Early Detection of Glaucoma Using Machine Learning

This study addresses the development of a machine learning model for the early detection of glaucoma from color fundus photographs (CFPs), aiming to mitigate one of the leading causes of irreversible blindness. By leveraging a large dataset of over 110,000 annotated images, we created and validated algorithms to enhance the accuracy and generalizability of glaucoma detection, capitalizing on the power of convolutional neural networks (CNNs).

Group-3

Addressing the Challenge of Glaucoma

Glaucoma Prevalence

Glaucoma is one of the leading causes of blindness because it frequently goes undetected until it has already caused significant visual loss. Early diagnosis is crucial to prevent the progression of the disease and preserve eyesight.

AI-based Screening

Previous research studies have explored the use of CFPs for artificial intelligence-based glaucoma screening, with promising results. However, the performance of AI models often declines when tested on different datasets, highlighting the need for robust and generalizable algorithms.

1

2

3

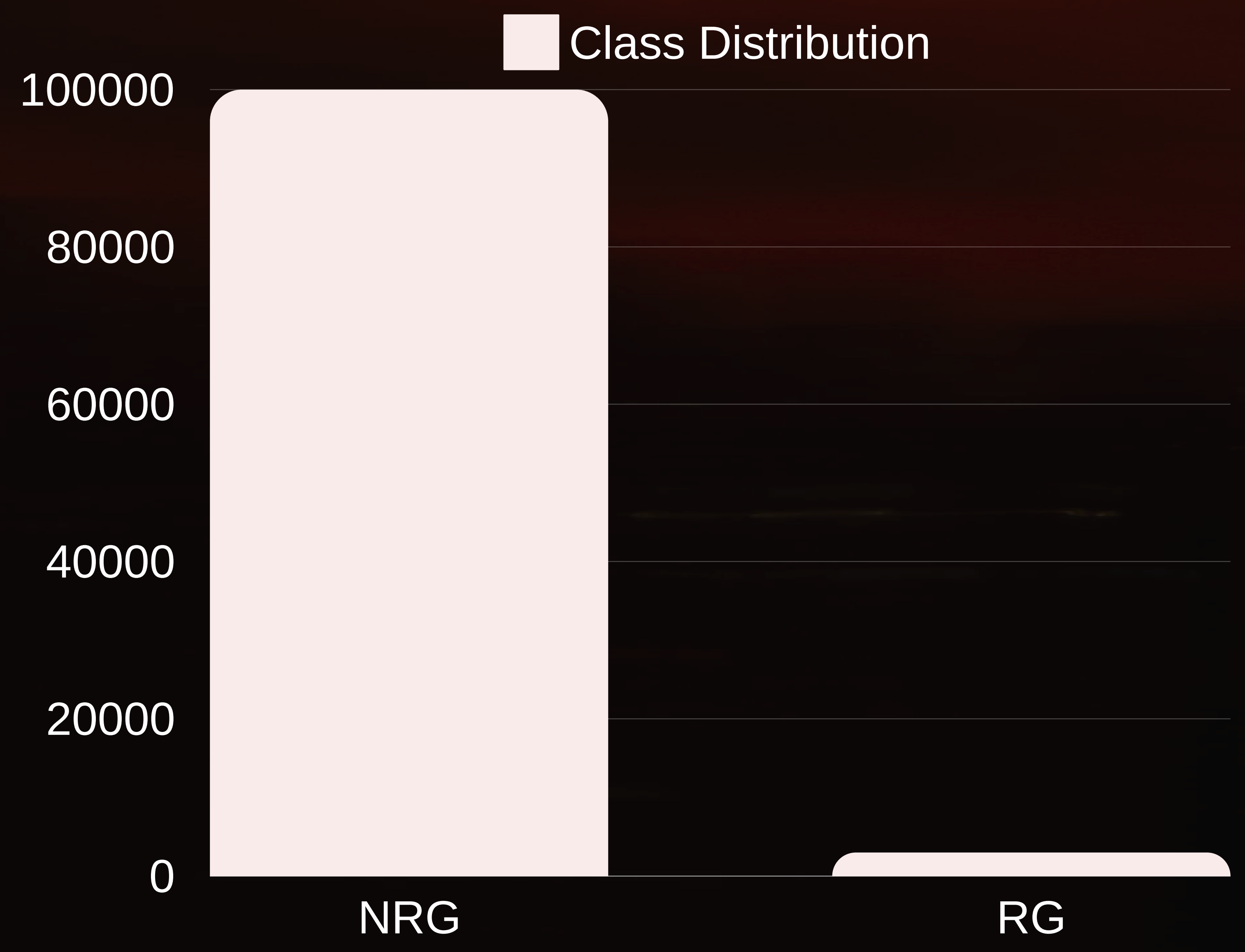
