INSIGHT: AI-BASED SCREENING TOOL FOR EYE DISEASES USING A NOVEL MULTIMODAL FUSION TECHNIQUE

Ananya Raghu, Anisha Raghu

ABSTRACT

Age-related Macular Degeneration (AMD), Glaucoma, Diabetic Retinopathy (DR), Macular Edema, and Myopic Degeneration affect 280 million people worldwide and are leading causes of blindness. Early screening for these diseases is essential, yet access to medical care remains limited in low- and middle-income countries (LMICs). We develop InSight, an AI-based app that is the first solution to combine patient metadata with fundus images for the simultaneous and accurate diagnosis of five eye diseases. InSight features a three stage pipeline: (1) real-time image quality assessment, (2) disease diagnosis model and (3) DR grading model to assess severity. Our disease diagnosis model incorporates three key innovations: (a) A novel fusion technique (MetaFusion) combining clinical metadata and images for improved accuracy, (b) a pre-training method leveraging supervised and self-supervised loss functions and (c) multi-task model to simultaneously predict 5 diseases. Trained and tested on BRSET, a dataset of lab captured fundus images, the quality checker model achieves near-100% accuracy in filtering low-quality fundus images. The MetaFusion multi-task model consistently outperforms models using only images by 6% in average balanced accuracy (BA) due to multimodal fusion and pre-training, achieving AUCs of 0.99 (DR, Myopic Fundus, AMD), 0.98 (Macular Edema), and 0.92 (Glaucoma), surpassing the previous reported value of 0.87 for DR. The multitask model achieves comparable performance to individual task-specific models, reducing complexity by 5x. To validate InSight on smartphone based retinal fundus images, we trained the models on a joint dataset of BRSET and mBRSET images, where mBRSET is a challenging dataset containing smartphone retinal camera fundus images for DR and Macular Edema. The Metafusion multi-task model outperforms models using only images by 4% in BA for mBRSET. In addition, the model trained on a joint (mBRSET+BRSET) dataset performs almost as well as models trained on BRSET and mBRSET respectively, demonstrating the robustness of the disease diagnosis model on a range of imaging conditions.T

***Index Terms*—** multimodal fusion, pretraining, multitask, diabetic retinopathy, macular edema, myopic fundus, glaucoma, age-related macular degeneration, fundus

1. INTRODUCTION

Age-related macular degeneration (AMD), glaucoma, diabetic retinopathy (DR), macular edema, and myopic degeneration affect over 280 million people worldwide and are leading causes of blindness. Early screening for these diseases is essential, yet access to medical care remains limited in low- and middle-income countries (LMICs). There is a significant shortage of doctors in LMICs, which contain over 90 percent of the world's blind population. We develop InSight, a multistage medical screening tool to combine patient metadata with fundus images to accurately diagnose five eye-related diseases simultaneously on a mobile phone.

1.1. Background/Key Terms

Diabetic Retinopathy (DR) is a complication of diabetes caused by high blood sugar levels, which leads to the damage of blood vessels in the eye. Over time, these blood vessels may leak or swell, and in more severe cases, may lead to the growth of abnormal blood vessels in the retina. This leakage can cause macular edema, a condition where fluid accumulates in the macula, the part of the retina responsible for sharp central vision. Macular edema can cause blurred or distorted vision and, if left untreated, can cause significant vision loss. Diabetic retinopathy is the leading cause of blindness, affecting over 100 million people globally [1].

In addition to diabetic retinopathy and macular edema, eye health can also be affected by glaucoma. Glaucoma is caused by increased fluid build-up in the eye, which leads to increased interocular pressure and ultimately damage of the optic nerve. This condition causes gradual vision loss, and progresses without noticeable symptoms in its early stages. This makes regular eye exams and screening especially important for Glaucoma, which is one of the leading causes of irreversible blindness that is undertreated globally. In LMICs, where access to medical screening is often limited, as many as 35% of people with Glaucoma are already blind as a result [2].

Age Related Macular Degeneration (AMD) is a metabolic disorder affecting the retina, leading to the release of waste products that are not properly broken down. This results in the formation of drusens and other changes of the fundus. Over time, these changes can cause progressive damage to the

macula—the central part of the retina responsible for sharp, detailed vision. As a result, AMD can lead to significant central vision loss, making it difficult to read and recognize faces. An estimated 19.8 million Americans, aged 40 and older, had age-related macular degeneration with 1.49 million experiencing vision-threatning late-stage AMD - this prevalence increased with age, especially among those 85 and older [3].

Myopic fundus is a severe form of myopia that can lead to rapid vision loss and increase the risk of blindness. It occurs due to the excessive stretching of the eye, which weakens the retina. While the damage is irreversible, early detection and proper management can help slow its progression and reduce the risk of severe vision impairment, making medical screening very important.

