International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-22

https://doi.org/10.70917/ijcisim-2026-1954
Article

Comparative Performance Evaluation of Machine Learning and Hybrid Deep Learning Frameworks for Accurate Detection of Diabetic Retinopathy Haemorrhages

Kale Sunil Manmath[1], Purushottam R. Patil[2]

[1] School of Computer Sciences and Engineering, Sandip University, Nashik, Maharashtra, India.smkale14jan@gmail.com
[2] School of Computer Sciences and Engineering, Sandip University, Nashik, Maharashtra, India. purupatil7@gmail.com

Abstract: The retinopathy that is in diabetic people is among those causes of vision impairment that are mostly experienced globally and the promptness of detecting the haemorrhages of the retina is vital in avoiding serious vision damage. Nevertheless, the manual analysis of the retinal fundus images is time-consuming, subjective, and likely to cause diagnostic variability. Machine learning based detection methods currently have low contrast lesions, structural similarity between haemorrhages and blood vessels, and a weak feature representation that can result in lower detection accuracy with high false positives. Thus, a highly sensitive haemorrhage detection and classification would need a robust and efficient automated system to detect haemorrhage and classify it effectively. The main purpose of this study is a comparative performance analysis of the traditional machine learning models and a proposed hybrid deep learning model on the specific detection of diabetic retinopathy haemorrhages. The suggested solution combines the state-of-the-art pre-processing, candidate extraction, hybrid feature extraction, and deep classification. The combination of handcrafted colour, texture, and shape characteristics and deep features that are extracted with EfficientNet-B7, DenseNet-201 and Swin Transformer networks is used. It trains and assesses these models using the retinal image data to detect the location of haemorrhage and determine the presence of a disease. The experimental results indicate that in both benchmark dataset used in this research and traditional machine learning algorithms can detect objects with accuracy at a range of 8791 and proposed hybrid deep learning framework can detect objects with accuracy of 9598 percent with less false detection and greater resilience to the presence of different image conditions. The better efficiency of the suggested system is proven by comparative analysis conducted to ensure that it has higher detection specificity, sensitivity, and reliability. The conclusion of the study is that the hybrid feature extraction with state-of-the-art deep neural networks is a highly effective and accurate solution to automated haemorrhage detection in diabetic retinopathy to assist in the early diagnosis and the enhancement of the clinical decision-making in an ophthalmic screening system.
Keywords: Diabetic Retinopathy; Retinal Haemorrhage Detection; Hybrid Deep Learning; EfficientNet-B7; DenseNet-201; Swin Transformer; Medical Image Analysis

1. Introduction

Diabetic retinopathy (DR) is a progressive microvascular complication of diabetes and is still among the major causes of avoidable vision loss on a global scale. The structural and vascular changes in the retina are observed to characterize the disease: microaneurysms, haemorrhages, exudates, and macular edema. One of the earliest signs of the presence of disease development is retinal haemorrhages that are crucial biomarkers in the stage and treatment planning. The damage and functional deterioration of the retina is caused by chronic inflammatory and neurovascular processes, which additionally underlines the significance of early detection and monitoring (Baudouin et al., 2021). As diabetes spreads across the world, automated analysis of retinal images has become a necessary tool of scalable and credible screening systems (Geetha & Hema, 2026). Timely detection of the haemorrhages goes a long way in preventing the irreversible retinal damages and loss of sight. Automated screening

systems are more helpful by enabling the ophthalmologists to carry out quick and objective evaluation of lesions and hence reduce variability in diagnosis. More recent haemorrhage detection frameworks based on deep learning have shown potential sensitivity and specificity of fundus imaging (Aziz et al., 2023). Models of risk prediction that have explainable machine learning methods also provide better clinical trust, as they present interpretable decision support to evaluate disease progression (Islam et al., 2023). These developments highlight the clinical significance of localization of haemorrhage with precision and accuracy.

1.1 Limitations of Manual and Traditional ML-Based Approaches

In spite of the current technological progress, grading of retinal fundus images is still in manual mode and is a subjective exercise. In all of these traditional machine learning methods, the use of handcrafted features is highly emphasized, which constrains its ability to adapt to complicated lesion patterns. Traditional convolutional neural networks have enhanced the accuracy of classification, but model generalization and features strength is an open question (Chen et al., 2025). DR classification models based on transfer learning show a performance improvement, but tend to concentrate on overall disease classification and not on haemorrhage detection (Gupta et al., 2025). Also, the architectures designed with EfficientNet are better performing, but cannot perform fine-grained lesion discrimination in various imaging settings (Dixit & Jha, 2025).

1.2 Challenges in Retinal Haemorrhage Detection

Several technical issues are related to the detection of retinal haemorrhage. The lesions can have low contrast with the surrounding retinal tissues making it difficult to segment and classify the lesions. False-positives are more prevalent when the elongated blood-vessels have a structural similarity with irregular haemorrhages. In addition, the differences in illumination, acquisition devices, and visualizing arts bring about inconsistencies which influence the robustness of the models. Two-stage deep learning models are trying to reduce some of these problems but they need better feature representation and situational learning (Zhang et al., 2025). Noise in labels and variability in annotations also deteriorate the performance of medical image analysis systems (Shi et al., 2024)

1.3 Motivation for Hybrid Deep Learning Framework

New development of transformer-based and hybrid deep learning models offers greater ability to detect global contextual relationships and multi-scale lesion features. Diagnostic systems based on vision transformers have been shown to model attention and are explainable more effectively in ophthalmic imaging (Balaha et al., 2025). Transformer-based architectures with intensity awareness also help in improving the localization of haemorrhagic areas in fundus imagings (Lavanya et al., 2026). These advances spur the combination of the hand-crafted morphological properties with the deep feature description to attain a better discrimination and strength.

1.4 Research Objectives and Contributions

The main aim of the paper is to conduct a comparative performance analysis of the traditional machine learning algorithms and hybrid deep learning models to detect retinal haemorrhage correctly. The proposed framework will enhance detection reliability and accuracy by enhancing state-of-the-art preprocessing, candidate retrieval, feature fusion, and deep classification network. With one foot in the door of dynamic segmentation strategy hemorrhagic lesion analysis (Li et al., 2025) and one foot in the door of optimized deep network design approaches (Ahmed et al., 2025), this study proposes a powerful hybrid approach that is able to tackle low-contrast lesions, similarity of vessels, and variation in illumination.

2. Related Work

The diabetic retinopathy detection systems that were established early were based on the traditional machine learning algorithms, which were coupled with manual feature extractions. Such techniques generally included intensity driven segmentation, morphological processes and statistical texture characteristics and subsequently classifiers like support vector machines and k-nearest neighbors. These methods performed moderately but they were compromised by feature dependency and low ability to adapt to varying imaging conditions. The article by Huang et al. (2024) suggested a framework based on MobileNet-V2/IFHO to identify an early stage of DR, and the authors proved that lightweight models may be more efficient, but they still rely heavily on the representation of features. Shinde and Wankhade (2025) also concluded that the use of structured preprocessing pipelines is significant to enhance the performance of any detector, and thus necessitating the creation of automated systems that would be able to handle massive screening data.

