DOI: https://doi.org/10.15276/aait.09.2026.16
UDC 004.8:616.379-008.64-073.756.8

Diabetic retinopathy diagnosis using deep neural networks

Yelizaveta I. Kabanova[1)]
ORCID: https://orcid.org/0009-0001-2692-5066; yelizavetakabanova@gmail.com
Nataliia V. Kuznietsova[1)]
ORCID: https://orcid.org/0000-0002-1662-1974; natalia-kpi@ukr.net. Scopus Author ID: 56412465200
Kateryna O. Ivanko[1)]
ORCID: http://orcid.org/0000-0002-3842-2423; ivanko-ee@lll.kpi.ua. Scopus Author ID: 55819298100
Vishwesh Kulkarni[2) 3)]
ORCID: https://orcid.org/0000-0002-22858652; vishwesh.kulkarni@kcl.ac.uk. Scopus Author ID: 7201425741
[1)] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, Beresteiskyi Ave. Kyiv, Ukraine
[2)] King's College London, building, street Strand, London, WC2R 2LS, United Kingdom
[3)] SUSTech-King's School of Medicine, building, street Shenzhen, 518 055, China

ABSTRACT

Relevance: is driven by the high prevalence of diabetic retinopathy as one of the leading causes of vision loss worldwide and the limited availability of ophthalmological diagnostic resources. In medical applications, predictive uncertainty assessment is critically important, since erroneous yet confident decisions may lead to serious clinical consequences. Despite significant advances in deep learning-based DR diagnostic systems, insufficient attention has been paid to the systematic interaction between data preprocessing strategies, class imbalance handling, and predictive uncertainty estimation, all of which directly affect model reliability. **Aim and objectives:** the aim of this study is to develop and experimentally validate an automated DR stage classification framework that achieves high diagnostic performance while systematically evaluating predictive reliability and uncertainty. **Methods:** the proposed framework integrates computer vision and deep learning methods based on convolutional neural networks with transfer learning and incorporates an adaptive attention mechanism that dynamically regulate channel and spatial attention strength according to feature variability. The approach also incorporates two preprocessing pipelines, class-balancing techniques, and uncertainty estimation through stochastic inference without modifying the training procedure. A systematic experimental study was conducted to evaluate different combinations of preprocessing and balancing strategies under fixed architectural conditions. Model performance was evaluated using standard classification metrics and an original integral reliability indicator, which jointly accounts for classification quality, prediction confidence, and predictive uncertainty. **Results:** combining advanced preprocessing, targeted balancing strategies, adaptive attention mechanisms, and uncertainty estimation significantly improves both classification effectiveness and predictive reliability. Proposed configuration achieved the highest scores, while uncertainty-aware models exhibited lower predictive variance and higher confidence, particularly in challenging or ambiguous cases. **Conclusions:** the proposed framework provides a structured methodology for reliability-aware DR classification and contributes an integrated evaluation approach that enhances the practical applicability of deep learning systems in clinical decision support.

Keywords: Diabetic retinopathy; fundus images; deep learning; adaptive attention; image recognition; uncertainty estimation

For citation: **Kabanova Y. I., Kuznietsova V. N., Ivanko K., Kulkarni V. "Diabetic retinopathy diagnosis using deep neural networks". *Applied Aspects of Information Technology*. 2026; Vol.9 No.2: 220–233. DOI: https://doi.org/10.15276/aait.09.2026.16**

1. INTRODUCTION

Diabetic Retinopathy (DR) is one of the leading causes of preventable blindness worldwide, with an estimated 2.6 % of global blindness attributable to diabetes-related retinal damage [1]. Without early diagnosis and timely treatment, DR can progress to irreversible vision loss, making early detection essential for effective treatment.

Diabetic retinopathy develops as a result of progressive damage to retinal blood vessels caused by chronic hyperglycemia, leading to vascular leakage, ischemia, and neovascularization. Typical retinal manifestations include microaneurysms, hard exudates, hemorrhages, and macular edema, which are visible in fundus images and form the basis for automated DR classification [2].

In clinical practice, the ophthalmologists manually evaluate fundus images using specialized lenses and illumination techniques. This procedure is time-consuming and requires substantial effort, knowledge, and clinical experience. Consequently, automated computer-aided diagnosis systems based on deep learning have gained increasing attention as tools to support large-scale DR screening.

Recent advances in deep learning, particularly convolutional neural networks (CNNs) have demonstrated remarkable performance in DR classification tasks, often reaching or exceeding expert-level accuracy [3]. However, the reliability of such systems strongly depends on several factors. Firstly, fundus images exhibit substantial variability

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

in illumination, contrast, noise, artifacts, and resolution, making preprocessing a crucial component of robust diagnostic pipelines. In addition, publicly available DR datasets exhibit severe class imbalance, with the majority of samples belonging to no DR, while DR stages are underrepresented [4]. This imbalance dramatically reduces model sensitivity to minority classes and limits model generalization in real-world clinical settings

Most existing studies [5], [6], [7] primarily focus on improving network architectures, integrating attention mechanisms, or designing specialized loss functions. While these techniques are important, considerably less attention has been paid to systematically analysing how preprocessing strategies and class balancing techniques interact when the backbone architecture and loss function are fixed. Moreover, most works evaluate models exclusively using accuracy-based metrics, implicitly assuming that confident predictions are also reliable. In medical applications, however, this assumption is particularly dangerous, as incorrect but confident predictions may lead to severe clinical consequences. Another important limitation of existing DR classification systems is the lack of explicit uncertainty estimation.

The aim of this study is to develop an automated DR classification framework that not only achieves high accuracy but also systematically evaluates predictive reliability and uncertainty, addressing key limitations of current CNN-based systems.

2. LITERATURE REVIEW AND PROBLEM STATEMENT

Early studies on the automated detection of diabetic retinopathy relied on handcrafted features combined with classical machine-learning classifiers. Methods based on Support Vector Machine (SVM) or texture descriptors demonstrated promising results but were limited by their dependence on manual feature engineering and poor generalization across datasets [8], [9].

With the emergence of deep learning, especially after the introduction of convolutional neural networks, the performance of automated diagnosis improved dramatically. Studies have shown that an Inception-v3-based model can surpass the sensitivity and specificity of ophthalmologists [10].

Subsequent research explored alternative pretrained architectures such as ResNet, Inception, DenseNet, and EfficientNet for DR classification, achieving progressively improved performance [10].