Each disease has specific fundus indicators, as elaborated in Table 1.

Disease	**Fundus Indicators**
Diabetic Retinopathy (DR)	Darkened fovea, hemorrhage, thrombosis
Glaucoma	Increased cup-to-disc ratio
Macular Edema	Fluid buildup around the macula, hard exudates (yellowish lipids and proteins)
Myopic Fundus	Visible tessellations (appearance of large blood vessels at the back of the eye)
Age-Related Macular Degeneration (AMD)	Hard exudates around the macula, drusens (yellow deposits in the retina), subretinal hemorrhage

Table 1. Fundus Indicators for Various Eye Diseases

Key features of the fundus, showcasing both normal anatomical structures and potential disease biomarkers, are shown in Figure 1 and Figure 2 for visual reference. Figure 1 illustrates components of the healthy fundus, including the optic disc, optic cup, macula, fovea, arteries, and veins, which are important for visual function. Figure 2 highlights possible pathological abnormalities, such as hard exudates and myopic tessellation, which may indicate retinal degeneration, as well as a possible hemorrhage, a biomarker for disorders like diabetic retinopathy. These features are essential for diagnosing and monitoring eye-related diseases.

2. BACKGROUND

2.1. Prior Work

Currently, many AI-based approaches for the detection of eye diseases have focused on single disease detection. [4], [5] [6], [7] have focused solely on the detection of one eye disease - diabetic retinopathy, age related macular degeneration, myopic degeneration, and glaucoma, respectively. In addition, while many approaches have applied deep learning transfer models for disease detection, the standard ImageNet-based pretraining was used for these models [8] [9]. ImageNet based pretraining is a common practice in computer vision, but has notable limitations when applied to medical image analysis. The domain discrepancy between images from ImageNet and fundus images can limit the effectiveness of features learned from ImageNet when transferred to other tasks [10]. In addition, the features learned from ImageNet images may not align well with specific characteristics of medical images, which can lead to suboptimal performance. Recently, RETFoundMAE [11] has revolutionized fundus based image analysis by applying MAE techniques to build a foundation model. However, this model is huge and is un-suited for applications on mobile devices. Finally, many approaches for disease detection have used only images, rather than integrating multiple data sources. Combining imaging data with clinical records, genetic information, and patient demographics can provide a more comprehensive understanding, potentially aiding with disease detection.

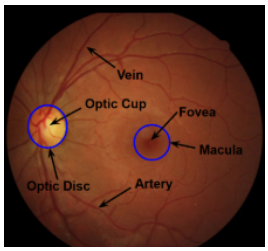


Fig. 1. Annotated fundus with key anatomical structures

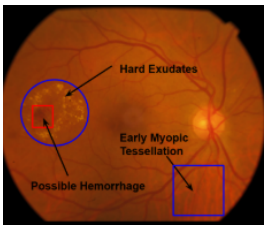


Fig. 2. Annotated fundus with select abnormalities

2.2. Novelty

We develop InSight, a medical screening tool for detection of multiple eye diseases on resource constrained mobile devices, using both images and patient metadata. The novelties of our method include:

- Three-stage pipeline consisting of an image quality checker, fundus disease detector and a dedicated DR grading model providing real time feedback at the time of image capture.
- mobile app that can diagnose five eye diseases, without needing connectivity to the internet.
- Novel fusion method (MetaFusion) for combining images and patient metadata for diagnosis.
- Novel pre-training method combining supervised and self-supervised losses to leverage multiple datasets of fundus images to improve model accuracy.
- *Single multi-task* model to predict all five diseases efficiently with high accuracy.

3. MATERIAL AND METHODS

3.1. Datasets

To develop InSight, we primarily make use of the **Brazilian Multilabel Opthamological Datset (BRSET)** [4]. This dataset contains over 16000 fundus images from 8524 Brazilian patients, with clinical metadata and pathological/anatomical classification parameters. Patient age, gender, diabetes diagnosis, duration of diabetes diagnosis are the clinical metadata of interest. BRSET also contains disease diagnosis information for a variety of eye diseases, including diabetic retinopathy, glaucoma, macular edema, age-related macular degeneration, and myopic fundus. Information about the proportion of patients with each of these diseases in the dataset is shown in Table 2.

Eye Diseases	**BRSET Patients With Disease (%)**
Diabetic Retinopathy	6.4%
Glaucoma	19.7%
Macular Edema	2.5%
AMD	2.2%
Myopic Fundus	1.6%

Table 2. Statistics of BRSET dataset

We also use the **Mobile Brazilian Multilabel Ophthalmological Dataset (mBRSET)** [12] dataset for our experiments. This dataset contains 5164 images captured on smartphone-based retinal fundus cameras from 1291 patients containing labels for DR and Macular Edema. This dataset is used to test the performance of InSight on smartphone captured fundus images.

