# DR-MobiCB-Onto: A Modular Framework for Lesion-Level Diabetic Retinopathy Detection and Ontology-Driven Semantic Reasoning

*Ali* Alkwzahy[1*]

[1]Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq

**Abstract.** Diabetic Retinopathy (DR) is a visual impairment caused by long-term diabetes that damages the blood vessels in the retina. It is one of the main causes of blindness worldwide, and early detection is very important to prevent progression. Therefore, existing Convolutional Neural Network (CNN)-based models have improved lesion detection, but lack semantic interpretability. While ontology-driven systems enable structured reasoning, they depend on precise lesion-level inputs, and bridging these paradigms enhances diagnostic reliability. Existing DR models have limitations in lesion-level accuracy, vessel segmentation, and semantic interpretation. To address these limitations, this research proposes a unified framework, namely, DR-MobiCB-Onto model comprising a modular DR detection pipeline integrating MobileNetV3 and Convolutional Block Attention Module (CBAM), namely DR-MobiCB with domain-specific ontology-driven reasoning for efficient lesion detection and semantic interpretation. Preprocessing to enhance the image includes bilateral filtering, CLAHEU, and Z-score normalization. The Extended Adaptive Density-Based Spatial Clustering (EADBSC) method is used to segment the Thick Blood Vessels (TBVs) in the enhanced image, and finally, lesions were detected using DR-MobiCB with dilated convolutions. Evaluation was performed on the Messidor and IDRiR datasets, obtaining 97.4% and 96.8% accuracy and AUC scores of 0.987 and 0.981, respectively.

## 1 Introduction

Diabetes is a disease characterized by continuing metabolic syndrome, and it disrupts the ability of the body to supply the hormone insulin, often leading to systemic complications [1][2]. Diabetic Retinopathy (DR) involves microvascular damage in both type 1 and type 2 diabetes [3], inflammation, and neurodegeneration in the retina, and remains one of the leading causes of limited adult vision globally. According to the 2019 report from the International Agency for the Prevention of Blindness (IAPB), approximately 463 million people worldwide are affected, equating to approximately 1 in 11 individuals with limited vision and diabetic retinopathy [4][5][6]. Early diagnosis is crucial to avoid irreparable vision loss; however, the current diagnostic approach is based on centralized imaging datasets, which implies privacy issues in the context of regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) [6].

These methods depend on the physical clinical examination fundus images, where eye specialists check features like the cup-to-disc ratio, rim thickness, and structural changes. However, this process is subjective, slow, and needs expert skill, which can sometimes cause misdiagnosis [7]. Recent developments in machine learning and medical imaging have transformed DR diagnosis through automated systems that utilize deep learning, image segmentation, and feature extraction, thereby improving detection speed and diagnostic precision [8]. However, these techniques usually need a lot of labeled data and powerful computing, and they don't always work well across different imaging environments, making them less accessible in real-world settings [9]. Existing DR models have limitations in lesion-level accuracy, vessel segmentation, and semantic interpretation. To address these limitations, this research proposes a unified framework, namely, the DR-MobiCB-Onto model.

The contribution of the proposed framework is as follows:

1. The proposed framework introduces DR-MobiCB, a modular pipeline integrating MobileNetV3 with Convolutional Block Attention Module (CBAM) optimized for lesion-level diabetic retinopathy detection. This architectural design enhances interpretability and supports targeted optimization, whereas dilated convolutions enable multi-scale feature extraction, which is critical for identifying microaneurysms and exudates.
2. A domain-specific ontology layer is embedded within the framework to enable structured semantic reasoning over lesion types and spatial relationships. This component addresses the

* Corresponding author: ali.alkwzahy.iu@gmail.com

semantic limitations of conventional CNN-based models, facilitating clinically meaningful interpretation and improving diagnostic transparency.

3. The framework incorporates a composite preprocessing strategy, such as bilateral filtering, CLAHEU, and Z-score normalization, to enhance the retinal image quality and preserve vessel boundaries. Thick Blood Vessels (TBVs) were segmented using the Extended Adaptive Density-Based Spatial Clustering (EADBSC) method, which adapts to local density variations and ensures continuity and precision in vascular mapping.

The remaining part of this research organized as follow: section 2 represent related works, section 3 denotes the proposed methodology, Result and discussion present in Section 4 and Section 5 presents the conclusion of this research.