Imaging Techniques

Imaging techniques such as color fundus photographs (CFPs) and optical coherence tomography (OCT) play a vital role in the early detection of glaucoma, enabling healthcare professionals to identify the disease at its earliest stages.

Developing a Robust Glaucoma Detection Model

1 Dataset

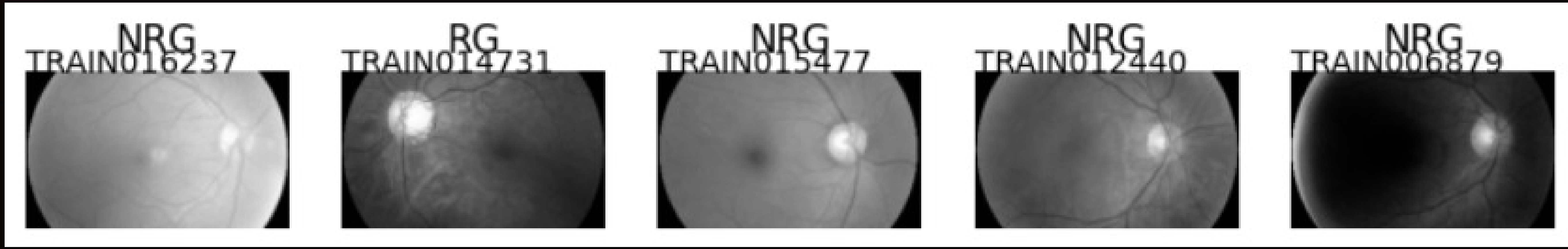
We utilized a dataset of over 110,000 carefully annotated fundus images obtained from approximately 60,000 individuals undergoing glaucoma screening. Each image was labeled as either "referable glaucoma" (RG) or "no referable glaucoma" (NRG), and the RG images were further annotated with up to ten additional labels associated with various glaucomatous features.



Developing a Robust Glaucoma Detection Model

2 Preprocessing

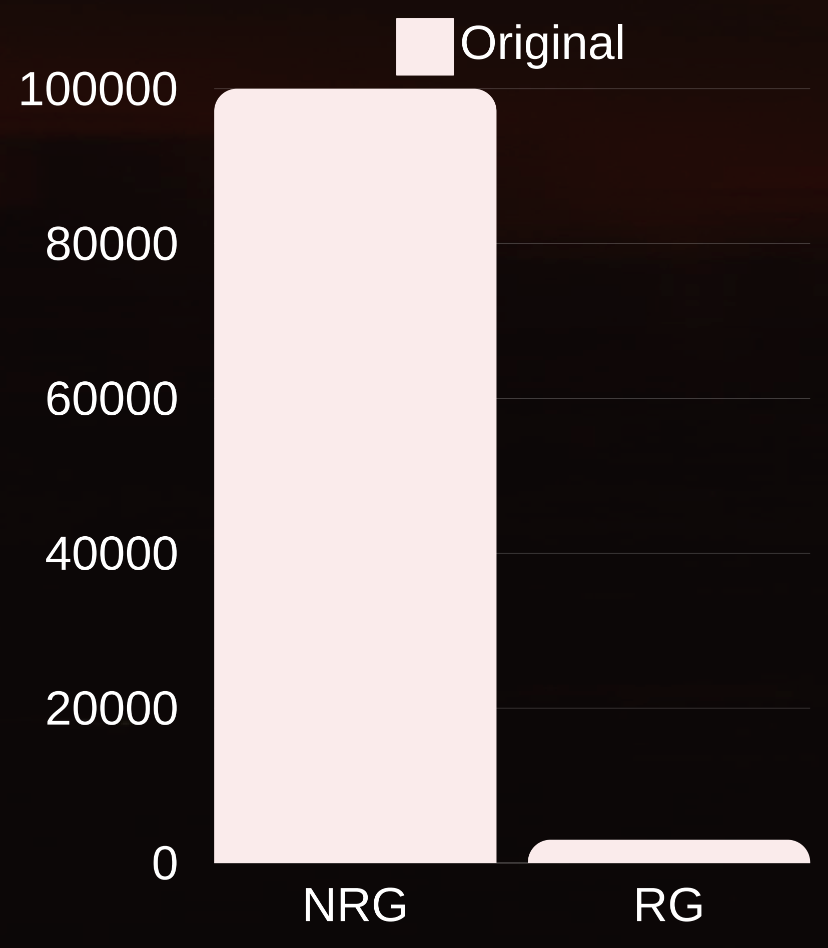
We implemented a novel cropping technique to focus on the central areas of the eye, crucial for glaucoma analysis. They also optimized the grayscale conversion process to emphasize textural and structural changes rather than color variations, enhancing computational efficiency while preserving diagnostic relevance.

