Multimodal Macular Degeneration Classification using OCT Imaging

Martin Goessweiner
Department for Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
mgosswei@andrew.cmu.edu

Jiayi Liu
Department for Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
jiayili5@andrew.cmu.edu

Jainam Modh
Department for Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
jmodh@andrew.cmu.edu

Jonathan Tang
Department for Biomedical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
jtang3@andrew.cmu

***Abstract*—Age-related macular degeneration (AMD) is one of the leading causes of vision loss worldwide, with early diagnosis remaining a significant clinical challenge. In this work, we develop a multimodal classification framework that integrates optical coherence tomography (OCT) imaging and structured clinical data to distinguish between stages of AMD, including early, intermediate, geographic atrophy, scarring, and wet AMD. OCT images are processed using a pretrained vision encoder (RETFound), while clinical patient information tabular data is modeled using a gated TabTransformer that captures contextual interactions between clinical features. Although a biologically informed preprocessing pipeline was developed using shearlet-based edge detection and retinal flattening, it could not be fully deployed due to data set size and computational constraints. We implement a cross-attention fusion strategy to integrate both modalities and classify the AMD stage using a BERT-style decoder. The image-only model achieved 92.45% accuracy at the volume level via majority voting, while the gated TabTransformer reached 46.3% accuracy on clinical data alone. The multimodal system achieved 79.7% validation accuracy at the B-scan level after just 10 epochs, demonstrating the complementary value of combining imaging with clinical metadata. This study highlights the potential of multimodal deep learning for improving diagnostic accuracy in retinal disease.**

***Clinical Relevance*—This paper presents a multimodal deep learning approach for the classification of various stages of AMD using OCT images and electronic health record information. The model achieves a 79.7% accuracy in classifying AMD in OCT B-scans.**

I. Introduction

The macula is the central region of the retina responsible for fine visual tasks such as reading and facial recognition. Damage to this area leads to significant vision impairment, even if the surrounding retinal regions remain intact.

Age-related macular degeneration (AMD) is a leading cause of central visual loss in the elderly. It affects the macula via photoreceptor degeneration, damage to the retinal pigment epithelium (RPE), and changes in the choroid. AMD progresses in early, intermediate and late stages, with late AMD manifesting as either neovascular (wet) AMD or geographic atrophy (GA), the advanced form of dry AMD. GA is irreversible and typically results in central vision loss within 2 – 3 years of diagnosis [1].

Recent estimates show that roughly 200 million people were affected globally by 2020, with this number expected to rise to 288 million by 2040 due to an aging population [2]. Risk factors include age, smoking, hypertension, and genetic predisposition, particularly in genes regulating inflammation and lipid metabolism [2], [3].

Emerging evidence links AMD to systemic conditions such as Alzheimer's disease, cardiovascular disease, and increased all-cause mortality [2], [4]. These associations may arise from shared mechanisms that involve oxidative stress, inflammation, or vascular dysfunction, potentially offering new diagnostic windows - particularly in early disease.

Despite its prevalence and systemic relevance, early AMD detection remains challenging. The condition is often asymptomatic, and hallmark lesions like drusen can go unnoticed until substantial damage has occurred [1]. Optical coherence tomography (OCT) enables noninvasive, high-resolution imaging of retinal and choroidal structures, but minor early changes and variable image quality make consistent diagnosis harder [5].

Although OCT is the diagnostic gold standard, interpretation is labor-intensive and subjective. Deep learning (DL) approaches have shown promise for automating OCT analysis. For example, Lee et al. achieved 93% accuracy in classifying scans as AMD or normal using EMR-labeled data [6], yet performance lags behind experts for more detailed classification, such as distinguishing early vs. late or wet vs. dry AMD.

This reflects a broader challenge in retinal AI: inconsistent ground truth. Expert graders often disagree on fine-grained stages, and many public datasets rely on EMR codes rather than standardized grading [7], limiting classifier utility in nuanced clinical settings.

We want to formally thank P. Sang Chalacheva, Ph.D., for her continued support with this project and her very informative lectures.

