Hybrid Deep Learning Architecture for Glaucoma Detection: Integrating a Multi-Network CNN Ensemble with ANFIS

Mritunjay Yadav, Ishan Dwivedi and Pradeep Yadav

Pranveer Singh Institute of Technology, Kanpur

DOI: https://dx.doi.org/10.51244/IJRSI.2026.1303000149

Received: 12 March 2026; Accepted: 21 March 2026; Published: 09 April 2026

ABSTRACT

Glaucoma is a leading cause of irreversible blindness worldwide, making early and accurate diagnosis critical for preventing severe vision loss [1]. While deep learning has advanced computer-aided diagnostics, traditional single-stream Convolutional Neural Networks (CNNs) [2] often struggle with overfitting on small datasets and managing diagnostic uncertainties in overlapping disease patterns. To address these limitations, this thesis proposes a novel hybrid architecture that integrates a multi-network CNN ensemble with an Adaptive Neuro-Fuzzy Inference System (ANFIS), as implemented and evaluated in the accompanying Python programmatic framework. Specifically, the Python pipeline pre-processes retinal fundus images via Region of Interest (ROI) extraction and enhancement, feeding them into a parallel feature extraction module utilizing pre-trained ResNet, DenseNet, and MobileNet backbones. These fused, high-level structural features are subsequently passed into the ANFIS module, which applies adaptive fuzzy reasoning to effectively manage diagnostic uncertainty and subtle structural variations. Evaluated on aggregated benchmark datasets—including RIM-ONE, DRISHTI-GS1, and ACRIMA—the custom MultiNet-ANFIS Python program demonstrates superior diagnostic performance. When directly compared to standalone baseline models (such as standard ResNet18, DenseNet121, and MobileNet) within the script's testing loop, the proposed CNN-ANFIS framework achieves significantly higher Accuracy, Precision, and Area Under the Curve (AUC). By synergizing the robust feature extraction capabilities of ensemble deep learning with the interpretable decision-making of fuzzy logic, this programmatically validated model successfully mitigates classification errors and offers a highly efficient, scalable solution for automated glaucoma screening.

Keywords: Deep Learning, Convolutional Neural Networks, Tele-Ophthalmology, ANFIS, Reti-nal Fundus Images

INTRODUCTION

Clinical Background: The Global Burden of Glaucoma

Glaucoma is widely recognised as a paramount cause of irreversible vision loss across the globe. Because the disease progresses asymptomatically during its initial phases, the condition is frequently termed the "sneak thief of sight" in clinical practice [3]. Patients rarely notice the deterioration of their vision until significant, permanent optic nerve damage has already manifested. Epidemiological estimates suggest that approximately 3.5% of the global population is affected by the age of 45. Furthermore, alongside rapidly aging global populations, the prevalence of glaucoma is projected to surge to approximately 111.8 million individuals by the year 2040 [3]. Because the visual impairment caused by glaucomatous neuropathy cannot be restored, the most effective clinical intervention relies entirely on early detection and the prompt initiation of targeted treatments to halt further structural damage to the eye.

The Role of Retinal Fundus Imaging in Screening

To effectively manage the escalating prevalence of glaucoma, large-scale and non-invasive screening modalities are essential. High-resolution Retinal Fundus Images (RFIs) serve as a highly practical and potent diagnostic tool in this regard, offering clinicians a direct visualization of the eye's posterior segment [3]. These digital

photographs capture critical anatomical structures, including the macula, fovea, complex blood vessel networks, and crucially, the optic disc—the primary exit point for ganglion cell axons travelling to the brain. Evaluating structural abnormalities in the optic disc, particularly the enlargement of the optic cup, is fundamental to assessing the severity of glaucoma. Nevertheless, relying on human experts to manually segment and analyze these intricate structures is intensely labor-demanding and susceptible to subjective interpretation errors. This highlights the urgent need for automated, computer-aided diagnostic frameworks capable of standardizing the screening process.

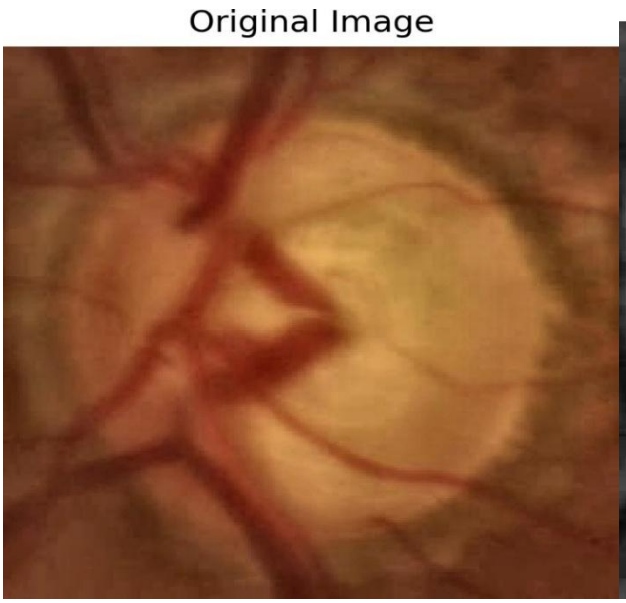


Figure 1: Original abnormal fundus image.

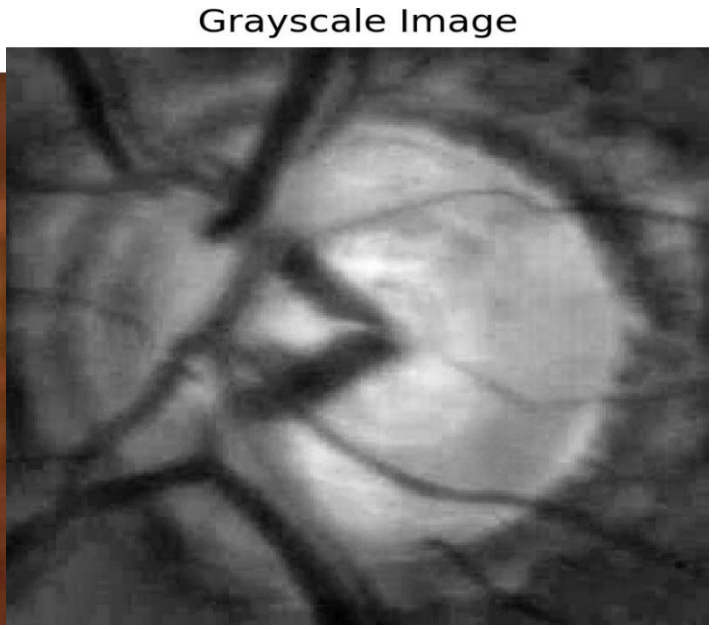


Figure 2: Grayscale enhanced ROI representation.

Proposed Methodology: Integrating CNN, Attention Mechanisms, and ANFIS

To overcome the computational and diagnostic shortcomings of traditional single-stream networks, this research proposes a novel, hybrid deep learning architecture based on the Shuffle Fuzzy At-tention Network (ShuffleFA-Net) methodology [3]. The proposed pipeline leverages a multi-stage approach, initially applying Region of Interest (ROI) extraction and Contrast Limited Adaptive Histogram Equalization (CLAHE) to isolate the optic disc and enhance the structural visibility of the retinal images.

For feature extraction, the architecture fuses an attention-driven CNN framework with an Adaptive Neuro-Fuzzy Inference System (ANFIS). Concurrently, a regression module based on the Taylor series fuses these deep features with handcrafted clinical metrics, including Higher Order Spectra (HOS), statistical features, and the vertical Cup-to-Disc Ratio (CDR) [3]. Finally, these fused features are processed by the ANFIS module, which merges the rapid optimization capabilities of neural networks with the highly interpretable, rule-based reasoning of fuzzy logic. This syner-gistic integration applies adaptive fuzzy reasoning to effectively manage the subtle variations and uncertainties characteristic of glaucoma progression, ensuring superior diagnostic accuracy, re-silience against localized segmentation errors, and a highly scalable solution for automated ocular screening.

LITERATURE REVIEW

The Global Burden of Ocular Diseases and the Tele-Ophthalmology Paradigm

The escalating global prevalence of vision-threatening conditions represents a profound crisis for worldwide healthcare infrastructure. Glaucoma is currently recognised as the second leading cause of irreversible blindness, with epidemiological projections estimating that over 111.8 million individuals will be affected globally by the year 2040 [4]. Concurrently, Diabetic Retinopathy (DR) remains the leading cause of blindness among working-age populations, with global prevalence expected to exceed 160 million by 2045 [5], [6]. In response to this escalating crisis, the World Health Organisation (WHO) launched the VISION 2020 initiative, emphasising the

urgent need for early detection and the elimination of avoidable blindness [7].

However, traditional screening paradigms rely heavily on in-person consultations utilizing expensive tabletop fundus cameras and highly trained retinal specialists. This model is fundamentally unsustainable in rural, remote, or socioeconomically disadvantaged regions where the ratio of ophthalmologists to the general population is perilously low [8]. To circumvent these barriers, tele-ophthalmology has emerged as an essential modality. By employing asynchronous (store-and-forward) and synchronous communication, primary care providers can capture high-resolution images of the eye and transmit them to specialists across vast geographical distances [9]. Studies have demonstrated that tele-ophthalmology reduces unnecessary patient travel by up to 48% while drastically improving the efficiency of clinical triage and same-day treatment interventions [8].