## 2 Related works

Table 1 presents a literature review on the existing DR detection models. This briefly explains the methodology diversity across different CNN-based, hybrid, and ensemble approaches, specifying dataset usage, architectural strengths, and interpretability limitations.

**Table 1.** Literature review of state-of-art method to detect the DR.

| Method Used | Datasets | Brief Methodology | Strength in Architecture | Limitation in Architecture |
|---|---|---|---|---|
| Transfer learning-based CNN using VGGNet-16 [10] | IDRiD, MESSIDOR, Kaggle EyePACS | Applied optic disc segmentation, edge detection, and normalization before classifying fundus images using softmax prediction | Preserved spatial granularity through fixed-size kernel filters | Deep convolutional stack reduced interpretability due to abstraction and dropout-induced sparsity |
| Custom CNN with 5 convolutional layers [11] | APTOS 2019, Messidor2, IDRiD | Preprocessed images with edge detection and optic disc segmentation, followed by softmax-based five-class classification | Enabled adaptive feature learning with Grad-CAM visualization | High parameter count and augmentation complexity hindered deployment in constrained environments |
| Hybrid pipeline: DWT + ANN segmentation + Gabor filter + Random Forest + Deep CNN[12] | Messidor | Combined biologically inspired preprocessing and feature selection with CNN optimized via Chicken Swarm Algorithm | Modular integration of wavelet and Gabor filters enhanced feature diversity | Sequential multi-stage design compounded feature transformations, reducing transparency |
| DBN + WKELM with Gannet Optimization [13] | MESSIDOR, DIARETDB1, IDRID | Used DBN for dimensionality reduction and WKELM for classification based on lesion features, tuned via GO algorithm | Multi-scale wavelet kernels and GO-based tuning improved convergence and generalization | Handcrafted lesion features and multi-phase tuning limited adaptability across imaging modalities |
| OVO-based ensemble of binary CNNs (MobileNet, InceptionV3, DenseNet121) [14] | APTOS 2019, IDRiD, Messidor-2, DDR | Decomposed five-class grading into ten binary classifiers trained through transfer learning and aggregated through voting | Simplified decision boundaries and enhanced interpretability through pairwise modeling | Required inter-model calibration and introduced dependency across classifiers |

Despite significant advancements in DR detection using Convolutional neural Networks (CNNs), ensemble classifiers, and segmentation architectures, the existing models exhibit limitations in terms of modular interpretability, lesion-level precision, and semantic reasoning. Most classification pipelines depend on massive CNN backbones that lack contextual awareness of retinal anatomy, whereas segmentation models frequently struggle with boundary preservation and multi-lesion discrimination, especially in early-stage DR. Therefore, to address these limitations, the DR-MobiCB-Onto framework is proposed, and a detailed explanation is presented in the next section.

## 3 Materials and method

This section explains the datasets used, dataset preprocessing to enhance the image before segmentation, and methodology explaining the segmentation of fundus images, lesion-level DR detection, and ontology interpretation, as shown in Figure 1.

