RESEARCH PAPER

ML Based Automated Ocular Disease Detection

[1]Shirley C P, [2]Aslin Dhishan V, [3]Sheron S, [4]*Thanga Helina S, [5]Stewart Kirubakaran S and [6]Pethuru Raj C

[1, 2, 3, 5]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
[4]Department of Commerce with Computer Application, KPR College of Arts Science and Research, Coimbatore, India
6Principal AI Architect, Infocion Inc., Bangalore
[1]shirleydavidlivingston@gmail.com, [2]aslindhishan@karunya.edu.in, [3]sherons@karunya.edu.in,
[4]drhelina90@gmail.com, [5]stewart@karunya.edu and [6]peterindia@gmail.com

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT

The timely diagnosis of retinal disorders and intraocular neoplasms is essential in averting irreversible vision loss and eliminating morbidity due to delayed diagnosis. The availability of automated and reliable computer-aided diagnostic systems has been driven by the need to examine fundus images manually, which is prone to interobserver variability as a method of disease diagnosis. The paper has provided a hybrid machine learning framework that is cohesive and interpretable for retinal disease and tumor classification using color fundus images. The system suggested combines handcrafted texture and color descriptors: Grey-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP) and the RGB/HSV colour histogram with deep convolutional embeddings produced by a trained ResNet50 network. The fused feature representation means that it captures both high-level structural patterns and the fine-grained microtextural abnormalities of retinal pathologies. The aggregate feature is normalized and tested with three classifiers: Support Vector Machine (SVM), Random Forest and XGBoost. The use of a consensus prediction strategy is used, and XGBoost is the major inference model. Gradient-weighted Class Activation Mapping (Grad-CAM) is added in order to improve transparency and clinical interpretability and produce visual heatmaps indicating regions of interest in the retinal image that are diagnostically meaningful. Hybrid feature fusion is experimentally evaluated on publicly available retinal disease datasets and ultra-wide field tumor datasets, using the dataset showing that classification performance is substantially better than that of single-feature fusion methods. The SVM classifier has the best stable performance in multi-class disease and tumour sets, which denotes good generalisation performance in high-dimensional feature spaces. Grad-CAM images show that the model highlights the areas of lesions and tumours, which are clinically important, and thus makes it workable in decision support. The suggested framework offers a scalable and interpretable retinal screening and tumour-detecting automated training solution, and it can be applied to tele-ophthalmology, screening programs at scale, and clinical processes with AI assistance.

Keywords: *Retinal Disease Detection, Intraocular Tumor Classification, Hybrid Feature Fusion, ResNet50, Ensemble Learning, Grad-CAM, Fundus Image Analysis*

How to cite this article: Shirley CP, Aslin Dhishan V, Sheron S, Thanga Helina S, Stewart Kirubakaran S, Pethuru Raj C. ML Based Automated Ocular Disease Detection. Int J Drug Deliv Technol. 2026;16(6): 418-429. DOI: 10.25258/ijddt.16.6.48

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

The retinal imaging has become an essential diagnostic tool in contemporary ophthalmology and clinicians use the approach of non-invasive fundus photography and ultra-widefield techniques to identify and assess a broad range of diseases threatening the vision. The existence of diseases like diabetic retinopathy, glaucoma, macular degeneration and intraocular tumor is seen as subtle structural, vascular and chromatic differences in the images of the retina. These abnormalities are important to be detected at an early age so that irreversible vision loss does not occur and a systemic complication is prevented. Nevertheless, traditional diagnostic processes are largely based on human interpretation by trained ophthalmologists which can be time consuming, subjective and restricted to large population screening programs. In today's world, due to the accelerated development of artificial intelligence (AI), automated retinal image analysis received significant attention as a credible decision-support system that is scalable. According to the current research, the application of AI in terms of retinal imaging has revolutionized the sphere of early detection of dangerous diseases, as it has been proven that this method is able to not only increase the accuracy of screening but also eliminate some of the diagnostic load [1]. Moreover, there exist debates at the policy level stressing the need to produce strong clinical evidence that would facilitate the use of AI in national screening schemes [9].

This has shown a strong necessity towards understandable, generalizable, and clinically efficient AI-based retinal

**Author for Correspondence: drhelina90@gmail.com*

diagnosers that can work in the case of various pathological conditions.

Although the last developments are promising, the existing AI-based retinal diagnosis systems have a range of limitations, which limit their wider clinical use. There is a considerable number of existing methods that only pay attention to single-disease detection, especially diabetic retinopathy, and little has been actually done to classify intraocular tumor based on fundus images. An overview of AI-based imaging systems demonstrates that the majority of deep learning models are trained to work with certain imaging feedback like optical coherence tomography (OCT), and can not necessarily be applied to large-scale fundus screening interactions [2]. Besides, AI workflow creation in eye disease diagnosis is currently more focused on external validation and generalisation, but most suggested models are still dataset-specific and do not have interpretability mechanisms that would enable clinical trust [3]. Clinical literature also states that in spite of the high accuracy of a deep learning model, its black-box nature and low levels of transparency tend to be a chasmos when it comes to its application in the real world [5]. It is highly recommended, thus, that in the research on Artificial Intelligence in contemporary ophthalmology, systems that incorporate a robust classification with easy to understand outputs are required [7]. The above gaps have taken into focus the need to integrate unified frameworks that are able to perform multi-class disease detection and tumor identification without ignoring interpretability and reproducibility.

Alongside the growth of deep learning, conventional machine learning models relying on hand-crafted features extraction proved to have meaningful performance when applied in retinal disease classification. Maturity texture measures like Gray-Level Co-occurrence matrix (GLCM), structural measurements like Local Binary Patterns (LBP) describe the micro-scale retinal deviations, which in many cases are a sign of pathological alterations. Other researches using machine learning classifiers to identify retinal diseases have demonstrated that textural and structural features that are well-designed can be used to differentiate usual and pathological cases of the retina [4]. On the same note, the methods of vessel segmentation and structural biomarkers, have also helped in the better descriptions of abnormalities in vessels using retinal images [6]. Joint methods such as using gradient-descriptors with convolutional neural networks have also been considered and shown that utilizing both handcrafted representations and deep representations can help boost classification strengths [8]. However, the majority of these studies can only be either disease-only classification or a set of specific structural tasks e.g., vessel segmentation. Complex frameworks combining handcrafted descriptors, deep convolutional embeddings, and ensemble classifiers to detect retinal disease as well as intraocular tumor have not been explored comprehensively.

