International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

Detection of Retinitis Pigmentosa Using Hybrid Deep Learning Architecture

Kondabathula Meghana[1], K. Dhivya[2], Muntha Raju[3], Macha Balacharan[4], Shaik Arif[5]*
[1,4,5]UG Scholar, Dept. of CSE, Nalla Malla Reddy Engineering college, Hyderabad, Telangana, India
[2,3]Assistant Professor, Dept. of CSE, Nalla Malla Reddy Engineering college, Hyderabad, Telangana, India
Emails: *kmeghana293@gmail.com[1], divya.cse@nmrec.edu.in[2], raj.jntu9@gmail.com[3], machacharangoud@gmail.com[4], shaikarif04966@gmail.com[5]*

Abstract
Retinitis Pigmentosa (RP) is a hereditary vision disorder where the photosensitive cells of the retinal degenerate thus leading to blindness. RP victims commonly experience night vision issues, then constrict side eye vision, which may result in tunnelling eyesight and even total blindness. The early detection is necessary as it assists in the planning of treatment and prevents further development of the infection. This paper presents a new and improved method utilizing a Vision Transformer (ViT-B16), that analyses the whole picture of retina to more accurately discover patterns, along with EfficientNet-B4, a CNN algorithm that retrieves crucial information from pictures. It is recommended to use a multi-modal images training that supports Ultra-Wide Field (UWF), Fundus Autofluorescence (FAF), including colour fundus pictures. The rotation plus brightness modification procedure is carried out by data augmentation to expand the quantity of dataset in order to aid the algorithm in learning more efficiently. Gradient weighted class activation mapping(Grad-CAM) was applied to identify the regions of retinal surface which made the biggest contributions to algorithm's forecasting. Additionally, RP effectiveness was separated into Early, Moderate, and Severe levels with a stage estimation method utilizing Grad-CAM activation intensity. Through testing and validation, the architecture exhibits exceptionally high precision, F1-Score, and accuracy. It provides a scalable and effective method of detecting Retinitis Pigmentosa through significantly increased accuracy through the use of multi-modal image training. The previous RP identification availed by this method allows appropriate medication in good time that improves the overall quality of life of the affected individuals.

Keywords: *EfficientNet-B4, Grad-CAM, Hybrid deep learning architecture, Multi modal image training, Vision Transformer (ViT-B16), Retinitis Pigmentosa.*

1. Introduction
Retinitis Pigmentosa is a challenging disorder to diagnose properly because it is a hereditary disorder that destroys light-sensitive cells. It is a method that determines RP with the combination of artificial intelligence classifiers and temporal data that are taken with mfERG, unlike earlier studies that involved quantitative analysis of amplitudes and latencies based on multifocal electroretinograms [1]. Minor fundal changes complicate their detection within a short time. Deep learning models, including Xception, Inception V3, and Inception Resnet V2, have been examined in order to identify them at the very start [2]. RP Structural scanning is not always sufficient to predict the extent of sight. Studies on the capability of deep learning algorithms to predict vision loss in mixed optical scans such as OCT and laser fundus scans are underway [3]. The resemblance to other retinal diseases makes retinitis pigmentosa condition diagnosis difficult and troublesome to establish with ease and ease through fundus pictures. SE-RestNet system promotes the ability of RP to detect by removing irrelevant information and identifying important pathways [4]. The proposed RPS-Net is an effective division method which is adapted to identify dark areas in fundus images and directs the categorization phase to identify RP [5]. To evaluate combined sensitivity and effectiveness in the diagnosis, rigorous inspection and systematic review of AI results in RP identification in many different types of images are conducted to evaluate diversity, prejudices, and the strength of evidence in favour of AI RP diagnosis [6]. The outcomes of progressive blindness may be the result of numerous genetic

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

eye disorders such as retinitis pigmentosa, which is usually difficult to diagnose. A deep learning model with multiple inputs can distinguish between normal cell and genetic eye disease using IR and coloured fundus image as input [7]. To achieve the correct phasing needed to comprise the required scheduling, medication multifocal electroretinogram P1 amplitude maps were cross-linked with grayscale visual fields maps [8]. In studies and reports on how hereditary eye diseases such as Retinitis Pigmentosa and Stargardt are reviewed and analysed, studies without external validation generally showed worse outcomes. To attain the mass acceptance, real-world generalizability, transparent datasets, standardized reporting, and prospective validation are needed [9]. In combination with measures of visual field sensitivity, Ultra Wild Field captures a broad area of the retina. Instead of classification, a deep learning CNN model got trained to predict the picture attributes into continuous visual field sensitivity values to regress [10]. 34 male patients with X-linked RPGR-related RP were used in the dataset. Parameters of 3D OS, including OS thickness, EZ area, and OS volume, were recreated, and boundaries of EZ and RPE were established through all B-scans in the volume with segmentation algorithm (with manual correction) of deep learning [11]. The U-Net is used together with a sliding window CNN refinement network in the development of a deep learning model to segment retinal layer boundaries, including the ellipsoid zone and the RPE interfaces [12]. An interactive machine learning-driven systems biology system integrating interactome, proteomic and genomic data to map RP processes. The research provides RP with drug repurposing and therapeutic development customized [13]. The use of Fundus Autofluorescence images of RP is characterized by the training Convolutional Neural Network model on the basis of the macular functioning, which is able to identify regions of reduced and preserved macular sensitivity [14]. The diagnosis of the problem and prevention of severe visual impairment are related to timely detection. Deep Convolutional Neural Networks system identifies RP in retinal fundus images automatically and enhances the accuracy of diagnosis and allows extensive screening of RP [15]. The system suggests hybrid deep learning architecture to take one of the color Fundus, FAF, and Ultra-Wide Field images as input and combine EfficientNet-B4 and Vision Transformer (ViT-B16). The algorithm correctly labels images of the retina of the two types: clear and RP-affected and, in the latter case, it applies Grad-CAM visualization to highlight the areas of damage. The system defines the levels of illness as Early, Moderate or Severe depending on the activation intensity and spread.