3.2. Data Preprocessing

For certain metadata features such as hypertension, the information was not reported separately - instead it was stated along with other conditions in the Comorbidities column of the dataset, which contains self-reported patient clinical history. Data processing was done to extract hypertension occurence as a binary feature from the patient clinical history. In addition, some of the patients diagnosed with diabetic retinopathy were recorded as not having diabetes, which is incorrect since diabetic retinopathy is a complication of diabetes. Similarly, patients diagnosed with diabetic macular edema by definition have both diabetes and diabetic retinopathy. The data was processed so that these consistency conditions were enforced. We only consider metadata features that are common to BRSET and mBRSET.

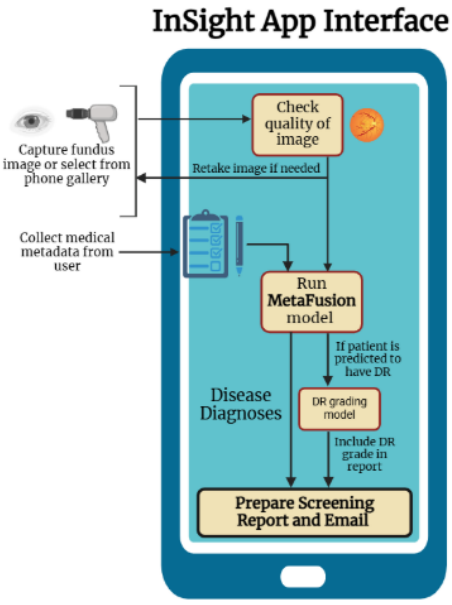


Fig. 3. InSight Pipeline

3.3. High Level InSight Pipeline

The InSight pipeline consists of three stages as shown in Figure 3, an image quality checker, a multimodal pretrained disease diagnosis model, and a diabetic retinopathy grading model. The pipeline follows these steps:

1. Image is captured or selected from the phone gallery.
2. The quality of the image is assessed by the image quality checker. If it is of bad quality (blurred, altered brightness), the user is prompted to retake the image.
3. When the image is of sufficient quality, it is passed to the disease diagnosis model.
4. The diagnosis for each disease is displayed to the user.
5. For diabetic retinopathy, the stage of progression of the disease is reported.
6. Finally, a medical screening report is provided and can be emailed to the patient.

We split the BRSET dataset into training, validation and test with a ratio of 0.5:0.25:0.25. We also perform experiments with a combined dataset of BRSET and mBRSET, where we adopt the same split ratio. We create an augmented test set by applying blur and illumination changes to the test set to increase its size by 2x.

3.4. Image Quality Checker Model

Our pipeline begins with an image quality checker, which is designed to detect low quality images that are passed as input for InSight medical screening. For training the image quality checker, we blur the images in the training set and use these along with original images.

In order to warn users of images that are of low quality, we perform data augmentations on BRSET images to create

equivalent low quality images. The albumentations package is used, with MotionBlur, GaussianBlur and MedianBlur to simulate blurring due to camera motion and Random Brightness Contrast was used to simulate different image exposure conditions. The CNN was trained as a binary classifier to classify these images as either high or low quality. The filtered images are then passed to the rest of the pipeline - Disease Diagnosis Model and Diabetic Retinopathy Grading Model.

3.5. Disease Diagnosis Model

The disease diagnosis model takes in fundus images and metadata and uses a novel fusion method to fuse images with metadata, followed by linear layers that are task specific. The model architecture is shown in Figure 4. For deployment on low end mobile devices, we choose resnet18 as the image backbone. Fusion of metadata with images is done using MetaFusion, described briefly below.

3.5.1. MetaFusion Algorithm

MetaFusion is a novel fusion method that combines intermediate representations (embeddings) prior to the final layer of a model, integrating clinical metadata with imaging data. The key ideas of the algorithm are:

1. Apply a correction term to image and metadata embeddings prior to concatenation using the idea of residual connections.
2. We pick the correction term to be of the form $f(x_1, x_2) = x_1 \odot s(x_1, x_2)$ where $s(x_1, x_2)$ is a similarity function that ranges from -1 to 1.
3. We choose the similarity function to be $s(x_1, x_2) = tanh(x_1 \odot Wx_2)$, inspired by the design of MetaBlock [13].