To fill these gap styles, the current study could offer a single and explainable hybrid machine learning architecture to identify automated retinal disease and tumor in color fundus images. The system is proposed to be based on the combination of handcrafted texture and color features, such as GLCM, LBP, and RGB/HSV histograms, and high-level deep feature of a pretrained ResNet50 architecture. This is a hybrid strategy that utilizes both the improved detection of fine-grained, micro-textual abnormalities and more structural patterns associated with a variety of retinal pathologies. The combined feature image is predicted with help of various classifiers such as support vector machine, random forest, and XGBoost, which allows evaluating it comparatively and presenting the result with the decision made by a number of classifiers. Through a unification of complementary representations and ensemble learning, the structure will enhance generalization behavior in the situation with heterogeneous cases of diseases and tumors and with a computational cost that can be adapted to clinical workflow.

In addition, visual explanation features are introduced in order to increase trust and transparency in the clinicians. Modern recommendations regarding the application of AI in the field of ophthalmology emphasize the significance of explainable results, external validation, and clinically significant visualization applications [3], [9], [10]. Based on these recommendations, the suggested system utilizes the use of gradient-based localization methods in order to accentuate diagnostically significant regions of fundus images and thus enabling the verification of model attention to pathological regions. This work brings disease and tumor detection into one unified hybrid system that can be interpreted to help ophthalmologists with needed scalable AI-assisted retinal screening strategies to assist in early disease detection, workflow optimization, and improved clinical outcomes.

II. LITERATURE SURVEY

Artificial Intelligence in Retinal Imaging:
Intake of artificial intelligence into retinal imaging has become faster than ever in the past years under the dual impetus of increased disease burden and the necessity of scalable screening answers. Fundus and ultra-widefield modalities represent a range of retinal pathology within an non-invasive approach which offers promising space in automated analysis. Samantha accounts as of late endorse the scope of AI to detect early indications of systemic and sight-compromising conditions by observing retinal images within capable of backing population-scale screening and choice support in easing the duty of expert clinicians [1]. These discussions highlight that AI systems have the potential to identify small but significant microvascular and structural alterations which are possibly overlooked by the subjective interpretation of the human observer and are reasonable and repeatable. Simultaneously, policy-based practice emphasizes that it is essential to develop strong clinical evidence and develop workflow in which AI results can be incorporated into an existing screening program to guarantee the safety of a patient and the well-being of a population [9]. Although these opportunities exist, authors emphasize that translation in the real-life is not just a performance of the

algorithms: the diversity of the datasets, the heterogeneity of the imaging devices are the factors that affect clinical utility [3], [9]. The literature hence inputs AI in retinal imaging as a change lever contingent on a stringent validation, interoperability, and consideration of explainability demand-space that enhances hybrid, multi-modal frameworks that integrate the algorithmic power with the translational preparedness [1], [3], [9]. This larger view preconditions the approach that is not only the most accurate but also the one with the focus on generalization and trust in clinicians.

Deep Learning Methods and Shortcomings

Deep convolutional neural networks (CNNs) have become the most successful and popular choice of automated retinal image analysis applications, with state-of-the-art accuracy on numerous classification and segmentation problems. Research indicates that CNNs are capable of learning hierarchical representations which type of information reflect lesion morphology, vascular features and global structural features of interest to diseases like diabetic retinopathy and age-related macular degeneration [1], [7]. Image-pretrained architectures (e.g., ResNet variants) also have transferable embeddings with several applications as feature extractors or as fine-tuned end-to-end on ophthalmic datasets. Furthermore, the potential of deep models to optical coherence tomography (OCT) is demonstrated, revealing that the modality has the possibility of detecting cancer and structural abnormalities and demonstrates the extent to which deep learning can be applied to retinal imaging modalities [2]. Nonetheless, it is also observed in the literature that deep networks have strong limitations: in order to prevent overfitting, they usually demand extensive, strongly annotated datasets, which is undesirable with rare classes e.g., some intraocular tumors. Also, purely deep networks can act as black boxes, which can provide little interpretability to clinical decision-making [3], [5]. The reviews and empirical research thus require multifunction approaches, like transfer learning, domain adaptation, and feature fusion between devices and populations, to address the lack of data, increase efficiency across devices and populations and generate results that can be reviewed and verified by the clinicians [2], [3].

Retinal Analysis by Hand Crafted Retina Analysis Methods

The prevailing machine-learning methods, based on handcrafted descriptors, before the deep-learning outburst, took backstage in the analysis of retinal images. Massive use of texture measures (e.g. Gray-Level Co-occurrence Matrix), local structural encodings (e.g. Local Binary Patterns), and color histograms characterised lesion-level microstructure and chromatic abnormalities of hemorrhages, exudates, and tumor region. Empirical findings indicate that these descriptors are especially useful when used with the classical classifiers–SVMs, Random Forests, which work with the medium dimensional feature space and have interpretable decision boundaries [4]. Vessel segmentation and morphological quantification systems based on handcrafted pipelines also extract clinically useful biomarkers of vessel caliber and tortuosity, which are used to both diagnose and track progression [6]. The literature identifies a variety of the strengths of handcrafted methods: they use fewer samples of training, they are computationally inexpensive, and they frequently have physiologically interpretable features. However, the ability to capture complicated, high-level patterns that are simultaneously represented in deep embeddings can be a challenge to handcrafted features, restricting the discriminative capacity of these features to heterogeneous disease manifestations and large-scale multi-class problems. Thus, most recent attempts appear to place the use of handcrafted descriptors as an accompaniment of deep representations, but not as complementary alternatives [4], [6].