2. Methodology

The algorithm technique of the hybrid deep learning framework applies several picture training to help in the most precise diagnosis of disease. Three scans- colour fundus, fundus autofluorescence (FAF) and ultra-wide field (UWF) scans images are trained on the architecture. By uploading a single picture of the aforementioned or trained pictures at one point, an individual is capable of determining the existence of the illness. The findings are shown in an intuitive web page that facilitates simple engagement and understanding. The hybrid architecture is the combination of CNN EfficientNet-B4 and Vision Transformer(ViT-B16) where EfficientNet-B4 extracts the features and spatial patterns such as optic disc shape, blood vessel structure, pigmentations, damaged areas of the retina and transformer then processes the extracted feature to learn global contextual relationships across the entire retina to identify how different retinal regions interact and changes due to RP. After prediction of RP stages are estimated using a stage estimation mechanism based on Grad-CAM activation intensity Fig.1).

2.1. Input Layer

The system accepts input of retinal pictures with Ultra-Wide Field (UWF), Color Fundus and Fundus Autofluorescence (FAF). The data of individual images helps the system in detecting core and outer eye defects which are correlated to Retinitis Pigmentosa (RP) by acquiring distinctive retinal features. It is guaranteed that the technology will help obtain the complete information about the eye through the adoption of several modalities, which will increase the overall accuracy of diagnostic findings.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

Figure 1 System Architecture

2.2. Data Preprocessing

Images are first taken through pre-model training processes augmentation techniques to increase productivity. These failures of a somewhat small dataset are solved by the usage of data enrichment methods like flipping and brightening. These changes reduce overfitting, augment the data set and improve the prediction capacity of the model through modelling the actual world changes in retinal images acquisition. The consequent outcome is the algorithm is more resilient and responsive to hidden information in the course of diagnosis.

2.3. Feature Extraction

It is a variant of CNN-based EfficientNet-B4 and Vision Transformer (VIT)-B16. ViT restores the global contextual associations in the image, but EfficientNet-B4 restores local and fine-grained retina features including pigmented patterns, vessels, and the regions of changes. This two-stage composite model has a strict local texture and global texture correlation learning to produce better accuracy and intelligible prediction of sickness.

EfficientNet-B4 Algorithm

- Convolution Operation

$$Yi, j, k = \sum M . \sum P$$

$$. \sum Q$$

$$Xi+p,j+q,m.Wp,q,m,k$$

$$m=1$$

$$p=1 \quad q=1$$

This creates feature maps Y by swiping learnable filters W over the input image X to compute local features.

Xi+p, j+q, m: Pixel value at position (i+p, j+q) in

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

channel m of the input image.
Wp, q, m, k: Weight of the convolutional filter at position(p, q), mapping input channel m to output channel. k.
Yi, j, k: Produce feature map for channel k at position (i,j).
M: The No. of input channels.
p, q: Convolutional kernel height and breadth.
i, j: The output feature map's spatial indices. k: The output channel index.

- Swish Activation

  f(x)= x. σ(x)=x/(1+e^{-X})
  x: Value entered into the activation function.
  $\sigma(x)$: Function that takes input from the range(0,1).
  f(x):The Swish activation's output.

Compared to ReLU, Swish enables greater feature learning and a smoother gradient flow.