Putting it all together, we define the MetaFusion block in Equation (1).

$$f(x_1, x_2) = x_1 \odot tanh(x_1 \odot (Wx_2)) \tag{1}$$

The MetaFusion Layer consists of multiple of these correction terms (denoted by f) in order to modify the image and metadata (Equations (2) and (3)).Equation (2) represents the modification of the image based on the metadatas, and Equation (3) represents the modification of the metadata embeddings based on the image.

$$X_1 = X_1 + \sum_{i=1}^{5} f(X_1, X_{2,\,i}) \tag{2}$$

$$X_{2,i} = X_{2,i} + f\,(X_{2,i}, X_1) \tag{3}$$

3.5.2. Multitask Classification

All five diseases are predicted simultaneously using a single model, making it computationally efficient for mobile deployement. The loss function for training is the sum of the cross entropy losses for each of the five tasks.

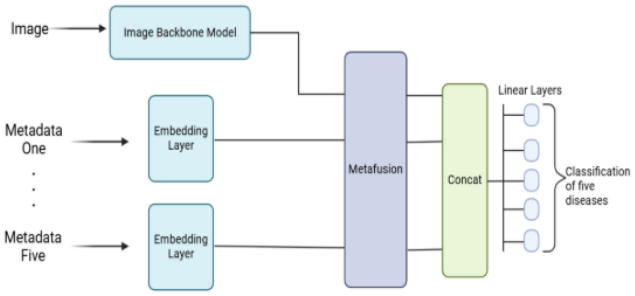


Fig. 4. Disease Diagnosis Model Architecture

3.5.3. Pretraining

To address limited availability of data, we make use of other existing datasets of individual eye diseases [14], [15], [16], [17]. We combine individual datasets for diabetic retinopathy and Glaucoma into a larger dataset of 130000 images. We develop a novel loss function that combines a supervised classification loss with an unsupervised image reconstruction loss, as shown in Figure 5 to ensure that pre-training does not overfit the model to Diabetic Retinopathy and Glaucoma classification.

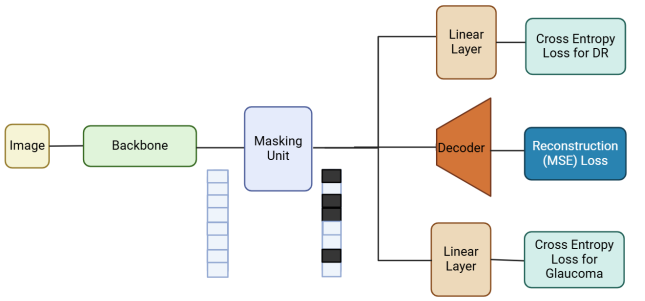


Fig. 5. The image is passed through a backbone to create an image embedding. A mask is then applied to the image embedding to remove parts of the embedding. At this point, three different losses are computed - cross entropy loss for diabetic retinopathy, cross entropy loss for glaucoma, and a reconstruction loss.

3.5.4. Diabetic Retinopathy Grading Model

For the DR grading model, we only consider images with diabetic retinopathy in the BRSET dataset for training, validation and testing. The model outputs whether the patient has mild

non-proliferative diabetic retinopathy, or severe proliferative diabetic retinopathy.

4. RESULTS

We evaluate the performance of the three stage pipeline using the following metrics:

- **Balanced Accuracy (BA):** Average of true positive rate and true negative rate: used to handle imbalanced datasets. BRSET is highly imbalanced (see Table 2)
- **Receiver Operating Characteristic (ROC):** Graph of true positive rate vs. false positive rate at various classification threshold settings
- **Area under curve (AUC):** area under ROC curve, measure of classifier performance

4.1. Overall improvements

We evaluate the end to end performance for disease detection on the augmented test set. Note that only images passed by the image quality checker are considered for evaluation of the disease diagnosis model in Figure 6.

Disease	Single Image Baseline: No image quality check	MetaFusion Model
Diabetic Retinopathy	0.79	0.96
Glaucoma	0.77	0.85
Macular Edema	0.83	0.97
Myopic Fundus	0.87	0.94
AMD	0.82	0.92

Fig. 6. Combining image quality checker with MetaFusion model and pre-training gives large balanced accuracy improvements

4.2. Disease Diagnosis model

In this section we analyze the performance of the disease diagnosis model in isolation, assuming that the image quality checker rejects all blurred images. We evaluate the impact of pre-training, fusion and multi-tasking below.

We consider a single image baseline model (Single Image Baseline) and compare the performance of this model with and without pre-training in Figure 7.

Eye Disease	Single Image Baseline	Single Image Baseline + Pretraining
Diabetic Retinopathy	0.86	0.87
Glaucoma	0.80	0.85
Macular Edema	0.89	0.92
Myopic Fundus	0.92	0.94
AMD	0.87	0.88

Fig. 7. Pretraining leads to balanced accuracy Improvements for all five diseases. The largest improvement was observed for Glaucoma, which makes sense as the model was pretrained extensively for Glaucoma classification. A large improvement was also observed for Macular Edema, suggesting that pretraining for the diabetic retinopathy classification task helps with the detection of severe DR (Macular Edema is a serious complication of diabetic retinopathy).