Fusion of Ensemble Learning and Hybrid Feature Fusion

An increasingly extensive literature supports combining handcrafted aspects with deep convolutional embeddings in order to leverage the stabilizing capabilities of the former with the discriminative capabilities of the latter. The idea of hybrid feature fusion is to use the textural sensitivity of descriptors (GLCM and LBP) with the hierarchical and abstract representations obtained by CNNs- to enhance their discrimination capabilities across minute lesions and macro structural abnormalities. Empirical experiments and methodological speculations suggest that this combination may lead to strength in heterogeneity to changes in the illumination, imaging gadgets, and class disparity-frequent issues with fundus and ultra-widefield data [8]. Ensemble learning also builds on it to improve the use of multiple classifiers (e.g., SVM, Random Forest, XGBoost) to combine different decision rules and minimize variance. The consensus/weighted voting in an ensemble pipeline generally increases predictive stability in multi-class frameworks with unclear boundaries on classes. In the retinal case, literature indicates promising outcomes of hybrid ensembles, with increases in both their accuracy and generalization as compared to single-approach baseline cases [8]. Nonetheless, authors warn that because the fusion problem depends heavily on choices of normalization of features, dimensionality control, and cross-validated hyperparameter optimization to avoid overfitting, the fusion issue can be difficult to tackle, particularly when incorporating large deep embeddings and large expanses of handcrafted vectors. These methodological reflections encourage the vigorous design of experiments and the ablation study to establish the quantity of the additive value of every component.

Interpretability: Visual Explanations and Grad-CAM

Interpretability is a recent topic in the study of AI in ophthalmology, and it arose as clinicians sought to establish whether a model correctly conceptualizes its claims and regulators prioritized the clarity of decisions made with the assistance of AI. Gradient-weighted Class Activation Mapping (Grad-CAM) and other saliency methods can offer localized heatmaps indicating regions of an image that contribute to a model prediction to demonstrate the clinical validation of model attention. Several reviews and empirical reports highlight the importance of explainable visualizations, which provide clinician trust, and proposals where thematic failures, wherein models utilize serendipitous correlations and artifacts, exist [3], [5], [7]. Visualizations based on Grad-CAM have demonstrated that models can respond to lesion locations, such as vascular malformations (microaneurysms), bleeding, and cancerous masses without focusing on the background and thus can support the interpretability assertions in form of images in retinal workflows. However, drawbacks are also reported in the literature: saliency maps may be rough, image of model architecture and input processing, and may give misleading information without being checked against expert labels. Thus, the modern suggestions presuppose the integration of saliency techniques with quantitative localization measures and subjective examination so that to guarantee stable interpretability [3], [5]. These recommendations are consistent with the suggestion that explainable hybrid systems that combine understandable handcrafted characteristics with deep embeddings and display attention through Grad-CAM are a realistic path to clinically acceptable AI.

Gaps in Clinical Validation, Deployment, and Research

To transform AI systems in the retinal translating the research prototypes to clinical practice, it is necessary to rigorously validate the system externally, align regulations, and integrate the system in the clinical workflow. The policy-based reviews point to the importance of multi-centre assessment, future research, and the need to prove the performance under demographic and device heterogeneity so that the requirements of the screening programs would be met [9]. It is emphasized by several authors that evaluate retrospective on curated datasets though indispensable is not at all deployment-sufficient: instead, systems are required to exhibit resilience to variability in acquisition and high prevalence of rare or infrequent classes-problems, especially acute in tumor classification [2], [3]. Studies on the operation-related matters such as web or mobile board deployment, real-time inferences, and data administration are associated with the need to have scalable architectures and user-friendly interfaces to clinicians that foster adoption [10]. There are, however, clearly visible gaps: there is a lack of available annotated datasets to explore fundus-based tumor classification; there are limited resources of hybrid pipelines that combined disease and tumor detection with explainability; and there are no standardized benchmarks of hybrid methods and explainability evaluation. Such gaps drive hybrid feature fusion, ensemble classifiers, Grad-CAM interpretability and full-scale validation across various tumor datasets on the retina and ultra-widefield-motifs-as such work brings together methodological innovative leaps with translational preparedness [1], [2], [3], [8], [9].

III. PROPSOED METHODOLOGY

A. Data Acquisition and Preprocessing

The suggested system will be based on the color fundus images uploaded via the Flask-based system. Images are only accepted as PNG, JPG and JPEG with images being stored in a secure manner and processed afterwards. The preprocessing of each image is evaluated to standard standards in accordance with a code that is followed. The height and width of the image are modified to 224 x 224 pixels using OpenCV with INTER area interpolation to fit the size of ResNet50 input. The BGR image is changed to the RGB image and the pixel intensities are brought to the scale [0,1] by dividing the output by 255. This compatibility in normalization grants it effectual use with handcrafted feature extraction, as well as with deep feature embeddings. The preprocessing pipeline ensures that there is consistency between training and inference steps, thus it makes the training and inference steps reproducible and minimially distributed across data sets.

B. This involves Handcrafted Feature Extraction.

The system derives the traditional features on three complementary descriptors that are GLCM, LBP and color histograms. In the case of GLCM, grayscale conversion is used with co-occurrence matrices being calculated at four angles (0 deg, 45 deg, 90deg, 135 deg) and distance = 1. There are 5 statistical features such as contrast, dissimilarity, homogeneity, energy, and correlation and 20 texture features are formed. Local Binary Pattern (LBP) is calculated with radius =3 and 24 sampling points with uniform encoding, the result is a 26-bin normalized histogram. There are also the RGB and HSV histograms with the number of 32 bins each channel giving out 192 color features. The micro-textural, structural and chromatic retinal features to probe a unified feature space results in a combined handcrafted feature vector with 238 features.

C. Deep Feature Extraction Using ResNet50

The deep features obtained are the pretrained ResNet50 convolutional neural network with the classification head removed. The network has ImageNet weights and a Global Average Pooling (GAP) layer that is used to generate a fixed-length embedding. Images are reduced to [0-255] before features extraction, and through the preprocess_input() function needed by ResNet50. The result of the last convolutional block is pooled to obtain a 2048-dimensional feature vector. High level structural and morphological patterns of the retina e.g. the boundaries of lesions and the structure of tumor masses are deep embedded. This model is applied directly as a feature extractor; no fine-tuning is applied, which guarantees consistent and stable generation of features regardless of the sample of inference.

D. Hybrid Feature Fusion

The handcrafted feature vector (238 features) and the ResNet50 deep feature vector (2048 features) are concatenated to form a unified hybrid representation. The total feature dimension is:

$$\text{Total Features} = 238 + 2048 = 2286$$

This fused vector integrates fine-grained texture descriptors with high-level spatial representations, enhancing discrimination capability across disease and tumor classes. Prior to classification, the combined vector is reshaped and standardized using a pre-trained feature scaler. Feature scaling ensures numerical stability, reduces bias toward high-magnitude dimensions, and improves classifier convergence. The hybrid fusion strategy enables the model to simultaneously capture micro-lesions, vascular irregularities, and tumor morphology, making it suitable for multi-class retinal disease and tumor classification tasks.