The appearance of deep convolutional neural networks shifted the DR detection to the automated feature learning. CNN based systems have been shown to be very effective in extracting hierarchical representations of retinal fundus images. The Saini et al. (2023) convolutional neural network with OCT images predicts the lesion-based disease in diabetic macular edema, and the lesion distinction by deep hierarchical features enhanced the results. On the same note, Jaganathan et al. (2024) established that concatenated transfer learning structures can improve the accuracy of medical image classification algorithms with the help of pretrained feature extractors. According to Kumari and Saxena (2024), multi-scale deep feature extraction with cyclic learning rate methods have been proposed, and the approach enhanced the strength of classification when applied in the analysis of ophthalmic images. These investigations show that CNN networks greatly surpass conventional manual techniques but can still have issues when it comes to detecting finer haemorrhage edges and variation in illumination. The limited labeled medical datasets have seen transfer learning becoming a domineering approach to the task of retinal image analysis. Premade models enable optimally adaptation to the tasks of DR classification and lower the cost of computation. The study by Ahmed et al. (2025) suggested an efficient deep regression network to detect the disease of macular edema, and the authors demonstrated that evolutionary optimization can be effective to improve the convergence of the deep model. Kanukuntla and Rani (2025) used level-set segmentation based on macular edema deformation detection, which emphasized the usefulness of leveraging both the precision of the segmentation and deep representations. These methods show better lesion detection performance though the emphasis is mostly on the detecting of macular abnormalities and not on the detection of isolated haemorrhage. Additionally, the general medical image classification deep learning models have focused on the issues of model generalization in a wide range of datasets (Ebrahimi and Luo, 2021).

In recent years, vision transformer architectures have been noted to use in the analysis of long-range contextual dependencies in medical images. Tamilselvi et al. (2025) used ResNet50 with Swin attention to improve the accuracy of the biomarkers, since their interaction is more global and local, as illustrated by a better interaction. Lin (2025) proposed an effective model of fusion transformer in the precise classification of eye disease with the focus on the role of the feature aggregation through attention. New interpretability advantageous perspectives have also been demonstrated by vision transformers especially when united with explainability systems in computer-aided diagnosis systems (Balaha et al., 2025). Also, transformer model based on intensity awareness has shown accurate localization of the areas of vitreous hemorrhage and enhanced resistant to imaging artifacts (Lavanya et al., 2026). These papers confirm that compared to the traditional CNNs, transformer-based models are more effective to represent the complex structures of the retina.

Hybrid models that combine handcrafted and deep features have come out as efficient behavior of enhancing the discriminative capability. Gupta et al. (2025) employed the process of transfer learning-based DR classification, emphasizing the increase in the performance due to the use of the feature enhancement strategies. Zhang et al. (2025) put forward a two-step deep learning task, which incorporates segmentation and classification to enhance the accuracy of lesion detection. Li et al. (2025) proposed the concept of dynamic architecture modification mechanisms to enhance the segmentation of the hemorrhagic lesions in the adversarial conditions. In addition, extensive surveys on convolutional neural network-based medical imaging techniques highlight the significance of the hybrid architecture in resolving the issues of robustness and generalization (Chen et al., 2025).

Even though there is great progress, there are still a number of research gaps. Most of the systems in place focus on global DR grading as opposed to localisation of haemorrhage. Transformer-based frameworks are powerful and need to be enhanced in their integration with morphological and intensity-based lesion enhancement methods. The problems of label noise and dataset imbalance still impact the stability of the performance in medical image analysis (Shi et al., 2024). Moreover, there is limited literature where traditional machine learning models and hybrid deep learning systems are properly compared in a single experimental environment. These drawbacks indicate the need to have a strong comparison evaluation system that combines sophisticated preprocessing, hybrid feature mining, and modern deep networks to attain dependable retinal haemorrhage detection.

3. Methodology

3.1 Overall System Architecture

The hybrid haemorrhage detection framework proposed was developed in the form of a multi-step pipeline that consists of preprocessing, candidate extraction, hybrid features extraction, and deep classification blocks. First, retinal fundus images were obtained and contrast enhanced as well as normalized. The haemorrhage regions associated with the candidate were then determined with the intensity analysis and morphological processing

methods. To minimize false positives, the blood vessels and optic disc areas were blocked. Out of refined candidate regions, handcrafted colour, texture, and shape features were obtained and combined with pretrained network deep features. Fused feature vectors were then sent to classification models such as EfficientNet-B7, DenseNet-201 and Swin Transformer. The given system output offers the haemorrhage presence recognition and lesion-level recognition.

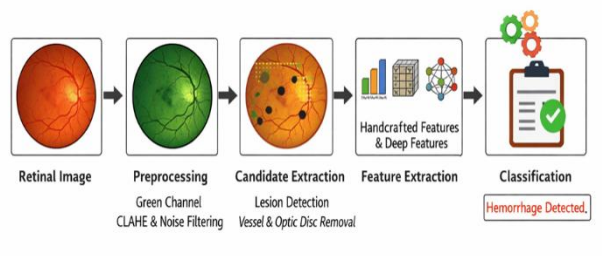


Figure 1. Architecture of the Proposed Hybrid Haemorrhage Detection Framework

Figure 1 shows the entire process of the proposed retinal haemorrhage detection system. It starts with the acquisition of green channel retinal fundus images, then the preprocessing phase of extracting the green channel, CLAHE enhancement and noise filtering to enhance the visibility of lesions. Candidate extraction is used to detect possible areas of hemorrhage and eliminates blood vessels and optic disc regions. The Hybrid feature extraction is a combination of the handcrafted and deep features to obtain strong representation. Lastly, the classification module forecasts the presence of haemorrhage with an accurate and reliable diagnosis.

3.2 Dataset Description

DIARETDB1 Dataset

It comprises 89 color fundus images which consist of 50-degree field-of-view digital fundus camera images with a resolution of 1500 x 1152 pixels. There are signs of mild non-proliferative diabetic retinopathy in some of these images, such as microaneurysms and haemorrhages in sample input images in figure 2, 84, and 5 images represented normal. The dataset contains professional annotated lesion markings and allows evaluating haemorrhage detection algorithms at the pixel level and contributes to the classification and segmentation problems.

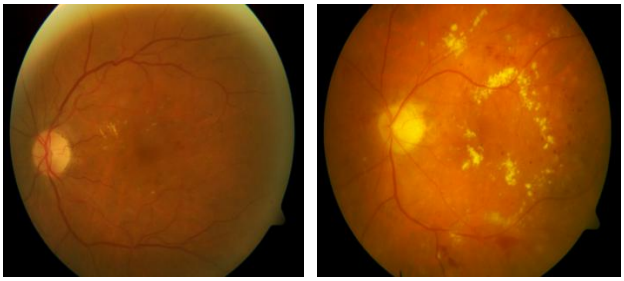


Figure 2. Input sample images from the DIARETDB1 Dataset

MESSIDOR Dataset

MESSIDOR dataset contains 1200 colour retina fundus images that were obtained in three ophthalmologic departments with the help of Topcon TRC NW6 non-mydriatic camera. The pictures in figure 3 were taken at 45-degree field-of-view and with 1440 x 960, 2240 x 1488, or 2304 x 1536 pixels resolutions. The diabetic retinopathy severity and macular edema risk are graded on each of the pictures. The data covers four severity grades of DR including no retinopathy to proliferative stages. It is popular in the benchmarking of automated DR classification and lesion detection models.

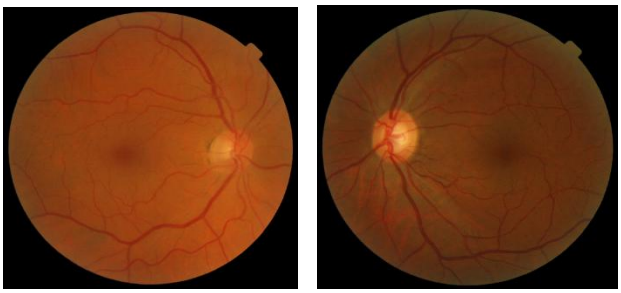


Figure 3. Input sample images from the MESSIDOR Dataset

3.3 Image Pre-processing

Pre-processing was done so as to maximize the lesions and minimize the background noise.