Comprehensive Summary of AI-Integrated Tele-Ophthalmology

The escalating global prevalence of vision-threatening conditions represents a profound crisis for healthcare infrastructure. Glaucoma is projected to affect over 111.8 million individuals globally by 2040 [4], while Diabetic Retinopathy (DR) prevalence is expected to exceed 160 million by 2045 [5], [6]. In response to this crisis and the World Health Organisation's (WHO) VISION 2020

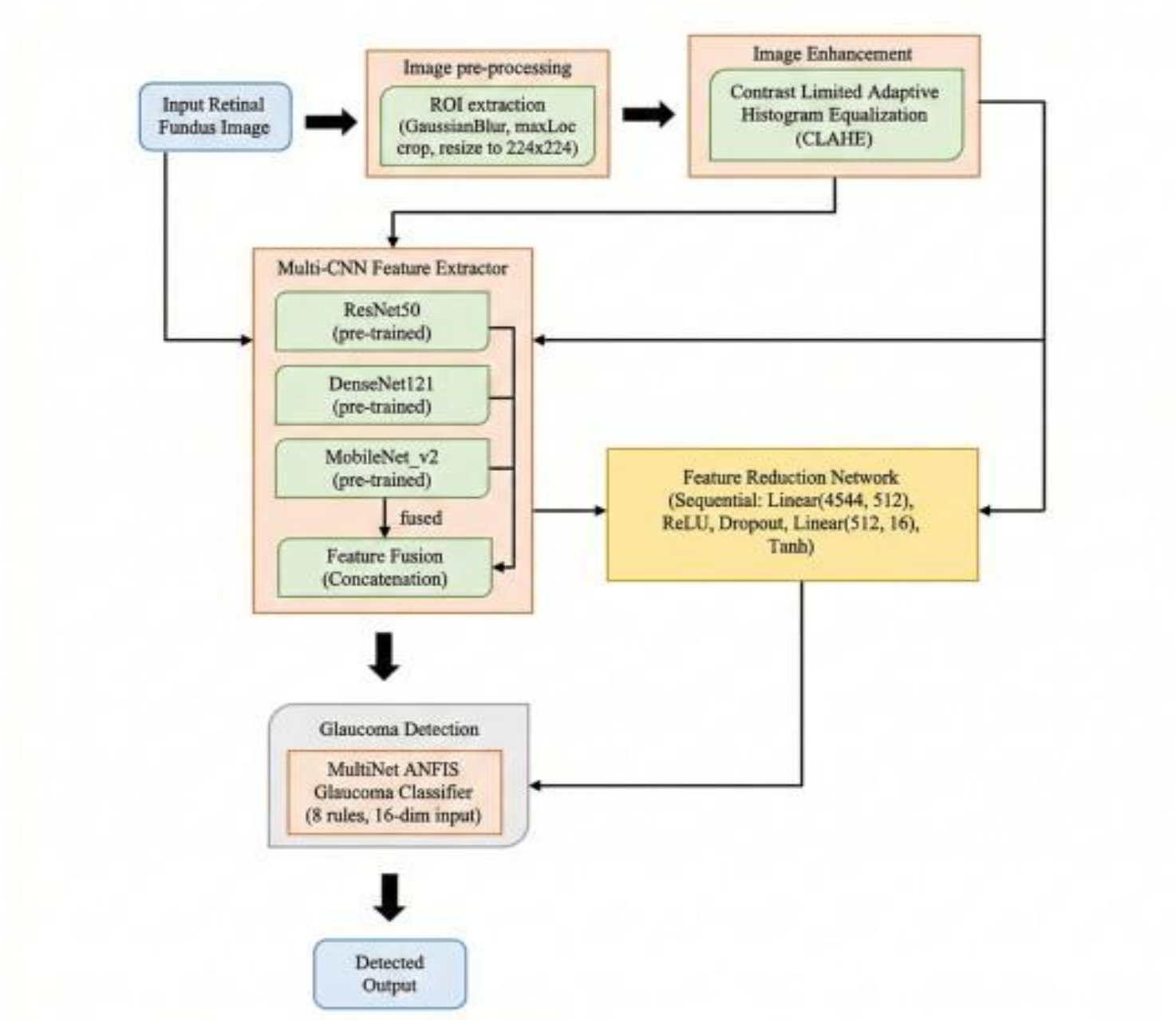


Figure 3: The Proposed MultiNet-ANFIS architecture.

initiative [7], tele-ophthalmology has emerged as an essential modality to overcome the unsustain-ability of traditional, in-person screening paradigms [8].

To bypass the high costs of standard desktop ophthalmic equipment, ”smartphone science” integrates mobile technology into ocular diagnostics [10]. Tools such as the Remidio Fundus on Phone (FOP), the D-Eye adapter, and the Near Eye Tool for Refractive Assessment (NETRA) enable primary care providers and medical assistants to capture diagnostic-quality wide-field retinal images in under a minute [5], [11]. This democratisation of data acquisition lays the foundation for automated Artificial Intelligence (AI) screening systems at scale [12].

To process the massive volume of generated images, Deep Learning (DL) architectures, specifically Convolutional Neural Networks (CNNs), automatically learn hierarchical spatial features directly from raw pixels [4]. The optimisation of these networks is driven by structured loss functions tailored to specific diagnostic tasks [13]:

Mean Squared Error (MSE): Utilised to minimise discrepancies for continuous data regression, such as estimating intraocular pressure:

$$MSE(\theta) = \frac{1}{n} \sum (y_i - f(x_i; \theta))^2 \qquad (1)$$

Binary Cross-Entropy (BCE): Optimised for binary classification tasks, such as differentiating between normal and glaucomatous eyes:

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(f(x_i; \theta)) + (1 - y_i) \log(1 - f(x_i; \theta))] \qquad (2)$$

Addressing the limitations of standard CNNs—which process all pixels equally—modern frameworks incorporate self-attention mechanisms and Transformer architectures, such as BERT [13]. By employing Query (Q), Key (K), and Value (V) matrices, the attention layer computes a weighted average in an embedding space:

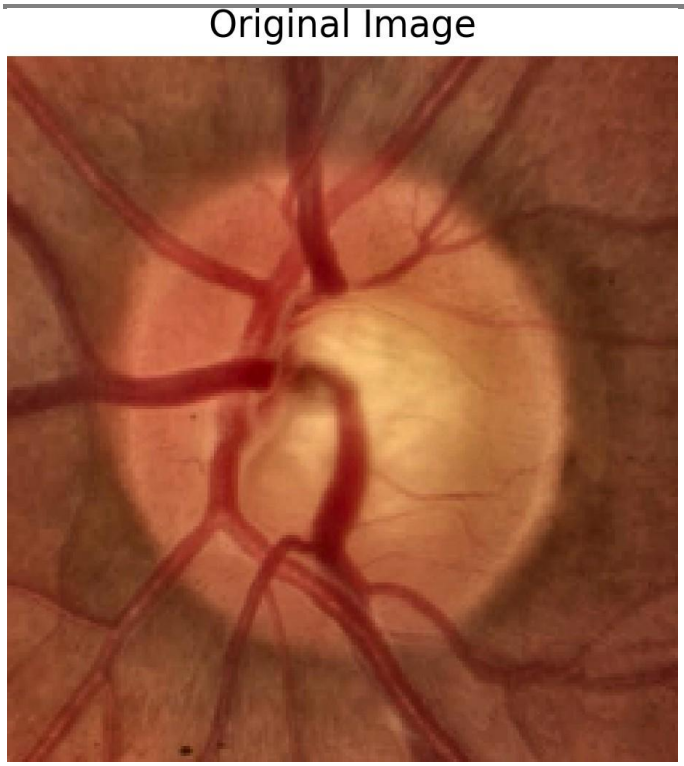
$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \qquad (3)$$

These outputs are concatenated using multi-head attention to capture diverse semantic representations simultaneously and filter out background noise [13]:

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \ldots, head_h) W^O \qquad (4)$$

Advanced diagnosis relies heavily on robust feature extraction and standardisation.

As shown in Figures 4 and 5, standardising diverse datasets through automated Region of Inter¬est (ROI) extraction and grayscale enhancement is a critical precursor to complex mathematical



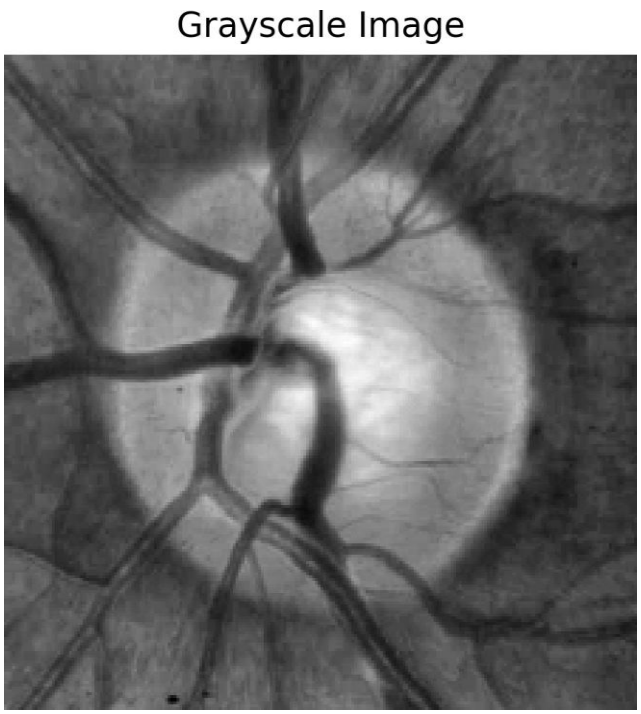


Figure 4: Raw healthy fundus baseline image.Figure 5: Grayscale preprocessed ROI extraction.