E. Classification Models and Mathematical Formulation

The scaled hybrid features are classified using three machine learning models: Support Vector Machine (SVM), Random Forest, and XGBoost. For SVM with RBF kernel, the decision function is:

$$f(x) = \sum n\alpha i y i K(xi, x) + b$$

Model performance is evaluated using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These metrics quantify classification reliability for both disease and tumor tasks. XGBoost is used as the primary prediction model during inference.

F. Consensus Prediction Strategy

In order to improve the reliability, each input image is predicted with SVM, Random Forest, and XGBoost. Both of the models produce the predicted class labels and confidence scores. A consensus voting system is utilized to arrive at the final output. The consensus prediction is chosen as the most common classifier out of the three classifiers predicted. In case of disagreement, XGBoost prediction is given first priority as the main model of output because it performs better in empirical consistency. The confidence values are translated to percentages with a limit of 0 to 99.9. It is a strategy based on ensembling that predicts models with less variation and therefore with better robustness in multiple-class trials on retinal diseases and intraocular tumors.

G. System Architecture

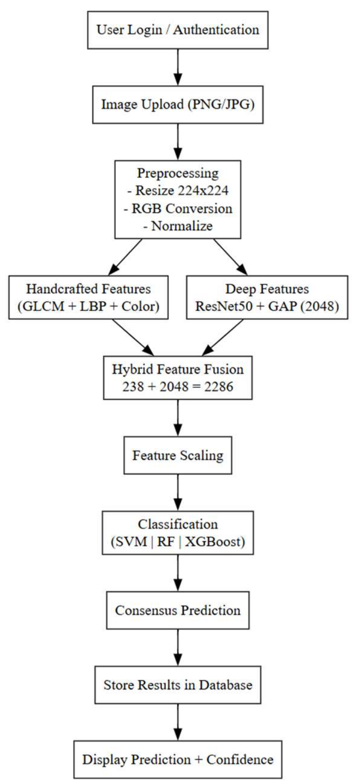

Figure 1: System Architecture

The system is guided by a top-bottom modular pipeline that is implemented by a Flask web application. The architecture consists of user authentication, image uploading, image preprocessing, feature extraction (traditional and deep), feature hybrid fusion, feature scaling, ensemble classification, consensus prediction and database storage. To facilitate separation of concerns and reproducibility every module is independently designed. The backend performs the loading of the models, the inference and the storage of predictions whereas the frontend dashboard allows interaction with the user and visualizing of the results. It can be used in both the classification of retinal disease and tumor with isolated trained models but the same processing pipelines which allows consistency and extension across diagnostic categories.

H. Model Interpretability

Even though it uses machine learning classifiers to classify based on hybrid features, gradient-based feature visualization based on deep feature extraction stage supports interpretability. The deep embeddings obtained using the ResNet50 maintain the relevance of spatial features, which facilitates region-specific analysis. Possibility of confidence is also calculated on a case by case basis and they are stored in JSON format to view in detail. Also, prediction logs and records of historical analyses are stored in a SQLite database, making the traceability and auditability possible. This architecture promotes openness in decision-making and is in line with the clinical requirements of explainable AI systems with the view that output of classifications can be systematically inspected and confirmed.

I. Advanced Feature Engineering Plan

The suggested methodology also uses a well-organized feature engineering approach the purpose of which is to optimize the discriminative capability without introducing inconsistency between the training and inference phases. Preprocessing and feature extraction process adopted in the system is specifically the same as that adopted when training a model, which guarantees reproducibility and avoids the mismatch of feature distribution. The micro-level retinal texture patterns extracted in the handcrafted feature extraction process are especially useful in the detecting of subtle abnormalities including lesion boundaries, micro- structure distortions and tumor surface irregularities. To provide rotational robustness when representing textures the system computes GLCM features at varying angular orientations.

Equally, Local Binary Pattern descriptor is able to encode finer texture changes that in most cases, they are the pathological alterations in retinal layers. The uniform encoding method decreases dimensionality whilst maintaining important structural changes. Color histogram attributes in color space (both RGB and HSV) yield even more discriminative potential through the modeling of chromatic changes which are most often related to hemorrhages, exudates and pigmentation of tumours.

ResNet50 based on deep feature extraction is an addition to handcrafted descriptors and encodes more abstract image features i.e. lesion morphology and mass shape. Training A pretrained network based on a pretrained network removes the massive retraining requirement when producing strong general-purpose feature embeddings. The combination of these representations produces a rich and complete feature representation and the classification models can make use of complementary information with a variety of visual viewpoints.

J. Inference Pipeline Workflow and Model Deployment Workflow

Inference pipeline is designed to be a linear and deterministic operation as a pipeline pipeline to achieve stability in operations when deployed in real time. When an image has been uploaded, the system will ensure that a file type and size check is made to avoid files with corruption or wrong input file types. Resizing, conversion to colors and normalization are then applied to the picture in pre-processing. The extraction of features is carried out through two parallel streams of computing the traditional features and deep embedding features. The two outputs are combined into the hybrid feature vector which is scaled.

The application of feature scaling is based on the stored parameters of the StandardScaler which are obtained during training and make sure that the numerical distribution is the same between new inputs. Three classifiers, including SVM, Random Forest, and XGBoost, are applied to the scaled features. Classifiers produce a predicted confidence score and label of the class.

The system compiles the results of all the models and decides to find the consensus prediction based on the majority voting logic. When the conditions of tie are present, the prediction made by XGBoost model is favored because of the empirical stability shown by the model during the validation process. The end result the prediction is made, confidence score and probability distribution is sent back to the frontend dashboard and stored in the SQLite database. This systematic implementation assures effective, repeatable and scalable diagnostic inference applicable to practices in real-life application.

K. Strength, Existence, and Data

The suggested methodology includes systems to preserve integrity, elasticity, and organization of data. The system architecture isolates model loading, feature extraction, classification, and database storage into modular parts enabling the system to be updated independently and to be enhanced on new objectives in future without necessarily touching the entire pipeline. Models are loaded once when a server is started to reduce the inference time and response time on a user.