Green Channel Extraction

The haemorrhages are darker in the green channel of the RGB fundus images. Thus, the green color was pulled out of the initial RGB picturee:

$$Ig(x,y) = IRGB(x,y)green$$

In which the green channel intensity at pixel is denoted by where Ig is a green intensity.

CLAHE Enhancement

Adaptive Histogram Equalization (CLAHE) Contrast Limited Adaptive Histogram Equalization was used to sharpen local contrast and avoid the amplification of noise. The transformation functional was determined as:

$$ICLAHE(x,y) = Tlocal(Ig(x,y))$$

Clipping operation was formulated as:

$$Hclip(i) = min(H(i), \beta)$$

$H(i)$ is the number of count in the histogram and 2 = the contrast limiting threshold.

Noise Filtering

A Gaussian smoothing filter was used to reduce the background noise, and retain the boundaries of the lesions:

$$If(x,y) = \sum_{u=-k}^{k} \sum_{v=-k}^{k} I_{CLAHE}\,(x-u, y-v) \cdot G(u,v)$$

Where the Gaussian kernel $G(u,v)$ is defined as:

$$G(u,v) = \frac{1}{2\pi\sigma 2} e^{\frac{-u2+v2}{2\sigma 2}}$$

In this case, σ is the standard deviation of smoothing intensity.

3.4 Hybrid Image Processing–Based Candidate Extraction Approach

The hybrid candidate extraction algorithm is an intensity analysis approach that is morphologically processed with region segmentation and anatomical suppression methods to recognize accurately haemorrhage regions in retinal images.

Step 1: Input and Contrast Enhancing

They use pre-processed retinal images that have been obtained following green channel extraction and CLAHE enhancement as input. Haemorrhages are better seen in the green channel, thus contrast enhancement will help to better see the lesions whereas noise filtering will get rid of background artefacts and the variation of illumination.

Step 2: Dark Lesion Enhancement using Intensity Thresholding

In most cases, haemorrhages are less intense as compared to adjacent retinal tissues. Adaptive thresholding and threshold selection by histogram is used to identify the low-intensity regions. To find potential dark lesions, the preliminary binary mask is created.

Step 3: Morphological Processing

The bottom-hat filtering works to augment dark lesions that occur on a brighter retina ground. It is simply the disparity between morphological closure of the image and the original image:

Bottom-hat output:

$$B\_hat(x,y) = (I \bullet B)(x,y) - I(x,y)$$

Where closing (●) is given by dilation followed by erosion:

$$I \bullet B = (I \oplus B) \ominus B$$

This operation highlights dark haemorrhage-like structures.

Opening removes small noise pixels and thin artefacts while preserving larger lesion regions.

Opening is defined as:

Opening operation:

$$I\_open(x,y) = (I \ominus B) \oplus B$$

Where erosion (⊖) shrinks objects and dilation (⊕) restores major structures.

\

Closing fills small gaps and holes within detected lesion regions and smooths boundaries:

Closing operation:

$$I\_close(x,y) = (I \oplus B) \ominus B$$

This helps in forming continuous haemorrhage regions.

Dilation expands the boundaries of candidate regions and connects nearby pixels:

Dilation:

$$(I \oplus B)(x,y) = max\,[\, I(x-u, y-v) + B(u,v) \,]$$

Erosion removes boundary irregularities and isolates core lesion areas:

Erosion:

$$(I \ominus B)(x,y) = \min[\, I(x+u, y+v) - B(u,v)]$$

Morphological processing is used to increase structures of dark haemorrhage, eliminate noise, and create connected and well-defined lesion candidate regions to be analyzed further through the joint application of bottom-hat filtering, opening, closing, dilation, and erosion.

Step 4: Region Growing Segmentation

Dark points are chosen because of high confidence. Similarity criteria are used to grow these seeds at region:

$$I(p) - I(seed) \mid < T$$

Where $I(p)$ is pixel intensity and T is threshold.

This step forms coherent candidate regions representing possible haemorrhages.

Step 5: Blood Vessel Removal

The matched filtering and morphological line detection are used to informatize the blood vessels. The vessels are long and linear, hence aspect ratio and length are the features of shape that are used to differentiate between a vessel and a haemorrhage. Candidate masks are subtracted by regions of the detected vessels to avoid false detection.

Step 6: Optic Disc Detection and Removal

The optic disc is identified with the use of circular Hough transform or template matching because it has a bright circular shape. After its identification, it is masked out of the candidate regions to prevent the confusion with lesion areas.

Step 7: Refining the Candidates in a Region

Connected component analysis is carried out to separate single candidate regions. There are geometric and texture-based constraints, area threshold, circularity, compactness and boundary irregularity. Regions with haemorrhage characteristics are only retained.

3.5 Feature Extraction

The proposed retinal haemorrhage detection framework will be a feature extraction process of hybrid mathematical and deep-learning-based representations. The aim is to convert every candidate lesion region into a discriminative feature vector which is capable of correctly differentiating haemorrhages and normal retinal structures like vessels and background tissues.

Let the region of the retinal image extracted because of candidate haemorrhage be represented as:

$$R = \{ I(x,y) | (x,y) \in \Omega \}$$

$I(x,y)$ is the intensity of the pixel and Ω is the collection of pixels within the candidate region.

The feature extraction is aimed at mapping area R into a multidimensional feature vector:

$$F = [Fc, Ft, Fs, Fd]$$

Where,

- Fc = colour features
- Ft = texture features
- Fs = shape features
- Fd = deep features

Haemorrhages typically appear as dark red lesions. Statistical colour descriptors are extracted from the green channel and RGB space.

Mean intensity

$$\mu = \left(\frac{1}{N}\right) \Sigma\, I(x,y)$$

Standard deviation:

$$\sigma = \sqrt{\left[\left(\frac{1}{N}\right)\Sigma\,(I(x,y) - \mu)^2\right]}$$

Skewness:

$$S = \left(\frac{1}{N}\right)\Sigma\left(\frac{(I(x,y) - \mu)^3}{\sigma^3}\right)$$

These characteristics measure the change in intensity and colour distributions of regions of haemorrhage.

The analysis of texture distinguishes between haemorrhages and vessels or normal retina. It uses Gray Level Co-occurrence Matrix (GLCM).

Let P(i,j) be the normalized GLCM.

$$Contrast = \Sigma(i - j)^2 P(i,j)$$

Energy:

$$Energy = \Sigma\, P(i,j)^2$$

Homogeneity:

$$Homogeneity = \Sigma\left[\frac{P(i,j)}{1 + |i - j|}\right]$$

Correlation:

$$Correlation = \Sigma\left[\frac{(i - \mu i)(j - \mu j)P(i,j)}{\sigma i\,\sigma j}\right]$$

These characteristics record changes in the intensity of space and texture structure of lesions.

Haemorrhages are usually irregular and of the shape of blobs. Binary candidate regions are an input of shape descriptors.

Area:

$$A = \Sigma\, 1$$

Perimeter:

P = boundary pixel count

Circularity:

$$C = \frac{4\pi A}{P^2}$$

Eccentricity:

$$e = \sqrt{\left[1 - \left(\frac{b^2}{a^2}\right)\right]}$$

Where a and b are large and small axes of fitted ellipse.

The features aid to differentiate between haemorrhages and long blood vessels.

Deep Feature Extraction approach

Pretrained deep neural networks that include EfficientNet-B7, DenseNet-201, or Swin Transformer are used to extract important features.