In order to understand the impact of fusing metadata with images, we do the following comparisons: We compare a model that uses only imaging data (with pre-training) with a model that uses only metadata (by concatenating the metadata embeddings) with the MetaFusion model that fuses information from both models in Figure 8.

Eye Diseases	Single Image Baseline + Pretraining	Metadata Only	MetaFusion
Diabetic Retinopathy	0.87	0.94	0.96
Glaucoma	0.85	0.61	0.85
Macular Edema	0.92	0.86	0.97
Myopia	0.94	0.50	0.94
AMD	0.88	0.72	0.92

Fig. 8. For certain diseases like Diabetic Retinopathy and Macular Edema, the metadata-only model does well, with balanced accuracies of 0.94 and 0.86 respectively. This is expected as diabetes is an important predictor of both diseases. For Myopia, Glaucoma, and AMD, the single image baseline with pre-training outperforms the metadata-only model since the fundus image features contain important information. MetaFusion combines information from both modalities and shows the best performance for all diseases.

The ROC of the disease diagnosis model is shown in Figure 9.The improvements due to combining pre-training and fusion are shown in Figure 10. We also compare the multi-task model with five individual models, each tuned for one disease in Figure 11.

4.3. DR Grading model

The DR grading model is able to achieve a relatively high BA of 0.85 due to the cascaded architecture of the InSight pipeline as shown in Figure 12.

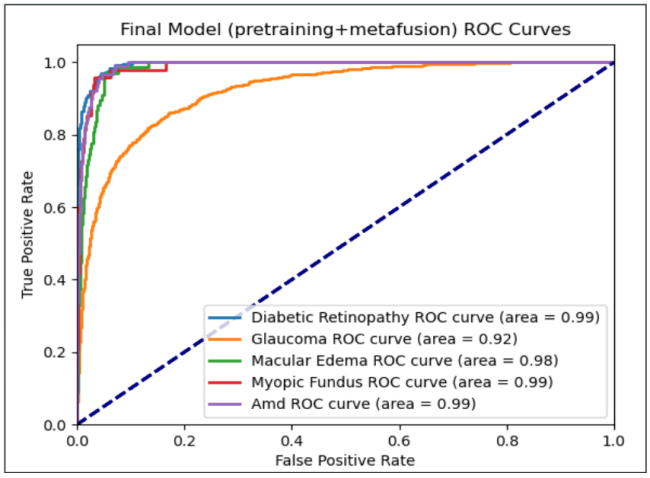


Fig. 9. Our MetaFusion model achieves AUCs of 0.99 (DR, Myopic Fundus, AMD), 0.98 (Macular Edema), and 0.92 (Glaucoma). Notably, the AUC of 0.99 for DR outperforms the result of 0.87 in [4]

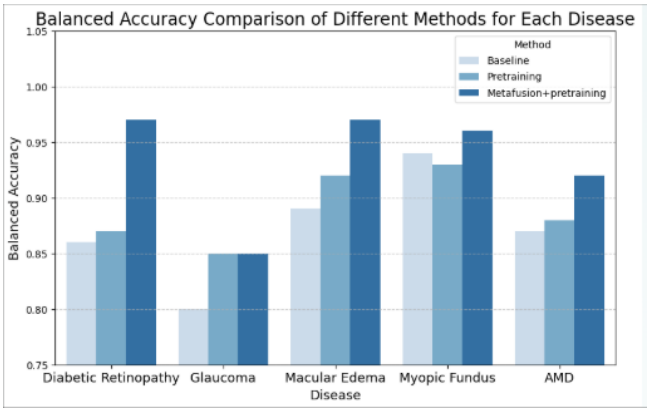


Fig. 10. Combined MetaFusion with pretraining consistently has the highest balanced accuracy for all five diseases. For Diabetic Retinopathy, Macular Edema and AMD, we observe large improvements in balanced accuracy with the addition of Metafusion to the pretraining method. For Glaucoma, pretraining significantly improves balanced accuracy, with fusion playing an insignificant role.

Eye Disease	MetaFusion: 5 individual models for each disease	MetaFusion: 1 model predicting all diseases
Diabetic Retinopathy	0.95	0.96
Glaucoma	0.85	0.85
Macular Edema	0.95	0.97
Myopic Fundus	0.94	0.94
AMD	0.92	0.92

Fig. 11. Single task and Multitask performance comparisons were also conducted with the MetaFusion method. Both demonstrated comparable balanced accuracy for all five diseases, showing how our multitask classification method achieves high performance at a much lower complexity.

