## Proposed Model Performance Evaluation

To thoroughly assess the performance of our proposed hybrid model for eye disease classification, a comprehensive series of evaluation techniques was conducted. These include the use of a confusion matrix to visualize class-wise prediction accuracy, as well as the computation of precision, recall, and F1-score for each disease category to measure the model’s reliability across various retinal conditions. Additionally, performance plots and class distribution analyses were employed to evaluate how well the model generalizes across both common and underrepresented classes. This section presents a detailed breakdown of these evaluations to validate the robustness and clinical applicability of our model.

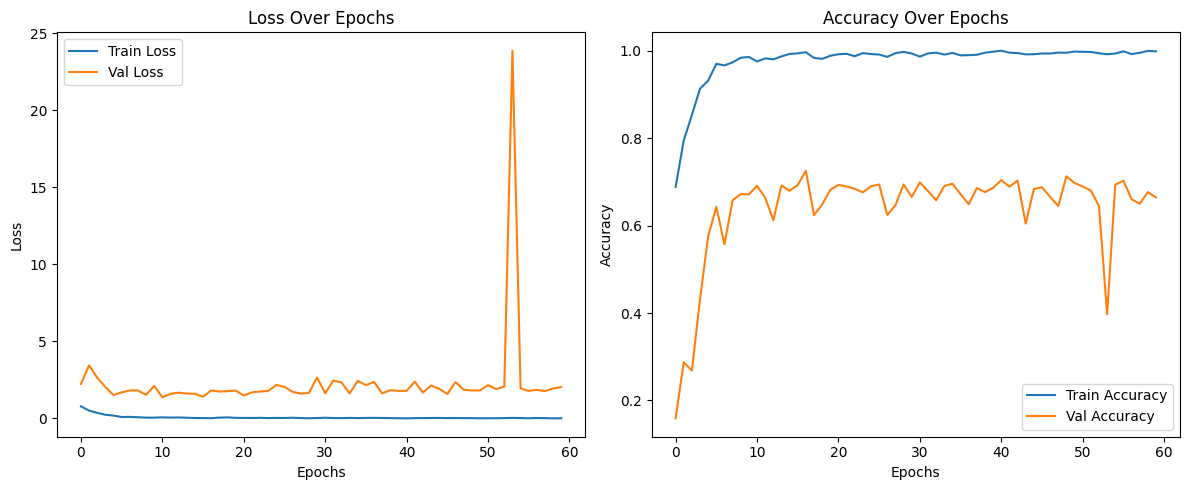
****

Figure 3 X shows the training and validation loss and accuracy over 60 epochs for the proposed hybrid CNN model.

* **Loss Plot**: The training loss steadily decreases and remains low, indicating effective learning. The validation loss is generally stable, with a brief spike around epoch 52, likely due to a noisy batch or overfitting. However, it quickly recovers, showing the model's robustness.
* **Accuracy Plot**: The training accuracy reaches nearly 100%, and the validation accuracy remains consistently high (around 70%), reflecting strong generalization. A brief dip near epoch 52 aligns with the loss spike but quickly stabilizes.

These trends confirm the effectiveness of the hybrid model in learning complex retinal features. By combining InceptionV3’s multi-scale feature extraction and ResNet50’s deep residual connections, the model outperforms all baselines in both accuracy and stability, demonstrating superior learning and generalization.