More recent research has focused on enhancing model performance through architectural refinements and specialized loss functions. Attention mechanisms, including channel-wise, spatial, and hybrid approaches, have been widely adopted to improve the localization of pathological retinal features such as microaneurysms, hemorrhages, and exudates. Modules such as Convolutional Block Attention Module (CBAM) have been shown to enhance feature representation by guiding the network's focus toward clinically relevant regions [11], [12], [13]. Additionally, segmentation-based approaches (e.g., U-Net) have been used to extract detailed retinal structures, including blood vessels, exudates, hemorrhages, and microaneurysms, improving lesion-level analysis [7], [14].

In parallel, studies have shown that data preprocessing and class balancing are critical for robust performance. Weighted cross-entropy, Focal Loss, and data-level balancing strategies improve sensitivity to minority classes [15]. Recent studies also emphasize the importance of curated and annotated datasets to ensure reliable model training and reduce noisy labels [16].

Despite these advances, two important gaps remain. First, few studies systematically analyse the interaction between preprocessing pipelines and class-balancing strategies under controlled architectural conditions. Secondly, uncertainty estimation is rarely incorporated. Existing works predominantly rely on deterministic predictions and accuracy-based metrics, which provide limited insight into model reliability. Although some studies [17], [18] have implemented Bayesian deep learning methods, such as Monte Carlo Dropout and Variational Bayesian layers, providing epistemic uncertainty estimations that improve clinical trustworthiness, they generally do not systematically analyse the influence of data preprocessing pipelines, class-balancing strategies, or attention mechanisms on predictive reliability. Moreover, uncertainty metrics in these studies are typically reported independently of classification performance, making it difficult to assess practical clinical applicability.

This work addresses these gaps by combining adaptive attention mechanisms with Monte Carlo Dropout-based uncertainty estimation and introducing an integrated reliability metric tailored for medical image classification. It also evaluates multiple preprocessing and class-balancing configurations using a DenseNet121 backbone, enabling a systematic assessment of both classification performance and reliability on real DR

datasets. The scientific contributions of this work are as follows:

– a comprehensive experimental framework enabling isolated analysis of preprocessing and class-balancing strategies under fixed architectural conditions;

– a systematic comparison of four preprocessing and class-balancing configurations using a fixed DenseNet121 backbone;

– an adaptive modification of the Convolutional Block Attention Module (CBAM), where the strength of channel and spatial attention is dynamically adjusted based on feature variability;

– the integration of Monte Carlo Dropout to estimate epistemic uncertainty without modifying the training procedure;

– the introduction of an original Reliability Score (RS) metric that jointly accounts for classification performance (QWK), prediction confidence, and predictive uncertainty.

Results demonstrate that uncertainty-aware models with adaptive attention mechanisms can provide more reliable and clinically meaningful predictions.

3. RESEARCH AIM AND OBJECTIVES

The aim of this study is to develop and experimentally validate an integrated and uncertainty-aware methodology for diabetic retinopathy (DR) stage classification from fundus images. The study develops a methodology that improves prediction accuracy, stability, and reliability through a comprehensive analysis of data preparation, deep learning models, and uncertainty assessment. The study also conducts a comparative evaluation of preprocessing and class-balancing strategies, assessing models using classification metrics and uncertainty indicators.

The objectives of the research are:

1) analyse the current state of research and modern approaches to DR diagnostics;

2) investigate the characteristics of fundus images and available datasets;

3) develop an approach to image preprocessing to improve data quality;

4) conduct an experimental study to evaluate the impact of different data preparation strategies on classification results;

5) develop a deep learning model for automatic classification of DR stages based on convolutional neural networks;

6) propose an integrated metric for the comprehensive assessment of classification quality and predictive reliability;

7) analyse the experimental results, compare the proposed model with existing methods, and assess the practical applicability of the approach for automated medical diagnostics.

The object of the research is the classification of diabetic retinopathy stages from fundus images.

The subject of the research is deep learning models and methods for DR stage classification, including approaches to enhancing the reliability and stability of their predictions.

4. MATERIALS AND METHODS

In this study, the problem of automatic classification of diabetic retinopathy (DR) stages based on retinal fundus images is addressed using deep learning methods, in particular, convolutional neural networks. The input data are colour images of the retina, each of which corresponds to a certain stage of the disease. The result is a predicted class of diabetic retinopathy, which can be used by medical specialists to support clinical decision-making.

This study follows a systematic and iterative approach to data preparation, which involves consistent improvement of image processing, class balancing, and evaluation of their combined impact on model performance.

Fig. 1 depicts the specially developed methodology for this study. The proposed framework starts from dataset acquisition and image preprocessing, followed by class balancing and data splitting. Subsequently, multiple model configurations are evaluated using a fixed CNN backbone with different attention mechanisms, loss functions, and training strategies. The models are trained and validated under consistent conditions, and their performance is assessed using classification and uncertainty-related metrics. The final model configuration is selected based on comparative experimental evaluation. All obtained results are further used for analysis and selection of the most effective model.

This approach provides transparency, repeatability, and the ability to track the contribution of each method to the result of training the model.

4.1. Dataset

To investigate the influence of the performance of the models, the APTOS 2019 Blindness Detection (APTOS 2019 BD) dataset was used. This open-source dataset, available on Kaggle [19], contains 3,662 colour fundus images captured at Aravind Eye Hospital (India). Each image is provided in RGB format, accompanied by a unique identifier and a diagnostic label representing DR severity, annotated by expert ophthalmologists.

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

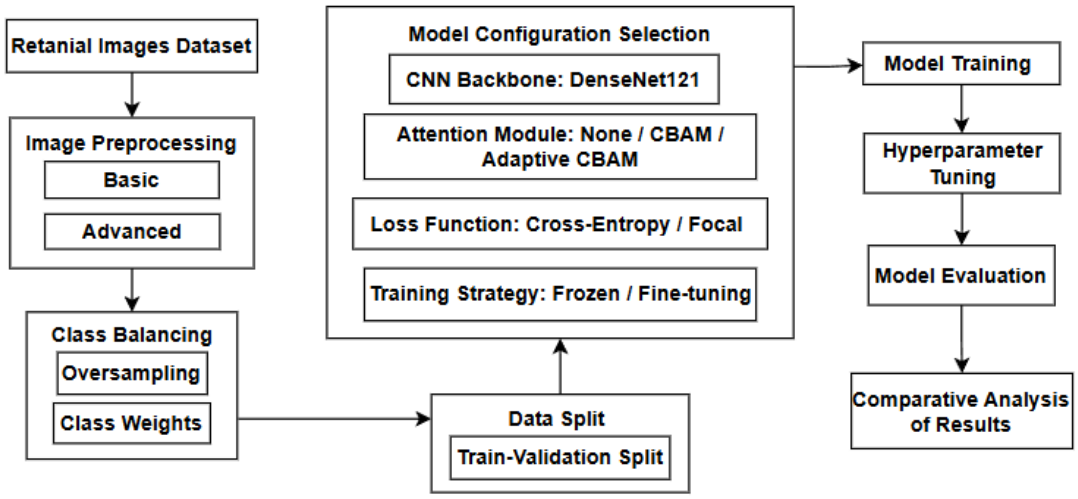


Fig. 1. **Developed methodology for DR classification**
Source: **compiled by the authors**