Moreover, most DL models treat OCT as raw pixel input, disregarding anatomical priors like retinal layer boundaries. Yet studies show that preprocessing—such as noise reduction, contrast enhancement, or segmentation—can significantly improve classification performance [8]. For example, integrating edge detection or contour-based segmentation has yielded accuracies exceeding 96%, surpassing models trained on raw images. Luo et al. also found that selecting effective edge detectors improved retinal boundary delineation and reduced downstream error propagation [9].

Still, many systems are limited to binary tasks (e.g., AMD vs. normal) or trained on narrow datasets. Few generalize across AMD subtypes or stages while incorporating physiologically meaningful preprocessing.

To address this, we propose a multimodal classification pipeline that combines biologically informed image preprocessing with modern deep learning. We apply a complex shearlet transform to OCT scans, exploiting its directional and multiscale sensitivity to extract retinal boundaries. Compared to wavelets or Canny filters, shearlets offer optimal detection of curved anatomical features [10], [11], enabling us to flatten and align the retina across patients to improve consistency and model interpretability.

We then use RETFound, a foundation model pretrained on over 700,000 OCT scans [12], for multi-class AMD classification. To incorporate systemic context, we encode diagnostic metadata using TabTransformer, a transformer architecture optimized for tabular data [13].

Finally, we integrate both modalities using a cross-attention fusion module that enables joint reasoning over anatomical and clinical features. This multimodal approach improves classification accuracy and bridges the gap between raw imaging data and patient-level understanding.

II. Method

In this study, we design and train a multimodal AMD classification system by combining a pre-trained image encoder and a gated TabTransformer for clinical tabular data. Each encoder is first trained independently, then frozen to preserve learned representations, and finally integrated into a unified fusion model whose classification head is trained end-to-end.

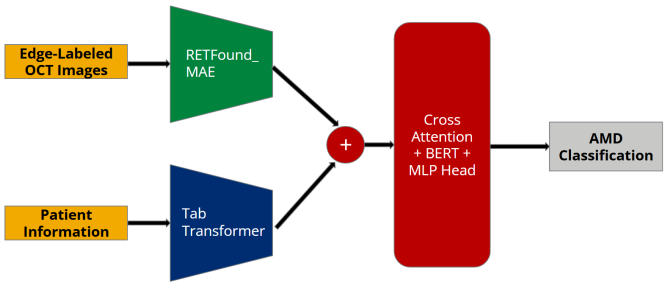

Fig. 1. Overall model architecture summary for AMD classification using multimodal fusion.

A. Dataset

The dataset utilized for model training and evaluation is the University of Pittsburgh Medical Center (UPMC) AMD-OCT dataset, generously provided by the Choroidal Analysis and Research Lab. This dataset comprises of 4,028 high-resolution OCT volumes collected from 91 patients diagnosed with AMD during routine clinical examinations across the UPMC health system. These volumes are distributed over both the left and right eyes for each patient and also longitudinally over multiple ophthalmology clinic appointments from 2007 to 2023. Each volume is composed of approximately 160 B-scans and varies depending on the type of OCT machine used for a combined total of 637,408 2D images. Ground truth clinical annotations for each volume are provided by a trained UPMC ophthalmologist for various AMD subtypes: Not AMD, Early-AMD, Intermediate-AMD, Geographic Atrophy (GA), Scarring, and Wet AMD. Figure 2 presents the distribution of the 4,028 labeled volumes across the diagnostic categories, providing an overview of class representation within the dataset.In addition to imaging data, the dataset is enriched with tabular clinical information such as International Classification of Diseases (ICD) diagnostic codes, patient demographics, and substance use for supporting comprehensive multimodal analyses.