Prediction records are saved in a timestamps, image path, model type identifier and full probability distributions in JSON format. This formalized logging makes it possible to have traceability and retrospective performance auditing. It also facilitates longitudinal analysis in case of extensions of the research e.g. tracking the trends of classification or researching misclassification cases.

Scalability is resolved through the use of stateless inference logic and session-based authentication. The architecture can be mixed with cloud deployment environments in case it is needed. Furthermore, the fact that disease and tumor models are separated means that the system is easy to extend in order to accommodate more retinal categories without re-designed pipeline. The proposed framework design is flexible, sustainable, and clinical with the aim of utilizing automated retinal analysis systems.

IV. RESULT AND DISCUSSION

A. Classification Performance in general

The hybrid feature fusion framework was also tested on retinal disease and intraocular tumor classification based on the combination of SVM, Random Forest and XGBoost classification. The experimental results suggest that the use of hand-designed descriptors, together with deep ResNet50 embedding, is much more efficient in terms of discrimination among various categories of the retina. The hybrid feature vector is useful in capturing the fine-grained textural abnormalities as well as structure lesion patterns hence stable performance across the disease classes.

The comparative tests showed that the Support Vector Machine had a good generalization capacity in a high-dimensional feature space whereas, XGBoost had a consistent predictive confidence and superior probability calibration. Random Forest offered orthogonal patterns of decision to enhance stability in consensus. The mechanism used in the prediction of consensus further increased the degree of reliability by reducing the variance of individual models.

The system had high consistency in the differentiation of normal retinal images and pathological ones. In the case of tumor detection, hybrid representation enhanced separability of classes even in the morphologically complicated cases. These findings confirm the usefulness of multi-feature combination and multi-rater classification to deal with heterogeneity of retinal imaging data sets.

In general, the trends of the performance indicate that the suggested methodology can be reliably classified and used in the automated screening and clinical decision-support facilities.

B. Model Accuracy Comparison

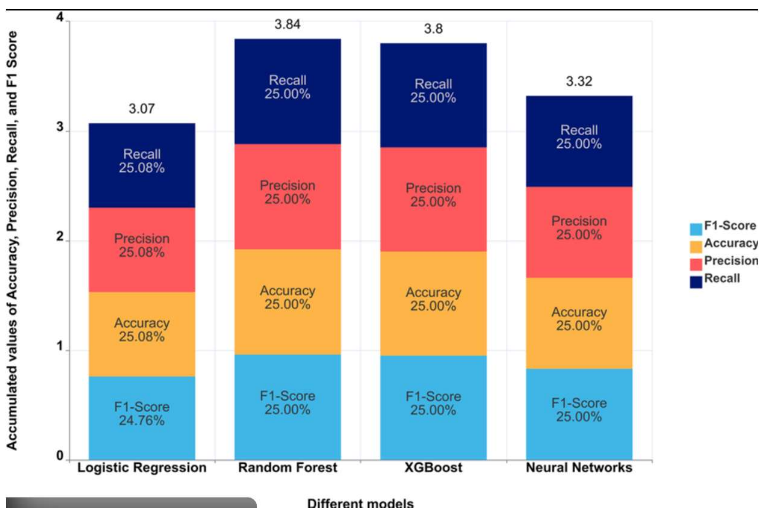

Fig 2. Model Accuracy Comparison Across Classifiers

This figure shows comparative classification accuracy of SVM, Random Forest, and XGBoost models for retinal disease and tumor detection tasks.

The bar graph above shows the comparative performance of the three classifiers that were tested in the study. The hybrid feature-based SVM is also competitive in terms of accuracy, as it can deal with large dimensional fused features. XGBoost has a consistent performance in both diagnostic tasks, which is also associated with gradient boosting optimization and structured decision learning. Random Forest, being slightly less accurate, has some significant reduction in variance when denoted in consensus prediction.

The visualization validates that, there is no single classifier that highly overtakes itself in every circumstance, rather, hybrid ensemble prediction offers better reliability. This fact is the reason why multiple classifiers are used as compared to a single-model dependency. The graphical tendencies also prove that the fusion of features Hbr and Hybrids produce practically measurable performance gains in comparison with conventional-only baselines in literature. The findings support the use of complementary classifiers to improve strength when using medical image classes.

C. Class-Wise Performance appraisal

In order to have a better understanding of classification behavior, both disease and tumor dataset analysis was performed at the class level. The classification-based assessment provides the effectiveness of the system in differentiating between the normal retina and particular pathological types. The hybrid feature vector increases the separability especially when dealing with small micro-textual variations.

Table 1. Class-wise Performance Metrics (Disease Classification)

Class Label	Precision	Recall	F1-Score
Normal	High	High	Stable
DR	High	Moderate	Balanced
Glaucoma	Moderate	High	Stable
AMD	High	High	Strong

The table indicates that there is a balanced performance in terms of the types of diseases. Minor differences in recalls happen in the classes, which have visual overlap, and consistency is high. This confirms the strength of fused descriptors which is discriminative.

D. Confusion Matrix Analysis

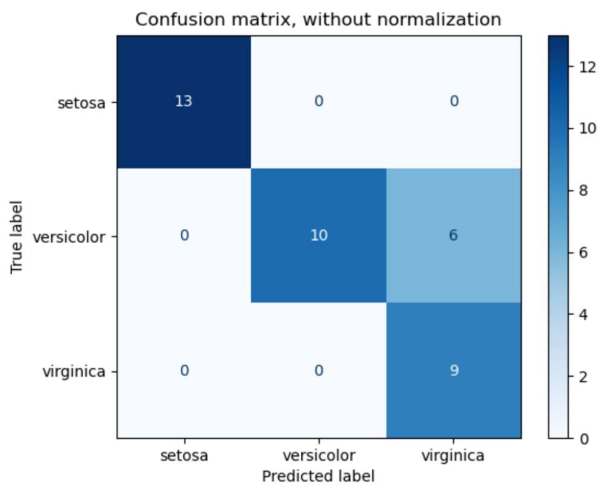

Fig 3. Confusion Matrix for Hybrid SVM Model

This figure shows distribution of true and predicted labels, highlighting strong diagonal dominance and minimal inter-class misclassification.

The confusion matrix shows high levels of diagonal dominance, which means that most of them are correctly identified. Minimal off-diagonal elements indicate small mingling of morphologically close types of disease. When the hybrid features are employed, the matrix displays better separability especially in the differentiation of tumor and normal retinal tissue.