The target variable consists of five classes (Fig. 2):

1. No DR – 0;
2. Mild DR – 1;
3. Moderate DR – 2;
4. Severe DR – 3;
5. Proliferative DR – 4.

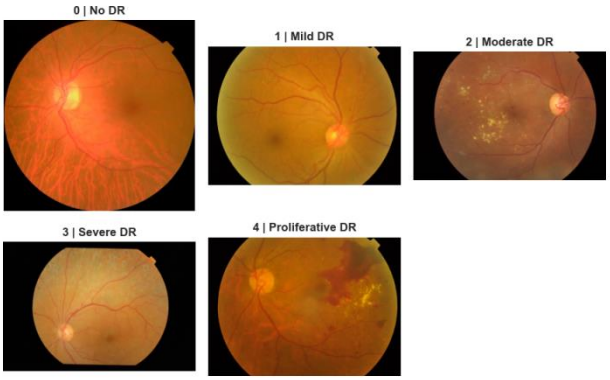


Fig. 2. **Samples of fundus images from APTOS 2019 Dataset for each DR stage**
Source: **compiled by the authors**

Exploratory data analysis (EDA) was performed to evaluate dataset quality, including:

– data completeness and integrity;
– presence of noise, artifacts, or uneven illumination;
– accuracy of class labels;
– class distribution.

The analysis revealed significant class imbalance, with nearly 50% belonging to class 0, while classes 3 and 4 were underrepresented (Fig. 3). Image sizes also varied considerably, ranging from 358×474 to 2848×4288 pixels.

Exploratory analysis also revealed variations in image brightness and contrast, the presence of artifacts, and noise, highlighting the necessity of robust preprocessing pipelines to ensure reliable model training.

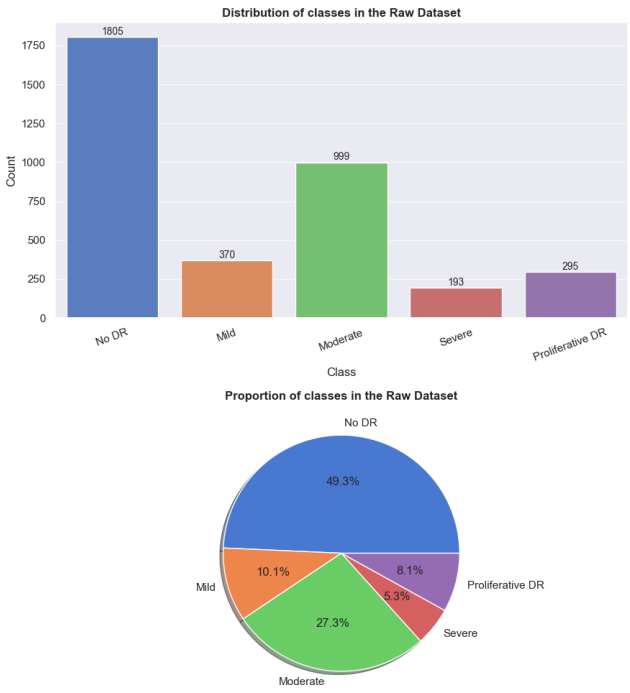


Fig. 3. **Distribution of DR classes in the APTOS 2019 Dataset**
Source: **compiled by the authors**

Overall dataset limitations included:

– severe class imbalance;
– variability in contrast and illumination;
– presence of noise and artifacts;
– heterogeneity in image sizes and orientation.

To address these challenges, preprocessing techniques were designed to standardize image

quality, and class-balancing methods were applied to mitigate the impact of dataset imbalance.

4.2. Data preprocessing

Preprocessing of medical images is one of the most important stages in building automatic pathology recognition systems. In the case of diabetic retinopathy diagnosis, the quality of the retinal image directly affects the effectiveness of deep models, as visual features – blood vessels, microaneurysms, haemorrhages, exudates – determine the level of damage. Preliminary data analysis showed that most images in the APTOS 2019 dataset have uneven lighting, black frames around the fundus, glare, and colour variations, which are due to the shooting conditions and equipment.

Two preprocessing techniques were implemented in the study:

1) *basic preprocessing* includes resizing to 224×224 pixels, normalization, and the removal of black corners from images using an algorithm based on pixel-intensity thresholding [20]. A binary mask of informative regions is constructed, rows and columns containing non-zero values are identified, and the image is cropped to the minimal bounding rectangle that preserves all colour information;

2) *advanced preprocessing* follows the Ben Graham method [21] and includes cropping and resizing, normalization, implementation of Gaussian blurring, linear combination of original and blurred images to enhance retinal vessels, and contrast enhancement using CLAHE.

The basic approach (A) reduces noise and black areas, while the Ben Graham processing (B) improves the visibility of small details and contrast.

Fig. 4 shows a comparison between basic and advanced preprocessing results.

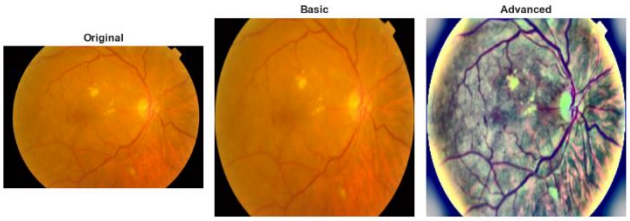


Fig. 4. **Comparison of basic and advanced preprocessing techniques**
Source: **compiled by the authors**

4.3 Data Augmentation

When working with medical imaging data, the number of samples for class is often highly imbalanced: healthy cases typically dominate the dataset, while pathological examples are significantly underrepresented. This imbalance not only biases the learning process but also increases the risk of model overfitting.