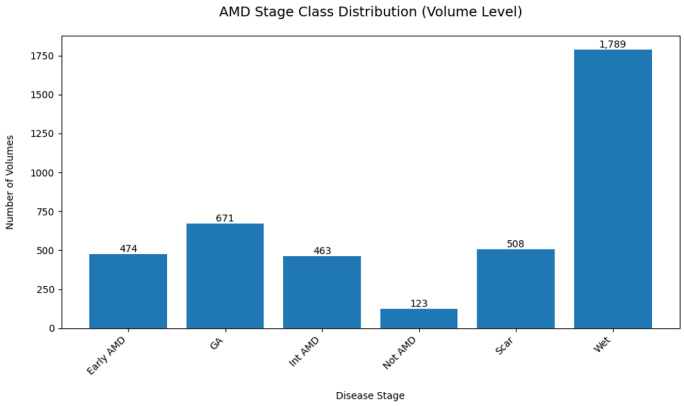

Fig. 2. Clinically annotated label distribution of the AMD-OCT dataset for multi-class AMD classification. Note the significant imbalance in data size between "Not AMD" vs. "Wet" labels.

B. Image Encoder

1) Pre Processing: Although the final models were trained on raw OCT images due to time constraints, we developed a custom preprocessing pipeline to enhance the visibility of the retinal structure for downstream classification. This pipeline used a complex shearlet transform for multiscale edge detection on OCT B scans, targeting boundary layers of the retina.

The images were transformed with two scales and shear levels [1,2], then reconstructed using only second-level coefficients to emphasize curved/sheared retinal features. An edge map was computed from the edge measure β:

$$\beta = \frac{\sum_{odd} - \sum_{even} - n_{\text{scales}} \cdot T}{n_{\text{scales}} \cdot \max \sum_{odd} + \epsilon} \tag{1}$$

where T and ϵ control edge thickness and noise suppression. The result was fixed at [0,1], thresholded at 0.8, and the dominant edge in each column was fit with a fourth-order polynomial. The columns were then vertically rolled to flatten the retina and center it across the samples. An example is shown in Fig. 3.

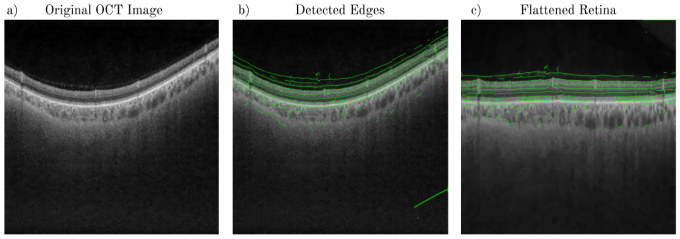

Fig. 3. Preprocessing pipeline. a: original image, b: edge detection and filtering, c: retina flattened.

Since the only available shearlet library (pyShearLab) lacked GPU support [14], we reimplemented it in PyTorch, including directional filters, multiscale decomposition, and coefficient normalization. This yielded a 5× speedup over the original CPU version. However, due to the scale of the dataset, pre-processing could not be completed in time and was excluded from the training. The pipeline remains functional for future use, but is still projected to take ~ 7 days to process all images compared to the originally projected $\sim 35 - 40$ days.

2) Model Selection - RETFound: RETFound is a foundational model composed of a masked autoencoder (MAE) trained on 736,442 OCT B-scans using self-supervision [7]. When used to perform prognosis classification of Wet AMD, the model obtained an AUROC score of 0.799 (95% CI 0.796, 0.802). We extracted the pretrained encoder from the model to use in our AMD classification system - a large vision Transformer (ViT-large) with 24 Transformer blocks and a resulting embedded vector size of 1,024. As this model only evaluates on individual B-scans, we were unable to develop a 3D volume-level classification architecture that establishes appropriate positional encoding compatible with backpropagation.

3) Image-Only Classification: Prior to multimodal fusion, baseline performance was established when training only on OCT images. Each 2D B-scan was encoded into a 1024-dimensional feature vector by using RETFound encoder with frozen weights, followed by a MLP classifier consisting of a 128-dimensional projection, ReLU activation, dropout, and a final output layer for six-class prediction. Only the classification head was trained using the following settings:

- **Epochs:** 10
- **Batch Size:** 128
- **Optimizer:** Adam (lr 1×10^{-4}, weight decay 1×10^{-4})
- **Scheduler:** `ReduceLROnPlateau` on validation loss (factor 0.5, patience 3)
- **Validation:** After each epoch, evaluate on held-out volumes (80:20 volume-level split), tracking loss and accuracy.