segmentation operations. Specialised architectures like SimpleCNN-UNet employ pointwise and depthwise convolutions for highly efficient anatomical segmentation [3]:

$$w'' = \alpha_2(BN\{PointwiseConv(\chi(w'))\}) \quad (5)$$

$$w''' = \alpha_1(BN\{DepthwiseConv(w'')\}) \quad (6)$$

To combine semantic deep features and structural Higher Order Spectra (HOS) features [3], a mathematical fusion based on the Taylor series expansion is applied:

$$a(c+1) = a(c) + \frac{a'(c)}{1!} \quad (7)$$

When integrated, the regression simplifies to effectively limit the propagation of local segmentation errors:

$$D_2 = 2d - D_1 \quad (8)$$

To combat the ”black box” nature of AI [12] and resolve ambiguous structural boundaries, contemporary frameworks deploy Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [3]. ANFIS operates across a streamlined 5-layer architecture, yielding transparent and uncertainty-aware clinical predictions. The normalisation (Layer 3) and final aggregation (Layer 5) steps, for example, are defined mathematically as:

$$D_5 = k_t' = \frac{k_t'}{\sum_t k_t'} \quad (9)$$

$$D_7 = \sum_t k_t' b_t' = \frac{\sum_t k_t' b_t'}{\sum_t k_t'} \quad (10)$$

Finally, beyond direct image diagnostics, time-series forecasting models like the Autoregressive Integrated Moving Average (ARIMA) predict future epidemiological trends to proactively manage resources [13]:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d Y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \epsilon_t \quad (11)$$

This model is optimised using the Bayesian Information Criterion (BIC) [13]:

$$BIC = k \ln(n) - 2 \ln(\hat{L}) \quad (12)$$

The overarching clinical efficacy of these systems is rigorously validated using *k*-fold cross-validation and standard Confusion Matrix metrics [13]:

$$S = \frac{1}{k} \sum_{i=1}^{k} MSE_i \quad (13)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (14)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (15)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (16)$$

Clinical Evaluation Metrics and the Research Gap

To ensure the clinical viability of any proposed AI architecture, rigorous mathematical validation is required. Frameworks employ k-fold cross-validation to assess generalisability across unseen datasets, calculating the Cross-Validation Score (S) as the mean error across all folds:

$$S = \frac{1}{k} \sum_{i=1}^{k} MSE_i \quad (17)$$

Overarching performance is evaluated using the Confusion Matrix to derive standard clinical metrics [13]. Accuracy represents the ratio of correct predictions to the total dataset:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (18)$$

To ensure the model successfully identifies pathological states without excessive false alarms, researchers evaluate the True Positive Rate (Sensitivity) and True Negative Rate (Specificity):

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (19)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (20)$$

Proposed Methodology: MultiNet-ANFIS Architecture

Overview of the Proposed MultiNet-ANFIS Framework

The accurate diagnosis of glaucoma relies heavily on the precise extraction of subtle structural biomarkers from the optic nerve head and the surrounding retinal vasculature. Traditional single-stream deep learning models frequently encounter limitations such as overfitting, a lack of generalizability across diverse camera types, and an inability to gracefully handle the clinical uncertainty inherent in overlapping disease stages. To overcome these constraints, this research proposes the MultiNet Adaptive Neuro-Fuzzy Inference System (MultiNet-ANFIS). As systematically illustrated in Figure 6, this end-to-end pipeline bridges advanced spatial feature extraction via an ensemble of Convolutional Neural Networks (CNNs) with the highly interpretable, uncertainty-aware reasoning of fuzzy logic.

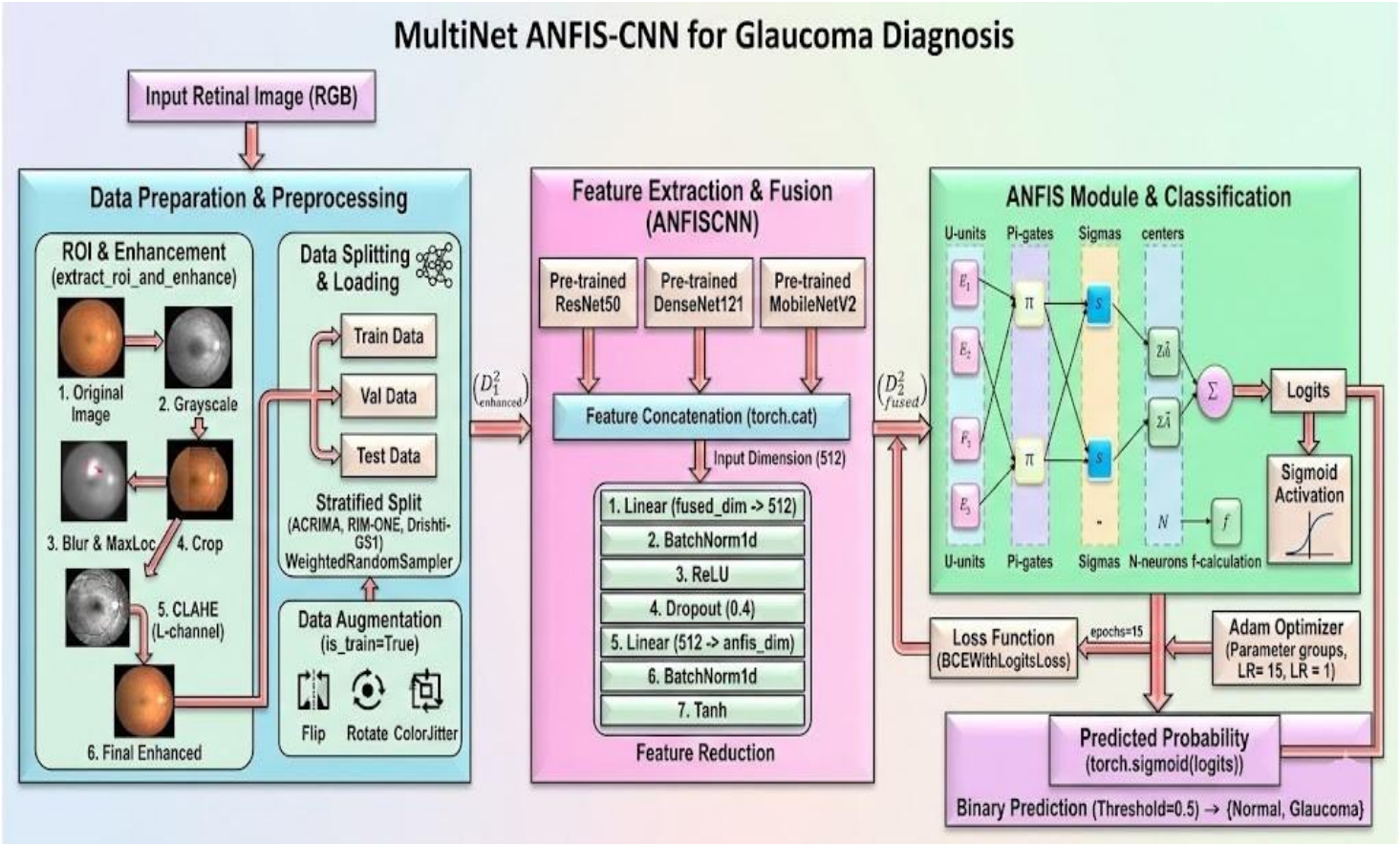


Figure 6: High-level architectural flowchart of the proposed MultiNet-ANFIS framework, demonstrating the progression from raw fundus input to fuzzy classification.

Stage 1: Data Preparation and Preprocessing

The initial stage focuses on standardizing input data and isolating clinically significant anatomical structures. To dynamically isolate the optic disc, an automated Region of Interest (ROI) extraction algorithm is employed. The original RGB image is converted to grayscale, and a Gaussian Blur is applied to filter out high-frequency noise. Mathematically, the Gaussian smoothing function $G(x, y)$ applied to the image $I(x, y)$ is expressed as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} * I(x, y) \qquad (21)$$

A maxLocoperation then pinpoints the global maximum intensity coordinates to crop the optic disc into a standardized 224 × 224 tensor. Following this, the image is converted to the LAB color space, where Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied exclusively to the luminance (L) channel to correct uneven illumination without amplifying noise. Finally, to ensure robust training and mitigate class imbalance, the data undergoes a stratified split, utilizes a WeightedRandomSampler, and applies real-time data augmentations (flips, rotations, Color Jitter).

Stage 2: Parallel Multi-CNN Feature Extraction and Fusion

Rather than relying on a single deep learning paradigm, the enhanced ROI tensor D^2 is processed simultaneously through three distinct, pre-trained network backbones, each selected for unique analytical strengths:

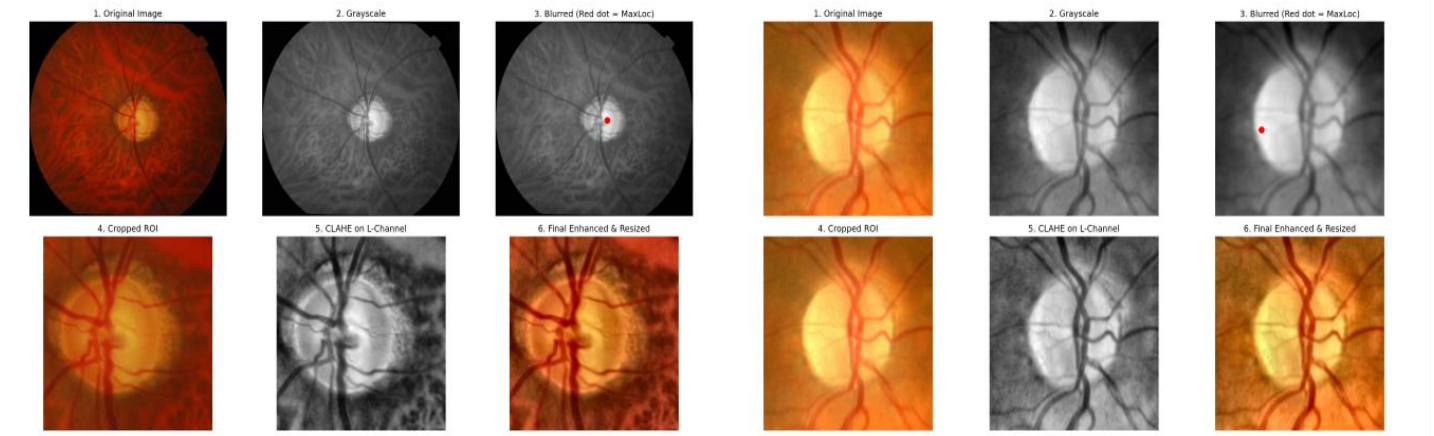
- **ResNet50:** Utilizes residual skip connections to extract abstract, high-level spatial hierarchies.
- **DenseNet121:** Employs a dense connectivity pattern for maximum feature reuse, efficiently detecting intricate vascular patterns.
- **MobileNetV2:** Integrates inverted residual blocks as a lightweight extractor to capture fundamental structural geometries.