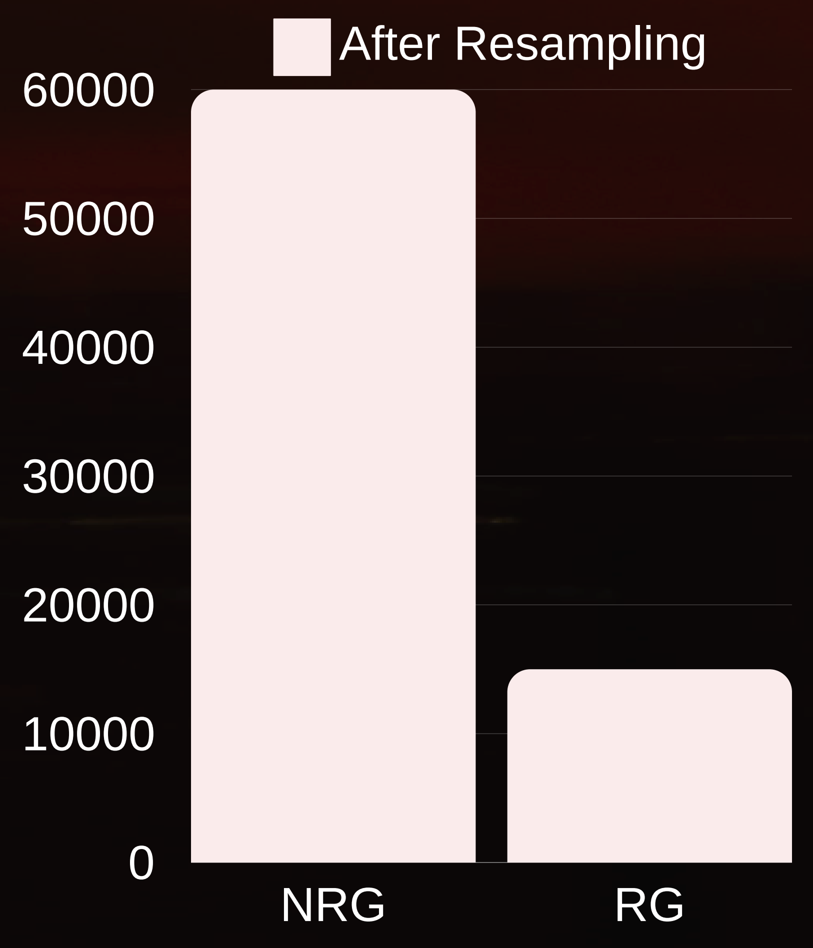


Developing a Robust Glaucoma Detection Model

3 Class Imbalance

To address the significant class imbalance in the dataset, the researchers applied resampling techniques to balance the classes. By increasing the number of glaucoma images and reducing non-glaucoma images, they created a more representative training set, crucial for training unbiased models sensitive to detecting glaucoma.





Exploring Model Architectures

1 Custom Architectures

We experimented with various custom convolutional neural network (CNN) architectures tailored for glaucoma detection, adapting dense layers and modifying the overall structure to optimize performance and generalizability.