C. Tabular Encoder

1) Data Preprocessing: We collect nine categorical clinical fields and three continuous variables: age at visit, visual acuity (VA), and ICD-10 primary diagnosis codes. Categorical fields are integer-encoded with an `OrdinalEncoder`, mapping unseen values to a reserved token. Continuous features undergo the following steps:

- **ICD-10 Frequency (icd_freq):** Count occurrences of each ICD-10 code across the full dataset, then apply min–max normalization to scale counts into [0,1].
- **Visual Acuity (va_continuous):** Convert Snellen fractions to logMAR units to linearize acuity measures, then standardize to zero mean and unit variance.
- **Age at Visit:** Standardize to zero mean and unit variance.

After preprocessing, we form a token sequence of length 12 (9 categorical embeddings + 3 continuous projections), each embedding dimension $d = 64$.

a) Feature Selection: Table I lists the five key clinical variables and the rationale for including each.

TABLE I
SELECTED TABULAR FEATURES WITH JUSTIFICATION.

Feature	Reason
Age	A 2022 study showed that approximately 1 in 10 Americans aged 50+ have early AMD, rising to 3 in 10 at age 80+ [15]
Gender	Early AMD was more frequent in men than in women (p = 0.030) [16]
Smoking	Current smokers have a two- to three-fold increased risk of developing AMD vs. non-smokers [17]
Alcohol	Moderate to heavy alcohol consumption is linked to higher risk of early AMD [18]
Illicit Drug	Illicit drug abuse is associated with retinal microvascular occlusions, talc retinopathy, and maculopathy—leading to vision loss and potentially confounding AMD assessment [19]
Visual Acuity	Wet AMD typically causes more severe VA loss due to choroidal neovascularization [20]

2) Model Selection - TabTransformer: TabTransformer is a self-attention based architecture specifically designed for tabular data modeling [13]. It transforms the initial embeddings of categorical features into contextualized embeddings by capturing feature interactions across all columns. Such contextual embeddings improve robustness to missing and noisy values, support efficient end-to-end training, and enable semi-supervised pre-training to leverage unlabeled data. Extensive experiments demonstrate that TabTransformer matches or outperforms state-of-the-art tree-based ensemble methods (e.g., GBDT) on various benchmarks. Given its superior accuracy, interpretability, and stability on heterogeneous tabular datasets, we adopt TabTransformer to process our clinical tabular data.

3) Gating Mechanism: Within each of the $L = 4$ transformer blocks, after multi-head self-attention and residual normalization, we compute a gate vector:

$$\mathbf{g} = \sigma\big(\mathbf{W}_g\mathbf{H} + \mathbf{b}_g\big) \in [0,1]^{B\times M\times d}$$

and perform element-wise modulation:

$$\widetilde{\mathbf{H}} = \mathbf{H} \odot \mathbf{g},$$

before feeding into the feed-forward network. This gating step filters out noisy feature interactions and enhances robustness on heterogeneous clinical data.

4) Tabular-Only Pretraining: The gated TabTransformer encoder is pretrained on the processed tabular data with the following settings: token embeddings of dimension 64, four transformer blocks each with four attention heads, and an output head producing six logits corresponding to AMD stages. These hyperparameters are derived from Mei et al. [21], but with depth and heads halved to suit our smaller dataset and reduce overfitting.

Training uses a `WeightedRandomSampler` for class balance, a batch size of 64, and the Adam optimizer (initial learning rate 1e-3). We apply a `StepLR` scheduler that halves the learning rate every 40 epochs, train for 200 epochs, and monitor performance on a stratified validation split. During each epoch, we record both training and validation loss to ensure the model is converging appropriately and to detect any signs of overfitting early.

D. Multimodal Fusion and Training

We explored a hybrid early-fusion strategy: tabular and image embeddings are fused via cross-attention before being passed to a shared decoder.