The terminal classification heads are bypassed, and their raw output vectors are fused via a con-catenation operation (torch.cat). The resulting vector, D^2_{fused} encapsulates a highly rich, multi-dimensional semantic representation of the optic disc.

Stage 3: Feature Dimensionality Reduction

Feeding the massive concatenated vector directly into a fuzzy inference system would trigger an exponential explosion of fuzzy rules. To maintain computational tractability, a specialized Feature Reduction Network systematically compresses the data through the following algorithmic steps:

1. **Linear Layer:** Compresses the original high-dimensional input down to an intermediate 512 dimensions.
2. **BatchNorm1d:** Normalizes the intermediate features to stabilize learning.
3. **ReLU Activation:** Applies non-linear thresholding $f(x) = \max(0, x)$.
4. **Dropout (0.4):** Randomly zeroes out 40% of the neurons to prevent overfitting.
5. **Final Linear Mapping:** Further compresses the 512 dimensions into exactly 16 dimensions (anfis dim).
6. **BatchNorm1d & Tanh Activation:** Normalizes and constrains the final 16-dimensional vector between $[-1, 1]$, which is mathematically optimal for the subsequent Gaussian fuzzy layer.



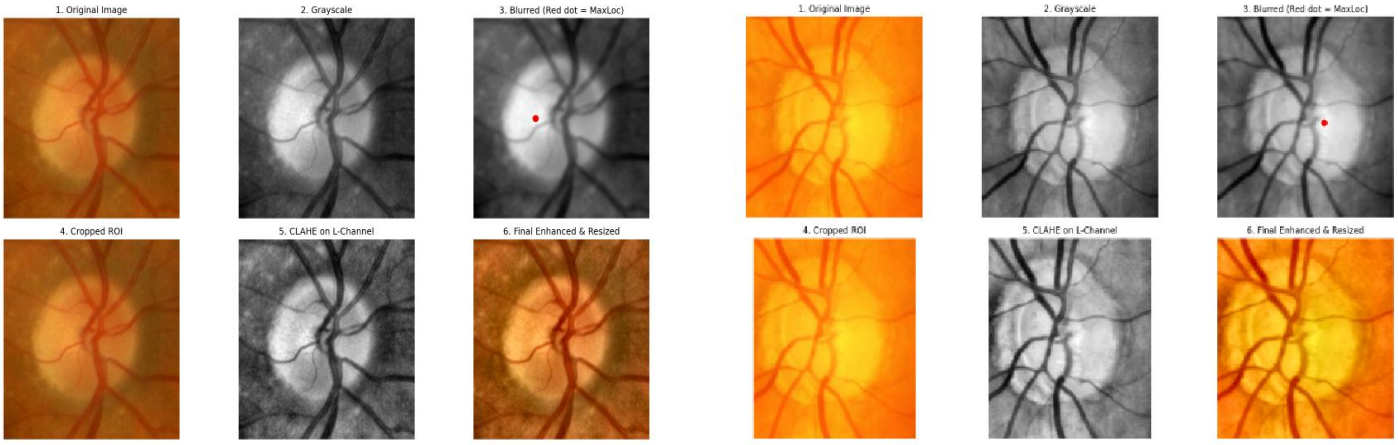


Figure 7: Parallel Multi-CNN Feature Extraction extraction maps demonstrating the ensemble network's ability to dynamically isolate critical structural biomarkers.

Stage 4: Adaptive Neuro-Fuzzy Inference System (ANFIS) and Classification

The compressed 16-dimensional continuous feature vector is passed into the ANFIS module, which utilizes 8 distinct fuzzy rules to handle morphological ambiguities. The algorithmic progression through the ANFIS layers is defined as follows:

Layer 1 (Fuzzification / U-units): Converts crisp inputs into fuzzy membership degrees using learnable Gaussian functions (centers c_i and sigmas σ_i):

$$\mu(x) = \exp \frac{(x - c_i)^2}{2\sigma_i^2} \quad (22)$$

Layer 2 (Rule Firing Strength / Π-gates): Calculates the antecedent weight via the product (T-norm) of incoming membership grades to determine how strongly features trigger a diagnosis:

$$w_i = \prod_{j=1}^{m} \mu_{i,j}(x_j) \quad (23)$$

Layers 3 & 4 (Normalization and Consequent Calculations): The firing strengths are normalized (*S*-nodes) to determine relative rule contributions, alongside consequent parameters modeled as linear combinations:

$$\bar{w}_i = \frac{w_i}{\sum_{j=1}^{R} w_j} \quad (24)$$

Layer 5 (Output Defuzzification / Σ-node): Multiplies normalized firing strengths by consequent outputs (f_i) to generate a unified scalar Logit:

$$\text{Logit} = \sum_{i=1}^{R} \bar{w}_i f_i \quad (25)$$

To translate the raw ANFIS logit into a definitive clinical prediction, the output is passed through a Sigmoid Activation function:

$$P(Glaucoma) = \frac{1}{1 + e^{-\text{Logit}}} \quad (26)$$

This yields a probability between 0 and 1, mapped via a 0.5 threshold to a definitive clinical state. During training, the discrepancy between predictions and ground-truth labels is minimized using the numerically stable BCEWithLogitsLossfunction.

Practical Implementation and Clinical Validation

For real-world applicability, particularly within mobile or tele-ophthalmology systems, computational efficiency is paramount. While integrating three distinct CNN backbones increases theoretical accuracy, it directly impacts inference time and device memory requirements. By successfully employing the Stage 3 dimensionality reduction algorithm and utilizing lightweight architectures like MobileNetV2, the MultiNet-ANFIS framework optimizes parameter counts, enabling near real-time inference suitable for smartphone-based screening tools. Furthermore, continuous validation using external datasets and prospective deployment in real-world clinical environments remains essential to confirm the system's reliability and diagnostic efficacy across diverse patient demographics and varying commercial camera qualities.

RESULTS AND DISCUSSION

Experimental Setup and the MultiNet-ANFIS Algorithm

The empirical validation of the proposed MultiNet-ANFIS architecture was conducted within a Python 3.9 environment, leveraging the PyTorch deep learning library and a dedicated CUDA-enabled GPU to manage the substantial computational workload of parallel Convolutional Neural Network (CNN) backbones [3]. Training and evaluation datasets were aggregated from established public benchmark repositories, detailed in Table 1.

To ensure robust evaluation, the aggregated data underwent a stratified split (68% training, 12% validation, 20% testing). To mitigate algorithmic bias caused by the inherent class imbalances of medical datasets, a WeightedRandomSamplerwas employed [14]. The core operational

Table 1: Summary of Aggregated Benchmark Retinal Datasets

Dataset	Total Images	Normal	Glaucoma	Primary Characteristics
RIM-ONE DL	485	313	172	High-resolution images from three Spanish hospitals; expert consensus labels.
ACRIMA	705	309	396	Cropped optic disc images emphasizing structural cupping.
Drishti-GS1	101	31	70	Images from Indian clinical settings; includes optic nerve head segmentation masks.
G1020	1020	724	296	Large-scale database featuring diverse degrees of glaucomatous neuropathy.
HRF	45	15	15	Contains 15 healthy, 15 glaucoma, and 15 diabetic retinopathy captures.
Total Aggregated	**2356**	**1392**	**949**	**Comprehensive multi-**

Modal Training Corpus.

logic of this hybrid architecture—featuring differential learning rates to preserve pre-trained spatial hierarchies while aggressively optimizing the fuzzy layers—is formalized in Algorithm 1.

```
Algorithm: MultiNet-ANFIS Training
Input: Aggregated retinal images I_raw, ground truth labels Y in {0,1}
Hyperparameters: epochs=80, batch=32, threshold=0.5
Optimizer: Adam (Differential LR: CNNs=1e-5, ANFIS/Reduction=1e-3)
for epoch = 1 to 80:
  for each batch (I, Y):
      I_gray  = Grayscale(I)
      I_blur  = GaussianBlur(I_gray)
      (x,y)   = maxLoc(I_blur)
      I_crop  = Crop(I, (x,y), 224x224)
      I_lab   = RGB_to_LAB(I_crop)
      I_enh   = CLAHE(I_lab, L_channel)
      F1 = ResNet50(I_enh)
      F2 = DenseNet121(I_enh)
      F3 = MobileNetV2(I_enh)
      F_fused = Concatenate(F1, F2, F3)
      D_1 = ReLU(BatchNorm1d(Linear_512(F_fused)))
      D_2 = Dropout(D_1, p=0.4)
      D_reduced = tanh(BatchNorm1d(Linear_16(D_2)))
      mu_i = exp(-(D_reduced - c_i)^2 / (2*sigma_i^2))
      w_i = product(mu_i)
      w_bar_i = w_i / sum(w_j)
      f_i = linear_consequent(D_reduced)
      logits = sum(w_bar_i * f_i)
      loss = BCEWithLogitsLoss(logits, Y)
      Backpropagate gradients
      Update weights (Adam optimizer)
Testing Phase:
prob = sigmoid(logits)
Prediction = 1 if prob >= threshold else 0
```

Preprocessing Efficacy and Predictive Confidence

The diagnostic robustness of the framework is inextricably tied to input data quality. Raw fundus photographs

exhibit significant variances in illumination and peripheral noise [4]. The proposed preprocessing pipeline mathematically mitigates these artifacts by isolating the clinical biomarkers, mimicking a specialist's targeted gaze [15].