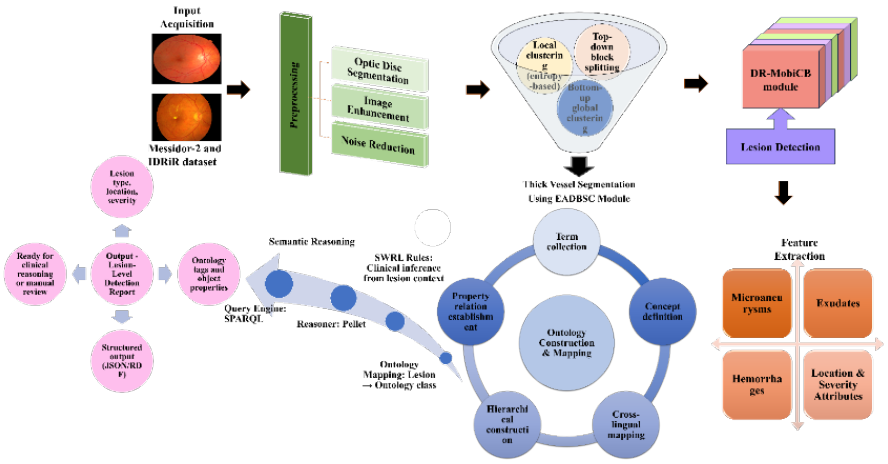


**Fig. 1.** The architecture of DR-MobiCB-Onto model.

### 3.1 Dataset

The two popular datasets, MESSIDOR 2 [15] and IDRiD [16], were employed in this research. MESSIDOR dataset contains 1200 retinal fundus images divided into four classes (0–3) based on DR severity. Class 0 represents healthy eyes with no disease, while Classes 1, 2, and 3 represent increasing levels of severity. The dataset is imbalanced across these classes; therefore, the vast majority of Class 0 images occupy nearly half and the There are only a few Class 1 images, which may cause model bias. The IDRiD dataset includes 413 training retinal images and about 103 testing images. These pictures are of one eye (a left eye or a right eye), and there is no age information presented. The IDRiD is a retinal image dataset of the Indian population that contains 516 train-set images, each with image-level DME grade annotations of normal (healthy), non-CSME, and CSME classes. The MESSIDOR-1 and MESSIDOR-2 datasets contain 1200 and 1748 retinal images, respectively, and include image-level annotations of DME severity grades. MESSIDOR-1 offers 3-class DME grades, but the MESSIDOR-2 dataset only offers screening (normal vs. DME) annotation of DME classification.

### 3.2 Data preprocessing

Retinal fundus images vary in resolution from 1440×960 to 2304×1536 pixels, requiring standardization before model ingestion. Bilateral filtering is applied to reduce capture noise and smooth homogeneous regions while preserving the edge details essential for lesion localization, especially around microaneurysms and hemorrhages. Contrast enhancement was performed using CLAHEU (Contrast Limited Adaptive Histogram Equalization with Unsharp Masking). CLAHEU improves the local contrast on small tiles and limits the noise amplification in uniform areas. Unsharp masking sharpens the lesion boundaries by subtracting the blurred version of the image. To emphasize low-intensity features, negative transformation highlights dark lesions, log transformation enhances subtle features by compressing the dynamic range, and power-law transformation adjusts the brightness and contrast using gamma correction. Optic disc removal prevents the misclassification of bright anatomical regions. The optic disc is segmented by identifying the brightest pixel cluster and applying circular masking to exclude it.

This step helps to differentiate hard exudates and cotton wool spots from non-pathological bright areas. The images were resized to 512×512 or 128×128 pixels using interarea interpolation, ensuring compatibility with CNNs while preserving structural integrity. Pixel intensities are normalized using Z-score scaling, standardizing values to a 0–1 range by dividing them by 255, which reduces bias and improves training convergence. Optional Canny edge detection delineates vessel boundaries and lesion contours through gradient computation, non-maximum suppression, and dual-threshold hysteresis. To address class imbalance and improve generalization, data augmentation was applied during training. Techniques to simulate lesion diversity across DR grades include rotation, flipping, shifting, cropping, and Gaussian scale-space augmentation. Preprocessing efficacy was validated using PSNR and SSIM. CLAHEU and bilateral filtering consistently yielded PSNR > 40 dB and SSIM > 0.99, indicating high fidelity and structural preservation.

When a negative transformation is applied to fundus images, they are represented as Equation (1),

$$s = L - l - r \tag{1}$$

The transformation of the log is defined as Equation (2),

$$s = clogr + 1 \tag{2}$$

Where $s$ and $r$ are the pixel values of initial input and final output images, respectively, while constant is represented by $c$, which can be expressed as Equation (3),

$$s = cr^{\gamma} \tag{3}$$

The process of sharpening involves subtracting a softened copy of the image from the original.

Sharpening with an unsharp filter makes the image clearer and improves segmentation accuracy.

### 3.3 Proposed modular DR detection pipeline integrating MobileNetV3 and convolutional block attention module (DR-MobiCB) methodology

The proposed DR-MobiCB-Onto framework integrates a modular diabetic retinopathy (DR) detection pipeline with ontology-based semantic reasoning. It comprises three core stages: vessel segmentation using EADBSC, lesion-level detection using MobileNetV3 enhanced with a Convolutional Block Attention Module (CBAM), and semantic interpretation through a domain-specific ontology aligned with biomedical standards.