- **Cross-Attention Fusion (Early Fusion):** The 1024-dim image feature (from RETFound MAE) and the 64-dim tabular feature (from the gated TabTransformer) are each linearly projected to a 768-dim space. A single multi-head cross-attention block then uses the tabular projection as the query and the image projection as key/value. The result is added residually and normalized to yield a fused 768-dim vector.
- **BERT Decoder:** We concatenate the fused vector with the original tabular projection into a sequence of length 2 and feed these embeddings into a pretrained BERT-Base model (12 layers, 12 heads, hidden size 768) via its `inputs_embeds` API, using an all-ones attention mask.
- **Classification Head:** The BERT output for the first token (`[CLS]`) is passed through an MLP—$\mathrm{Linear}(768, 128)$ $\rightarrow$ ReLU $\rightarrow$ $\mathrm{Dropout}(0.2)$ $\rightarrow$ $\mathrm{Linear}(128, 6)$—to produce logits for the six AMD stages.

a) Training Protocol: Only the cross-attention fusion block, BERT parameters, and MLP head are trained (both encoders frozen). Modified settings compared to image-only training:

- **Batch Size:** 16
- **Validation:** After each epoch, evaluate on held-out volumes (50% split).

III. Results

A. Image Encoder Only

a) B-scan Level Performance: Training for classification solely on OCT B-scan images was conducted on a dataset of 637,408 slices (3,222 training volumes, 806 validation volumes) using an 8:2 volume-level split. After 10 epochs, the model reached 84.59% training and 80.05% validation accuracy, with corresponding losses of 0.4292 and 0.5484.

b) Volume Level Performance: To evaluate performance at the volume level, we aggregated slice-wise predictions using majority voting. This approach yielded a volume accuracy of **92.45%** (3724/4028 correct) across the entire dataset. Subsequently, we calculated a confidence level for each volume prediction, and these results are summarized in Figure 4.

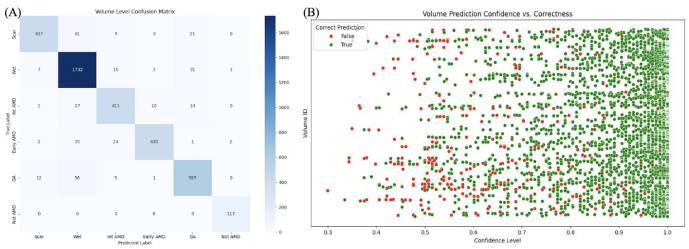

Fig. 4. (A) Confusion matrix of volume-level predictions for the image-only model after majority voting across 2D slices. (B) Scatter plot of prediction confidence vs. correctness at the volume level. Each point represents a volume, with color indicating prediction accuracy.

B. Tabular Encoder Only

a) Training Loss and Convergence: Figure 5 (A) shows the training loss curve over 200 epochs. The loss decreases steadily and plateaus after around 150 epochs, indicating that the gated TabTransformer has converged on the tabular data.

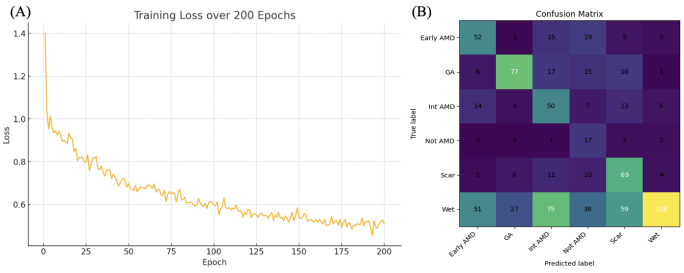

Fig. 5. (A) Training loss curve for the gated TabTransformer. Loss stabilizes after 150 epochs, demonstrating convergence. (B) Confusion matrix for the gated TabTransformer on the test set. Most misclassifications occur into the *Wet* category, and the fewest into *Not AMD*, reflecting remaining class imbalance.

b) Test Accuracy: On the held-out test set (806 samples), the tabular encoder achieved an overall accuracy of 46.3%.

c) Key Classification Performance: The model performs best on *Wet* AMD (highest precision), and worst on the *Not AMD* class due to its small sample size. Although we applied a weighted sampler and loss, residual imbalance still leads to fewer correct predictions for underrepresented classes.

C. Multimodal System

a) Fusion Training Results: After freezing both the image and tabular encoders, we trained the fusion and classification heads for 10 epochs (sampling 50% of batches each epoch). The fusion head achieved a best validation accuracy of 79.66%.