2 Pre-trained Models

In addition to custom architectures, we also explored the integration of pre-trained models, such as VGG16, to leverage the feature extraction capabilities of these well-established networks and further enhance the model's performance.

3 Evaluation Metrics

We prioritized the area under the curve (AUC) as the primary evaluation metric, as it serves as a robust measure of the model's ability to differentiate between glaucoma and non-glaucoma cases, and it is insensitive to changes in the class distribution

Task 1 Model Architecture

```
base_model = VGG16(include_top=False, input_tensor=(120, 200, 1))

for layer in base_model.layers:
    layer.trainable = False

x = Flatten()(base_model.output)
x = Dense(128, activation='relu')(x)
x = Dropout(0.1)(x)
x = Dense(128, activation='relu')(x)
x = Dropout(0.1)(x)
output = Dense(1, activation='sigmoid')(x)

model = Model(inputs=base_model.input, outputs=output)
model.compile(optimizer='adam', loss='binary_crossentropy',
              metrics=['auc']
history = model.fit(train_generator, validation_data=val_generator,
              epochs=30, callbacks=[es, mc], batch_size=64)
```

Layer	Output Shape
input layer	120, 200, 1
vgg16	3, 6, 512
flatten	9216
dense	128
dropout (0.1)	128
dense	128
dropout (0.1)	128
dense	1

Task 2 Model Architecture

```
model = Sequential([
    layers.Conv2D(32, (3, 3), activation='relu', input_shape=(120, 200, 1)),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Flatten(),
    layers.Dense(256, activation='relu'),
    layers.BatchNormalization(),
    layers.Dropout(0.1),
    layers.Dense(10, activation='sigmoid')
])

model.compile(optimizer='adam', loss='categorical_crossentropy',
              metrics=['accuracy', 'Precision', 'Recall', 'AUC'])

history = model.fit( train_generator, validation_data=val_generator,
              epochs=30, callbacks=[mc] )
```

Layer	Output Shape
input layer	120, 200, 1
conv2d	118, 198, 32
max_pooling	59,99,32
conv2	57,97,64
max_pooling	28,48,64
flatten	86016
dense	256
batch_normalization	256
dropout (0.1)	256
dense	10

Comprehensive Model Evaluation

ROC Curve and Confusion Matrix

We visually represented our results using a confusion matrix to show the breakdown of correct and incorrect predictions, and a ROC curve to evaluate the model's diagnostic ability at various thresholds.

Recall and F1-Score

Recall was used to check how well the model identified all glaucoma cases, while the F1-score helped balance precision and recall, ensuring both aspects were equally important.

Accuracy and Precision

We evaluated the model's performance using a range of metrics, including accuracy to gauge the overall correctness of the model and precision to assess the accuracy of positive glaucoma detections.