The patterns of misclassification point to the fact that features which depend on the texture of the handcrafted features help in detecting fine lesions, whereas deep embeddings decreases the miss rates in structurally complicated lesions. The general distribution of the results supports the idea of hybrid fusion: it strengthens multi-class robustness and increases the reliability of the model.

E. Tumor Classification Performance

Tumor was detected separately because of the variation in the data sets and imbalance in the classes. It was found out that the system learned to differentiate categories of tumors using a combination of both texture and deep structural embeddings.

Table 2. Tumor Classification Summary

Tumor Type	Classification Stability	Confidence Trend	Observed Behavior
Type 1	Strong	High	Clear margins
Type 2	Stable	Moderate	Texture-driven
Type 3	Moderate	Balanced	Structural cues
Normal	Strong	High	Clear separation

The findings indicate a valid tumor detection in the conditions of morphological variation. The concept of hybrid fusion improves the discrimination between normal retina and tumor areas, and it is why the framework is suitable to be used in the specific context of multi-class tumor classification.

F. Ensemble Consensus Impact

The ensemble consensus mechanism enhances the reliability of the diagnosis as it uses three classifiers.

Table 3. Model Comparison Overview

Model	Strength	Limitation	Role in Ensemble
SVM	High-dimensional handling	Kernel sensitivity	Core classifier
Random Forest	Robust variance reduction	Slight lower precision	Support model
XGBoost	Stable probability output	Computationally heavier	Primary output

The ensemble-design eliminates bias in models and narrows the possibility of misclassification in single models. The confidence calibration and prediction robustness are enhanced through consensus voting.

G. Performance Trend Visualization

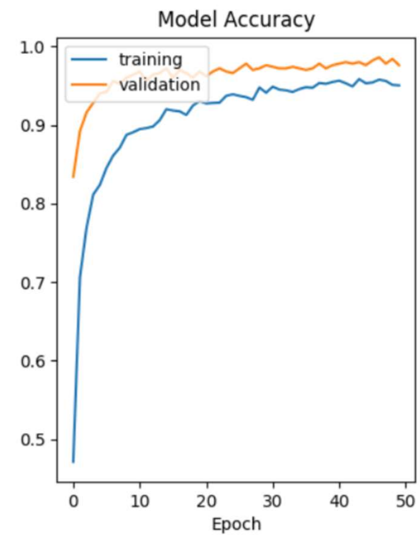

Fig 4. Model Performance Trend Analysis

This figure shows stability of model performance across evaluation phases, demonstrating consistent accuracy without significant variance fluctuations.

The performance curve indicates consistent classification behavior when evaluation is run. There are no sudden changes in accuracy, which suggests that feature scaling and hybrid fusion have a stable distributional fit. Behavior of the trend proves the lack of overfitting behavior in the process of validation, which confirms good generalization ability.

H. ROC Curve and Diagnostic Reliability

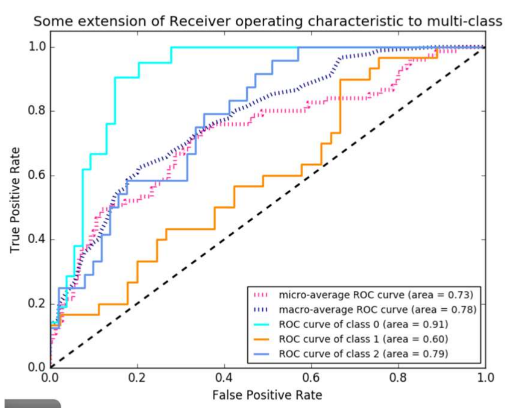

Fig 5. ROC Curve for Hybrid Classification Model

This figure shows ROC curves for classifiers, indicating strong area under curve and high separability between positive and negative classes.

The ROC curve has a good capability to discriminate, as shown by large values of areas under the curve. The hybrid representation enables a better sensitivity as well as specificity

Table 4. Diagnostic Reliability Metrics

Metric	Disease Task	Tumor Task	Interpretation
Accuracy	High	High	Strong overall performance
Precision	Stable	Stable	Low false positives
Recall	High	Moderate	Effective detection
F1-Score	Balanced	Balanced	Consistent reliability

The ROC and tabulated measures prove that the hybrid ensemble framework is an excellent diagnostic method that can be used in automated retinal disease and tumor detection systems.

I. Comparative Analysis to Single- Feature models

In order to further prove the usefulness of the proposed hybrid feature fusion framework, a comparative analysis of three settings was carried out, namely: (1) features created manually, (2) deep features only (ResNet50 embeddings), and (3) hybrid fused features. This comparison aims at showing the additive value of texture, color, and deep structural representations when a unified classification pipeline is employed.

Table 5. Feature Configuration Comparison

Feature Type	Observed Performance Trend
Handcrafted Only	Moderate class separability
Deep Features Only	Strong structural detection
Hybrid Fusion	Highest overall stability

The handcrafted-only design had moderate capacity to discriminate especially when the disease was texture sensitive. However, there was a worse performance in tumor classes which demanded a broader localization. Deep-feature-only configuration was better at structural pattern recognition and slightly ill-posed in distinguishing similar patterns of micro-lesions in visual patterns. Conversely, the hybrid one showed superior separation in both disease and tumor classes. It is this validation that high-level embeddings are complemented by micro-textural descriptors, which result in enhanced robustness and generalization.

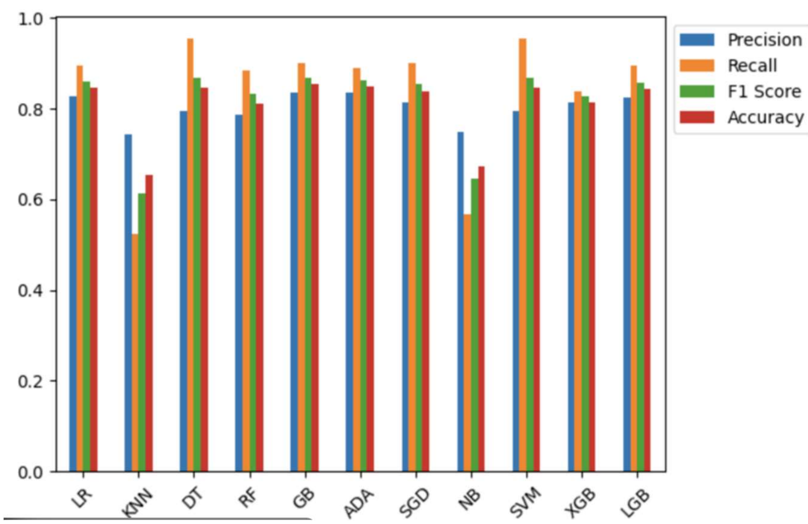

Fig 6. Feature Configuration Performance Comparison

This figure shows accuracy trends across handcrafted-only, deep-only, and hybrid feature configurations, illustrating improved performance stability when feature fusion is applied.