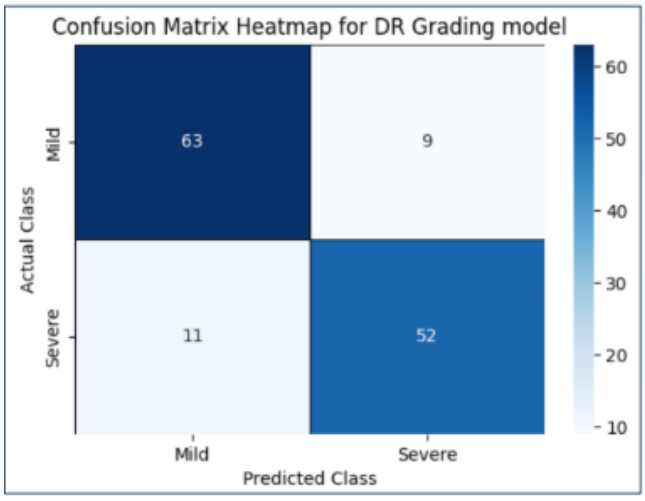


Fig. 12. The Diabetic Retinopathy Grading model achieves a balanced accuracy of 0.85.

4.4. Extension to Smartphone captured images (mBRSET)

We also report results on the models trained on both BRSET and mBRSET.

- The image quality checking model generalizes well for both BRSET and mBRSET and has close to 100% accuracy
- DR grading has a balanced accuracy of 80% on mBRSET, with the BRSET BA is unchanged.

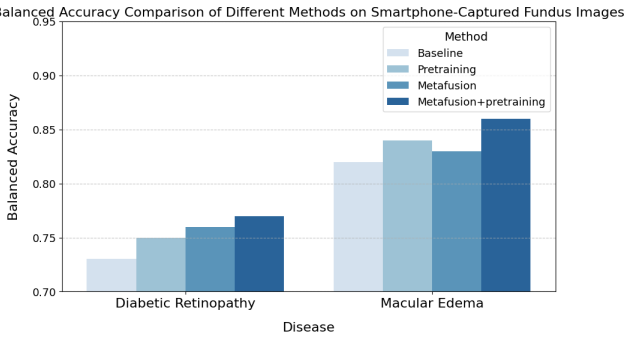


Fig. 13. For a model trained on the joint dataset, combined MetaFusion and pretraining shows improvements of 4% in BA for both DR and macular edema on mBRSET.

5. DISCUSSION

To better understand the importance of metadata, we look at the correlation between metadata and disease labels. We note that there is a high correlation between diabetic retinopathy and macular edema with diabetes and duration of diabetes, as shown by the larger correlation coefficients in the heatmap. This outcome is expected, as both diabetic retinopathy and

Disease	Trained on BRSET Tested on BRSET	Trained on BRSET and mBRSET Tested on BRSET
Diabetic Retinopathy	0.96	0.96
Glaucoma	0.97	0.97
Macular Edema	0.94	0.93
Myopic Fundus	0.85	0.83
AMD	0.92	0.90

Disease	Trained on mBRSET Tested on mBRSET	Trained on BRSET and mBRSET Tested on mBRSET
Diabetic Retinopathy	0.79	0.79
Macular Edema	0.87	0.86

Fig. 14. Our disease diagnosis model generalizes well across smartphone and lab captured fundus images. A model trained on a joint (mBRSET+BRSET) dataset performs almost as well as models trained on BRSET and mBRSET respectively.

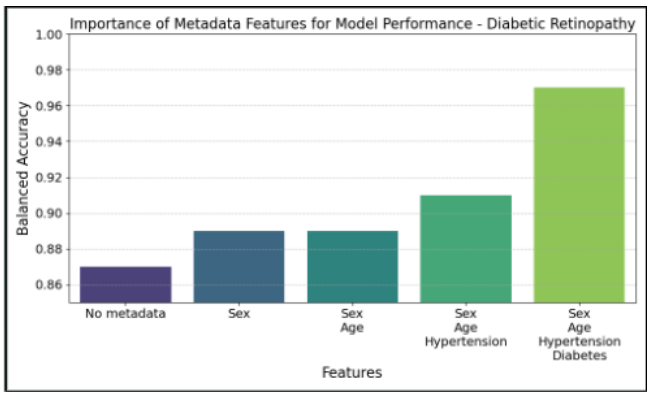


Fig. 15. Here we can observe the increasing trend in balanced accuracy with the addition of each metadata for diabetic retinopathy prediction. When adding certain metadata, such as Age, we hardly see an improvement, but when Diabetes is added, there is a drastic increase in balanced accuracy. This aligns with the correlation heatmap findings in Figure 16, which highlight diabetes and diabetes duration as the most influential factors.