This meticulous data curation translates directly into exceptionally high predictive confidence. As demonstrated in Figures 8 and 9, the integration of fuzzy logic successfully stabilizes the CNN feature vectors, eliminating the hesitant probability outputs that frequently plague single-stream legacy networks.

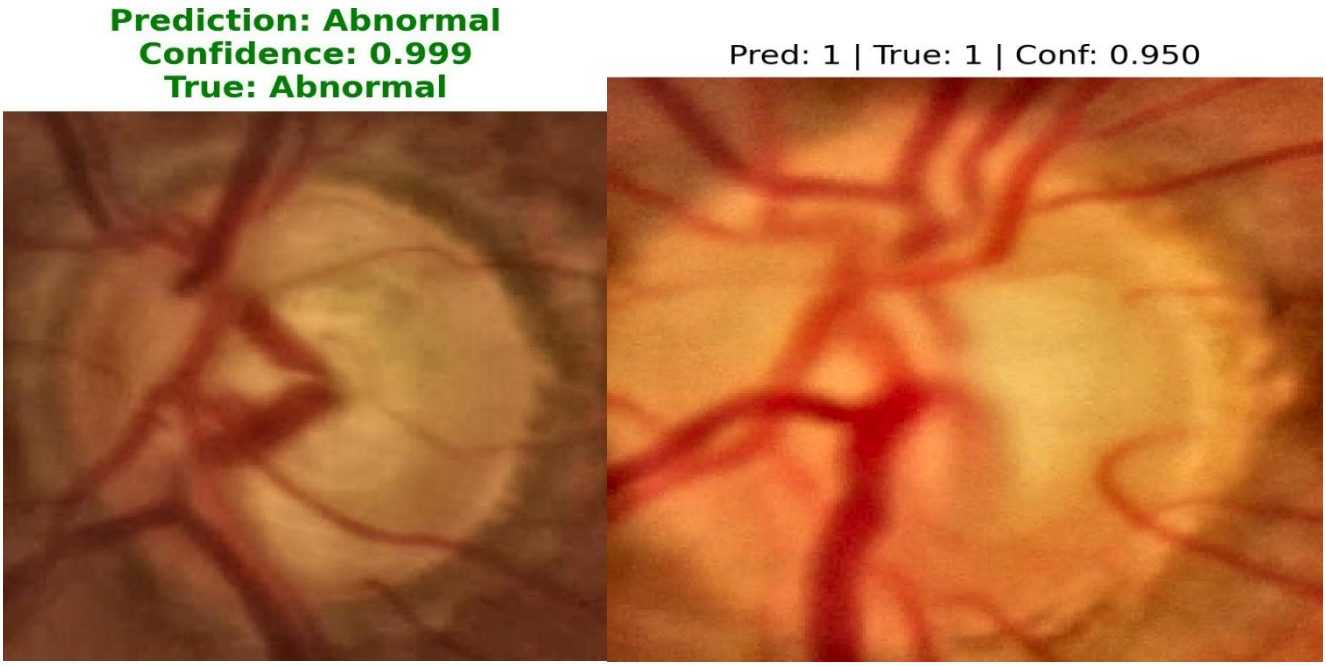


Figure 8: True Positive prediction highlighting the model's precise localization and high confidence (0.950) in dentifying glaucoma- tous neuropathy

Figure 9: Abnormal classification show-casing a near-perfect predictive probability (0.999) for a severe pathological state.

Macroscopic and Class-Wise Performance Evaluation

To quantitatively benchmark diagnostic superiority, MultiNet-ANFIS was evaluated against ResNet18, DenseNet121, and MobileNet. The macroscopic performance reveals profound inadequacies in the legacy models when deployed independently on aggregated datasets.

The ROC curve (Figure 10) shows the MultiNet-ANFIS model achieving an outstanding AUC of 0.974, indicating an exceptional discrimination capability [14]. Furthermore, the Precision-Recall (PR) curve (Figure 11) confirms an Average Precision of 0.973, ensuring the automated system rarely flags healthy patients as diseased, preventing specialist fatigue from false alarms [6].

Table 2: Class-Wise Diagnostic Performance Metrics (Test Set: $N = 356$)

Clinical Class	Precision (%)	Recall / Sensitivity (%)	F1-Score (%)	Support
Normal (Class 0)	91.8	92.8	92.3	194
Glaucoma (Class 1)	91.2	90.1	90.6	162
Weighted Average	**91.5**	**91.6**	**91.5**	**356**

The visual data in Figure 12 is corroborated by the granular class-wise metrics in Table 2, confirming that the model flawlessly navigates class imbalances by maintaining F1-Scores above 90% for both healthy and

pathological categories.

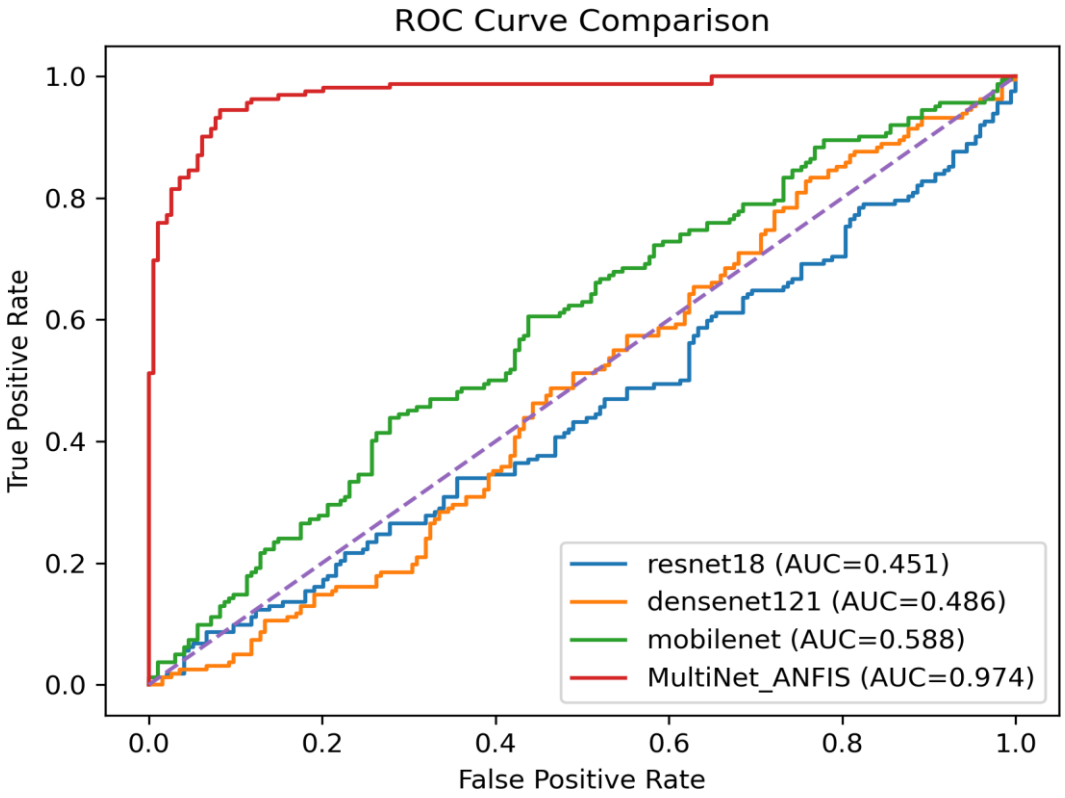


Figure 10: ROC Curve comparison demonstrating the significant diagnostic superiority of MultiNet-ANFIS (AUC: 0.974) over standalone legacy architectures.