IV. Discussion

A. Model Development and Generalization

Initial experiments with a limited dataset (173 volumes) resulted in pronounced overfitting, with training accuracy exceeding 95% but test accuracy falling below 40%, underscoring the model's inability to generalize. Upon scaling the dataset to 4028 volumes, both training and testing accuracies surpassed 80%, with volume-level accuracy reaching 92.45%. These improvements emphasize the critical importance of dataset size in achieving robust performance.

To ensure a fair evaluation of generalizability, we adopted strict volume-level data partitioning, preventing any overlap between training and test sets. This approach better simulates real-world clinical deployment and mitigates the risk of overestimated performance due to data leakage.

B. Gating Improves TabTransformer Performance

Under identical preprocessing and training settings, the baseline TabTransformer achieved a test accuracy of 44.78%, whereas the gated variant improved this to 46.28%. The learnable gating layers effectively suppress noisy or uninformative feature interactions, resulting in more discriminative embeddings and a measurable boost in classification performance.

While the tabular-only model reached approximately 46% accuracy, combining OCT images with structured clinical data led to a substantial improvement, with accuracy approaching 80%. This highlights the complementary nature of the two modalities, where limitations in one can be mitigated by informative cues in the other.

C. Dimensionality Constraints and Design Trade-offs

While clinical diagnosis inherently depends on 3D anatomical understanding, our current encoder processes individual 2D B-scan slices. Developing an end-to-end 3D model is technically challenging, particularly in designing differentiable aggregation functions that support effective gradient backpropagation across slices. This is because volume-level prediction requires encoding all 2D slices and aggregating their features—often through operations such as top-k selection, thresholding, or hard attention, which are inherently non-differentiable and disrupt the flow of gradients during training. Given these constraints, we adopt a practical alternative: independent slice-wise classification, followed by majority voting for patient-level predictions. While this sacrifices some spatial continuity, it enables scalable training and delivers competitive diagnostic accuracy.

D. Limitations and Future Directions

Despite weighted sampling and loss, the *Not AMD* class remains under-represented and suffers lower recall. Future work should acquire or synthesize additional *Not AMD* samples to better balance the dataset. Our multimodal fusion was constrained to just 10 epochs by resource limits; extending training duration and exploring alternative fusion strategies (e.g. late fusion, multi-stage fine-tuning) may yield further gains.

To illustrate that fusion training has not yet fully converged, we include the training and validation loss curves over the 10 epochs (Figure 6). Notice that the validation loss is still decreasing at epoch 10, indicating room for further improvement with prolonged training.

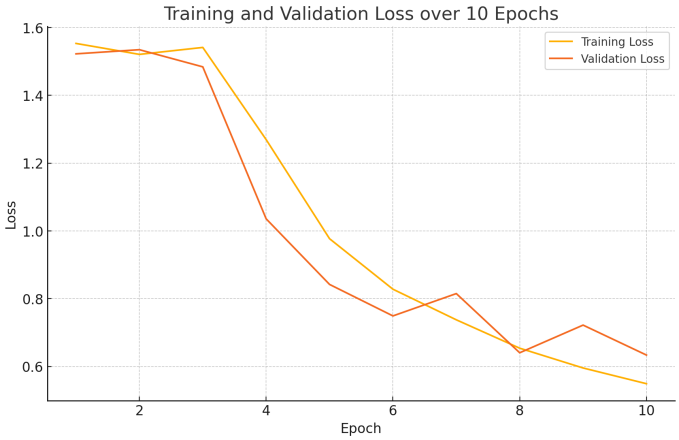

Fig. 6. Training (yellow) and validation (red) loss curves for the multimodal fusion over 10 epochs. Validation loss continues to decline at the final epoch, suggesting the model has not fully converged.

Future work will also focus on improving model interpretability and clinical trust. This includes: (i) quantitatively evaluating the influence of specific anatomical regions on predictions, and (ii) exploring threshold calibration strategies aligned with clinical decision-making principles. Expanding the framework to incorporate longitudinal patient histories or additional imaging modalities (e.g., fundus photography) may further improve diagnostic robustness and generalizability.