By introducing data augmentation, the model is trained on a wider set of visual variations, which reduces the risk of overtraining and improves the ability to recognize pathologies in new, unknown fundus images.

In this study, data augmentation was applied at two levels. First, manual augmentations were used during the class balancing stage to oversample underrepresented classes and achieve a uniform class distribution. This oversampling was performed on the entire dataset before splitting into training and validation subsets. The applied transformations used for oversampling are summarized in Table 1.

Table 1. **Data augmentation transformations for class balancing**

Transfor-mation	Value	Effect
Flip	Horizontal	Eye symmetry invariance
Brightness change	±20	Robustness to illumination variations
Rotation	15	Simulation of camera angle variations

Source: **compiled by the authors**

Second, augmentation was applied during model training to further increase data diversity and improve generalization performance. These transformations were applied after splitting the dataset to only to the training subset to further increase data diversity and improve generalization. These transformations are summarized in Table 2.

Table 2. **Data augmentation transformations during model training**

Transformation	Value
Flip	Horizontal
Brightness change	±20
Rotation	10
Width and height shift	10
Zoom	10

Source: **compiled by the authors**

Although oversampling for class balancing was applied before splitting, further augmentation during model training was applied only to the training set. leaving the validation data untouched. While this strategy improves class balance and data diversity, it may introduce a potential risk of sample similarity between subsets. This limitation is acknowledged and considered in the interpretation of the reported results.

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

4.4. Class Balancing

Class imbalance can bias model predictions toward majority classes [22]. To mitigate this, two complementary strategies were employed:

1) *oversampling:* minority classes were upsampled using targeted data augmentation, resulting in a balanced dataset with 1805 images per class. Applied transformations are described in the previous section;

2) *class-weighted loss*: higher weights were assigned to underrepresented classes following:

$$w_k = \frac{N}{N_k \cdot n}, \quad (1)$$

where *N* denotes the total number of samples, N_k is the number of samples for class *k*, and *n* is the total number of classes.

The class-weighted loss assigned inversely proportional weights to underrepresented classes, ensuring that rare but clinically critical DR stages contribute more to the loss function. Whereas oversampling increased the frequency of minority samples to a level comparable to the dominant class.

Four experimental setups were evaluated by combining two preprocessing strategies with two class-balancing methods (Table 3).

Table 3. **Experimental combinations**

Experiment	Preprocessing	Class Balancing
A0	Basic	Class Weights
A1	Basic	Oversampling
B0	Advanced	Class Weights
B1	Advanced	Oversampling

Source: **compiled by the authors**

4.5. Data Split

The effectiveness of any machine learning model directly depends on the correctness of dividing the data into subsets that perform different functions during model training and evaluation.

The dataset was split into training and validation subsets using an 80/20 ratio. The resulting split sizes for each experimental configuration were as follows:

– A0, B0: Training – 2929 images; Validation – 733 images;

– A1, B1: Training – 7220 images; Validation – 1805 images.

4.6. Model Selection

A pre-trained *DenseNet121* [23] was chosen as the backbone. DenseNet121 consists of 121 layers organized into four dense blocks separated by transition layers, followed by Global Average Pooling and a fully connected classification layer. This architecture is suitable for imbalanced datasets and offers efficient parameter utilization.

Focal Loss [24] was used to address class imbalance:

$$L_{FL}(y, \hat{y}) = -\sum_{i=1}^{C} y_i (1 - \hat{y}_i)^{\gamma} log\ \hat{y}_i\ , \quad (2)$$

where C is the number of classes, y is the truth label, $\hat{y}_i$ is the predicted probability of class i, and $\gamma > 0$ is the focusing parameter that reduces the influence of well-classified examples. In the context of the diabetic retinopathy classification, focal cross-entropy allowed us to compensate for the imbalance between classes and increased the sensitivity of the model to rare cases.

Convolutional Block Attention Module [25] was applied to compute sequential channel and spatial attention:

$$F' = M_c(F) \otimes F, F'' = M_s(F') \otimes F', \quad (3)$$

where F denotes the input feature map, F' is the feature map after channel attention, F'' is the final feature map after spatial attention, M_c is the channel attention, and M_s is the spatial attention, and $\otimes$ denotes element-wise multiplication. CBAM allowed the model to enhance informative features and suppress background noise. Channel attention helped the network learn the relationships between different features, while spatial attention helped in learning the relationships between different areas of the image.

Adaptive Convolutional Block Attention Module was proposed as an author's modification of CBAM, in which the strength of both channel and spatial attention was dynamically adjusted based on the statistical properties of the extracted feature maps. Unlike the standard CBAM, where attention weights are applied uniformly, the proposed approach used feature variance as a proxy for information diversity. This allowed the model to strengthen attention for informative representations and attenuate it for low-variance or noisy features.

This adaptive mechanism helped improve feature selection, focusing on retinal vessels and lesions characteristic of different DR stages, and contributed to improved robustness to variability in image quality without introducing a significant computational overhead.

Uncertainty estimation was performed using *Monte Carlo Dropout* [26], which approximates Bayesian inference by performing multiple stochastic forward passes during inference and

generating distributions of predictions for each image. Metrics such as prediction variance, entropy, and mean confidence were computed to assess predictive reliability. These values enabled the analysis of model stability and behaviour under uncertain or noisy input conditions.

4.7. Training

All models were trained using transfer learning with ImageNet-pretrained DenseNet121 as the backbone. Training was performed in two stages. In the first stage, the backbone network was frozen, and only the classification layers were trained using the Adam optimizer with a learning rate of 1×10^{-4}. In the second stage, fine-tuning was conducted by unfreezing the top layers of the backbone and continuing training with a reduced learning rate of 1×10^{-5}.

To improve training stability and reduce overfitting, Batch Normalization layers, dropout-based regularization, and L2 weight regularization were applied within the classification head. All models were trained under identical conditions to ensure fair comparison: input image size of 224 × 224, batch size of 32, and 30 training epochs.

4.8. Evaluation Metrics

To evaluate model performance, the following metrics were used [27].

Accuracy:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

Precision and Recall:

$$Precision = \frac{TP}{TP + FP}, \quad (5)$$

$$Recall = \frac{TP}{TP + FN}, \quad (6)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative values, respectively.

F1-score as the harmonic mean of precision and recall:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \quad (7)$$

Area Under the ROC Curve (AUC) measures the area under the Receiver Operating Characteristic curve and reflects the trade-off between true positive rate and false positive rate across different classification thresholds.