#### *3.3.1 Vessel segmentation and lesion detection*

Following pre-processing, the EADBSC method is used to segment TBV, as shown in Figure 1. Segmentation is the process of splitting an image into different components with the intention of modifying the image presentation and simplifying the image in a manner that makes it more consistent and easier to analyze. The EADBSC Technique was used to cut TBV. In this section, present the EADBSC is designed to overcome the issue of linear dependency observed in EADBSC and similar density-based clustering techniques It also improves boundary settings and performance when working with large data. The EADBSC begins with a top-down approach, where it splits the data into data blocks using a data block splitter. Local clustering is performed using a block of data. The extended ADBSC executes a hierarchical cycle first, and this cycle includes an image block splitter divides the image into clusters of blocks, followed by neighborhood grouping within individual block, to obtain the final clustering results, the E ADBSC executes a basic global grouping procedure, in which the image block merger is completed as a significant component. The image block splitter measures the mean information content per data block and uniformly splits the block. The entropy information is formally expressed as shown in Equation (4).

$$S(Y) = \sum Q(y)T(y) = -\sum Q(y)log_c Q(y) \quad (4)$$

Where, $y$ is a random variable in $Y$, and $Q(y)$ is the probability of $y$ occurring. The image block splitter is assumed to divide the TBV ROI, where $q$ represents in Equation (5).

$$q = \left[\sqrt{\frac{M}{f}}\right] \quad (5)$$

Here, $f$ denotes the average expected number of points within the image. Under the standard coordinate framework, the upper and lower boundaries that define each image block are represented by Equations (6) and (7), respectively.

$$h(u_j) = min_j + \alpha_j u_j \quad (6)$$

$$h_u(u_j) = min_j + (\alpha_j + 1)u_j \quad (7)$$

Where $u_j$ denotes the data block index and $h(u_j)$ and $h_u(u_j)$ are the lower and higher bounds, correspondingly. To create the image splitter much efficiently and effectively, a parameter that blocks division out of the ROI area of the TBV to a predefined number of sub-blocks is halted, and is expressed in Equation (8):

$$\alpha_s = \eta.\varepsilon_{FATHER} + s + 1 \quad (8)$$

Where $\varepsilon_{FATHER}$ is shown as the index of the ROI region of the TBVs are considered. Local clustering applies density-based clustering within each image block and is denoted in Equation (9).

$$e_1(c) = min_j + \alpha p(c_i) \quad (9)$$

Here, $\alpha$ represents the bias of the image splitter block, and $e_1(c)$ denotes the lower bound. The image focus count within the block scope is represented as Equation (10).

$$Ep_{new} = E \times Ep_{new} \quad (10)$$

EADBSC carries out global clustering to generate the final output by merging related local clusters that are confined to individual image blocks. This integration ensures a cohesive representation across the entire image. Using EADBSC, the thick blood vessel (TBV) region of interest (ROI) is effectively segmented from the fundus image and forwarded to the classifier for further analysis.

Lesion detection was performed using the DR-MobiCB module, as shown in Figure 1, which combines MobileNetV3 with Convolutional Block Attention Module (CBAM). MobileNetV3 is a lightweight CNN that applies depth-wise separable convolutions and inverted residual blocks to reduce computation while maintaining accuracy. It is suitable for deployment in low-resource environments, and mobile platforms are integrated into the MobileNetV3 backbone to improve the feature representation. It applies channel attention, followed by spatial attention, allowing the model to focus on important lesion regions. This helps in identifying small and scattered features such as microaneurysms, hemorrhages, and exudates. Dilated convolutions were used depth-wise in DR-MobiCB to increase the receptive field without adding extra parameters. This allows the model to capture contextual information around the lesions, which is important for distinguishing between different DR grades. The model was trained using annotated fundus images, and lesion-level features were extracted from enhanced and segmented inputs. The combination of MobileNetV3, CBAM, and dilated convolutions enables the efficient and accurate detection of DR-related lesions.