V. Conclusion

We have presented a multimodal classification framework for age-related macular degeneration (AMD) that integrates OCT image features from a pretrained RETFound encoder with structured clinical data embedded via a gated TabTransformer. By leveraging independent pretraining and weight freezing, each modality contributes complementary strengths: the image encoder captures fine-grained retinal morphology, while the gated TabTransformer adaptively filters and models clinical feature interactions.

Our tabular-only model achieved a test accuracy of 46.3%, outperforming the baseline TabTransformer (44.8%) due to the inclusion of learnable gating mechanisms. While modest in isolation, this performance was substantially enhanced through multimodal fusion. Specifically, cross-attention-based integration within a BERT-style decoder enabled the combined system to achieve a peak validation accuracy of 79.7% at the B-scan level.

At the volume level, the system achieved a classification accuracy of 92.45%, with each prediction accompanied by a calibrated confidence score—aligning with the probabilistic reasoning commonly used in clinical decision-making.

Overall, these findings highlight the strong potential of multimodal AI frameworks to enhance ophthalmic diagnostics, offering a promising pathway toward intelligent, data-driven diagnostic support in real-world ophthalmic care.

VI. Acknowledgments

We would like to sincerely thank Dr. Jay Chhablani, MD; Sandeep Chandra Bollepalli, Ph.D; Sharat Chandra; and the rest of the Choroid Analysis and Research (CAR) Lab for their help with dataset acquisition and technical guidance. Also, this project would not have been possible without the support of professor P. Sang Chalacheva, Ph.D.; teaching assistant Nishanth Arun, and the rest of our classmates in 42-687 Projects in Biomedical AI course.

VII. Author Contributions

Jainam maintained communication with the CAR lab for dataset acquisition. Jainam and Jonathan preprocessed the labelled annotations and tabular data. Martin provided significant literature review and developed the custom image preprocessing pipeline. While Jainam designed the overal model framework for training and evaluation, Jiayi implemented image-based classification and Jonathan implemented tabular and multimodal classification. Lastly, Jiayi summarized and analyzed the final results for reporting. In total, all four members contributed about equally (20-30% each) to the project.

References

[1] C. D. Regillo, L. M. Nijm, D. L. Shechtman, P. K. Kaiser, P. M. Karpecki, E. H. Ryan, M. S. Ip, E. Yeu, T. Kim, M. R. Rafieetary, and E. D. Donnenfeld, "Considerations for the identification and management of geographic atrophy: Recommendations from an expert panel.," *Clin Ophthalmol*, vol. 18, pp. 325–335, 2024.

[2] P. Mitchell, G. Liew, B. Gopinath, and T. Y. Wong, "Age-related macular degeneration.," *Lancet*, vol. 392, pp. 1147–1159, Sep 2018.

[3] L. S. Lim, P. Mitchell, J. M. Seddon, F. G. Holz, and T. Y. Wong, "Age-related macular degeneration.," *Lancet*, vol. 379, pp. 1728–1738, May 2012.

[4] S. C. Yeung, Y. You, K. L. Howe, and P. Yan, "Choroidal thickness in patients with cardiovascular disease: A review.," *Surv Ophthalmol*, vol. 65, pp. 473–486, Jul-Aug 2020.

[5] S. R. Singh, K. K. Vupparaboina, A. Goud, K. K. Dansingani, and J. Chhablani, "Choroidal imaging biomarkers.," *Surv Ophthalmol*, vol. 64, pp. 312–333, May-Jun 2019.

[6] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for the classification of oct images of normal versus age-related macular degeneration.," *Ophthalmol Retina*, vol. 1, pp. 322–327, Jul-Aug 2017.