- Compound Scaling

$d = \alpha\phi, w = \beta\phi, r = \gamma\phi$
d: Network depth (layer count).
w: Width (number of channels per layer).
r: Resolution (input image size).
α, β, γ: Depth, breadth, and resolution scaling coefficients.
ϕ: Compound coefficient governing total model size(EfficientNet-B4 use $\phi = 4$).
This scaling strikes a balance between performance and model complexity.
Vision Transformer(ViT-B16) Algorithm

- Patch Embedding

o xp∈ $RN\times (P\hat{}2\cdot C)$
xp: A matrix of picture patches that have been flattened.
$N = H\cdot W/P2$: Number of patches (for example, N = 196 for a 224×224 image and 16×16patches).
P: Patch size (with P = 16 for ViT-B16).
C: Channel count (usually 3 for RGB).
$P2\cdot C$: The quantity of values in each patch Every patch is handled sequentially, much like a token.
o $z0 = [x\text{class}; xp1\ E; xp2\ E; \ldots\ ; xpN\ E]$
+ Epos
xclass: A learnable categorization token that is appended to the sequence is called class. xpiE: The patch i is linearly projected using the embedding matrix E.
Epos: In order to maintain spatial order, positional encoding was implemented.
z0: Transformer's initial input sequence
By doing this, the image is ready for transformer processing.

- Multi-Head self attention

Attention(Q,K,V)= softmax ($QKT/\sqrt{dk}$) V
$Q = XWQ$: Query matrix based on weight WQ and input X.
K = XWK: Key matrix from input X and weight WK.
V = XWV: Value matrix from input X and weight WV.
dk: Key vector dimensions (for scaling). Softmax: Adjusts attention scores to normal. Attention (Q, K, V): A weighted sum of values determined by how close the keys and queries are. Detection of diffuse RP patterns across the retina is done by this mechanism.

- Feed Forward Network

FFN (x) = GELU ($xW1 + b1)W2 + b2$
x: Vector input token.
$W1$,$W2$: Learnable weight matrices.
$b1$, $b2$: Terms of bias.
GELU: Gaussian Error Linear Unit, or activation function.
Following attention, this improves the token representations.

- Classification Token Output

FViT = zLclass ∈ Rd.
zLclass: The classification token's final output following L transformer layers.
FViT: A global feature vector that is utilized in classification.
d: Dimension of the embedding of the token. The image's global context is summed up in this vector.

2.4. Fusion Layer

In order to generate a single representation, ViT and EfficientNet-B4 features are merged in the fusion step. This level allows the framework to enjoy the complementary capabilities of both the structures through combining local and global information. Striking the right balance between contextual information provided by transformers and especially the information structure provided

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

by CNNs leads to a better classification accuracy of the complex retinal images.

2.5. Classification Layer

The combination of the data is transmitted to one of the layers of neurons that are completely connected, which, in turn, generates the probability values of the
two classes, Healthy and Retinitis Pigmentosa. The probability scores are used to derive intelligible posterior likelihoods with the aid of a Softmax activation function. The approach further allocates a degree of certainty to every estimation, which assists in clinical understanding and demonstrates the trustworthiness of the decision made by the algorithm. The system uses Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize the parts of the retina that affect the choice of the algorithm, which increases interpretability and clinical credibility. The heatmap presents the area of damage in the RP cases visually. The extent and spread of these activations define the estimated sickness phase, which is computed using rule based approach as early, moderate or severe.

2.6. Output Layer

Lastly, the output layer provides the user in a web page with the predicted state of the disease, a confidence value, and the Grad-CAM representation with stage estimation. When there was the condition, the predicted phase is also shown in the algorithm so that physicians can make their judgment informed. The dashboard provides the ease of communication between the users and the trained model and helps to easily upload photos, graph, and interpret the results.

3.Discussions

This table compares the performance of three different image classification techniques, confirming that the Hybrid architecture with multimodal picture input is more effective than two well-known techniques. All measures showed consistently high scores for the suggested hybrid model: Accuracy (0.96), F1 Score (0.96), Precision (0.97), and Recall (0.94). The comparative models, on the other hand, showed several flaws. The ResNet-152, Grad CAM method had a significantly lower Recall (0.76), which suggested a high percentage of missed positive cases. Concurrently, RPS-Net showed less Precision (54.05), which considerably reduced its F1 Score (61.54) even if its Accuracy was reasonable that is 0.93 (Table 1).

Table 1 Comparison Table of the Systems

Methodologies used	Precision	Recall	F1 Score	Accuracy
Hybrid architecture with multimodal image training	**0.97**	**0.94**	**0.96**	**0.96**
ResNet-152, Grad CAM	0.89	0.76	-	0.85
RPS-Net	54.05	-	61.5	0.93

The above findings strongly and tightly indicate that the most dependable and clinically significant performance is obtained by integrating multimodal image training with a hybrid architecture, effectively balancing the critical trade-off between detecting actual positive cases and reducing false alarms.