Quadratic Weighted Kappa (QWK):

$$QWK = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}, \quad (8)$$

where $O_{i,j}$ is the observed confusion matrix, E_{ij} is the expected matrix under random agreement, and w_{ij} is the quadratic weight defined as:

$$w_{ij} = \frac{(i-j)^2}{(C-1)^2}. \quad (9)$$

Among these metrics, QWK was chosen as the main comparison metric, as it considers the ordinal nature of the classes and the different error weights between neighbouring stages of the disease.

5. RESEARCH RESULTS

Experiments A0 and B0, which employed class-weighted loss without oversampling, demonstrated moderate classification performance. In these settings, the introduction of Focal Loss consistently improved Quadratic Weighted Kappa (QWK). Specifically, in experiment A0, the use of Focal Loss increased QWK from 0.77 (baseline) to 0.82, accompanied by both improvements in Precision and Recall. In experiment B0, the baseline model already achieved a relatively high QWK value of 0.80 due to advanced preprocessing.

However, the integration of attention mechanisms (CBAM and Adaptive CBAM) in experiments A0 and B0 did not lead to substantial improvement.

A substantial performance improvement was observed in ***experiments A1 and B1***, where oversampling and data augmentation were applied. Compared to A0, experiment A1 exhibits marked increases across all major evaluation metrics, with QWK values exceeding 0.90 for attention-based models. Within this group, the combination of Focal Loss and CBAM showed the highest QWK, indicating improved sensitivity to minority DR stages.

The best overall results were obtained in experiment B1, which combines advanced preprocessing with oversampling. The model incorporating Adaptive CBAM and MC Dropout reached a QWK of 0.92, along with high Precision, Recall, and AUC values. These results indicate superior classification performance across both frequent and underrepresented disease stages.

Fig. 5 presents the mean values of key performance metrics across all experimental configurations. The results demonstrate that oversampling-based experiments (A1 and B1)

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

consistently outperform their class-weighted counterparts (A0 and B0) across all evaluated metrics.

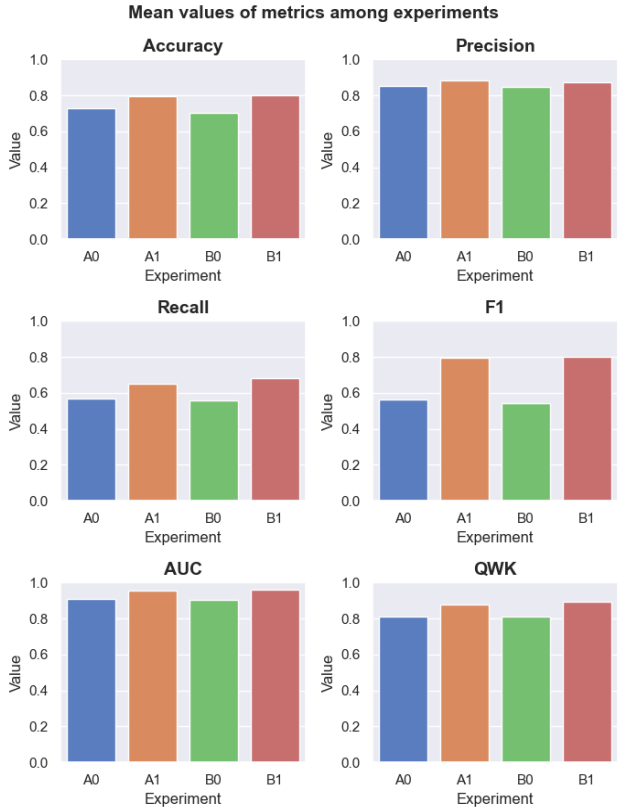


Fig. 5. **Mean metrics across all experiments**
Source: **compiled by the authors**

Fig. 6 shows the confusion matrix for the best-performing model from experiment B1. The matrix illustrates strong classification performance across all DR stages, including minority classes.

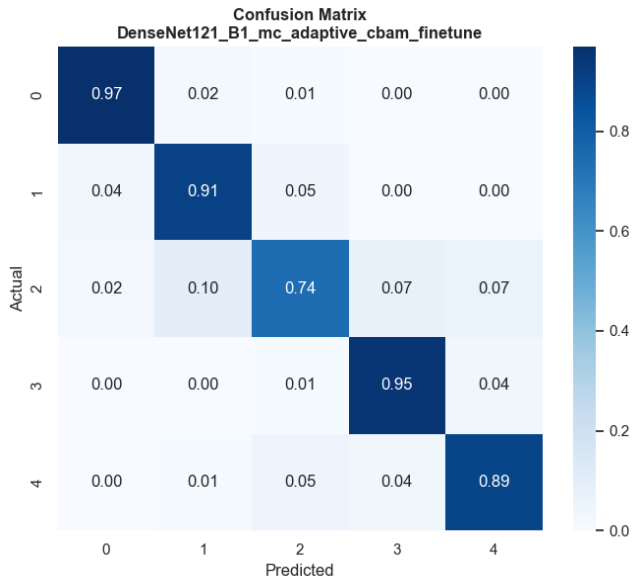


Fig. 6. **Confusion matrix for the best-performing model of experiment B1**
Source: **compiled by the authors**

Fig. 7 illustrates ROC curves for each class (0-4) of the B1 configuration.

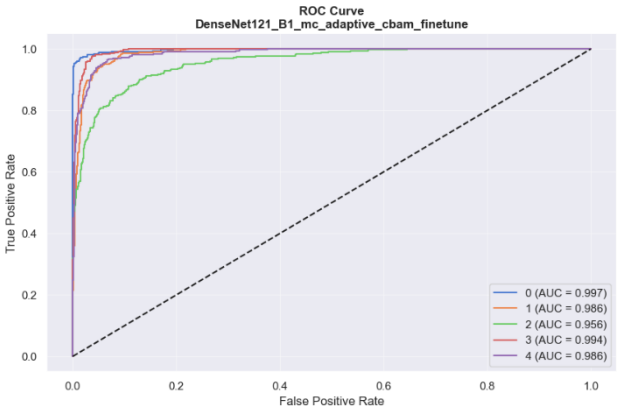


Fig. 7. **ROC curves for each DR class for the best-performing model of experiment B1**
Source: **compiled by the authors**

To estimate predictive confidence, epistemic uncertainty was estimated for models employing Monte Carlo Dropout during training. Experiments A1 and B1 exhibited lower mean entropy and reduced predictive variance compared to configurations without oversampling. Among all experiments, the B1 configuration demonstrated the lowest entropy and the highest average prediction confidence.