### *3.3.2 Ontology-based semantic reasoning*

Ontology-based reasoning is used to provide a structured interpretation of the lesion-level outputs generated by the DR-MobiCB module. Although CNN models such as MobileNetV3 with CBAM can detect lesions such as microaneurysms, hemorrhages, and exudates, they do not explain the clinical significance of these findings. Ontology reasoning helps bridge this gap by mapping the detected features to standardized medical concepts and applying rules to infer diabetic retinopathy (DR) severity. The reasoning module uses domain-specific ontologies, such as the Diabetes Mellitus Treatment Ontology (DMTO) and Chinese Diabetes Medical Ontology (CDMO). These ontologies contain classes of lesion types, anatomical regions, and diagnostic categories. For example, hard exudates are represented as DMTO_0000914 and hemorrhages are linked to CDMO_0000123. Each class includes object properties, such as $hasLocation$, $hasSeverity$, and $isPartOf$, which describe the relationships between lesions and retinal structures as shown in the Equation (11).

Detected lesions from DR-MobiCB were annotated using these ontological classes. Semantic Web Rule Language (SWRL) rules were applied to infer the DR grade. For instance, a rule may state that

$$hasLesion(?x, Hemorrhage) \wedge nearMacula(?x) \rightarrow suggestGrade(?x, ModerateNPDR) \quad (11)$$

This means that if a hemorrhage is found near the macula, the system suggests moderate no proliferative DR. The reasoning engine uses a Pellet reasoner to process these rules and generate inferences. SPARQL queries are used to retrieve lesion relationships and severity levels from the ontology. This structured approach ensures that lesion detection is not only numerical but also clinically meaningful. Ontology integration also supports the interoperability. By aligning with SNOMED CT and NCIT terminologies, the system can generate standardized outputs suitable for electronic health records and clinical decision-support systems. This makes the DR-MobiCB-Onto framework not only accurate, but also explainable and usable in real-world medical settings.

## 4 Results and discussions

The SPOT-Route framework was developed in Python 3.10. YOLOv8 was used for real-time passenger detection using Ultralytics and OpenCV. Semantic annotation employed SOSA/SSN ontologies using RDFLib and Apache Jena. Scheduling was optimized using Google OR-Tools (CP-SAT solver), supported by a Statistical Data Component (SDC) and a Real-Time Computer Vision Component (RTCVC). Experiments were carried out on a system with an Intel i7 CPU, 32GB RAM, and NVIDIA RTX 3080 GPU using simulated transit data and onboard video streams. Evaluation focused on scheduling responsiveness, detection accuracy, and semantic consistency. The DR-MobiCB-Onto framework was tested for lesion-level diabetic retinopathy classification using two public retinal datasets. It used modular attention and ontology-based reasoning to enhance lesion detection and interpretation. Performance was measured using sensitivity, accuracy, AUC-ROC, precision, and F1-score, derived from standard classification metrics. The model showed strong results in both classification and segmentation tasks, maintained low computational overhead, and demonstrated reliable generalization in limited-resource settings.

### 4.1 Performance analysis

**Table 2.** Performance analysis to compare the segmentation models on Messidor-2 and IDRiD in DR-MobiCB-Onto framework.

| Model | Dataset | Accuracy (%) | AUC | Recall | Precision | F1-score |
|---|---|---|---|---|---|---|
| Attention U-Net | Messidor-2 | 93.10 | 0.962 | 89.45 | 91.20 | 90.31 |
| | IDRiD | 92.50 | 0.955 | 88.90 | 90.10 | 89.49 |
| ResUNet++ | Messidor-2 | 94.20 | 0.968 | 90.10 | 91.75 | 90.92 |
| | IDRiD | 93.30 | 0.960 | 89.80 | 91.00 | 90.39 |
| TransUNet | Messidor-2 | 95.10 | 0.973 | 91.20 | 93.00 | 92.09 |
| | IDRiD | 94.40 | 0.969 | 90.50 | 92.10 | 91.29 |
| Proposed (DR-MobiCB-Onto) | Messidor-2 | 97.40 | 0.987 | 94.43 | 95.98 | 96.85 |
| | IDRiD | 96.80 | 0.981 | 93.87 | 95.57 | 95.88 |