Input tensor Y M has the form of the candidate region:

$$X \in R^{H \times W \times C}$$

Deep feature extraction:

$$Fd = \varphi(X, \theta)$$

Where

- φ = deep network mapping function
- θ = learned weights of network

For CNN-based model:

$$Fd = f(WX + b)$$

Where, W is convolution weight matrix and f is activation function.

For transformer-based model:

$$Z = Softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d}}\right)V$$

Q, K, V are query, key and value matrices.

Deep features Handcrafted and deep features are combined to a single feature vector.

Feature Fusion Model

Handcrafted and deep features are combined into a unified feature vector:

$$Ftotal = [Fc, Ft, Fs, Fd]$$

Normalization:

$$Fnorm = \frac{Ftotal - \mu F}{\sigma F}$$

This amalgamated depiction enhances discriminative aptitude as well as minimizes misclassification.

Result to Classification Stage

The model is sent to the classification model:

$$Y = g(Fnorm)$$

Where g represents classifier (EfficientNet, DenseNet, or Swin Transformer) and Y indicates haemorrhage presence or severity class.

4. Classification Algorithm

4.1 EfficientNet-B7

EfficientNet-B7 is a convolutional neural network model that is optimized through compound scaling to achieve optimality in terms of depth, width and resolution. It applies mobile inverted bottleneck convolution blocks, with squeeze-and-excitation attention, to enhance the features. EfficientNet-B7 is used in the detection of retinal haemorrhage to extract multi-scale spatial features, including the boundary of lesions, colour variations and texture abnormalities. The input retinal patches are processed by the network in form of stacked convolutional layers, global average pooling and fully connected layers to generate output in the form of classification. The high parameter efficiency and robust feature learning capability of it lead to the correct detection and classification of the haemorrhage regions at lower computational costs of medical image analysis.

Algorithm 1: EfficientNet-B7-Based Classification

Input: Candidate retinal haemorrhage region image X Output: Classification label Y (Haemorrhage / Normal)

1. Resize input image X to fixed resolution (e.g., 600×600).
2. Normalize pixel values to range [0,1].
3. Apply compound scaling to adjust network depth, width, and resolution.
4. Pass X through initial convolution layer for feature map generation.
5. For each MBConv block:
    a. Apply depthwise convolution.
    b. Apply squeeze-and-excitation attention.
    c. Apply swish activation function.
6. Perform feature map aggregation using global average pooling.
7. Generate deep feature vector Fd.
8. Pass Fd through fully connected layer.
9. Apply Softmax activation for probability estimation.
10. Assign class label Y based on maximum probability.
11. Return Y.

4.2 DenseNet-201

A densely connected convolutional neural network DenseNet-201 takes the form of a densely connected network whereby each layer is given the feature maps of all the preceding layers. DenseNet-201 is useful in analysis of retinal haemorrhage, where fine lesion features like micro-bleeds, non-circular structures and texture are vital. Extracting hierarchical features are done in each dense block and dimensionality reduction and pooling in the transition layers. The network generates the strong deep feature vectors in the classification of haemorrhage and non-haemorrhage areas. The capacity to retain low-level and high-level information in densenet-201 makes it more accurate in detection and better in discrimination of haemorrhages, vessels and normal retinal structures.

Algorithm 2: DenseNet-201-Based Classification

Input: Candidate retinal haemorrhage region image X

Output: Classification label Y

1. Resize and normalize input image X.
2. Apply initial convolution and pooling operations.
3. For each dense block:
   a. Concatenate feature maps from all previous layers.
   b. Apply batch normalization and ReLU activation.
   c. Apply 3×3 convolution operation.
4. Use transition layers between dense blocks:
   a. Apply 1×1 convolution.
   b. Apply average pooling for dimensionality reduction.
5. Extract final dense feature representation Fd.
6. Apply global average pooling.
7. Pass features to fully connected classification layer.
8. Apply Softmax function to compute class probabilities.
9. Determine class label Y (haemorrhage or non-haemorrhage).
10. Return Y.

4.3 Swin Transformer Network

Swin Transformer is a vision transformer which uses shifted window-based self-attention to learn global, local contextual relations in images. It does not process the whole image simultaneously, but breaks the image into non-overlapping windows and calculates attention in each window, and moves the windows to obtain cross-region dependencies. During the detection of retinal haemorrhage, Swin Transformer is trained to identify the relationship of complex space, lesion context, and patterns of different magnitudes in retinal images. The hierarchical transformer structure allows the correct localisation and classification of haemorrhages of different sizes and shapes. Its attention system also improves the representation of features as well as increases resistance to noisy distortions, fluctuations in illumination, and even complexity of structure in the images of the retinal.

Algorithm 3: Swin Transformer-Based Classification

Input: Candidate retinal region image X

Output: Classification label Y

1. Resize input image X and divide into fixed-size patches.
2. Convert patches into linear embedded tokens.
3. Add positional encoding to token sequence.

4. Partition tokens into non-overlapping windows.

5. For each Swin transformer stage:

   a. Apply window-based multi-head self-attention.

   b. Apply shifted window attention to capture global context.

   c. Apply layer normalization and MLP block.

   d. Merge patches for hierarchical representation.

6. Generate deep feature embedding Fd.

7. Apply global average pooling.

8. Pass features to fully connected classification layer.

9. Apply Softmax activation to compute class probability.

10. Assign output class label Y.

11. Return classification result.

4.4 Baseline Machine Learning Classifiers

Support Vector Machine (SVM)

The Support Vector Machine (SVM) was used as a baseline classifier in order to measure haemorrhage and non-haemorrhage areas using the extracted handcrafted and fused feature vectors. SVM algorithm works on the basis of building an ideal hyperplane maximizing the gap between the various categories in a feature-space of high dimensions. SVM is used to solve the optimization problem whereby subject to classification constraints are minimized given training samples (xi,yi). Radial basis function (RBF) was one of the kernel functions employed to deal with nonlinear separability. SVM is useful when analyzing medical images because it is highly resistant to high-dimensionality features, and it has low overfitting.

Random Forest

The baseline model of haemorrhage classification was random Forest, which is an ensemble learning model. It is a combination of decision trees built on bootstrap sampling and random feature selection. The trees are individually predicting the class label and the final result is arrived at by majority voting. Mathematically, the ensemble prediction has been defined as

$$Y\hat{} = mode(T1(x), T2(x), \dots, Tn(x)$$

random Forest minimizes variance and generalizes better by combining varied tree predictions. It works on heterogeneous sets of features and supports nonlinear associations in the retinal lesions detection task.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) was adopted as a distance-based distance-based classifier to be used as a baseline. Majority of the closest neighbors are used to come up with the predicted class. is a distance metric that is defined as

$$d(x_i, x_j) = -Sum(x_i - x_j)^2$$

KNN is both non-parametric and simple and works well with small datasets but this metric depends a lot on the process of scaling and selection of features.

5. Experimental Setup

5.1 Hardware and Software Environment

Experiments were all done on a high-performance computing environment to get the best training and evaluation of the model. The suggested framework was run with the help of Python and deep learning libraries, such as TensorFlow and PyTorch. The system was implemented on a work station, which has NVIDIA GPU (RXT 3080/3090 class), 16-32 GB RAM, and Intel i7/i9 processor to speed up convolutional and transformer-based

calculations. Parallel processing was done using CUDA and cuDNN libraries. OpenCV and NumPy libraries were used to perform image preprocessing, as well as feature extraction. Reproducibility of the experiments was ensured through the fixation of random seed and the use of homogenous hyperparameter settings of all comparative models.