Fig. 8 presents the distributions of epistemic uncertainty, entropy, and the confidence-entropy ratio for the B1 experiment.

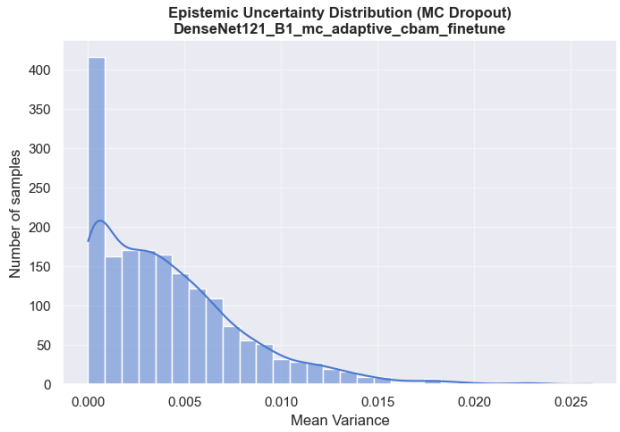
These results demonstrate that the configuration B1 has a greater stability of predictions and improved reliability in ambiguous or challenging cases.

To jointly evaluate classification accuracy and prediction reliability, an integrated metric Reliability Score (RS) was introduced:

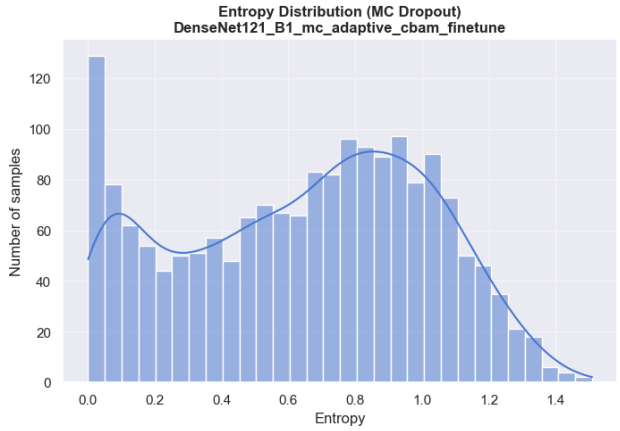
$$RS = \alpha ln(QWK + \varepsilon) + \beta(1 - H_{norm} + \varepsilon) + \gamma ln(C + \varepsilon), \quad (10)$$

where QWK denotes Quadratic Weighted Kappa, H_{norm} is the normalized mean entropy, C represents the mean prediction confidence, α, β, γ are weighting coefficients, and $\varepsilon > 0$ is a small constant. The logarithmic transformation was applied to stabilize the scale of QWK and confidence values and to reduce the dominance of extreme values.

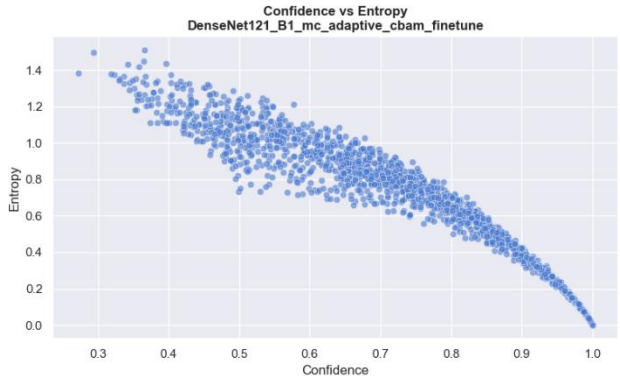
Fig. 9 compares Reliability Score values across all experiments. The B1 experiment achieved the highest RS value (0.61), while experiments without oversampling (A0 and B0) demonstrated substantially lower RS scores.



a



b



c

Fig. 8. **Distributions of:**
a) epistemic uncertainty; b) predictive entropy; c) confidence-entropy ratio for the best-performing model of experiment B1
Source: **compiled by the authors**

For qualitative analysis, Fig. 10 presents an example of classification of the same retinal image by the baseline model (A0) and the best-performing model (B1).

For the baseline model (A0), confidence corresponds to the maximum softmax probability obtained from a single deterministic forward pass. For the best-performing model (B1), confidence is computed as the maximum value of the mean predictive distribution obtained by averaging multiple stochastic forward passes using Monte Carlo Dropout.

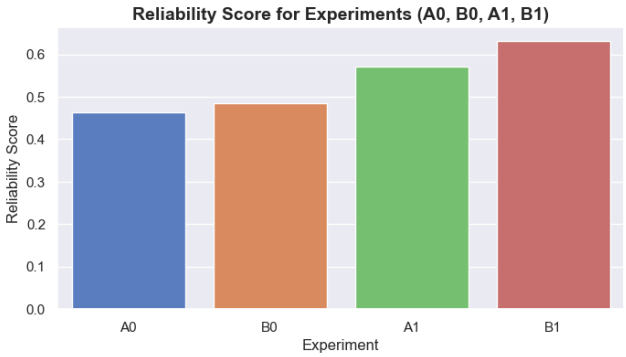


Fig.9. **Comparison of Reliability Score values across experimental configurations (A0–B1)**
Source: **compiled by the authors**

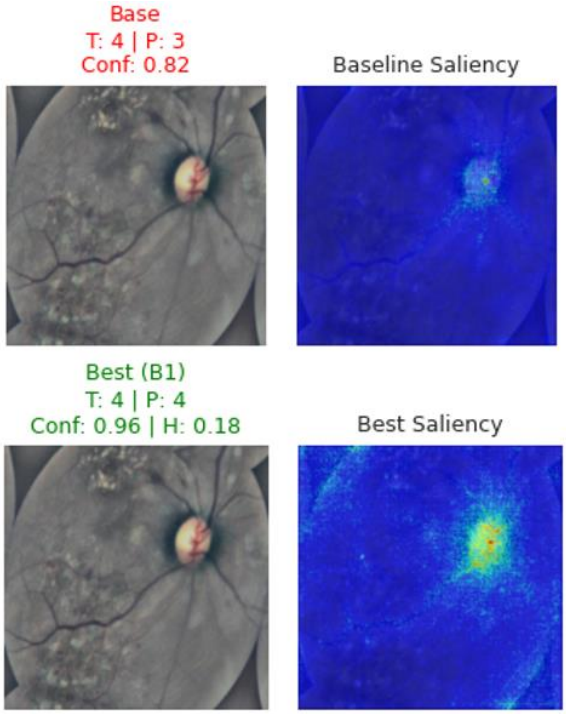


Fig.10. **Comparison of classification results for the same retinal image produced by the baseline model A0 and the best-performing model B1, including predicted class probabilities and saliency maps**
Source: compiled by the authors

As shown in the example above, the baseline model A0 incorrectly classified the image as class 3 with a confidence of 0.82, whereas the B1 model correctly identified class 4 with a confidence of 0.96.