4. Evaluation Metrics

The model's remarkable capacity to differentiate between cases of Retinitis Pigmentosa and healthy ones is demonstrated in Fig.2 below. It received excellent F1-scores of 0.94 (RP) and 0.97 (Healthy). All healthy samples were successfully identified by the model, which demonstrated perfect recall (1.00) and good precision (0.94) for the Healthy class (Support: 586). It had perfect precision (1.00) but a slightly poorer recall (0.89) for the RP class (Support: 319), missing roughly 11% of real RP situations. Overall, this model performs well in categorization. A deep examination of the classification model's performance on 905 test samples is given by this confusion matrix in Fig. 3. The program accurately detected 233 cases of Retinitis Pigmentosa (True Positives) and 636 healthy people (True Negatives). Importantly, the model showed flawless precision for the Retinitis Pigmentosa class, recording zero False Positives (FP=0), which means it never incorrectly identified a healthy

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

individual as having the condition. The key error type where the model's performance might be enhanced is represented by the 36 False Negatives (FN=36) in the matrix, which are actual cases of Retinitis Pigmentosa that the model misclassified as healthy.

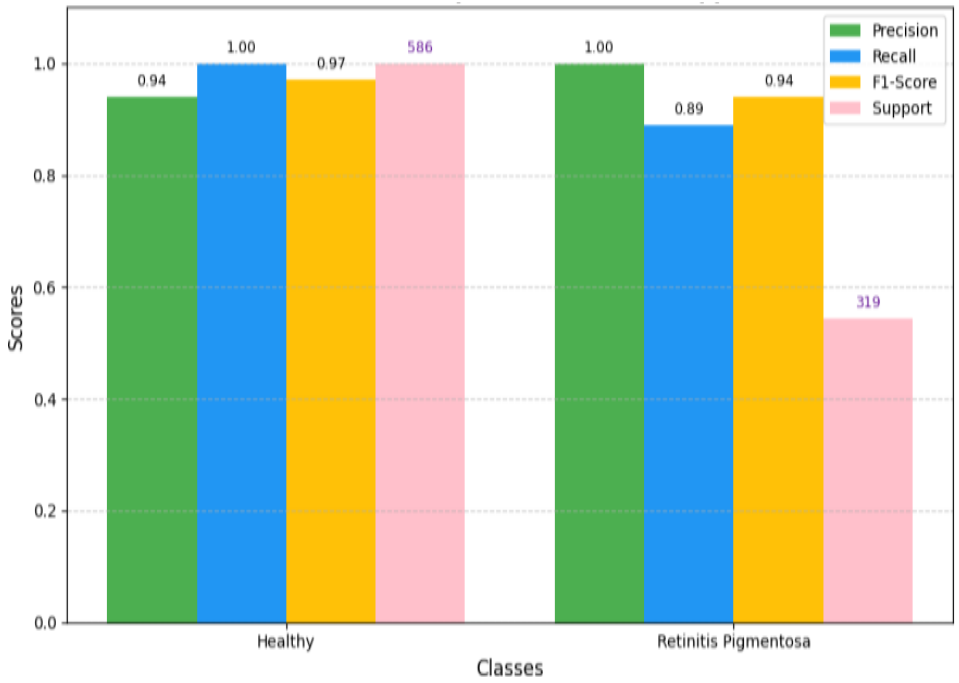

Figure 2 Performance Metrics of the System

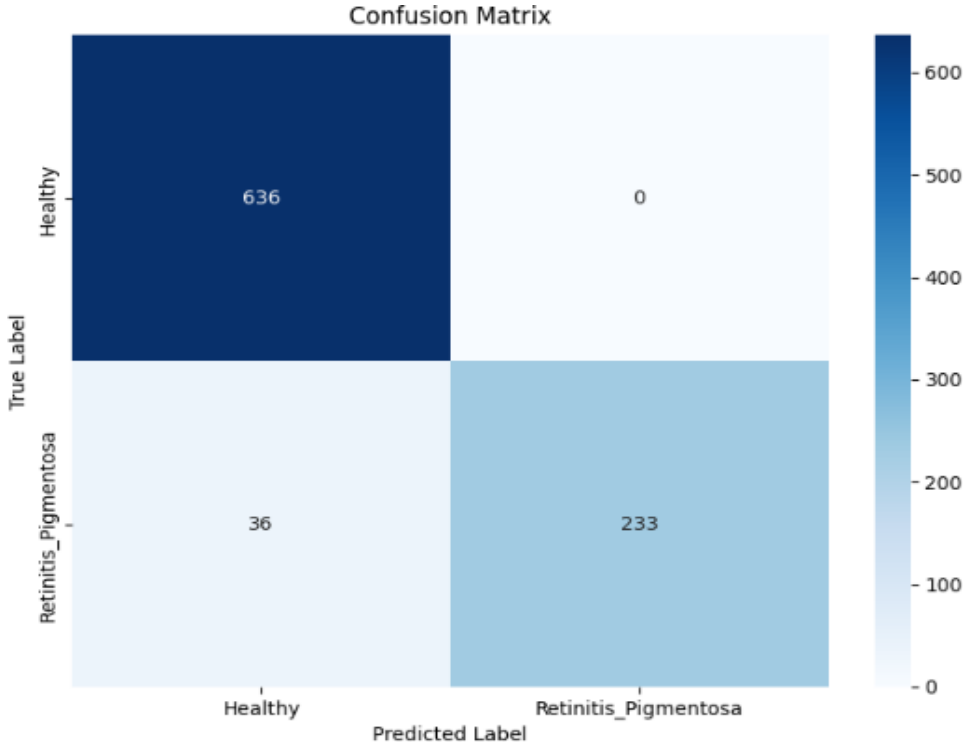

Figure 3 Confusion Matrix

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

From Table.2 clearly, the results indicate that the precision is very good in both the cases, recall is excellent in healthy case but lightly low in RP class, F1-score is also excellent in both cases, and accuracy is also very high. By seeing these metrics, we can say that the model provides a very promising and generalizable system.