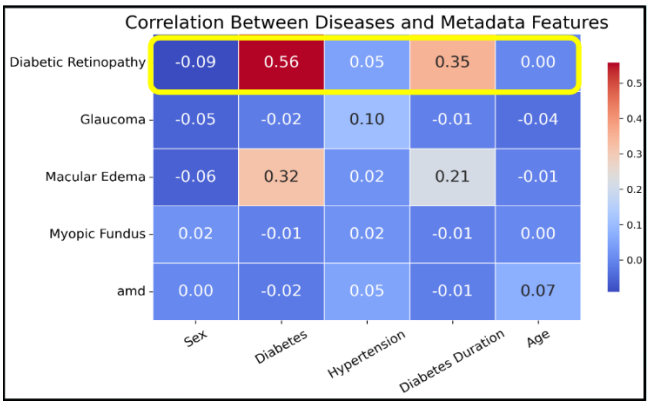


Fig. 16. Correlation Heatmap between Diseases and Metadata Features

macular edema are complications of diabetes and thus have a high correlation with these features. Other diseases, such as myopic fundus, show relatively very little correlation with the metadata. For AMD, we note that age is the most relevant metadata features as expected.

We compare the impact of individual metadata features on the performance of the diabetic retinopathy model in Figure 15.

It is important to verify that the model uses interpretable features in the fundus image for diagnosis. We conducted a saliency analysis for images of each disease, as shown in 17. In order to this, we implement a sliding window approach as described in [18].

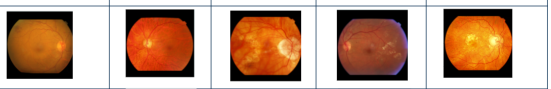
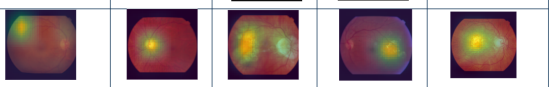
	Diabetic Retinopathy	Glaucoma	Myopic Fundus	Macular Edema	AMD
Original					
Saliency					
Fundus Image Salient Feature	Hemorrhage region	Optic disc / macula	Both the tilted crescent shape of optic disc and the tessellation patterns	Hard exudates surrounding the macula	Drusens and Exudates

Fig. 17. The model focuses on the correct fundus indicators for all diseases, providing interpretable explanations for the diagnosis, matching the fundus indicators explained in Table 1.

6. CONCLUSION

In this work, we developed InSight, a three-stage pipeline for accurate detection and diagnosis of five eye diseases. The image quality checker model effectively filters out low-quality images, improving the accuracy of downstream models. Our multitask disease diagnosis model, operating at five times lower complexity, achieves comparable performance to individual disease-specific models, making it a lightweight and efficient solution for app deployment. Our multistage pipeline consistently outperforms single-stage image-only models with an average improvements 11% in balanced accuracy.

Our model achieves AUCs of 0.99 (DR, Myopic Fundus, AMD), 0.98 (Macular Edema), and 0.92 (Glaucoma) on the BRSET dataset, surpassing the previous benchmark of 0.87 in [4].

In addition, InSight generalizes well to both smartphone fundus camera captured images and lab captured images. On the BRSET dataset, the disease diagnosis model shows an improvement of 6% in balanced accuracy across all five diseases compared to a model which relies only on images without pre-training enhancements. On the mBRSET dataset,

there is a 4% improvement in balanced accuracy for diabetic retinopathy and macular edema. In addition, the model trained on a joint (mBRSET+BRSET) dataset performs almost as well as models trained on BRSET and mBRSET respectively, demonstrating the robustness of the disease diagnosis model on a range of imaging conditions.

We are currently reaching out to ophthalmology researchers to further refine InSight by using datasets captured on smartphone fundus cameras with annotations for multiple diseases. We are also contacting smartphone-based retinal imaging system manufacturers to integrate our app to serve millions of people in LMICs.

We believe that our work establishes a foundation for AI-assisted eye disease screening, offering a scalable and accessible solution for early diagnosis and improved patient outcomes.

7. REFERENCES

[1] Zhen Ling Teo, Yih-Chung Tham, Marco Yu, Miao Li Chee, Tyler Hyungtaek Rim, Ning Cheung, Mukharram M. Bikbov, Ya Xing Wang, Yating Tang, Yi Lu, Ian Y. Wong, Daniel Shu Wei Ting, Gavin Siew Wei Tan, Jost B. Jonas, Charumathi Sabanayagam, Tien Yin Wong, and Ching-Yu Cheng, "Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis," *Ophthalmology*, vol. 128, no. 11, pp. 1580–1591, 2021.

[2] Victor H. Hu, "The global challenge of glaucoma," *Community Eye Health*, vol. 34, no. 112, pp. 61, January 2022.