**Table 3.** Performance analysis to compare classification models on both datasets.

| Model | Dataset | Accuracy (%) | AUC | Recall | Precision | F1-score |
|---|---|---|---|---|---|---|
| DenseNet121 | Messidor-2 | 85.42 | 0.964 | 85.42 | 85.42 | 85.42 |
| | IDRiD | 79.63 | 0.842 | 79.63 | 79.63 | 79.63 |
| InceptionV3 | Messidor-2 | 81.03 | 0.910 | 81.03 | 81.03 | 81.03 |
| | IDRiD | 85.19 | 0.852 | 85.19 | 85.19 | 85.19 |
| MobileNet | Messidor-2 | 73.00 | 0.870 | 72.92 | 72.92 | 72.92 |
| | IDRiD | 84.57 | 0.845 | 84.57 | 84.57 | 84.57 |
| Proposed (DR-MobiCB-Onto) | Messidor-2 | 97.40 | 0.987 | 94.43 | 95.98 | 96.85 |
| | IDRiD | 96.80 | 0.981 | 93.87 | 95.57 | 95.88 |

The performance analysis results are presented in Tables 2 and 3 presents a performance analysis of the segmentation and classification models across the Messidor-2 and IDRiD datasets. The proposed

EADBSC and DR-MobiCB modules consistently outperformed the baseline architectures in terms of accuracy, AUC, and F1-score, indicating enhanced lesion-level precision and generalizability.

These results validated the efficacy of attention mechanisms and modular clustering in capturing subtle retinal features.

## 4.2 Statistical analysis

This section compares the training setups and computational loads of each model, as shown in Tables 4 and 5 for both datasets. The comparison covers trainable parameters, FLOPs, optimizer type, and learning rate. DR-MobiCB demonstrates a good balance between performance and efficiency, making it suitable for low-resource environments.

**Table 4.** Statistical analysis for proposed DR-MobiCB-Onto using Messidor-2 dataset.

| Model | Trainable Params | FLOPs (GFLOPs) | Epochs | Batch Size | Optimizer | Learning Rate |
|---|---|---|---|---|---|---|
| DenseNet121 | ~9.98M | ~2.9 GFLOPs | 50 | 32 | Adam | $1e^{-4}$ |
| InceptionV3 | ~23.8M | ~5.7 GFLOPs | 50 | 32 | RMSProp | $1e^{-4}$ |
| MobileNet | ~7.2M | ~0.57 GFLOPs | 50 | 32 | SGD | $1e^{-3}$ |
| Proposed (DR-MobiCB-Onto) | ~9.2M (ensemble) | ~3.8 GFLOPs | 50 | 32 | AdamW | $1e^{-4}$ |

**Table 5.** Statistical analysis for proposed DR-MobiCB-Onto using IDRiD dataset.

| Model | Trainable Params | FLOPs (GFLOPs) | Epochs | Batch Size | Optimizer | Learning Rate |
|---|---|---|---|---|---|---|
| Attention U-Net | ~18.9M | ~2.1 GFLOPs | 100 | 16 | Adam | $1e^{-4}$ |
| TransUNet | ~105M | ~14.5 GFLOPs | 100 | 8 | AdamW | $3e^{-5}$ |
| ResUNet++ | ~12.3M | ~3.2 GFLOPs | 100 | 16 | SGD | $1e^{-3}$ |
| Proposed (DR-MobiCB-Onto) | ~18.5M | ~4.2 GFLOPs | 100 | 16 | AdamW | $5e^{-5}$ |

## 4.3 Ablation study

The ablation study isolates the contribution of specific components such as CBAM, dilated convolutions, and ontology reasoning within the proposed framework presented in Tables 6 and 7. Performance degradation in the baseline models underscores the necessity for modular attention and semantic integration. The modular integration of EADBSC segmentation and dilated convolutions consistently boosts both accuracy and AUC across datasets, whereas ontology interpretation offers the steepest gain in semantic alignment and decision confidence. The full DR-MobiCB-Onto pipeline demonstrated robust generalization across Messidor-2 and IDRiD, validating its architectural synergy.

**Table 6.** Ablation study of DR-MobiCB-Onto framework on Messidor-2 dataset.

| Model Variant | Accuracy ↑ (%) | AUC ↑ (%) |
|---|---|---|
| DR-MobiCB | 84.92 | 0.903 |
| DR-MobiCB + EADBSC Segmentation | 88.74 | 0.934 |
| DR-MobiCB + Dilated Convolutions | 90.56 | 0.950 |
| DR-MobiCB + Ontology Interpretation | 92.83 | 0.968 |
| Proposed (DR-MobiCB-Onto) | 97.40 | 0.987 |

