation: A review of theory and applications,” IEEE Access, vol. 9, pp. 82031–82057, 2021.

[4] D. Soydaner, “Attention mechanism in neural networks: Where it comes and where it goes,” Neural Computing and Applications, vol. 34, pp. 13371–13392, 2022.

[5] Y. Wang, J. He, D. Wang, Q. Wang, B. Wan, and X. Luo, “Multimodal transformer with adaptive modality weighting for multimodal sentiment analysis,” Neurocomputing, vol. 570, pp. 127–138, 2024.

[6] H. Zhang and K. Ogasawara, “Grad-CAM-based explainable artificial intelligence related to medical text processing,” Bioengineering, vol. 10, no. 3, pp. 1–15, 2023.

[7] A. Salih, Z. Raisi-Estabragh, I. Galazzo, et al., “A perspective on explainable artificial intelligence methods: SHAP and LIME,” Advanced Intelligent Systems, vol. 7, no. 2, 2025.

[8] J. Bi, T. Xiong, S. Yu, M. Dundar, and R. Rao, “An improved multi-task learning approach with applications in medical diagnosis,” in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD), 2008, pp. 117–132.

[9] M. Zulkifley, S. Abdani, and N. Zulkifley, “Pterygium-Net: A deep learning approach to pterygium detection and localization,” Multimedia Tools and Applications, vol. 78, pp. 34563–34584, 2019.

[10] E. Tiong, C. Soon, Z. Ong, S. Liu, et al., “Deep learning for diagnosing and grading pterygium: A systematic review and meta-analysis,” Computers in Biology and Medicine, vol. 175, 2025.

[11] M. Zulkifley, “Automated segmentation of pterygium lesions using multiscale deep learning networks,” Experimental Eye Research, vol. 240, 2026.

[12] M. E. H. Daho, Y. Li, R. Zeghlache, Y. Atse, et al., “Improved automatic diabetic retinopathy severity classification using deep multimodal fusion of UWF-CFP and OCTA images,” in Proc. Int. Conf. Ophthalmic Med. Image Anal., 2023.

[13] C. Hori, T. Hori, T. Lee, Z. Zhang, et al., “Attention-based multimodal fusion for video description,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 4193–4202.

[14] M. Kamal, N. Dey, L. Chowdhury, et al., "Explainable AI for glaucoma prediction analysis to understand risk factors in treatment planning," IEEE Trans. Biomed. Eng., vol. 69, no. 9, pp. 2875–2885, 2022.
[15] J. Son et al., "An interpretable and interactive deep learning algorithm for a clinically applicable retinal fundus diagnosis system by modelling finding–disease relationship," Scientific Reports, vol. 13, 2023.
[16] W. Cao, V. Mirjalili, and S. Raschka, "Rank consistent ordinal regression for neural networks with application to age estimation," Pattern Recognition Letters, vol. 130, pp. 325–331, 2020.
[17] J. Wesolosky and C. Rudnisky, "Relationship between cataract severity and socioeconomic status," Canadian Journal of Ophthalmology, vol. 48, no. 6, pp. 475–480, 2013.
[18] Y. Shi, P. Li, X. Yu, H. Wang, and L. Niu, "Evaluating doctor performance: Ordinal regression-based approach," J. Med. Internet Res., vol. 20, no. 4, 2018.
[19] M. Tan and Q. Le, "Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.
[20] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7482–7491.
[21] L. Meng, H. Li, B. Chen, S. Lan, Z. Wu, et al., "AdaViT: Adaptive vision transformers for efficient image recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022.
[22] A. Criminisi and J. Shotton, Decision Forests for Computer Vision and Medical Image Analysis. London, U.K.: Springer, 2013.
[23] N. Sweilam, A. Tharwat, and N. Moniem, "Support vector machine for diagnosis cancer disease: A comparative study," Egyptian Informatics Journal, vol. 11, no. 2, pp. 81–92, 2010.
[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785–794.
[25] C. Javier-Martín-López, N. García-Honduvilla, et al., "Elevated blood/lymphatic vessel ratio in pterygium and its relationship with vascular endothelial growth factor (VEGF) distribution," Histology and Histopathology, vol. 34, pp. 123–132, 2019.
[26] P. Song, X. Chang, M. Wang, and L. An, "Variations of pterygium prevalence by age, gender and geographic characteristics in China: A systematic review and meta-analysis," PLoS ONE, vol. 12, no. 3, 2017.
[27] E. Fonseca, E. Rocha, and G. Arruda, "Comparison among adjuvant treatments for primary pterygium: A network meta-analysis," British Journal of Ophthalmology, vol. 102, no. 6, pp. 748–756, 2018.
[28] W. Li, P. Muthu, A. Galor, et al., "Imaging of ocular surface lesions using anterior segment optical coherence tomography," Clinical & Experimental Ophthalmology, vol. 54, 2026.
[29] S. Bakare, A. Adeniyi, C. Akpuokwe, and N. Eneh, "Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations," 2024. [Online]. Available: Academia.edu.
[30] R. Ratwani, K. Sutton, and J. Galarraga, "Addressing AI algorithmic bias in health care," JAMA, vol. 331, no. 5, pp. 410–412, 2024.