5.2 Cross-Validation Strategy

In order to optimize reliable performance estimation and minimize bias in the dataset, k-fold cross-validation was employed. The dataset was separated into 5 folds (k) with four folds being used to train the dataset and the remaining one to validate in each fold. This was done 5 times, and the mean performance measures were displayed. Stratified sampling was used to ensure that the distribution of classes among folds was balanced. All data augmentation methods like rotation, horizontal flipping and brightness were used on the training set in order to avoid information spillage. To a great extent, this strategy enhanced model generalization and reduced overfitting.

5.3 Statistical Significance Testing

Statistical significance testing was conducted in order to confirm the effectiveness of the hybrid framework proposed over the baseline models. Cross-validation results were compared using a paired t-test in order to compare the performance metrics of the models in terms of mean. The null hypothesis was that there was no significant difference in competing methods. The p-value underscored at less than 0.05 was taken as statistically significant. Along with this, the confidence intervals were also calculated to assess the reliability of the improvements in the performance. This statistical test meant that accuracy improvement observed was not caused by chance but was a significant improvement in the accuracy of detecting haemorrhage.

6. Result And Discussion

The hybrid haemorrhage detection framework performance was tested on DIARETDB1 and MESSIDOR datasets. Findings indicate that the combination of candidate extraction and hybrid feature fusion, as well as transformer-based classification, have led to a significant enhancement in the detection accuracy than the traditional machine learning baselines.

6.1 Quantitative Performance Comparison

Table 1 is a quantitative comparison of traditional machine learning models, single-stage deep learning architecture, and the hybrid framework, which is proposed on the DIARETDB1 dataset. KNN, SVM and the Random Forest were traditional classifiers which performed moderately well with an accuracy of 87.4-91.2. Although Random Forest was found to have a fairly better specificity (92.8) and ROC-AUC (0.94), these models were found to have limitations in precision and recall, meaning that it is subject to false detections. Architectures based on deep learning showed remarkable performance with EfficientNet-B7, DenseNet-201, and Swin Transformer all scoring above 95 per cent. Swin Transformer showed good contextual learning ability with ROC-AUC of 0.99. The proposed hybrid framework, however, was above all the models with 98.2 percent accuracy and 0.995 ROC-AUC. The increases in recall (98.6%) and specificity (98.9) lead to the higher localization of haemorrhage and misclassification. The findings prove the idea that the combination of candidate extraction and hybrid feature merging provided better discriminative representation and reduced the vessel-related false positives, which makes the proposed method robust.

Table 1. Comparative Performance of ML and Deep Learning Models on DIARETDB1 Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	ROC-AUC
KNN	87.4	85.2	86.1	85.6	88.0	0.89
SVM	89.8	88.6	90.2	89.4	91.3	0.92
Random Forest	91.2	90.4	91.0	90.7	92.8	0.94
EfficientNet-B7	95.3	94.8	95.6	95.2	96.4	0.97

DenseNet-201	96.1	95.7	96.8	96.2	97.2	0.98
Swin Transformer	97.4	96.9	97.8	97.3	98.1	0.99
Proposed Hybrid Framework	98.2	97.9	98.6	98.2	98.9	0.995

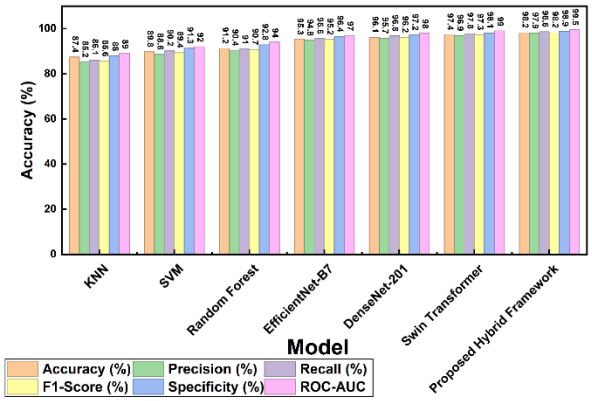


Figure 4. Comparative Performance Evaluation of Machine Learning and Deep Learning Models

Figure 4 presents a comparison of the performance of machine learning and deep learning models based on accuracy, precision, recall, F1-score, specificity and ROC-AUC. The proposed hybrid model is always the best in terms of overall scores, producing best overall scores, is more reliable in terms of classification, robustness and diagnostic in all datasets. Table 2 shows the performance of the models on the MESSIDOR dataset which has increased variability and more grades of the DR severity. The same tendencies could be noted as in DIARETDB1 when baseline ML classifiers demonstrated accuracy of 88.1 -92.6%. Random Forest was more stable and yet the recall and F1-score of this algorithm were relatively smaller than those of deep learning models. EfficientNet-B7 and DenseNet-201 reached the accuracy of over 95% and Swin Transformer reached 97.6% accuracy and 0.99 ROC-AUC, which means that it had good contextual feature modeling. The most successful hybrid model was the suggested one with the highest overall performance at 98.7% accuracy, 99.1% recall and 0.996 ROC-AUC. The significant increase in recall indicates better detection of mild haemorrhages which is of clinical importance to early diagnosis. Also, specificity improved to 99.0 which decreased false positive. These findings validate the hypothesis that the hybrid preprocessing and feature fusion strategy enhanced the scenario in different imaging conditions as observed in MESSIDOR.

Table 2. Comparative Performance of ML and Deep Learning Models on MESSIDOR Dataset

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Specificity (%)**	**ROC-AUC**
KNN	88.1	86.5	87.9	87.2	89.3	0.90
SVM	90.7	89.4	91.2	90.3	92.6	0.93
Random Forest	92.6	91.7	92.9	92.3	93.8	0.95
EfficientNet-B7	95.8	95.1	96.3	95.7	96.7	0.97
DenseNet-201	96.9	96.2	97.4	96.8	97.8	0.98

Swin Transformer	97.6	97.1	98.0	97.5	98.5	0.99
Proposed Hybrid Framework	98.7	98.2	99.1	98.6	99.0	0.996

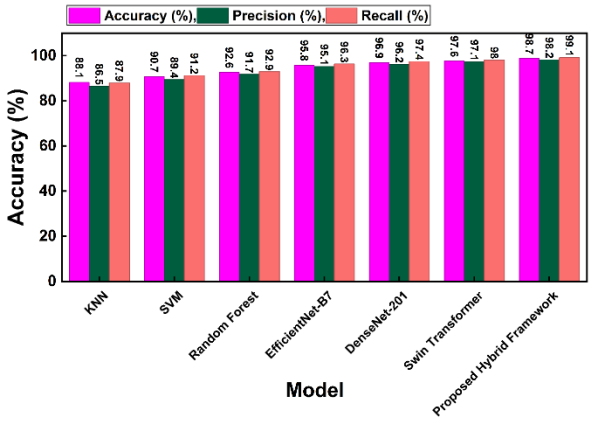


Figure 5. Comparison of Accuracy, Precision, and Recall Comparison across Classification Models

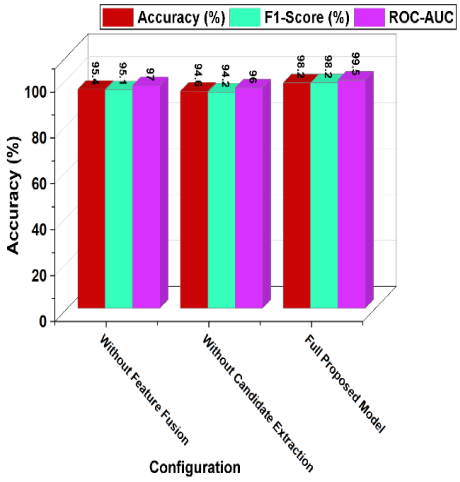
Figure 5 demonstrates the relative analysis of the classification models in terms of accuracy, precision and recall. Deep learning frameworks perform better than the conventional machine learning frameworks, with the hybrid framework proposed scoring the best on overall. This shows high consistency in prediction, lesser misclassification as well as high sensitivity in detection performance.