[3] David B Rein, John S Wittenborn, Zeb Burke-Conte, Rohit Gulia, Toshana Robalik, Joshua R Ehrlich, Elizabeth A Lundeen, and Abraham D Flaxman, "Prevalence of age-related macular degeneration in the US in 2019," *JAMA Ophthalmol.*, vol. 140, no. 12, pp. 1202–1208, Dec. 2022.

[4] Luis Filipe Nakayama, David Restrepo, João Matos, Lucas Zago Ribeiro, Fernando Korn Malerbi, Leo Anthony Celi, and Caio Saito Regatieri, "BRSET: A brazilian multilabel ophthalmological dataset of retina fundus photos," *medRxiv*, Jan. 2024.

[5] Ngoc Thien Le, Thanh Le Truong, Sunchai Deelertpaiboon, Wattanasak Srisiri, Pear Ferreira Pongsachareonnont, Disorn Suwajanakorn, Apivat Mavichak, Rath Itthipanichpong, Widhyakorn Asdornwised, Watit Benjapolakul, Surachai Chaitusaney, and Pasu Kaewplung, "ViT-AMD: A new deep learning model for age-related macular degeneration diagnosis from fundus images," *International Journal of Intelligent Systems*, vol. 2024, pp. e3026500, 2024.

[6] Ran Du, Shiqi Xie, Yuxin Fang, Tae Igarashi-Yokoi, Muka Moriyama, Satoko Ogata, Tatsuhiko Tsunoda, Takashi Kamatani, Shinji Yamamoto, Ching-Yu Cheng, Seang-Mei Saw, Daniel Ting, Tien Y Wong, and Kyoko Ohno-Matsui, "Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images," *Ophthalmol. Retina*, vol. 5, no. 12, pp. 1235–1244, Dec. 2021.

[7] Oleksandr Kovalyk, Juan Morales-Sánchez, Rafael Verdú-Monedero, Inmaculada Sellés-Navarro, Ana Palazón-Cabanes, and José-Luis Sancho-Gómez, "PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment," *Sci. Data*, vol. 9, no. 1, pp. 291, June 2022.

[8] Nour Eldeen M Khalifa, Mohamed Loey, Mohamed Hamed N Taha, and Hamed Nasr Eldin T Mohamed, "Deep transfer learning models for medical diabetic retinopathy detection," *Acta Inform. Med.*, vol. 27, no. 5, pp. 327–332, Dec. 2019.

[9] Nahida Akter, John Fletcher, Stuart Perry, Matthew P Simunovic, Nancy Briggs, and Maitreyee Roy, "Glaucoma diagnosis using multi-feature analysis and a deep learning technique," *Sci. Rep.*, vol. 12, no. 1, pp. 8064, May 2022.

[10] Laith Alzubaidi, Muthana Al-Amidie, Ahmed Al-Asadi, Amjad J. Humaidi, Omran Al-Shamma, Mohammed A. Fadhel, Jinglan Zhang, J. Santamaría, and Ye Duan, "Novel transfer learning approach for medical imaging with limited labeled data," *Cancers*, vol. 13, no. 7, pp. 1590, 2021.

[11] Yuhang Zhou, Michael A. Chia, Saskia K. Wagner, Murat S. Ayhan, Drew J. Williamson, Robrecht Struyven, Steven G. McDonagh, Xialei Liu, Xiaoxuan Zhang, Alastair K. Denniston, Pearse A. Keane, Cornelia Schmid, and Jeffrey de Fauw, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, pp. 156–163, 2023.

[12] Luis Filipe Nakayama, Lucas Zago Ribeiro, David Restrepo, Nathan Santos Barboza, Raul Dias Fiterman, Maria Luiza Vieira Sousa, Alexandre Durao Alves Pereira, Caio Regatieri, Fernando Korn Malerbi, and Rafael Andrade, "mBRSET, a mobile brazilian retinal dataset," June 2024.

[13] Andre G C Pacheco and Renato A Krohling, "An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3554–3563, Sept. 2021.

[14] Ascanipek, "Eyepacs, aptos, messidor diabetic retinopathy dataset," `https://www.kaggle.com/datasets/ascanipek/eyepacs-aptos-messidor-diabetic-retinopathy`.

[15] Deathtrooper, "Eyepacs airogs light dataset," `https://www.kaggle.com/datasets/deathtrooper/eyepacs-airogs-light`.

[16] Arnav Jain, "Glaucoma datasets," `https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets`.

[17] Andrew Mvd, "Ocular disease recognition (odir-5k) dataset," `https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k`.

[18] Britefury, "Image region-level saliency using vgg-16 conv-net," `https://github.com/Britefury/deep-learning-tutorial-pydata/blob/master/TUTORIAL%2003%20-%20Image%20region-level%20saliency%20using%20VGG-16%20conv-net.ipynb`.