**Table 7.** Ablation study of DR-MobiCB-Onto framework on IDRiD dataset.

| Model Variant | Accuracy ↑ (%) | AUC ↑ (%) |
|---|---|---|
| DR-MobiCB | 86.14 | 0.912 |
| DR-MobiCB + EADBSC Segmentation | 89.63 | 0.941 |
| DR-MobiCB + Dilated Convolutions | 91.47 | 0.957 |
| DR-MobiCB + Ontology Interpretation | 93.92 | 0.973 |
| Proposed (DR-MobiCB-Onto) | 96.80 | 0.981 |

## 4.4 Comparative analysis

**Table 8.** Comparative analysis for proposed DR-MobiCB-Onto using Messidor-2 and IDRiD datasets.

| Model | Dataset | Accuracy (%) | AUC | Recall (%) | Precision (%) | F1-score(%) |
|---|---|---|---|---|---|---|
| CNN-XAI [11] | Messidor-2 | 93.86 | NA | NA | NA | NA |
| | IDRiD | 91.18 | NA | NA | NA | NA |
| Ensemble model [14] | Messidor-2 | 85.41 | NA | 85.41 | 85.41 | 85.41 |
| | IDRiD | 85.00 | NA | 85.00 | 86.00 | 85.00 |
| Proposed | Messidor-2 | 97.4 | 0.987 | 94.43 | 95.98 | 96.85 |
| | IDRiD | 96.8 | 0.981 | 93.87 | 95.57 | 95.876 |

This Table 8 compares the proposed framework with state-of-the-art models, such as CNN-XAI [11] and ensemble model [14]. The DR-MobiCB-Onto pipeline exhibited marked improvements in lesion-level detection and interpretability, surpassing traditional classifiers in both datasets. The results indicate that semantic reasoning and modular design are critical for achieving clinically meaningful DR grading, particularly in early stage detection scenarios.

### 4.5 Discussion

The proposed DR-MobiCB-Onto framework shows improved performance compared to existing CNN and ensemble models in terms of lesion-level detection and semantic clarity. MobileNetV3 with CBAM enhancement detects small lesions, including microaneurysms and hemorrhages, and dilated convolutions allow larger contexts to be represented without increasing the model complexity. Additionally, the ontology layer adds clinical meaning to detected features, enabling structured interpretation aligned with DR. Evaluation on Messidor-2 and IDRiD datasets indicated an accuracy of 97.40% and 96.80%, respectively, which supports robustness. The contribution of ontology reasoning and CBAM is confirmed in ablation studies that show performance degradation, validating the contribution of the integration of each component. The fact that it is lightweight enables the model to be deployed in mobile and low-resource environments. However, this study had some limitations. The framework is based on quality-annotated datasets, and might not be robust to low-contrast or noisy images. Ontological reasoning is rule-driven and might not be able to identify complex or unusual lesion patterns. Multimodal integration (e.g., OCT or fundus metadata) is not provided, and real-time inference is not supported. Such gaps indicate that future research must consider adaptive reasoning, increased lesion coverage, and real-time implementation to enhance clinical relevance and scalability.

## 5 Conclusions

This research introduced DR-MobiCB-Onto, a modular framework for DR detection and semantic interpretation. The model combined MobileNetV3 with CBAM to provide an effective focus on lesions, dilated convolutions to provide efficient contextual enhancement, and EADBSC to provide efficient vessel segmentation. An ontology of the domain facilitates semantic reasoning, which maps the features identified to clinical concepts. Testing on the Messidor-2 and IDRiD datasets demonstrated accuracies of above 97.40% and 96.80% and AUCs of above 0.98, outperforming the existing CNN and ensemble models in detecting lesions in early stages. "The ontology layer improves transparency by connecting image features with DR detection, helping in clinical decision-making. Its modular and lightweight design makes it suitable for mobile use, especially in low-resource settings. The main improvements include better lesion focus, clearer semantics, and higher reproducibility. Ablation studies confirmed the importance of each module, especially CBAM and ontology reasoning. Future work will extend the ontology to cover more lesion types, add adaptive reasoning for rare cases, enable real-time inference, and support use in telemedicine platforms