[7] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, Y. Kihara, N. Allen, J. E. J. Gallacher, T. Littlejohns, T. Aslam, P. Bishop, G. Black, P. Sergouniotis, D. Atan, A. D. Dick, C. Williams, S. Barman, J. H. Barrett, S. Mackie, T. Braithwaite, R. O. Carare, S. Ennis, J. Gibson, A. J. Lotery, J. Self, U. Chakravarthy, R. E. Hogg, E. Paterson, J. Woodside, T. Peto, G. Mckay, B. Mcguinness, P. J. Foster, K. Balaskas, A. P. Khawaja, N. Pontikos, J. S. Rahi, G. Lascaratos, P. J. Patel, M. Chan, S. Y. L. Chua, A. Day, P. Desai, C. Egan, M. Fruttiger, D. F. Garway-Heath, A. Hardcastle, S. P. T. Khaw, T. Moore, S. Sivaprasad, N. Strouthidis, D. Thomas, A. Tufail, A. C. Viswanathan, B. Dhillon, T. Macgillivray, C. Sudlow, V. Vitart, A. Doney, E. Trucco, J. A. Guggeinheim, J. E. Morgan, C. J. Hammond, K. Williams, P. Hysi, S. P. Harding, Y. Zheng, R. Luben, P. Luthert, Z. Sun, M. McKibbin, E. O'Sullivan, R. Oram, M. Weedon, C. G. Owen, A. R. Rudnicka, N. Sattar, D. Steel, I. Stratton, R. Tapp, M. M. Yates, A. Petzold, S. Madhusudhan, A. Altmann, A. Y. Lee, E. J. Topol, A. K. Denniston, D. C. Alexander, P. A. Keane, and U. B. E. . V. Consortium, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156–163, 2023.

[8] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, and M. Masud, "Dl-cnn-based approach with image processing techniques for diagnosis of retinal diseases," *Multimedia Systems*, vol. 28, 08 2022.

[9] S. Luo, J. Yang, Q. Gao, S. Zhou, and C. A. Zhan, "The edge detectors suitable for retinal oct image segmentation.," *J Healthc Eng*, vol. 2017, p. 3978410, 2017.

[10] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A shearlet approach to edge analysis and detection," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 929–941, 2009.

[11] L. Xiaoming, X. Ke, Z. Peng, and C. Jiannan, "Edge detection of retinal oct image based on complex shearlet transform," *IET Image Processing*, vol. 13, no. 10, pp. 1686–1693, 2019.

[12] S. Pang, B. Zou, X. Xiao, Q. Peng, J. Yan, W. Zhang, and K. Yue, "A novel approach for automatic classification of macular degeneration oct images," *Scientific Reports*, vol. 14, no. 1, p. 19285, 2024.

[13] X. Huang, R. Arora, M. Sharma, K. Sun, S. Fabricant, and M. Zaidi, "Tabtransformer: Tabular data modeling using contextual embeddings," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5779–5791, 2020.

[14] S. Lock, "pyshearlab."

[15] Prevent Blindness, "Prevalence of age-related macular degeneration (amd)." Prevent Blindness web site, n.d. Retrieved 2025–03–25 from https://preventblindness.org/amd-prevalence-vehss/.

[16] M. Sasaki, S. Harada, Y. Kawasaki, and et al., "Gender-specific association of early age-related macular degeneration with systemic and genetic factors in a japanese population," *Scientific Reports*, vol. 8, p. 785, 2018.

[17] J. Thornton, R. Edwards, P. Mitchell, and et al., "Smoking and age-related macular degeneration: A review of association," *Eye*, vol. 19, pp. 935–944, 2005.

[18] C. Blank, "Study identifies alcohol use as potential risk factor for amd subtype." American Journal of Managed Care web site, 2021. Retrieved 2025–03–25 from https://www.ajmc.com/view/study-identifies-alcohol-use-as-potential-risk-factor-for-amd-subtype.

[19] J. Peragallo, V. Biousse, and N. J. Newman, "Ocular manifestations of drug and alcohol abuse," *Current Opinion in Ophthalmology*, vol. 24, no. 6, pp. 566–573, 2013.

[20] X. Zhang and T. Y. Y. Lai, "Baseline predictors of visual acuity outcome in patients with wet age-related macular degeneration," *Biomedical Research International*, vol. 2018, p. 9640131, Feb. 26 2018.

[21] Y. Mei, Z. Jin, W. Ma, Y. Ma, N. Deng, Z. Fan, and S. Wei, "Optimizing acute coronary syndrome patient treatment: Leveraging gated transformer models for precise risk prediction and management," *Bioengineering*, vol. 11, no. 6, p. 551, 2024.