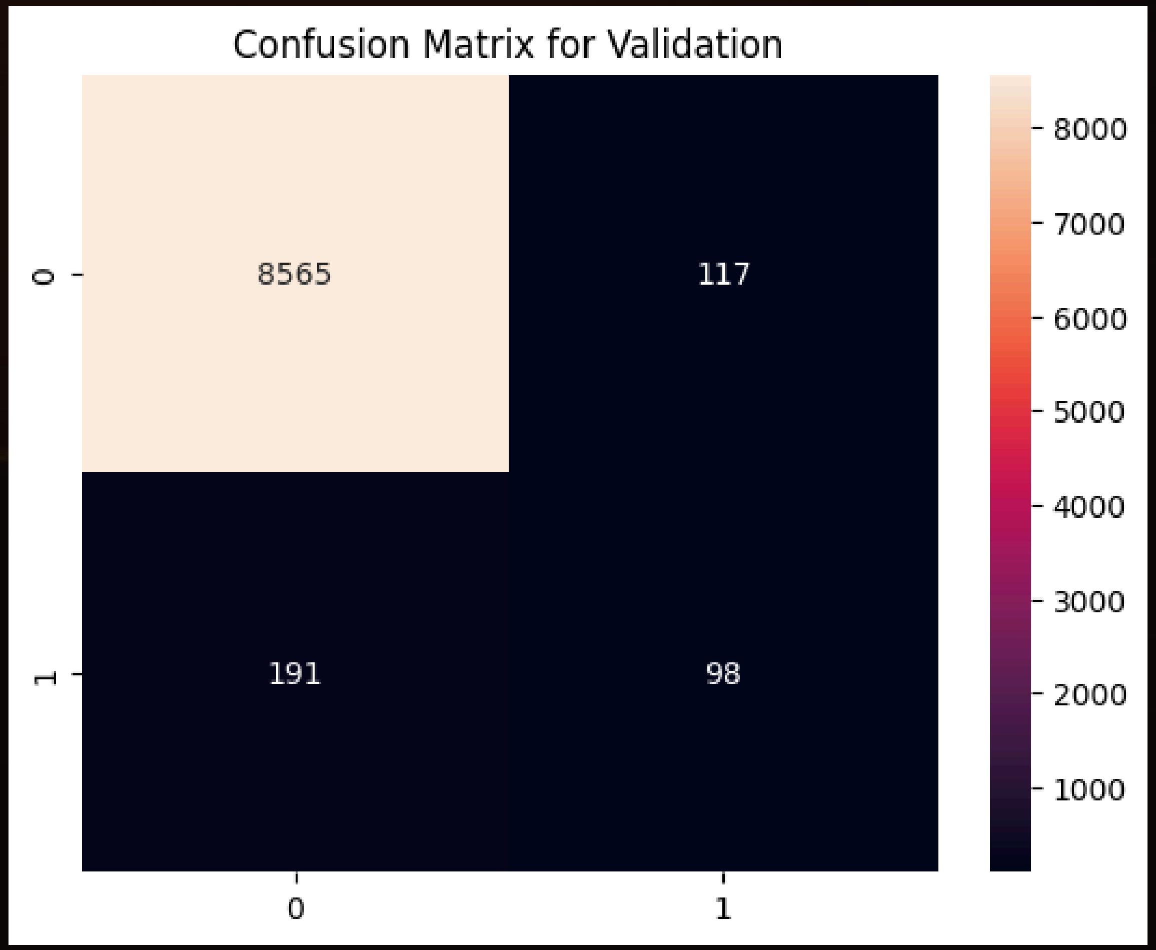
Sensitivity at 95% Specificity

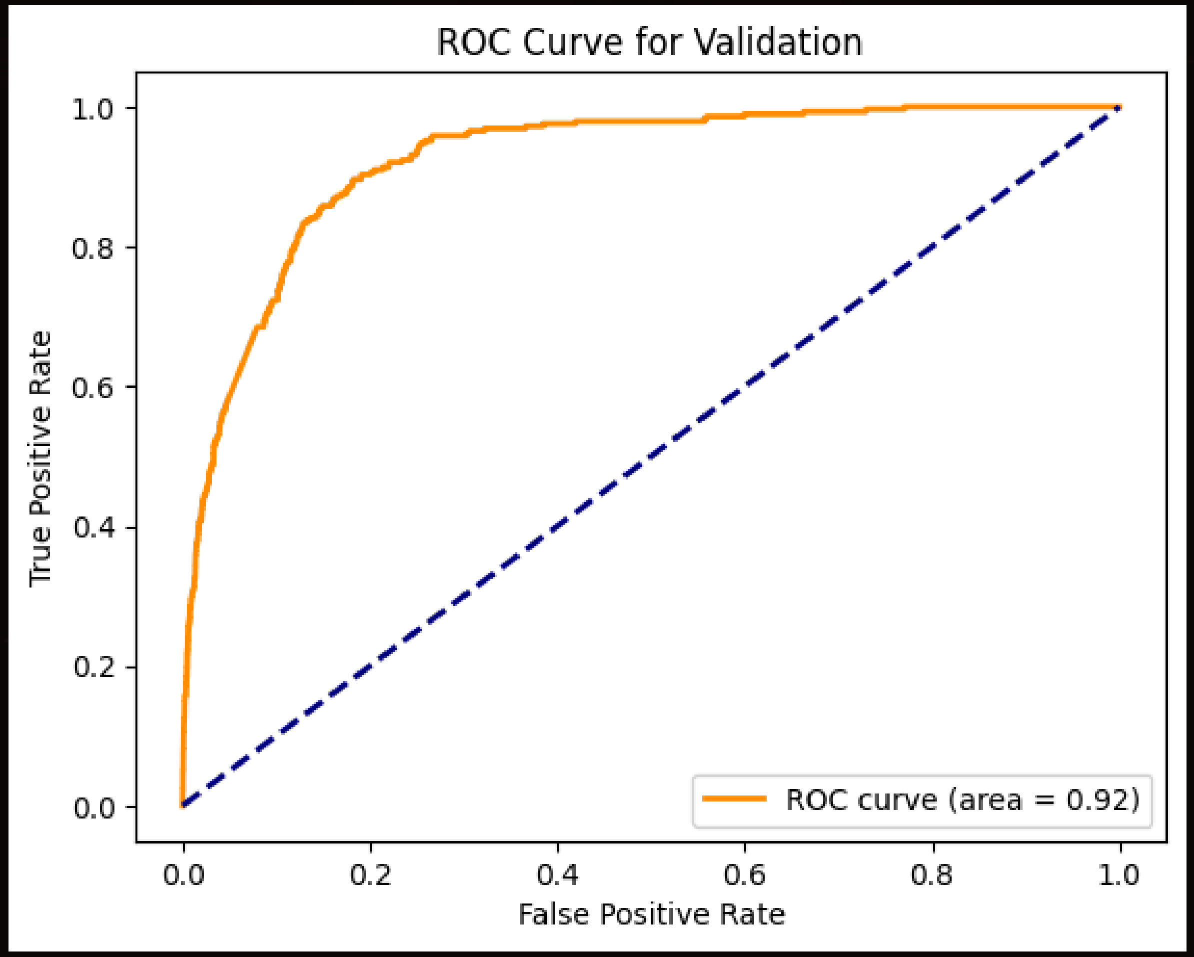
We determined a specific threshold that allowed the model to achieve 95% specificity as required by the challenge, ensuring the model minimized the risk of misdiagnosis and made it a reliable tool in clinical settings.