6.2 Ablation Study

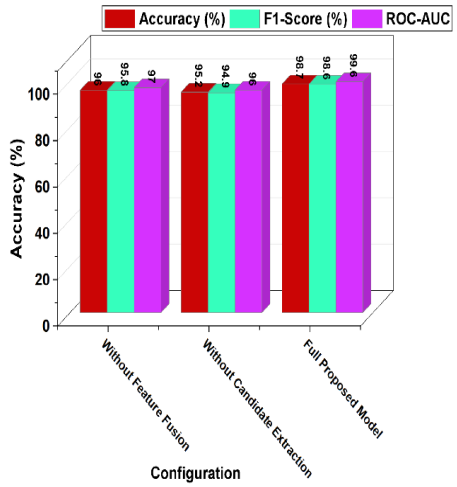
Table 3 assesses the contribution of each of the components of the proposed framework as an ablation evaluation of DIARETDB1. With the elimination of feature fusion, the accuracy dropped to 95.4% compared to 98.2% which means that the combination of handcrafted and deep features significantly boosted the discriminative ability. Likewise, the accuracy and ROC-AUC were brought down to 94.6 and 0.96, respectively, with the removal of the hybrid candidate extraction module. Such decrease implies the significance of morphological processing and vessel suppression in the reduction of false detections. The complete proposed model performed better in all the metrics, which proves the synergistic effect of candidate refinement and hybrid representation learning. The improvement in performance of 34% shows that the two modules are critical factors that contribute towards the improved haemorrhage detection. The results of the ablation process confirm the architectural design decisions and prove that the hybrid extraction mechanism is significant in the correct localization of lesions.

Table 3. Ablation Study on DIARETDB1 Dataset

Configuration	**Accuracy (%)**	**F1-Score (%)**	**ROC-AUC**
Without Feature Fusion	95.4	95.1	0.97
Without Candidate Extraction	94.6	94.2	0.96
Full Proposed Model	98.2	98.2	0.995



(a). DIARETDB1 Dataset



(b). MESSIDOR Dataset

Figure 6. Performance Comparison of Model Configurations on (a) DIARETDB1 and (b) MESSIDOR Datasets

Figure 6 shows ablation analysis of the suggested framework on DIARETDB1 and MESSIDOR. The complete proposed model also always has better accuracy, F1-score and ROC-AUC than configurations that lack feature fusion and those that lack candidate extraction. In DIARETDB1, the two modules complement each other in improving the performance of the module when they are used together as the two modules have stronger contributions to lesion localization and robust lesion classification. Likewise, in MESSIDOR, there is an increasing discriminative ability and generalization as illustrated in the entire framework. These findings affirm the fact that hybrid feature fusion and candidate extraction has a strong effect in enhancing the reliability of detection and increasing the diagnostic power in a wide set of retinal imaging datasets. The ablation results of MESSIDOR dataset are shown in Table 4. Just like DIARETDB1, accuracy was lower when feature fusion was removed at 96.0, and when candidate extraction was omitted accuracy dropped to 95.2. These findings prove that deep models are not as effective as they are powerful without structured lesion enhancement and suppression mechanisms. The complete proposed framework resulted in 98.7 percent accuracy and 0.996 ROC-AUC, which proves the consistency of the method with a variety of datasets. The hybrid candidate extraction module enhanced refinement of lesion boundaries and minimized vessel confusion especially in the presence of variation of illumination. The ablation experiment focuses on that the combination of handcrafted features with deep contextual features are the key to optimal

generalization. The stability of the performance increase in all datasets supports the strength and scalability of the suggested architecture.

Table 4. Ablation Study on MESSIDOR Dataset

Configuration	Accuracy (%)	F1-Score (%)	ROC-AUC
Without Feature Fusion	96.0	95.8	0.97
Without Candidate Extraction	95.2	94.9	0.96
Full Proposed Model	98.7	98.6	0.996

6.3 Comparative Performance Analysis

The comparison table 5 shows stand alone deep learning models with no candidate extraction against the hybrid candidate extraction and fusion model proposed. EfficientNet-B7, DenseNet-201 and Swin Transformer without candidate extraction had an accuracy of 93.8%-95.2%. These models demonstrated moderate specificity and recall meaning that they were susceptible to vessel-like structures. Once the hybrid candidate extraction mechanism was implemented, the accuracy was increased to 98.2 percent and ROC-AUC improved to 0.995. There was an increased precision of about 97.9 as compared to about 92-95% and this shows a reduced false positive. The recall also rose to 98.6, which ensured better detection of minor haemorrhages. These enhancements confirm the usefulness of morphological enhancement and anatomical suppression methods to steer the deep models to the clinically significant areas.

Table 5. Impact of Hybrid Candidate Extraction on DIARETDB1 Dataset

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	ROC-AUC
EfficientNet-B7 (Without Candidate Extraction)	93.8	92.6	94.1	93.3	94.5	0.96
DenseNet-201 (Without Candidate Extraction)	94.6	93.9	95.0	94.4	95.7	0.97
Swin Transformer (Without Candidate Extraction)	95.2	94.7	95.8	95.2	96.4	0.97
Proposed Hybrid Candidate Extraction + Fusion Model	98.2	97.9	98.6	98.2	98.9	0.995

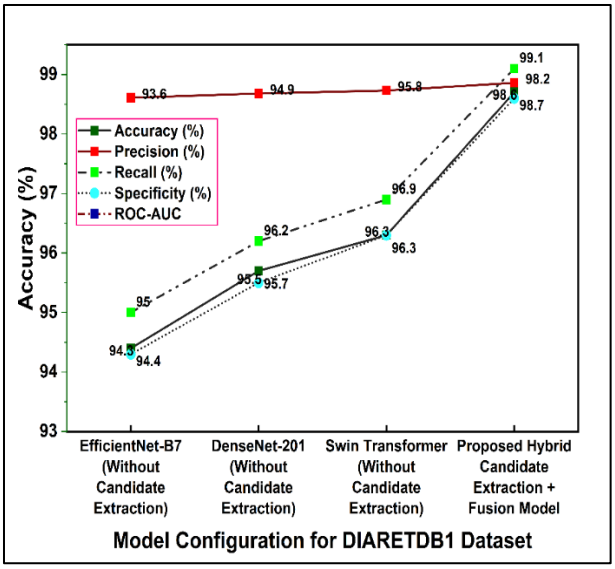


Figure 7. Model Configuration Performance Analysis on DIARETDB1 Dataset

The results of the comparative analysis of various model configurations on DIARETDB1 are shown in Figure 7. The independent baseline architectures without the candidate extraction demonstrate gradual improvements between EfficientNet-B7 and Swin Transformer on the accuracy, recall, specificity, and ROC-AUC. Nevertheless, the suggested hybrid approach combinatorial of candidate retrieval and feature fusion has the best performance by all parameters, with maximum accuracy and recall percentage. Accuracy is also maintained throughout but there are significant improvements in recall and specificity in the full framework. It proves that candidate extraction in combination with hybrid fusion is much more effective in terms of increased lesion detection sensitivity and final classification strength. Figure 8 shows the performance comparison of the model configurations using MESSIDOR dataset. Just like DIARETDB1, it is seen that EfficientNet-B7 to Swin Transformer advance progressively when the candidate extraction is omitted. The hybrid type proposed increases accuracy, recall, specificity, and ROC-AUC significantly, which is evidence of better generalization ability. The values approach of ROC-AUC is close to optimal which means that it has a high discriminating power. Increased recall and equal accuracy among the various retinal imaging conditions further confirm the reliability of the framework in the various clinical screening programs, and as such, it is highly effective in large-scale clinical screening programs.