1. Results

The below Fig.4 represents the training loss decreased from 0.12 to almost 0.0 for 15 epochs. The final period shows a steady increase in training accuracy approaching 0.99 and 1.0. A little over fitting is shown by the curve.

Table 2 Classification Report of the System

Class	Precision	Recall	F1-Score	Accuracy
Healthy	0.94	1.0	0.97	0.96
Retinitis Pigmentos a	1.0	0.89	0.94	0.96

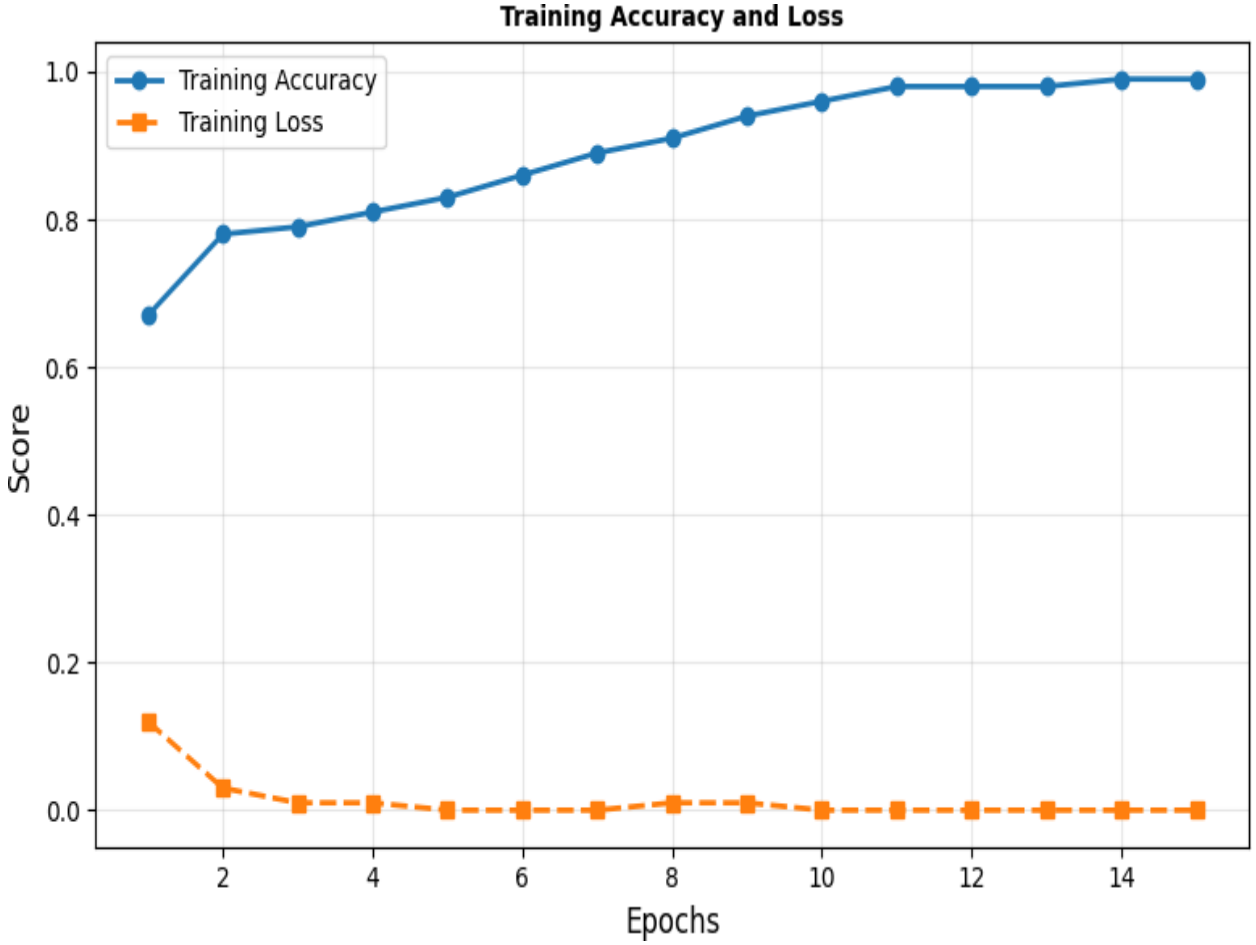

Figure 4 Training Metrics of the System

The below Fig.5 represents the testing loss decreased smoothly from 0.1 to 0.0. Excellent generalization was demonstrated by the test accuracy which is increased from 0.5 to 1.0. Across unseen data the model performed well with slight variation. The Fig.6 describes frontend displays a simple, user- friendly interface called “Retinitis Pigmentosa Detection”. User can upload one image(.png/.jpg) in colour fundus, FAF, UWF at a time. The “Analyse Image” button, which starts the model evalution is preminently displayed in the responsive design.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