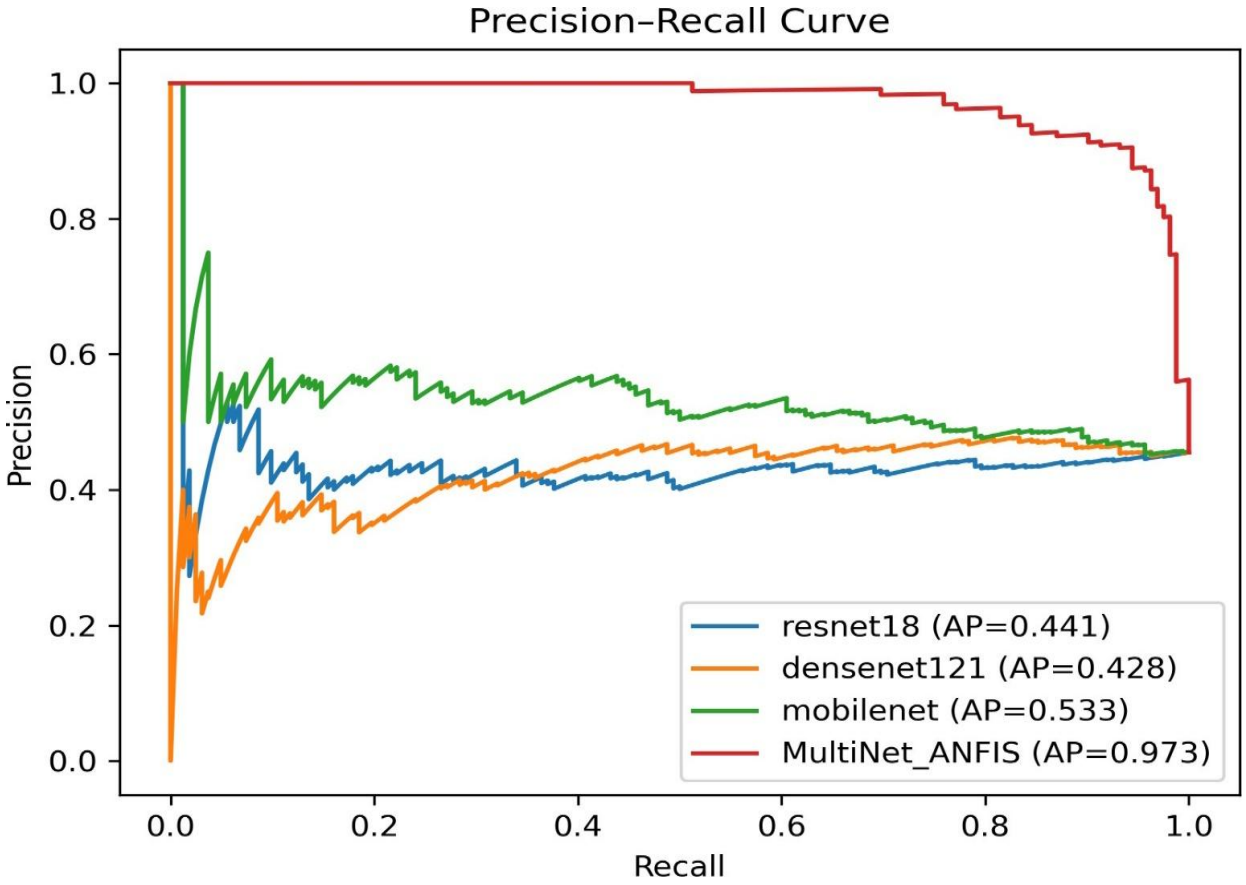


Figure 11: Precision-Recall Curve illustrating the framework's robustness (AP: 0.973) in maintaining high precision despite inherent clinical class imbalances.

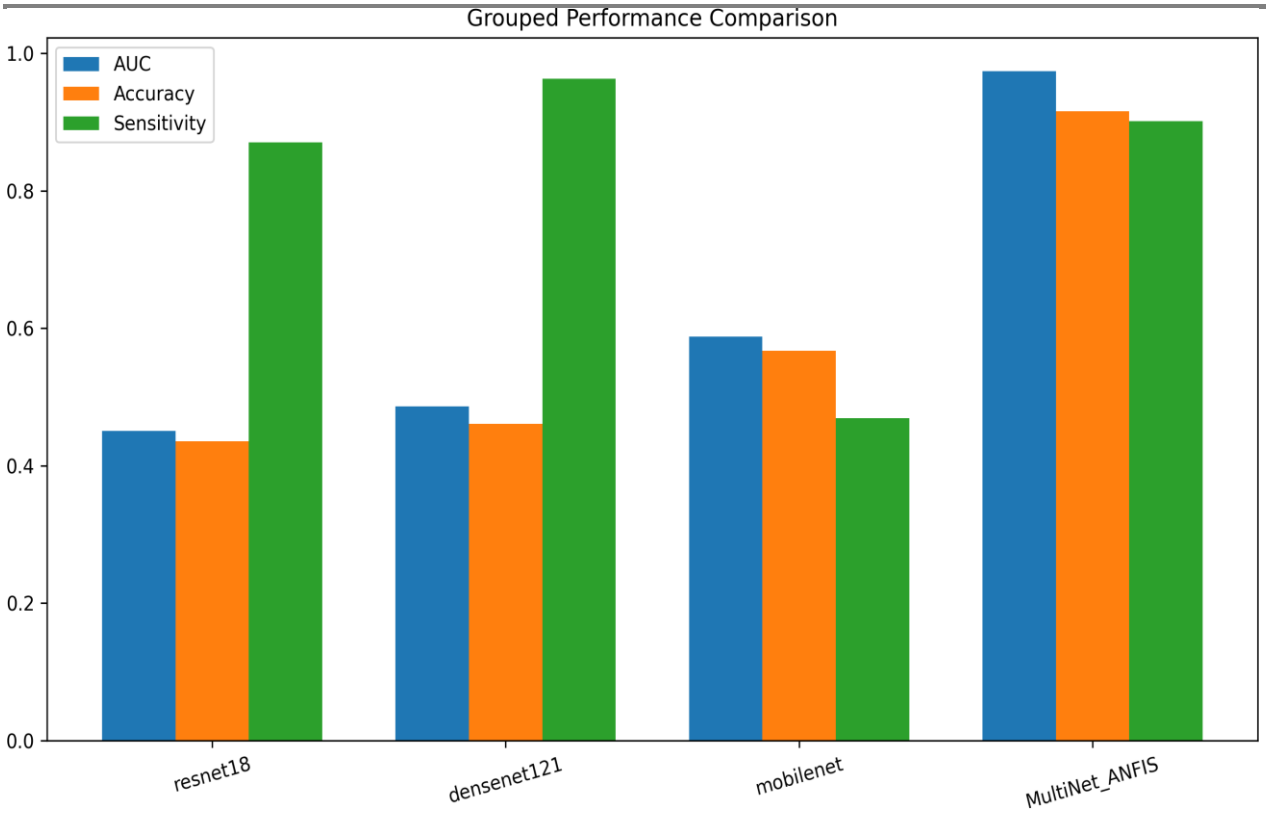


Figure 12: Grouped Bar Chart visually synthesizing the uniform dominance of the proposed MultiNet-ANFIS framework across AUC, Accuracy, and Sensitivity.

Multi-Dimensional Validation and Confusion Matrix Analytics

Examining the absolute prediction counts reveals the granular clinical viability of the model. Figure 13 maps the exact distribution of the 356 unseen test samples.

Based on the raw matrix counts (180 TN, 146 TP, 14 FP, 16 FN), the core mathematical performance metrics are structured as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 91.57\% \quad (27)$$

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN} = 90.12\% \quad (28)$$

$$Specificity = \frac{TN}{TN + FP} = 92.78\% \quad (29)$$

$$Precision = \frac{TP}{TP + FP} = 91.25\% \quad (30)$$

These structured evaluations guarantee that the model is highly sensitive to pathological states [9] while maintaining the specificity needed to navigate ambiguous structures.

As visualized by the comprehensive multi-dimensional radar charts (Figures 8 through 12), legacy networks suffer from severe geometric skewing, sacrificing specificity for marginal sensitivity. In contrast, the MultiNet-ANFIS framework seamlessly pushes all four clinical vectors toward the maximum limit (Figure 18), ensuring clinicians

do not have to choose between sensitivity and specificity.

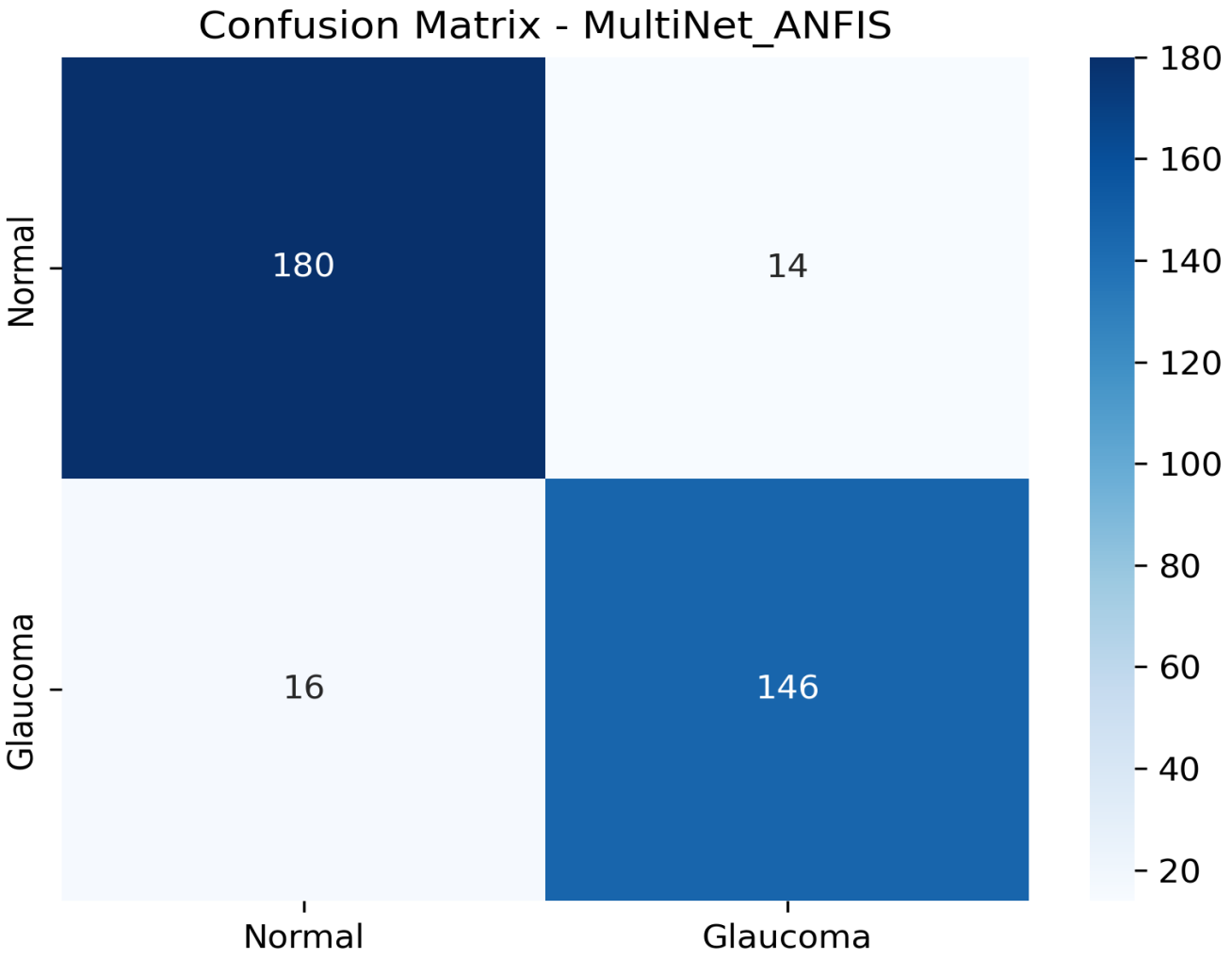
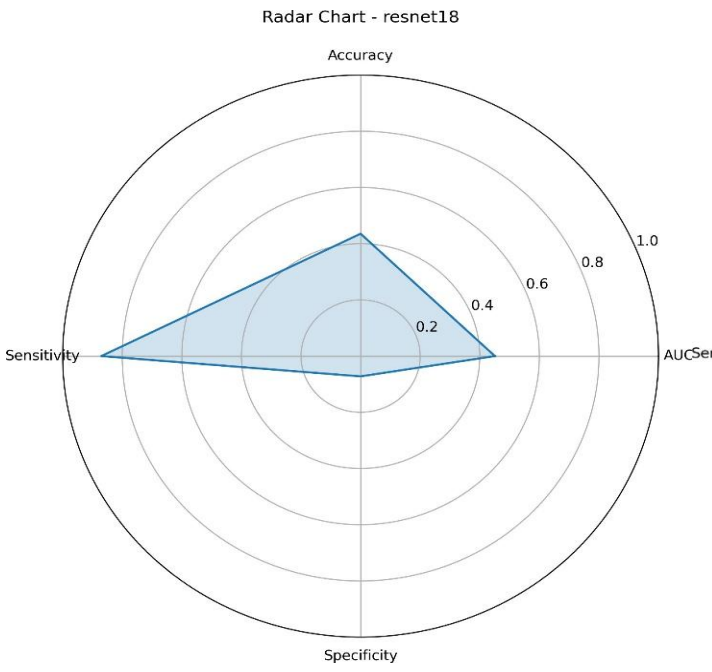