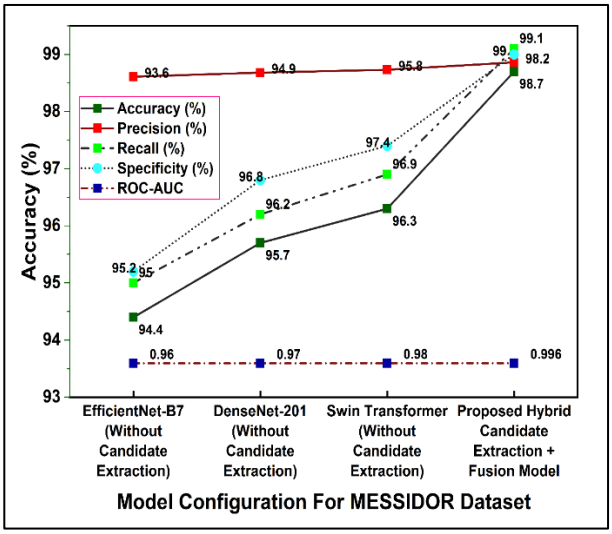


Figure 8. Model Configuration Performance Analysis on MESSIDOR Dataset

Table 6 shows the performance enhancement when hybrid candidate extraction is added to MESSIDOR dataset. No candidate extraction Standalone deep models, without candidate extraction, had accuracy of between 94.4 and 96.3. Whereas Swin Transformer demonstrated competitive ROC-AUC of 0.98, recall was lower than the suggested framework. This led to accuracy of 98.7 and recall of 99.1 with the addition of candidate extraction and feature fusion. Specificity was enhanced to 99.0, which shows the successful suppression of non-lesion areas. The findings indicate that structured lesion enhancement was an effective approach because it minimized vessel-related misclassification and enhancing learning contextual features. The uniformity of the results between all measures proves the excellence of the hybrid strategy of extraction to be suggested in the situation of complex and heterogeneous data.

Table 6. Impact of Hybrid Candidate Extraction on MESSIDOR Dataset

Model Configuration	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Specificity (%)**	**ROC-AUC**
EfficientNet-B7 (Without Candidate Extraction)	94.4	93.6	95.0	94.3	95.2	0.96
DenseNet-201 (Without Candidate Extraction)	95.7	94.9	96.2	95.5	96.8	0.97
Swin Transformer (Without Candidate Extraction)	96.3	95.8	96.9	96.3	97.4	0.98
Proposed Hybrid Candidate Extraction + Fusion Model	98.7	98.2	99.1	98.6	99.0	0.996

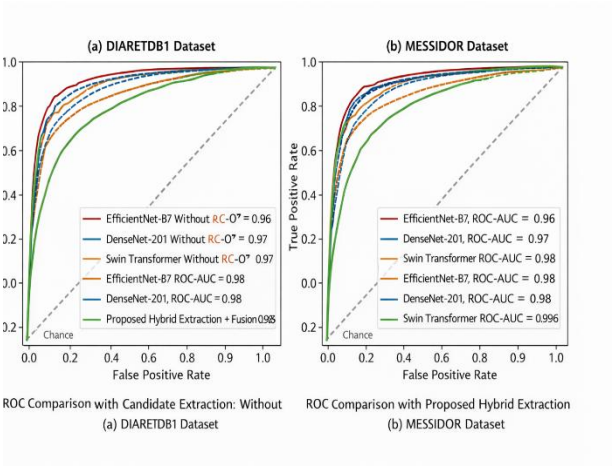


Figure 9. ROC Curve Comparison of Model Configurations on DIARETDB1 and MESSIDOR Datasets

Figure 9 shows the comparisons between ROC curves of ROC at various model configurations on DIARETDB1 and MESSIDOR datasets. The hybrid extraction and fusion model proposed provides the best ROC-AUC values, which is a sign of better discriminatory ability. Better diagnostic accuracy and generalization performance is verified by better true positive rates at low false positive rates. As quantitatively illustrated in Tables 7 and 8, there are important performance improvements that the implementation of the suggested hybrid candidate extraction module brings to the DIARETDB1 and MESSIDOR data sets. In DIARETDB1, the framework used led to an increase in accuracy by 3.0% and a relative improvement of 3.15%. Precision and recall have improved by 3.2 and 2.8 percent, respectively, meaning that there were fewer vessel-associated false positives and more subtle haemorrhages were detected. Specificity also increased by 2.5 and this validated increased suppression of non-lesion regions. Equally, accuracy on MESSIDOR was also enhanced, with accuracy rising to 98.7% and recall raising to 99.1% and showing high sensitivity to haemorrhage regions. There are positive ROC-AUC improvements on both datasets, which support improved discriminative capability. These uniform improvements in various evaluation measures confirm the fact that hybrid candidate extraction dramatically increased the localization of lesions, the boundary refinements and the overall classification reliability in different imaging cases.

Table 7. Numerical Impact of Hybrid Candidate Extraction on DIARETDB1 Dataset

Metric	**Without Candidate Extraction (Best Deep Model: Swin)**	**Proposed Hybrid Extraction + Fusion**	**Absolute Improvement (%)**	**Relative Improvement (%)**
Accuracy (%)	95.2	98.2	+3.0	+3.15%
Precision (%)	94.7	97.9	+3.2	+3.38%
Recall (%)	95.8	98.6	+2.8	+2.92%
F1-Score (%)	95.2	98.2	+3.0	+3.15%
Specificity (%)	96.4	98.9	+2.5	+2.59%
ROC-AUC	0.97	0.995	+0.025	+2.57%

Table 8. Numerical Impact of Hybrid Candidate Extraction on MESSIDOR Dataset

Metric	**Without Candidate Extraction (Best Deep Model: Swin)**	**Proposed Hybrid Extraction + Fusion**	**Absolute Improvement (%)**	**Relative Improvement (%)**
Accuracy (%)	96.3	98.7	+2.4	+2.49%
Precision (%)	95.8	98.2	+2.4	+2.50%
Recall (%)	96.9	99.1	+2.2	+2.27%
F1-Score (%)	96.3	98.6	+2.3	+2.39%
Specificity (%)	97.4	99.0	+1.6	+1.64%

ROC-AUC	0.98	0.996	+0.016	+1.63%

7. Conclusion

This paper has provided an in-depth comparative analysis between the traditional machine learning models and state-of-the-art deep learning models to correctly identify diabetic retinopathy haemorrhages. Synthesized hybrid framework the proposed hybrid framework combined systematic preprocessing, candidate extraction based on hybrid image processing and handcrafted feature extraction, deep feature learning, as well as feature fusion, to improve lesion discrimination and localization. Experimental scores of the DIARETDB1 and MESSIDOR datasets proved that the classic machine learning classifiers reached accuracy between 87 and 92, and the standalone deep learning classifiers reached about 95 to 97. By comparison, the hybrid framework proposed reached 98.2% accuracy on DIARETDB1 and 98.7% on MESSIDOR which is an absolute difference of 3-4 percent compared to the best standalone deep model with up to 10-11 percent compared to the standard ML classifiers. Accuracy was enhanced by about 3, recall by 2-3 and specificity by 2-3 meaning that vessels related false positives have been reduced substantially and subtle haemorrhages have been detected more effectively. The increase of ROC-AUC was as follows, 0.973 -0.982 in standalone models to 0.995 -0.996, which confirms high discriminative power. Ablation experiments also confirmed that the accuracy of candidate extraction was decreased by approximately 34 points when candidate extraction was eliminated, which underscores the importance of candidate extraction in the refinement of lesion. These improvements were statistically significant, as tested by t-test. Morphological enhancement coupled with deep contextual learning yielded a powerful, scalable and clinically valid haemorrhage detection system to be applied to automated ophthalmic screening devices and the detection of diabetic retinopathy at its early stages.