Analysis of the saliency map demonstrates that the B1 model focuses mainly on retinal vascular structures and local pathological areas characteristic

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

of diabetic retinopathy. In contrast, it is more difficult for the base model to focus on informative features due the absence of the attention module.

A comparative summary of the best-performing architectures within each experimental configuration is provided in Table 4, while the isolated impact of individual methodological modifications is analysed in Table 5.

6. DISCUSSION OF RESULTS

The results of this study demonstrate that the combined application of Focal Loss, oversampling, and Adaptive CBAM substantially improves classification performance on the imbalanced diabetic retinopathy dataset. As shown in Table 5, the introduction of Focal Loss alone already leads to a consistent improvement over the baseline configuration.

Oversampling emerged as a key contributing factor, consistently resulting in notable growths in QWK, F1-score, and AUC values compared to class-weighted loss alone. Importantly, the results indicate that the combination of oversampling with attention mechanisms significantly improves performance regardless of the preprocessing strategy employed. This effect is particularly evident when comparing configurations A1 and B1 (Table 4).

In contrast, when applied without oversampling, attention mechanisms alone do not guarantee performance improvements. In experiments A0 and B0, the inclusion of CBAM and Adaptive CBAM resulted in only marginal gains, suggesting that architectural enhancements are most effective when supported by sufficient data diversity. This observation highlights the importance of class balancing strategies on the data preprocessing stage.

Advanced preprocessing contributed to performance improvements when combined with oversampling, as evidenced by the superior results obtained in experiment B1. The integration of Adaptive CBAM enabled the model to emphasize

Table 4. **Metrics for the architectures within experiments**

Configuration	Preprocessing	Balancing	Architecture	Acc	QWK	F1	AUC
A0 – baseline	Basic	Class-weighted loss	DenseNet121	0.69	0.77	0.52	0.89
A0 – best	Basic	Class-weighted loss	Focal	0.74	0.82	0.59	0.91
A1 – best	Basic	Oversampling	Focal + CBAM	0.84	0.91	0.84	0.97
B0 – best	Advanced	Class-weighted loss	Focal + Adaptive CBAM + MC Dropout	0.71	0.82	0.55	0.90
B1 – best	**Advanced**	**Oversampling**	**Focal + Adaptive CBAM + MC Dropout**	**0.85**	**0.92**	**0.85**	**0.97**

***Source*: compiled by the authors**

Table 5. **Impact of individual methodological modifications on model performance**

Configuration	Compared to	Modification	ΔAcc	ΔQWK	ΔF1	ΔAUC
A0 – best	A0 – baseline	Focal Loss	↑0.05	↑0.05	↑0.07	↑0.02
A1 – best	A0 – best	Oversampling + CBAM	↑0.10	↑0.09	↑0.25	↑0.06
B1 – best	A1 – best	Advanced preprocessing + Adaptive CBAM + MC Dropout	↑0.01	↑0.01	↑0.01	–
B1 – best	B0 – best	Oversampling	↑0.14	↑0.10	↑0.30	↑0.07
B1 – best	A0 – baseline	Full pipeline	↑0.16	↑0.15	↑0.33	↑0.08

***Source*: compiled by the authors**

clinically relevant retinal regions, while Monte Carlo Dropout facilitated uncertainty-aware inference. Together, these components allowed the B1 configuration to achieve an optimal balance between classification accuracy and prediction reliability.

The proposed Reliability Score provided additional insight beyond conventional performance metrics by incorporating uncertainty information into the evaluation process. Although several configurations achieved comparable accuracy values, models trained without oversampling exhibited higher predictive entropy and lower reliability scores, which is particularly critical in medical decision-making scenarios. These results indicate that, despite comparable accuracy, models may be less stable and less confident in their predictions.

Qualitative analysis using saliency maps further confirmed that the best-performing model focuses on localized vascular and pathological structures characteristic of diabetic retinopathy. These findings suggest that the proposed framework improves not only quantitative performance but also model interpretability, which is critical for clinical decision-support systems.

Overall, the B1 configuration, which combines advanced preprocessing and oversampling demonstrates the best performance, achieving an optimal balance between Precision, Recall, AUC, QWK, and Reliability Score. This makes it a strong candidate for further development in clinical decision-support applications.

The obtained results emphasize the necessity of combining data balancing strategies, architectural adaptations, and uncertainty estimation to potentially develop robust and trustworthy medical image classification systems.

7. CONCLUSIONS

The aim of this study was to develop and experimentally validate an integrated and uncertainty-aware methodology for diabetic retinopathy stage classification, ensuring high diagnostic accuracy while systematically evaluating predictive reliability. The obtained results confirm that this objective has been achieved.

The study demonstrates that architectural improvements alone are insufficient when data imbalance and image quality issues are not properly addressed. While attention-based enhancements introduced certain improvements, their effectiveness increased significantly when combined with appropriate preprocessing and class-balancing strategies. The proposed framework substantially improved the baseline configuration across key classification metrics, confirming the importance of controlled experimental evaluation of data preparation strategies.

The scientific novelty of the work lies in the development of a structured reliability-oriented framework that integrates adaptive attention mechanisms, uncertainty estimation through stochastic inference, and an original integrated Reliability Score. Unlike conventional approaches that rely solely on accuracy-based indicators, the proposed methodology enables a joint assessment of classification quality, prediction confidence, and predictive uncertainty.

The results confirm that models with comparable classification accuracy may differ significantly in terms of predictive reliability. This finding emphasizes that accuracy alone is insufficient for evaluating medical image classification systems. The introduced integrated reliability indicator provides a more informative and clinically meaningful basis for comparing deep learning models in safety-critical diagnostic applications.

The practical relevance of the study lies in the development of an uncertainty-aware diagnostic framework suitable for automated clinical decision-support systems, particularly in environments with limited access to specialized ophthalmological expertise.