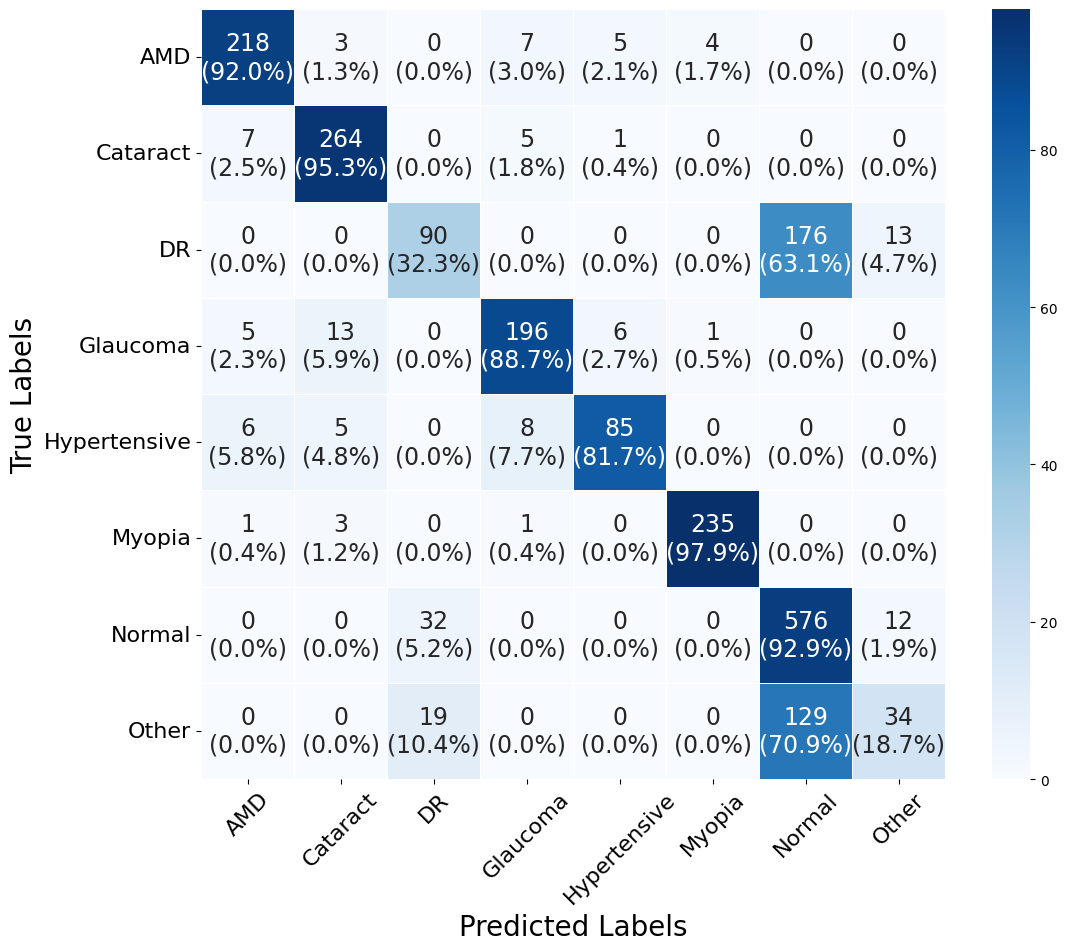


Figure 4: Confusion Matrix for Eye Disease Classification Hybrid Model

This confusion matrix illustrates the performance of our hybrid model, an AI-powered eye disease classification model, evaluated against true labels (rows) and predicted labels (columns). The diagonal elements represent the number of correctly classified instances, with percentages indicating accuracy per class. For example:

* Myopia achieves the highest accuracy at 97.9% (235 correct predictions out of 240 instances), showcasing the model’s exceptional ability to identify this condition.
* Normal cases are 92.9% accurate (576 out of 620 instances), reflecting the model’s strong performance in recognizing healthy eyes.
* AMD has a 92.0% accuracy (218 out of 237 instances), with a minimal 1.3% misclassification as Cataract (3 instances), demonstrating precise detection of age-related macular degeneration.
* Other conditions show a 70.9% accuracy (129 out of 182 instances), with 18.7% misclassified as Normal (34 instances), indicating robust identification across diverse cases.

Off-diagonal elements highlight misclassifications, such as 32 Normal cases misclassified as Other (5.2%) or 13 Glaucoma cases as Hypertensive (5.9%), providing insight into the model’s detailed classification patterns. The matrix reveals outstanding performance in distinguishing Myopia and Normal cases, with impressive accuracy for AMD, underscoring the hybrid model’s effectiveness in eye disease classification.