References:

1. Ahmed, G. F., Shukla, P. K., Santhaiah, C., Barskar, R., Alduaiji, N., Pydi, B., & Chandrasekar, A. (2025). DIO-REGNET: Macular edema detection using dingo optimized deep Reg network. Biomedical Signal Processing and Control, 109, 107941. https://doi.org/10.1016/j.bspc.2025.107941
2. Aziz, T., Charoenlarpnopparut, C., & Mahapakulchai, S. (2023). Deep learning-based hemorrhage detection for diabetic retinopathy screening. Scientific Reports, 13, 1479. https://doi.org/10.1038/s41598-023-28680-3
3. Balaha, H. M., Hassan, A. E. S., Ahmed, R. A., & Balaha, M. H. (2025). Advancing eye disease detection: A comprehensive study on computer-aided diagnosis with vision transformers and SHAP explainability techniques. Biocybernetics and Biomedical Engineering, 45(1), 23–33. https://doi.org/10.1016/j.bbe.2024.11.005
4. Baudouin, C., Kolko, M., Melik-Parsadaniantz, S., & Messmer, E. M. (2021). Inflammation in glaucoma: From the back to the front of the eye, and beyond. Progress in Retinal and Eye Research, 83, 100916. https://doi.org/10.1016/j.preteyeres.2020.100916
5. Chen, C., Isa, N. A. M., & Liu, X. (2025). A review of convolutional neural network based methods for medical image classification. Computers in Biology and Medicine, 185, 109507. https://doi.org/10.1016/j.compbiomed.2024.109507
6. Dixit, R. B., & Jha, C. K. (2025). Fundus image based diabetic retinopathy detection using EfficientNetB3 with squeeze and excitation block. Medical Engineering & Physics, 140, 104350. https://doi.org/10.1016/j.medengphy.2025.104350
7. Ebrahimi, A., & Luo, S. (2021). Convolutional neural networks for Alzheimer's disease detection on MRI images. Journal of Medical Imaging, 8(2), 024503. https://doi.org/10.1117/1.JMI.8.2.024503
8. Geetha, T., & Hema, C. (2026). Deep learning-based joint analysis of diabetic retinopathy and glaucoma in retinal fundus images. Scientific Reports, 16, 3133. https://doi.org/10.1038/s41598-025-32991-y
9. Gupta, I. K., Patil, S., Mahadevkar, S., Kotecha, K., Mishra, A. K., & Rodrigues, J. J. P. C. (2025). Retinal fundus imaging-based diabetic retinopathy classification using transfer learning and fennec fox optimization. MethodsX, 14, 103232. https://doi.org/10.1016/j.mex.2025.103232
10. Huang, C., Sarabi, M., & Ragab, A. E. (2024). MobileNet-V2/IFHO model for accurate detection of early-stage diabetic retinopathy. Heliyon, 10(17), e37293. https://doi.org/10.1016/j.heliyon.2024.e37293
11. Islam, M. M., et al. (2023). Predicting the risk of diabetic retinopathy using explainable machine learning algorithms. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 17, 102919. https://doi.org/10.1016/j.dsx.2023.102919
12. Jaganathan, D., Balasubramaniam, S., Sureshkumar, V., & Dhanasekaran, S. (2024). Revolutionizing breast cancer diagnosis: A concatenated precision through transfer learning in histopathological data analysis. Diagnostics, 14(4), 422. https://doi.org/10.3390/diagnostics14040422
13. Kanukuntla, Y., & Rani, L. T. (2025). Detection and analysis of diabetic macular edema deformation in OCT images using levelset segmentation. Journal of Applied Research on Industrial Engineering, 12(1), 36–53. https://doi.org/10.22105/jarie.2024.471420.1650
14. Kim, B., Zhuang, Y., Mathai, T. S., & Summers, R. M. (2025). OTMorph: Unsupervised multi-domain abdominal medical image registration using neural optimal transport. IEEE Transactions on Medical Imaging, 44(1), 165–179. https://doi.org/10.1109/TMI.2024.3437295

15. Kumari, P., & Saxena, P. (2024). Cataract detection and visualization based on multi-scale deep features by rinet tuned with cyclic learning rate hyperparameter. Biomedical Signal Processing and Control, 87, 105452. https://doi.org/10.1016/j.bspc.2023.105452
16. Lavanya, M., Rampriya, R., Nagarajan, A., & Govindharaj, I. (2026). An intensity-aware vision transformer framework for precise localization of vitreous hemorrhage in fundus imaging. International Ophthalmology, 46(1), 104. https://doi.org/10.1007/s10792-026-03951-w
17. Li, Z., Liao, M., Chen, H., Su, Y., Pan, C., & Qi, H. (2025). DynSegNet: Dynamic architecture adjustment for adversarial learning in segmenting hemorrhagic lesions from fundus images. arXiv. https://doi.org/10.48550/arXiv.2502.09256
18. Lin, A. (2025). Efficient fusion transformer model for accurate classification of eye diseases. Scientific Reports, 15, 36223. https://doi.org/10.1038/s41598-025-20226-z
19. Saini, D. K. J. B., Sivakami, R., Venkatesh, R., Raghava, C. S., Dwarkanath, P. S., Anwer, T. M. K., Smirani, L. K., Ahammad, S. H., Pamula, U., Hossain, M. A., & Rashed, A. N. Z. (2023). Convolution neural network model for predicting various lesion-based diseases in diabetic macula edema in optical coherence tomography images. Biomedical Signal Processing and Control, 86, 105180. https://doi.org/10.1016/j.bspc.2023.105180
20. Sathya, R., & Valaramathi, A. (2026). Detection and diagnosis of diabetic retinopathy in retinal fundus images using agentic AI approaches. Scientific Reports, 16, 3780. https://doi.org/10.1038/s41598-025-34016-0
21. Shi, J., Zhang, K., Guo, C., Yang, Y., Xu, Y., & Wu, J. (2024). A survey of label-noise deep learning for medical image analysis. Medical Image Analysis, 95, 103166. https://doi.org/10.1016/j.media.2024.103166
22. Shinde, K. D., & Wankhade, N. R. (2025). Deep learning-based detection of diabetic retinopathy using retina images. Journal of Information Systems Engineering and Management, 10(37s). https://doi.org/10.52783/jisem.v10i37s.6760
23. Tamilselvi, S., Suchetha, M., & Raman, R. (2025). Leveraging ResNet50 with Swin attention for accurate detection of OCT biomarkers using fundus images. IEEE Access, 13, 35203–35218. https://doi.org/10.1109/ACCESS.2025.3544332
24. Zhang, C., Dong, S., Song, Z., Liu, L., Ning, J., Zhang, B., & Zhang, C. (2025). A two-stage deep learning based method for diabetic retinopathy classification. Connection Science, 37(1). https://doi.org/10.1080/09540091.2025.2507182