Task 1 Model Evaluation

Metrics for Validation Data:

AUC	F1-Score	Recall	Precision	Accuracy	Threshold at Specificity 95.00%	Sensitivity at Specificity 95.00%
92.4%	0.388	0.339	0.455	96.5%	0.2718	0.685





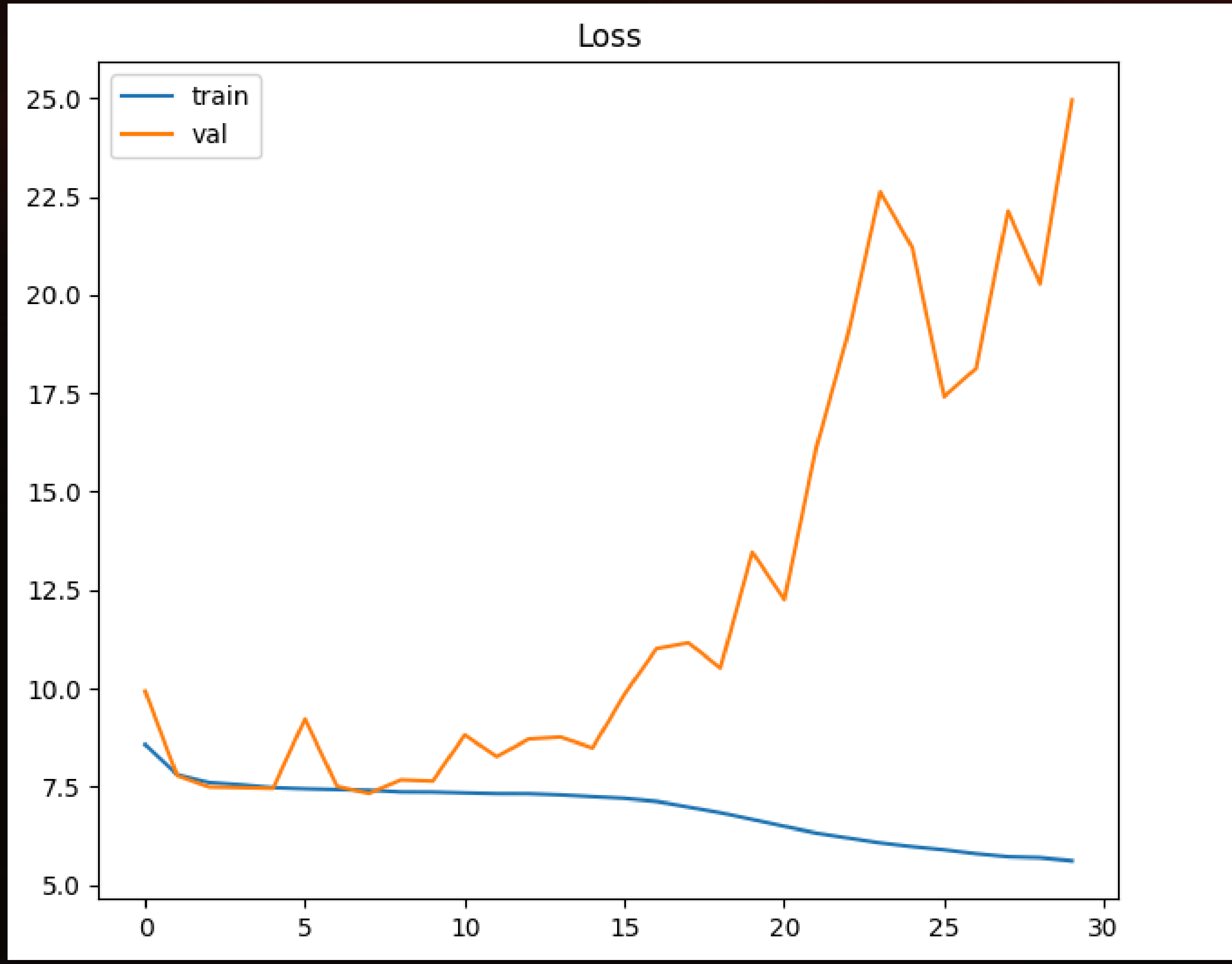
Task 2 Model Evaluation

Metrics for Train Data:

AUC	Recall	Precision	Accuracy	Loss
0.95	0.98	0.67	0.41	5.47

Metrics for Validation Data:

AUC	Recall	Precision	Accuracy	Loss
0.66	0.53	0.50	0.33	24.95



Promising Results and Potential Challenges

High AUC

The model achieved an impressive AUC of approximately 92% on both validation and test datasets, indicating its strong performance in differentiating between glaucoma and non-glaucoma cases.

Precision and Recall Concerns

However, the model's relatively low precision and recall rates suggest that while it is efficient at screening out non-glaucoma cases, it struggles to correctly identify true glaucoma cases among the positives it flags, indicating a need for further improvements.

Potential Overfitting in Task 2

The presence of high accuracy alongside lower precision and recall raises concerns about the model potentially overfitting the training data, a common challenge in machine learning that requires further investigation and refinement.

Enhancing the Model's Generalizability

Ensemble Techniques

Implementing ensemble methods, such as bagging, boosting, or stacking, could enhance model performance by combining predictions from multiple base models, mitigating the effects of overfitting and improving overall predictive accuracy.

Exploring Other Pre-trained Models

Investigating other pre-trained convolutional neural network (CNN) architectures, such as ResNet, Inception, or EfficientNet, could uncover models better suited for glaucoma detection, potentially revealing superior features for extracting glaucomatous characteristics from fundus images.

Validation on External Datasets

Testing the model's performance on diverse datasets acquired from different populations, imaging devices, and clinical settings can provide insights into its real-world utility and robustness across varied scenarios, validating its generalizability.

Clinical Validation and Integration

Collaborating with ophthalmologists and healthcare professionals for clinical validation of the model's predictions is essential before integrating it into clinical workflows, ensuring its efficacy and safety as a diagnostic tool for glaucoma screening.

Collaborative Efforts for Glaucoma Detection

Mohit	Data Preprocessing	Model Building	Model Training (Task 1)	Model Evaluation
Sidhant	Model Building	Model Training (Task 1)	Model Evaluation	Report Writing
Aryan	Handling Image Formats	Model Training (Task 1)	Model Training (Task 2)	Model Evaluation

By leveraging the collective expertise and contributions of each team member, we successfully developed and evaluated multiple machine learning models, ultimately choosing the best-performing model for further refinement and deployment in the glaucoma detection challenge.