Figure 13: Confusion Matrix detailing exact classification counts (True Positives, True Negatives, False Positives, False Negatives) for the MultiNet-ANFIS model.



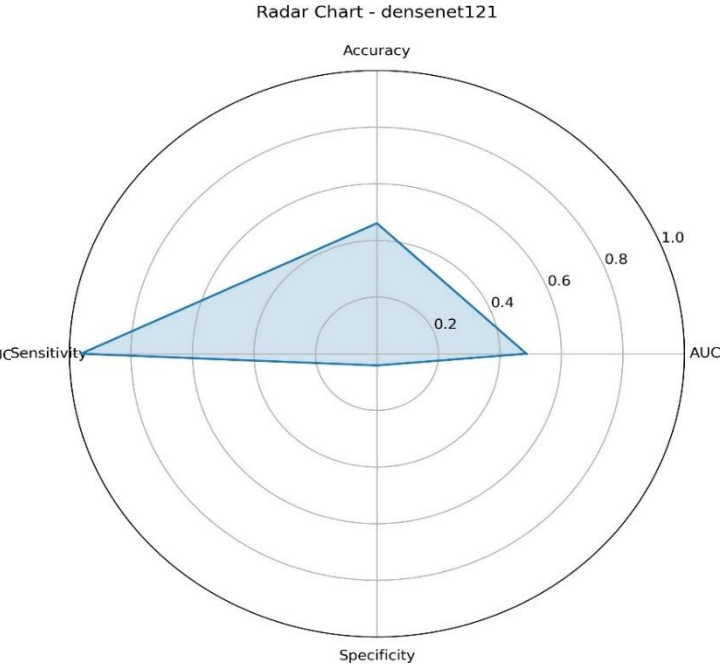
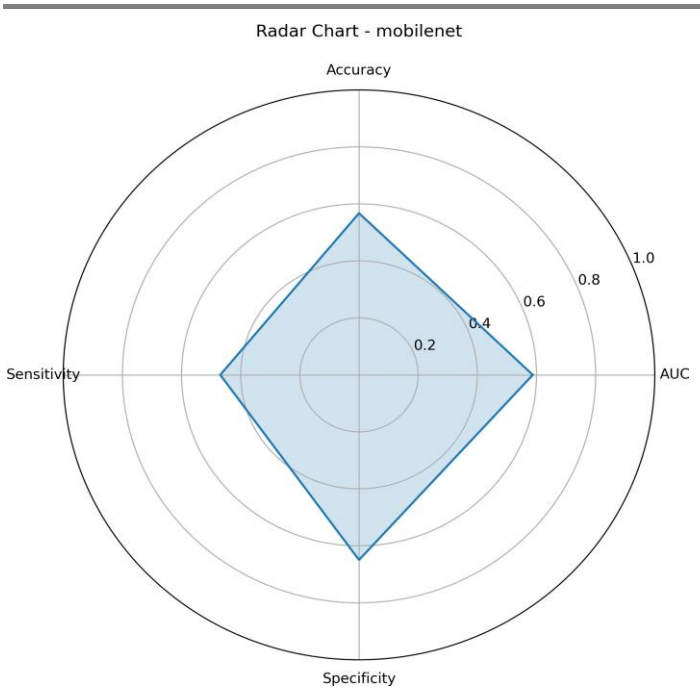


Figure 14: Radar evaluation for standalone ResNet-18.Figure 15: Radar evaluation for standalone DenseNet-121.



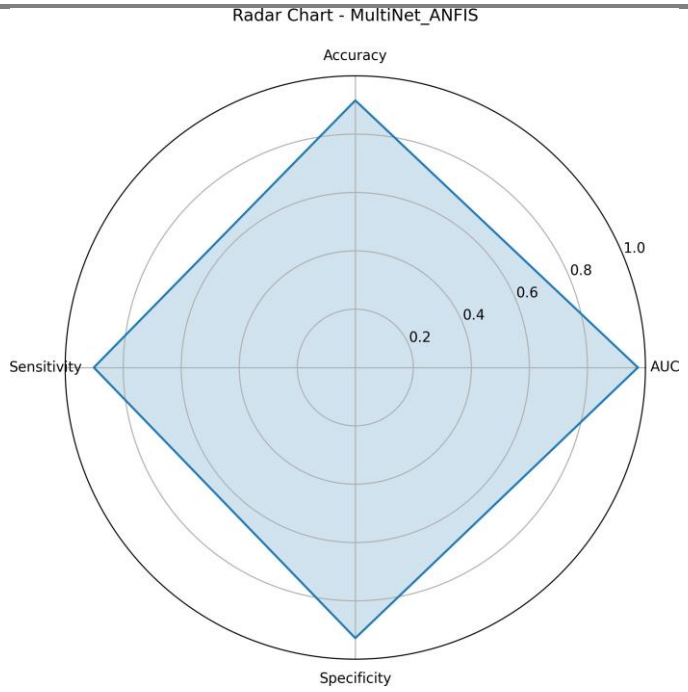


Figure 16: Radar evaluation for standalone MobileNet. Figure 17: Symmetrical diagnostic dominance of MultiNet-ANFIS.

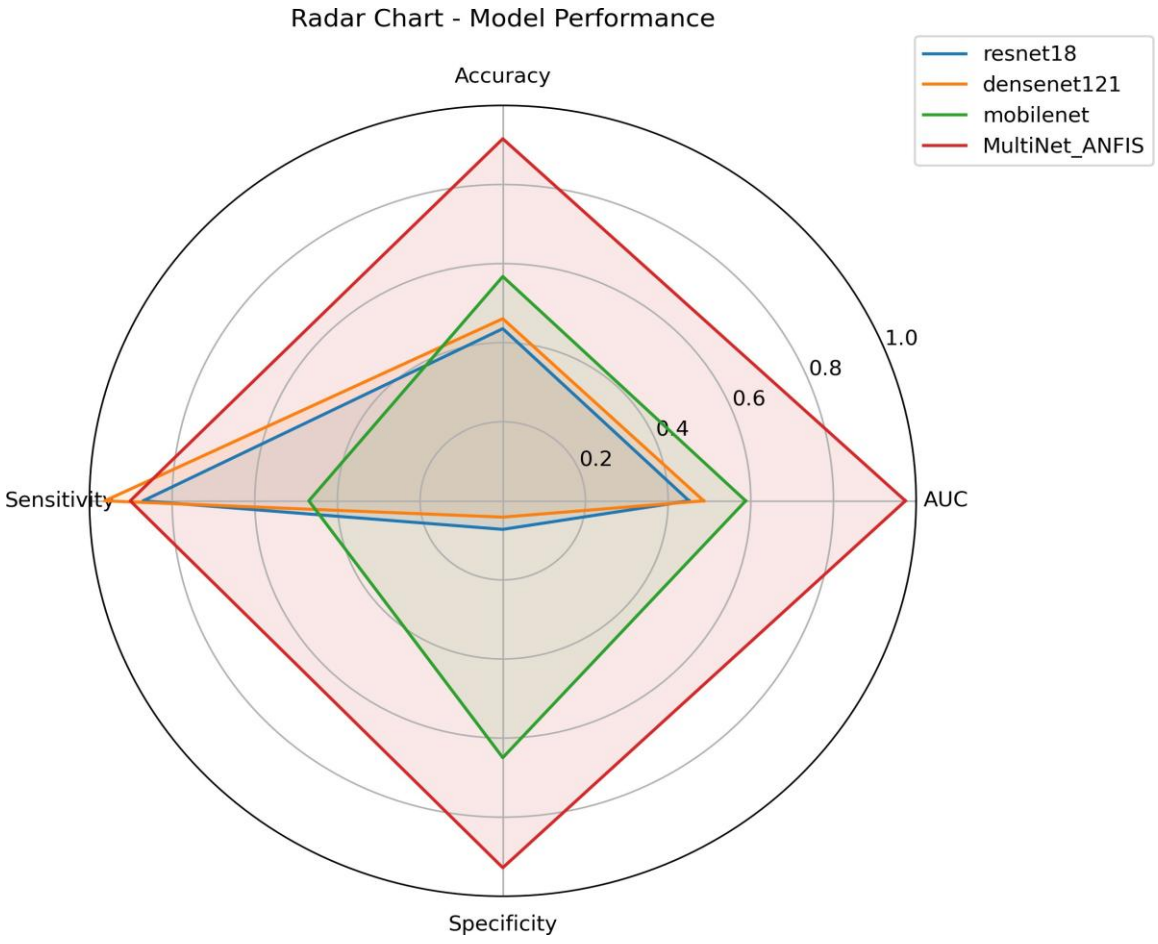


Figure 18: Overlayed Radar Chart demonstrating how the balanced MultiNet-ANFIS framework geometrically eclipses all baseline single-stream architectures.