## References

1. S.K. Henge, N.R. Viraati, M. Alhussein, A.S. Kushwaha, K. Aurangzeb, R. Singh, Detection of diabetic retinopathy using a multi-decision Inception-ResNet-blended hybrid model. IEEE Access **13**, 8988-9005 (2025). https://doi.org/10.1109/ACCESS.2024.3525154
2. K. Anitha, P.S. Prabha, K.S. Rekha, M.V. Prem, J.J. Amarnath, Detecting diabetic retinopathy using a hybrid ensemble XL machine model with dual weighted-kernel ELM and improved mayfly optimization. Expert Syst. Appl. **253**, 124221 (2024). https://doi.org/10.1016/j.eswa.2023.124221
3. M.S.B. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala, M.S. Waseem, A bi-directional long short-term memory-based diabetic retinopathy detection model using retinal fundus images. Healthc. Anal. **3**, 100174 (2023). https://doi.org/10.1016/j.health.2023.100174
4. P. Bidwai, S. Gite, B. Pradhan, H. Gupta, A. Alamri, Harnessing deep learning for detection of diabetic retinopathy in geriatric group using optical coherence tomography angiography-OCTA: a promising approach. MethodsX **13**, 102910 (2024). https://doi.org/10.1016/j.mex.2024.102910
5. M. Yinghua, Y. Heng, R. Amarnath, Z. Hui, Hard exudates segmentation in diabetic retinopathy using DiaRetDB1. IEEE Access **12**, 126486-126502 (2024). https://doi.org/10.1109/ACCESS.2024.3455433
6. R. Romero-Oraá, M. Herrero-Tudela, M.I. López, R. Hornero, M. García, Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading. Comput. Methods Programs Biomed. **249**, 108160 (2024). https://doi.org/10.1016/j.cmpb.2024.108160
7. O. Dib, A decentralized privacy-preserving framework for diabetic retinopathy detection using federated learning and blockchain. Results Eng. **26**, 105456 (2025). https://doi.org/10.1016/j.rineng.2025.105456
8. I. Govindharaj, A. Poongodai, D. Santhakumar, S. Ravichandran, R. Vijaya Prabhu, K. Udayakumar, S. Yazhinian, Enhanced diabetic retinopathy detection using U-shaped network and capsule network-driven deep learning. MethodsX **14**, 103052 (2025). https://doi.org/10.1016/j.mex.2024.103052
9. E. Ramezanzadeh, N. Shoeibi, A. Feizabadi, T. Banaee, M.H.B. Tussi, M. Tavakoli, Automated

detection of hard exudates in retinal fundus images for diabetic retinopathy screening using textural-based radon transform and morphology reconstruction. Biomed. Eng. Adv. **9**, 100180 (2025). https://doi.org/10.1016/j.bea.2025.100180

10. A. Jabbar, S. Naseem, J. Li, T. Mahmood, M.K. Jabbar, A. Rehman, T. Saba, Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas. Int. J. Comput. Intell. Syst. **17**, 135 (2024). https://doi.org/10.1007/s44196-024-00520-w
11. K.A. Alavee, M. Hasan, A.H. Zillanee, M. Mostakim, J. Uddin, E.S. Alvarado, I. de la Torre Diez, I. Ashraf, M.A. Samad, Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence. IEEE Access **12**, 73950-73969 (2024). https://doi.org/10.1109/ACCESS.2024.3405570
12. V. Thanikachalam, K. Kabilan, S.K. Erramchetty, Optimized deep CNN for detection and classification of diabetic retinopathy and diabetic macular edema. BMC Med. Imaging **24**, 227 (2024). https://doi.org/10.1186/s12880-024-01406-1
13. S. Krishnamoorthy, Y. Weifeng, J. Luo, S. Kadry, GO-DBN: gannet optimized deep belief network-based wavelet kernel ELM for detection of diabetic retinopathy. Expert Syst. Appl. **229**, 120408 (2023). https://doi.org/10.1016/j.eswa.2023.120408
14. S.J. Sidiq, T. Benil, A lightweight transfer learning-based ensemble approach for diabetic retinopathy detection. Int. J. Inf. Manag. Data Insights **5**, 100372 (2025). https://doi.org/10.1016/j.jjimei.2025.100372
15. The Messidor-2 dataset link:https://www.kaggle.com/datasets/mariaherrerot/messidor2preprocess?select=messidor-2 (accessed on 30th August 2025)
16. The IDRiD dataset link: https://www.kaggle.com/datasets/aaryapatel98/indian-diabetic-retinopathy-image-dataset (accessed on 30th August 2025)