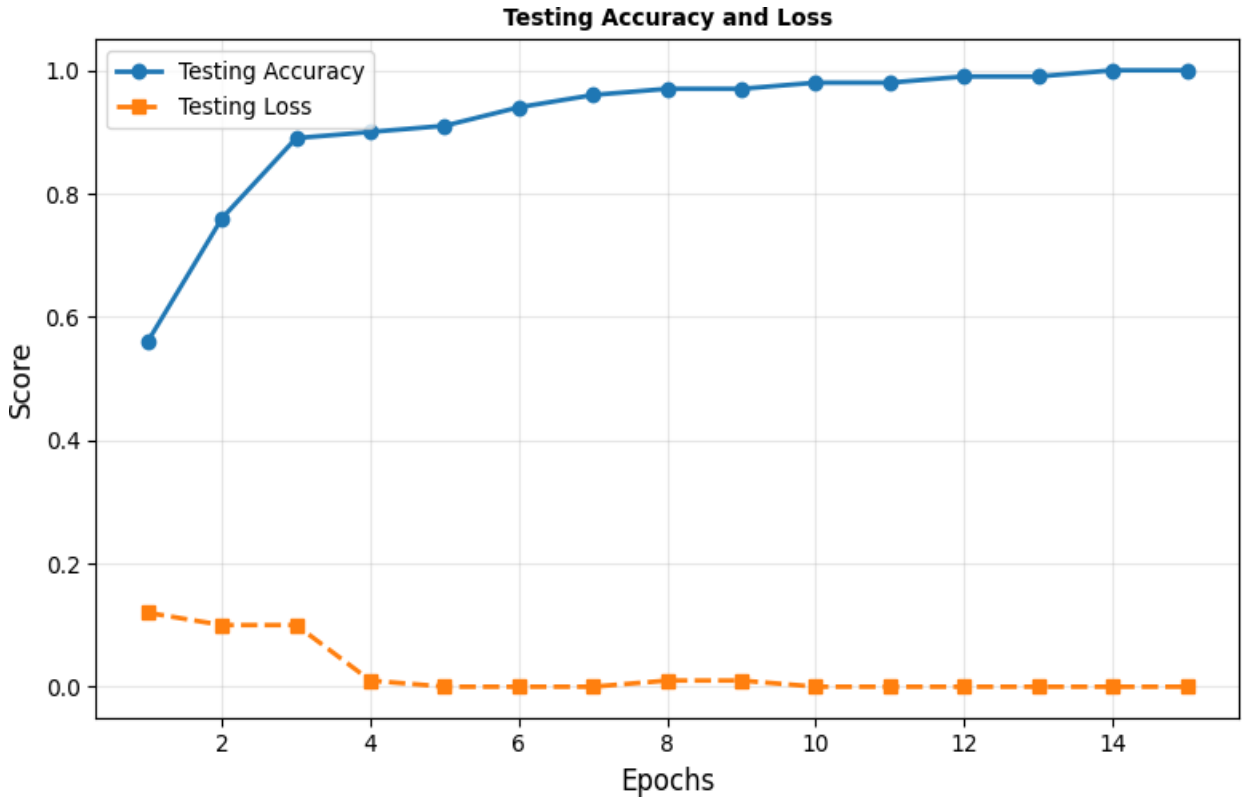

Figure 5 Testing Metrics of the System

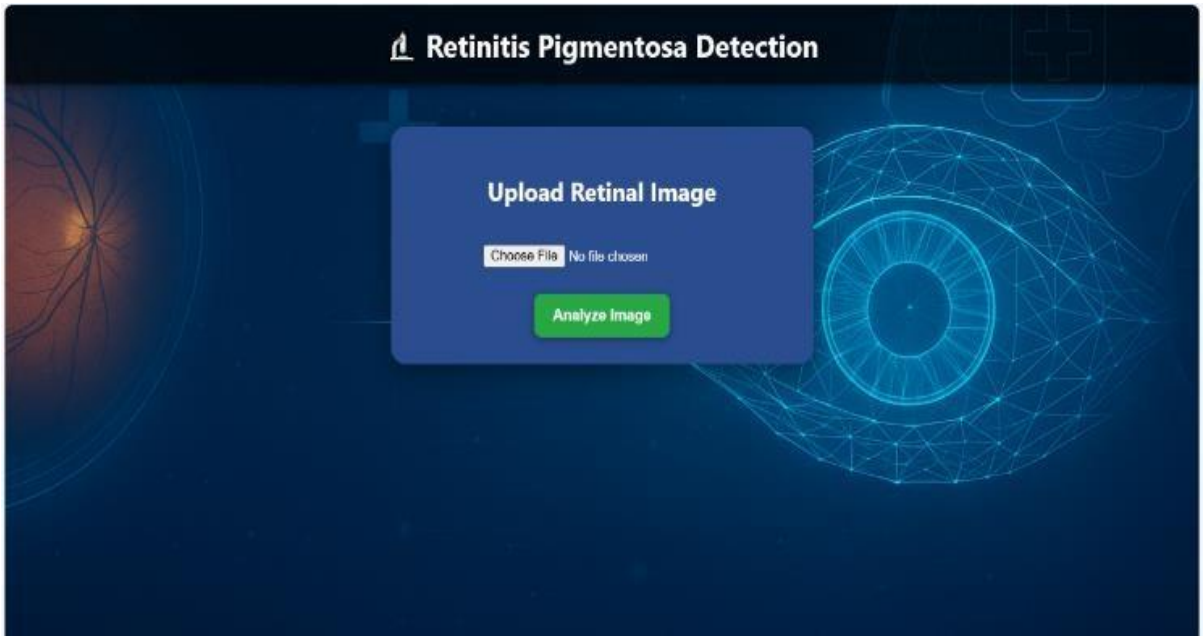

Figure 6 Input Data Acquisition

Fig.7: shows the Retinitis Pigmentosa Results when user submits the image. It displays the prediction (for example: Healthy/Retinitis Pigmentosa), confidence score (for example: 81.34%), stage also called as risk level along with the Grad-CAM heatmap highlighting the disease affected area.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 594-603
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0081

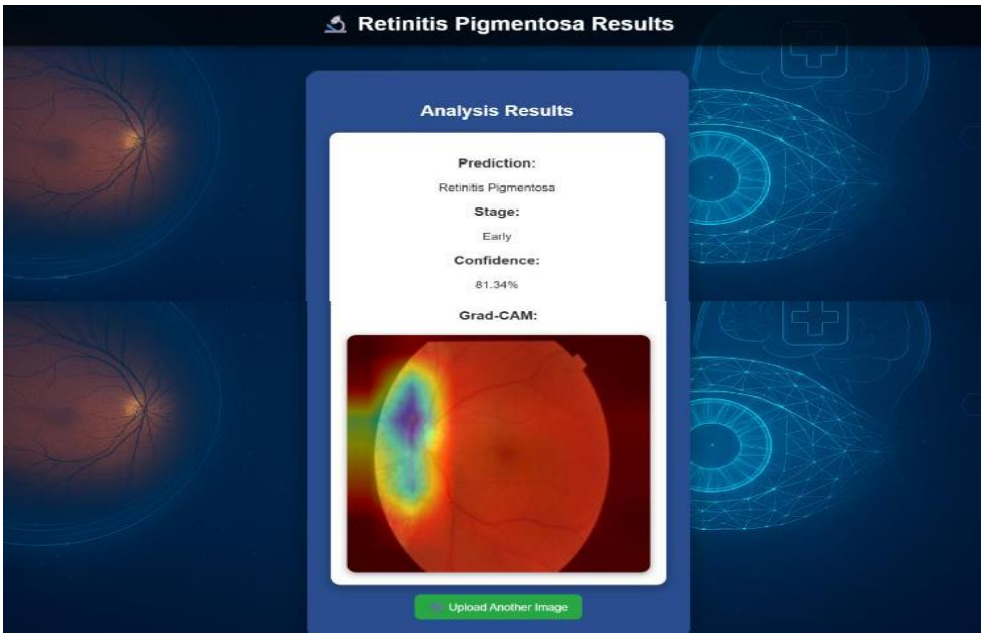

Figure 7 Result

Conclusion

The proposed hybrid deep learning model, which is a combination of EfficientNet-B4 and Vision Transformer (ViT) provides a precise and clear structure that automatically identifies and categorizes the Healthy and Retinitis Pigmentosa through multimodal retinal images, including Color Fundus, Fundus Autofluorescence (FAF), and Ultra-Wide Field (UWF) scans. Using ViTs context awareness of all images and EfficientNets local feature extraction, the system manages to classify all images into either healthy or unhealthy. Also, RP discovering demonstrates the level of confidence of the model, determines the stage of the disease (early, moderate or severe), and shows the affected areas of the retina with the help of a Grad-CAM visualization. All in all, this intelligent and intuitive AI system enhances the quality of clinical decision-making by helping ophthalmologists diagnose Retinitis Pigmentosa at an early stage and properly monitor it.

Future Scope

The project's future scope includes language support, real-time picture support to facilitate patient and physician use, integration of advanced algorithms and staged datasets to improve stage identification accuracy.