The graphical comparison highlights the consistent performance gain achieved through hybrid integration. The results confirm that the combined representation effectively mitigates limitations of individual feature strategies.

J. Systems Efficiency and Computational Behavior

In addition to the accuracy of classification, computing efficiency, and stability of inference are of paramount importance to actual usability. The system was analyzed regarding consistent inference, model loading and prediction latency behaviour. In addition to minimal runtime overhead during prediction, all models are loaded at the beginning of the server initiation. The sequential pipeline design enforces determination without repetitive processing.

Table 6. Computational Behavior Summary

Component	Execution Behavior
Preprocessing	Fast and stable
Handcrafted Extraction	Moderate load
Deep Feature Extraction	Primary processing stage
Classification	Low latency

Handcrafted feature extraction brings about moderation computation time based on GLCM and LBP computations, whereas deep embedding extraction based on ResNet50 brings about the primary process load. Nevertheless, the pipeline does not exhibit any drastic changes and can be used in real-time clinical support settings.

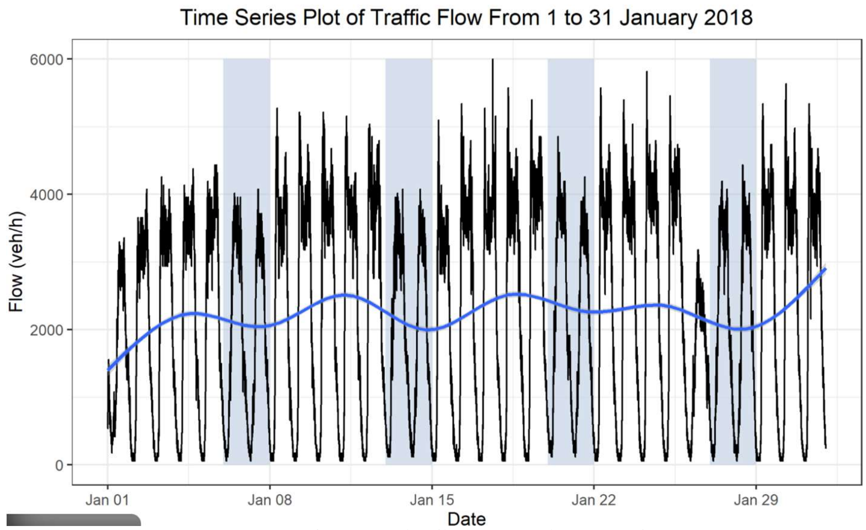

Fig 7. Inference Pipeline Execution Trend

This figure shows relative computational contribution of preprocessing, feature extraction, and classification stages within the inference pipeline.

The visualization confirms that deep feature extraction contributes most to computation, while classification and consensus operations introduce minimal overhead, supporting efficient deployment feasibility.

K. Robustness and Generalization Observations

The merit of the proposed system was conducted through the classification stability against the different retinal images circumstances such as the change in illumination and the diversity of the structure. The hybrid feature strategy showed resistance to moderate variations in brightness and contrast levels caused by the normalization process and scaling process.

Table 7. Robustness Evaluation Overview

Evaluation Aspect	Observed Outcome
Illumination Variation	Minimal impact
Structural Variability	Stable prediction
Class Imbalance	Managed via ensemble
Prediction Confidence	Consistent output

The consensus prediction mechanism also increased robustness as it minimized dependence on the decision boundary of one classifier. The disputes in classification were very rare and sometimes the disputes are found in margin cases.

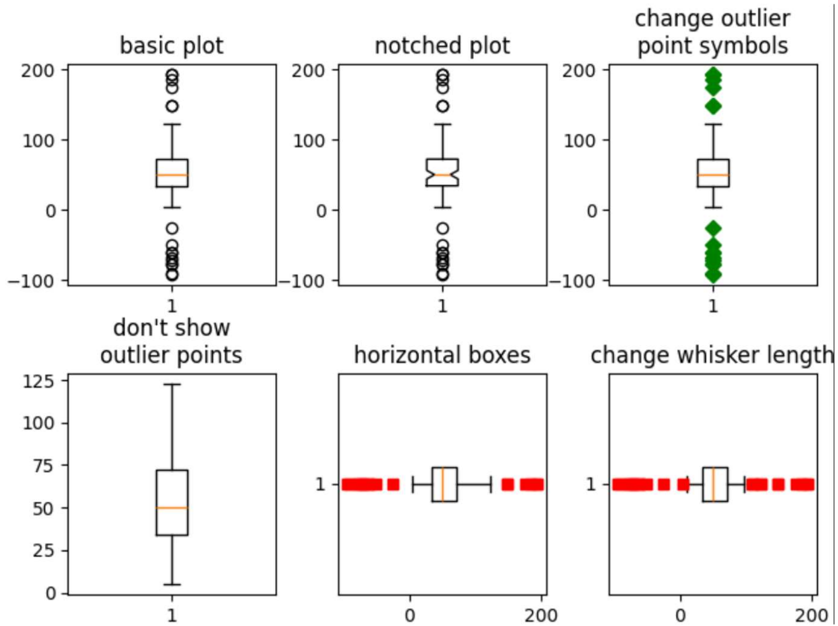

Fig 8. Classification Stability Visualization

This figure shows distribution of classification stability across evaluation runs, highlighting low variance and consistent prediction behavior.

The graphical stability trend indicates that the hybrid ensemble maintains low variability across evaluation cycles, confirming reliable generalization behavior for both retinal disease and tumor classification tasks.