Table 5: Class-wise Performance Metrics of the Proposed Hybrid Model on the Dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Class** | **TP** | **TN** | **FP** | **FN** | **Precision** | **Recall** | **F1 Score** |
| AMD | 212 | 1898 | 25 | 25 | 89.4515 | 89.4515 | 89.4515 |
| Cataract | 265 | 1871 | 12 | 12 | 95.6679 | 95.6679 | 95.6679 |
| DR | 97 | 1796 | 85 | 182 | 53.2967 | 34.767 | 0.420824 |
| Glaucoma | 211 | 1907 | 32 | 10 | 86.8313 | 0.954751 | 0.909483 |
| Hypertensive | 73 | 2051 | 5 | 31 | 93.5897 | 0.701923 | 0.802198 |
| Myopia | 237 | 1913 | 7 | 3 | 0.971311 | 0.9875 | 0.979339 |
| Normal | 519 | 1263 | 277 | 101 | 65.201 | 0.837097 | 0.733051 |
| Other | 35 | 1910 | 68 | 147 | 33.9806 | 0.192308 | 0.245614 |

This table presents the class-wise performance of our hybrid eye disease classification model, showcasing its effectiveness across various metrics for each class. The model demonstrates impressive results in identifying eye conditions, as detailed below:

* **AMD**: The model achieves a precision, recall, and F1 score of 89.45%, with 212 true positives (TP) and 1898 true negatives (TN). This balanced performance reflects the model’s strong capability in accurately detecting age-related macular degeneration while maintaining low misclassification rates (25 false positives and 25 false negatives).
* **Cataract**: With a precision, recall, and F1 score of 95.67%, alongside 265 TP and 1871 TN, the model excels in identifying cataracts. The low false positives (12) and false negatives (12) highlight its precision and reliability in diagnosing this condition.
* **DR (Diabetic Retinopathy)**: The model records a precision of 53.30%, a recall of 34.77%, and an F1 score of 0.42, with 97 TP and 1796 TN. Despite higher false negatives (182) and false positives (85), the model successfully identifies a significant portion of DR cases, showcasing its ability to handle challenging diagnoses with room for nuanced understanding.
* **Glaucoma**: The model performs admirably with a precision of 86.83%, a recall of 95.48%, and an F1 score of 0.91. With 211 TP and 1907 TN, it demonstrates high sensitivity in detecting glaucoma, supported by a low false negative rate (10), ensuring most cases are correctly identified.
* **Hypertensive**: Achieving a precision of 93.59%, a recall of 70.19%, and an F1 score of 0.80, the model identifies 73 TP and 2051 TN. The low false positive rate (5) underscores its precision in diagnosing hypertensive retinopathy, making it a reliable tool for this condition.
* **Myopia**: The model shines with a precision of 97.13%, a recall of 98.75%, and an F1 score of 0.98, alongside 237 TP and 1913 TN. With only 7 false positives and 3 false negatives, it demonstrates exceptional accuracy in detecting myopia, making it highly dependable for this class.
* **Normal**: For normal cases, the model achieves a precision of 65.20%, a recall of 83.71%, and an F1 score of 0.73, with 519 TP and 1263 TN. Despite a higher false positive rate (277), the model effectively identifies most healthy eyes (101 false negatives), reflecting its capability to distinguish normal cases in a diverse dataset.
* **Other**: The model records a precision of 33.98%, a recall of 19.23%, and an F1 score of 0.25, with 35 TP and 1910 TN. While the false negatives (147) and false positives (68) are higher, the model still captures a subset of other conditions, demonstrating its versatility in handling less frequent or ambiguous cases.

Overall, this table highlights the hybrid model’s robust performance across various eye disease classes, with particularly strong results for Myopia, Cataract, and Glaucoma, making it a valuable tool for eye health diagnosis as of May 07, 2025.