References

[1]. Bayram Karaman, Ayşe Öner, Ayşegül Güven, “Early Detection and Staging of Retinitis Pigmentosa Using Multifocal Electroretinogram Parameters and Machine Learning Algorithms, Physical and Engineering Sciences in Medicine,”2025. https://doi.org/10.1007/s13246-02501577-3.

[2]. Ta-Ching Chen, Wee Shin Lim, Victoria Wang, et al., “Artificial Intelligence–Assisted Early Detection of Retinitis Pigmentosa the Most Common Inherited Retinal Degeneration,” Journal of Digital Imaging, 2021. Springer: https://link.springer.com/article/10.1007/s10278-021-00479-6.

[3]. T. Y. A. Liu, C. Ling, L. Hahn, C. K. Jones, C. J. Boon, M. S. Singh, et al., “Prediction of Visual Impairment in Retinitis Pigmentosa Using Deep Learning and Fundus Images,” British Journal of Ophthalmology, 2022. PMC: https://www.ncbi.nlm.nih.gov/pmc/articles/PM C10579177/

[4]. Rubina Rashid; Waqar Aslam; Arif Mehmood, et al., "A Detectability Analysis of Retinitis Pigmetosa Using Novel SE-ResNet Based Deep Learning Model and

Color Fundus Images", 2024. IEEE: https://ieeexplore.ieee.org/abstract/document/10 440277

[5]. M. Arsalan, N. R. Baek, M. Owais, T. Mahmood, K. R. Park, "Deep Learning-Based Detection of Pigment Signs for Analysis and Diagnosis of Retinitis Pigmentosa," Sensors (MDPI), 2020. MDPI: https://www.mdpi.com/1424-8220/20/12/3454

[6]. A. M. Musleh, et al., "Diagnostic Accuracy of Artificial Intelligence in Detecting Retinitis Pigmentosa: A Systematic Review and Meta- Analysis," Survey of Ophthalmology, 2024. ScienceDirect: https://www.sciencedirect.com/science/article/a bs/pii/S0039625723001674

[7]. F. Jafarbeglou, H. Ahmadieh, F. Soleimani, et al., "A Deep Learning Model for Diagnosis of Inherited Retinal Diseases," Scientific Reports, 2025. Nature:https://www.nature.com/articles/s41598-025- 04648-3

[8]. Bayram Karaman, Ayşegül Güven, Ayşe Öner, Neslihan S. Kahraman, "Classification of Retinitis Pigmentosa Stages Based on Machine Learning by Fusion of Image Features of VF and MfERG Maps," Electronics (MDPI), 2025. MDPI: https:// doi.org/ 10.3390/ electronics14091867

[9]. S. Ashrafi, et al., "Diagnostic Accuracy of AI Models in Detecting Different Inherited Retinal Diseases: A Systematic Review and Meta- Analysis," Translational Vision Science & Technology (TVST), 2025. ARVO: https://tvst.arvojournals.org/article.aspx?articlei d=2810445

[10]. Daisuke Nagasato, Takahiro Sogawa, Mao Tanabe, et al., "Estimation of Visual Function Using Deep Learning From Ultra-Widefield Fundus Images of Eyes With Retinitis Pigmentosa," JAMA Ophthalmology, 2023. JAMA: https://jamanetwork.com/journals/jamaophthal mology/fullarticle/2801468

[11]. Y.-Z. Wang, W. Wu, D. G. Birch, et al., "Deep Learning–Facilitated Study of the Rate of Change in Photoreceptor Outer Segment Metrics in RPGR-Related X-Linked Retinitis Pigmentosa," Investigative Ophthalmology & Visual Science (IOVS), 2023. PMC: https://www.ncbi.nlm.nih.gov/pmc/articles/P MC10668621/

[12]. Y.-Z. Wang, et al., "A Hybrid Model Composed of Two Convolutional Neural Networks (CNNs) for Automatic Retinal Layer Segmentation of OCT Images in Retinitis Pigmentosa," Translational Vision Science & Technology (TVST), 2021. PMC: https://www.ncbi.nlm.nih.gov/pmc/articles/P MC8590180/

[13]. M. Esteban-Medina, et al., "The Mechanistic Functional Landscape of Retinitis Pigmentosa: A Machine Learning-Driven Approach to Therapeutic Target Discovery," Journal of Translational Medicine, 2024. BMC: https://translationalmedicine.biomedcentral.com/articles/10.1186/ s12967-024-04911-7

[14]. Taro Kominami, Shinji Ueno, Junya Ota, et al., "Classification of Fundus Autofluorescence Images Based on Macular Function in Retinitis Pigmentosa Using Convolutional Neural Networks," Japanese Journal of Ophthalmology, 2020. Springer: https://link.springer.com/article/10.1007/s10384-020-00752-1

[15]. S. San Lwin, et al., "Utilizing AI to Detect Retinitis Pigmentosa in Fundus Images," Investigative Ophthalmology & Visual Science (IOVS), 2025. ARVO: https://iovs.arvojournals.org/article.aspx?articl eid=2804664