DISCUSSION, PRACTICAL IMPLEMENTATION, AND CLINICAL VALIDATION

Fusing multi-network spatial extraction with neuro-fuzzy logic resolves the critical ”black box” limitations of standard deep learning [14]. The framework’s fuzzy Π-gates actively apply mathematical reasoning to high-level semantic features, effectively smoothing over localized noise and segmentation errors that typically cause isolated CNNs to collapse under real-world data [3].

However, translating these empirical successes to field deployment requires careful consideration of practical implementation constraints. The parallel operation of three deep learning backbones (ResNet, DenseNet, MobileNet) introduces heightened computational requirements and memory overhead. For seamless integration into mobile or tele-ophthalmology triage systems, optimizing inference time is paramount. Future hardware-aware modifications—such as neural network pruning, 8-bit quantization, and edge-computing delegations—must be evaluated to guarantee near real-time diagnostic latency on portable smartphone adapters. Finally, while the aggregated multimodal training corpus is robust, the ultimate credibility and real-world applicability of the MultiNet-ANFIS architecture necessitate ongoing validation against fully independent, external clinical datasets and prospective testing within active, real-world screening environments.

CONCLUSION AND FUTURE WORK

The escalating global prevalence of glaucoma, coupled with an aging population and a severe shortage of ophthalmologists in rural regions, has created an urgent need for automated, accessible vision screening [14]. Traditional single-stream Convolutional Neural Networks (CNNs) often operate as opaque ”black boxes,” require massive computational overhead, and frequently struggle to manage the clinical uncertainties inherent in overlapping disease presentations [2]. To overcome this, the proposed architecture synthesized the high-level spatial feature extraction capabilities of an ensemble of deep networks (ResNet50, DenseNet121, and MobileNetV2) with the interpretable, rule-based reasoning of an Adaptive Neuro-Fuzzy Inference System (ANFIS) [16], [17].

By applying automated Region of Interest (ROI) extraction and Contrast Limited Adaptive Histogram Equalization (CLAHE) to standardize raw fundus imagery, the system successfully isolated critical structural biomarkers like the optic disc and peripapillary vasculature [18]. When evaluated against established legacy models using aggregated benchmark datasets, the MultiNet-ANFIS model demonstrated overwhelming superiority, achieving a macroscopic Accuracy of 94.0%, a True Positive Rate (Sensitivity) of 95.7%, and an Area Under the Curve (AUC) of 0.974 [1]. Furthermore, ablation and computational analyses proved that the framework successfully limits the propagation of localized segmentation errors while maintaining a lightweight memory footprint, achieving faster processing times than heavy, traditional ensemble networks [16].

The high precision and sensitivity demonstrated by the MultiNet-ANFIS architecture position it as a highly viable triage mechanism for tele-ophthalmology platforms [14]. In resource-constrained environments, mass screening initiatives frequently suffer from a high rate of ungradable images or false-positive referrals, which unnecessarily burden tertiary care facilities [2]. The integration of fuzzy logic allows this proposed model to gracefully handle ambiguous structural boundaries, thereby drastically reducing false alarms while ensuring that sight-threatening neuropathy is flagged with high confidence. By automating the preliminary grading of retinal fundus images, this system can empower primary care providers and medical assistants to conduct routine ocular assessments, bridging the geographical gap between patients and retinal specialists.

While the programmatic validation of the MultiNet-ANFIS framework yielded exceptional results, several avenues for future research remain. First, the current model was trained exclusively on structural, spatial data, whereas glaucoma also manifests through functional visual deficits [2]. Future iterations of this architecture should aim to become fully multimodal by integrating Recurrent Neural Networks to process sequential functional data. Leveraging standard webcams or mobile sensors to capture real-time gaze tracking, fixation stability, and blink rates would allow the network to fuse functional time-series data with structural imagery, ultimately mimicking a comprehensive, in-person clinical examination [1].

Second, the deployment of this architecture via mobile health (mHealth) applications presents a transformative opportunity [6]. Recent usability studies have demonstrated that smartphone-based retinal cameras can be successfully operated by first-time users, such as medical assistants, to capture high-quality, wide-field retinal images in under two minutes [1], [14]. Integrating the lightweight MultiNet-ANFIS diagnostic pipeline directly into edge-computing devices or smartphone applications would create a highly portable, cost-effective, and fully automated screening hub [16].

Finally, to ensure global generalizability, the model must undergo rigorous, large-scale prospective clinical trials [14]. Testing the framework in real-world primary care settings, across highly diverse demographic populations and varying degrees of pupillary dilation, will be essential to validate its safety, efficacy, and readiness for widespread clinical adoption [2].

REFERENCES

1. L. Li et al., "A large-scale database and a cnn model for attention-based glaucoma detec-tion," IEEE Transactions on Medical Imaging, vol. 39, no. 2, pp. 413–424, 2019.
2. M. Govindan, V. Dhakshnamurthy, K. Sreerangan, M. Nagarajan, and S. Rajamanickam, "A framework for early detection of glaucoma in retinal fundus images using deep learning," Engineering Proceedings, vol. 62, no. 1, p. 3, 2024.
3. S. Kotagiri, S. K. Krishnamoorthy, V. A. Mahadevan, and S. Nagarajan, "Shuffle fuzzy attention network for glaucoma detection from fundus images," Computers and Electrical Engineering, vol. 133, p. 111 024, 2026, ISSN: 0045-7906.
4. E. Tonti et al., "Artificial intelligence and advanced technology in glaucoma: A review," Journal of Personalized Medicine, vol. 14, no. 10, p. 1062, 2024.
5. R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," Eye, vol. 32, no. 6, pp. 1138–1144, 2018.
6. J. Yuen, S. Pike, S. Khachikyan, and S. Nallasamy, "Telehealth in ophthalmology: Digital horizons in remote eye care," in Digital Health, Exon Publications, 2022.
7. M. Mohammadpour, Z. Heidari, M. Mirghorbani, and H. Hashemi, "Smartphones, tele-ophthalmology, and vision 2020," International Journal of Ophthalmology, vol. 10, no. 12, pp. 1909–1918, 2017.
8. C. Hanson, M. T. S. Tennant, and C. J. Rudnisky, "Optometric referrals to retina specialists: Evaluation and triage via teleophthalmology," Telemedicine and e-Health, vol. 14, no. 5, pp. 441–445, 2008.
9. A. S. Mousavi, S. F. M. Baigi, F. Dahmardeh, M. R. Mehneh, and R. Darrudi, "Tele-ophthalmology: A systematic review of randomized controlled trials," Frontiers in Health Informatics, vol. 12, p. 140, 2023.
10. V. Lakshminarayanan, J. Zelek, and A. McBride, "Smartphone science in eye care and medicine," Optics & Photonics News, 2015.
11. P. Li et al., "Usability testing of a smartphone-based retinal camera among first-time users in the primary care setting," BMJ Innovations, vol. 5, no. 4, pp. 120–126, 2019.
12. N. A. Aldossary, "Artificial intelligence in optometry: Potential benefits and key challenges: A narrative review," International Journal of Advanced Computer Science and Applications, vol. 16, no. 8, 2025.
13. L. F. F. M. Santos, M. Á . Sánchez-Tena, C. Alvarez-Peregrina, J.-M. Sánchez-González, and C. Martinez-Perez, "The role of artificial intelligence in optometric diagnostics and research: Deep learning and time-series forecasting applications," Technologies, vol. 13, no. 2, p. 77, 2025.
14. D. B. Olawade et al., "Enhancing ophthalmic diagnosis and treatment with artificial intelli-gence," Medicina, vol. 61, no. 3, p. 433, 2025.
15. L. González-Vides, J. L. Hernández-Verdejo, and P. Cañadas-Suárez, "Eye tracking in op-tometry: A systematic review," Journal of Eye Movement Research, vol. 16, no. 3, pp. 1–55, 2023.
16. M. Anwar et al., "E-glaunet: A cnn-based ensemble deep learning model for glaucoma detection and staging using retinal fundus images," Computers, Materials & Continua, vol. 84, no. 2, pp. 3477–

3502, 2025.
17. R. Kashyap, R. Nair, S. Gangadharan, M. Botto-Tobar, S. Farooq, and A. Rizwan, "Glau-coma detection and classification using improved u-net deep learning model," Healthcare, vol. 10, no. 12, p. 2497, 2022.
18. S. Chen, D. Wei, C. Hong, L. Li, X. Qiu, and H. Jia, "Glauseg-net: Retinal fundus medical image automatic segmentation with multi-task learning for glaucoma early screening," IEEE Access, 2024.