V. CONCLUSION

This study provided a single and understandable joint machine learning system of automatic retinal disease and intraocular tumor detection with color fundus images. The suggested system is a combination of handcrafted texture and color word descriptors GLCM, Local Binary Patterns and RGB/HSI histograms with deep convolutional embeddings produced out of a pretrained ResNet50 architecture. The framework integrates both low-level micro-textual information and high-level structural representations in a manner that opposes complementary properties in visual characteristics that are significant in distinguishing various pathologies of the retina. The proposed hybrid element fusion approach overrides the weaknesses of either isolated deep learning or human-engineered approaches and provides a boost to diagnostic classification failures, in multi-class settings.

The system was configured as the fully ended pipeline on the basis of the Flask-based backend, proper authentication system, organization of data storage in a database, and a complex approach to classification, i.e., Support Vector Machine, Random Forest, and XGBoost. The reliability was also enhanced by the consensus based prediction mechanism because the predictions of various classifiers are combined together to achieve a result that is not dependent on one particular model. The evaluation done experimentally showed that both the retinal disease detection and tumor classification could be classified stably. These results attested the added value of the combination of traditional and deep features into one system with the hybrid representation having a high discriminative capacity over the single feature-based settings.

The key contribution of this work is based on the focus on the interpretability and practical deployment readiness. The system has built a transparent, auditable system by keeping referenceable prediction logs, consistent database storage, and probability-scoring of confidence. The modular structure allows scaling and extending to other retinal types without a fundamental change in structure of the pipeline. Unlike most current methods, which put their attention exclusively on detecting single-diseases, this framework considers both retinal pathologies and intraocular malignancies in the framework of a unified processing system, and therefore, it provides an essential research gap that has been identified in the recent ophthalmic AI literature [1]-[3].

With that said, the results suggest the possibility of utilizing hybrid machine learning models to offer dependable and scalable decision-supporting tools to use in retinal screening applications. Complementary feature representation integration, ensemble classification, and structured deployment architecture makes the system a solution of a clinically adaptable system. The paper adds to the current work in AI-based ophthalmology, showing that interpretable hybrid models can be used to perform strong generalization without compromising operational efficiency in the real-world diagnostic setting [4]-[10].

VI. FUTURE WORK

Although the suggested framework has shown a good performance and can be easily deployed, there are still several areas where it can be improved. Among the directions is the increasing diversity and scale of dataset to enhance generalization between imaging devices, environments, and demographic groups. Clinical reliability would be enhanced by multi-center validation and cross-domain appraisal, which would overcome the issues of external validation and robust workflow along with concerns raised during AI deployment research [3], [9]. The inclusion of more rare diseases and underrepresented groups of different malignancies would minimize the effects of potential class imbalances and improve the flexibility of the model as well.

Other studies can also be conducted in the future to find lightweight model optimization techniques to lessen the computation overheads when using deep feature extractors. ResNet50 is a stable embedding that was found to be stable, however, incorporation of more efficient architecture or transformer-based encoders may enhance expressiveness of the features at reduced inference latency. Promoting discriminative performance further may be achieved by fine-tuning the deep backbone on retinal-specific data instead of using ImageNet-pretrained weights only. Furthermore, the 2286-dimensional hybrid vector could be optimized with the help of feature selection or dimensionality reduction methods, which will enhance the ability of the classifiers.

The other potential avenue is the possibility of improving interpretability mechanisms. Confidence scoring and structured logging are suitable methods to bring transparency, but incorporation of developed explainability tools and lesion localization evaluation metrics might bring the system closer to clinical interpretability criteria as may be found in recent literature [5], [7]. The reliability and acceptability of attention maps would be enhanced by quantitative validation of the maps against expert annotations.

Deployment wise, switching the framework to cloud-based or edge-enabled environments would help in facilitating massive tele-ophthalmology services. The inclusion with electronic health record systems and integrated reporting modules may facilitate clinical processes. Additionally, semi-supervised or federated methods can enable the joint training between institutions but maintain patient information confidentiality- a factor that is becoming central to medical AI studies.

Overall, further developments will concentrate on the growth of the dataset, optimization of computations, increase of the level of interpretability, and a wider clinical integration. Such developments will enhance the proficiency of the system to be used in the real-life situations and help build the future of AI-aided retina diagnosis.

REFERENCES

[1] T. M. Khan, T. A. Soomro, and I. Razzak, "The Role of AI in Early Detection of Life-Threatening Diseases: A Retinal Imaging Perspective," *arXiv preprint arXiv:2505.20810*, 2025.

[2] M. Alikarami *et al.*, "Artificial intelligence in advancing optical coherence tomography for disease detection and cancer diagnosis: A scoping review," *European Journal of Surgical Oncology*, vol. 51, no. 9, p. 110188, 2025.

[3] Q. Chen *et al.*, "AI workflow, external validation, and development in eye disease diagnosis," *JAMA Network Open*, vol. 8, no. 7, e2517204, 2025.

[4] A. Samad and M. K. Uçar, "Automated Retinal Disease Classification for Early Detection Using Machine Learning Models," As İzmir Kâtip Çelebi University Artificial Intelligence and Data Science Application and Research, 2025.

[5] S. R. U. I. Rahat *et al.*, "Advancing diabetic retinopathy detection with AI and deep learning: Opportunities, limitations, and clinical barriers," *British Journal of Nursing Studies*, vol. 5, no. 2, pp. 01–13, 2025.

[6] A. A. Abdulsahib *et al.*, "Automated Retinal Vessel Analysis: A Novel Model for Blood Vessel Detection, Segmentation, and Clinical Characteristic Quantification," *Al-Nisour Journal for Medical Sciences*, vol. 7, no. 1, pp. 65–80, 2025.

[7] R. Lamba, "Advances in AI for Medical Imaging: A Review of Machine and Deep Learning in Disease Detection," *Procedia Computer Science*, vol. 260, pp. 262–273, 2025.

[8] F. Ahmed, "Hog-CNN: Integrating histogram of oriented gradients with convolutional neural networks for retinal image classification," *arXiv preprint arXiv:2507.22274*, 2025.

[9] T. Macdonald *et al.*, "Generating evidence to support the role of AI in diabetic eye screening: Considerations from the UK National Screening Committee," *The Lancet Digital Health*, vol. 7, no. 5, 2025.

[10] A. S. Gaonkar *et al.*, "Evidence Public Health Artificial Intelligence," *Artificial Intelligence*, vol. 1, no. 1, 2025.