The proposed framework was evaluated solely on the APTOS 2019 dataset. Additionally, class balancing through oversampling was performed prior to dataset splitting, which may introduce a potential risk of similarity between training and validation samples. Although additional augmentation was applied only to the training subset, this configuration represents a methodological limitation and should be addressed in future work. External validation on other datasets such as EyePACS or Messidor was not performed. Therefore, future research should focus on validating the framework on larger and more diverse datasets to assess its generalization capabilities across different clinical conditions.

Further investigation of alternative uncertainty estimation methods, class balancing techniques, and adaptive attention mechanisms may also improve model robustness.

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

REFERENCES

1. Alyoubi, W. L., Abulkhair, M. F. & Shalash, W. M. “Diabetic retinopathy fundus image classification and lesions localization system using deep learning”. *Sensors*. 2021; 21 (11): 3704. DOI: https://doi.org/10.3390/s21113704.

2. Bandello, F., Zarbin, M. A., Lattanzio, R. & Zucchiatti, I. “Clinical strategies in the management of diabetic retinopathy: A step-by-step guide for ophthalmologists”. *Cham, Switzerland: Springer*. 2018. DOI: https://doi.org/10.1007/978-3-319-96157-6.

3. Alqahtani, A. S., Alshareef, W. M., Aljadani, H. T., Hawsawi, W. O. & Shaheen, M. H. “The efficacy of artificial intelligence in diabetic retinopathy screening: A systematic review and meta-analysis”. *International Journal of Retina and Vitreous*. 2025; 11 (1). DOI: https://doi.org/10.1186/s40942-025-00670-9.

4. Moannaei, M., et al. “Performance and limitation of machine learning algorithms for diabetic retinopathy screening and its application in health management: A meta-analysis”. *BioMedical Engineering OnLine*. 2025; 24 (1). DOI: https://doi.org/10.1186/s12938-025-01336-1.

5. Basarab, M. R. & Ivanko, K. O. “Investigation of fundus images for detection of diabetic retinopathy stage using deep learning”. *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*. 2023; 94: 49–57. DOI: https://doi.org/10.20535/radap.2023.94.49-57.

6. Ghosh, S. & Chatterjee, A. “Introducing feature attention module on convolutional neural network for diabetic retinopathy detection”. *arXiv*. 2023. DOI: https://doi.org/10.48550/arXiv.2308.02985.

7. Harisha, M. S. & Bhosale, A. A. “Detection and classification of diabetic retinopathy using deep learning algorithms”. *arXiv*. 2024. DOI: https://doi.org/10.48550/arXiv.2401.02759.

8. Priya, R. P. R. & Aruna, P. “SVM and neural network based diagnosis of diabetic retinopathy.” *International Journal of Computer Applications*. 2012; 41 (1): 6–12. DOI: https://doi.org/10.5120/5503-7503.

9. Gulshan, V., et al. “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs”. *JAMA*. 2016; 316 (22): 2402–2410. DOI: https://doi.org/10.1001/jama.2016.17216.

10. Minarno, A. E. “Classification of diabetic retinopathy based on fundus images using InceptionV3.” *International Journal on Informatics Visualization*. 2025; 9 (1): 23–28. DOI: https://dx.doi.org/10.62527/joiv.9.1.2155.

11. Basarab, M. R. & Ivanko, K. O. “Deep learning for the detection and classification of diabetic retinopathy stages”. *Microsystems, Electronics and Acoustics*. 2024; 29 (2). DOI: https://doi.org/10.20535/2523-4455.mea.309642.

12. Romero-Oraá, R., et al. “Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading”. *Computer Methods and Programs in Biomedicine*. 2024: 108160, https://www.sciencedirect.com/science/article/pii/S0169260724001561.
DOI: https://doi.org/10.1016/j.cmpb.2024.108160.

13. Bhati, A., et al. “An interpretable dual attention network for diabetic retinopathy grading: IDANet”. *Artificial Intelligence in Medicine*. 2024. p. 102782. DOI: https://doi.org/10.1016/j.artmed.2024.102782.

14. Manoj, S. H. & Bosale, A. A. “Detection and classification of diabetic retinopathy using deep learning algorithms for segmentation to facilitate referral recommendation for test and treatment prediction”. *arXiv*. 2024. – DOI: https://doi.org/10.48550/arXiv.2401.02759.

15. Mardianta, S., Supriyanto, C. & Wijaya, A. “Diabetic retinopathy detection based on convolutional neural networks with SMOTE and CLAHE techniques”. *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2504.05696.

16. Mihai, A. & Groza, A. “Explainable fundus image curation and lesion detection in diabetic retinopathy”. *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2512.08986.

17. Ramalingam, M., Riaz, Y., Rajamanoharan, P. & Dasanayaka, P. “Enhancing safety in diabetic retinopathy detection: Uncertainty-aware deep learning models with rejection capabilities”. *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2510.00029.

18. Akram, M., Adnan, M., Ali, S. F., et al. “Uncertainty-aware diabetic retinopathy detection using deep learning enhanced by Bayesian approaches”. *Scientific Reports*. 2025; 15: 1342. DOI: https://doi.org/10.1038/s41598-024-84478-x.

19. “APTOS 2019 blindness detection dataset”. *Kaggle*. – Available from: https://kaggle.com/c/aptos2019-blindness-detection. – [Accessed: Aug. 21, 2025].

20. Sikder, N., Chowdhury, M. S., Arif, A. S. M. & Nahid, A.-A. “Early blindness detection based on retinal images using ensemble learning”. *22nd International Conference on Computer and Information Technology (ICCIT)*. Dhaka, Bangladesh. 2019. p. 1–6. DOI: https://doi.org/10.1109/ICCIT48885.2019.9038439.

21. “Kaggle diabetic retinopathy detection competition report”. *University of Warwick*. 2015. – Available from: https://kaggleforum.messageattachments.storage.googleapis.com/88655/2795/competitionreport.pdf. – [Accessed: Oct. 14, 2025].

22. Altalhan, M., Algarni, A. & Alouane, M. T.-H. “Imbalanced data problem in machine learning: A review”. *IEEE Access*. 2025; 13: 13686–13699. DOI: https://doi.org/10.1109/ACCESS.2025.3531662.

23. Huang, G., Liu, Z., Van der Maaten, L. & Weinberger, K. Q. “Densely connected convolutional networks”. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017. p. 4700–4708. DOI: https://doi.org/10.1109/CVPR.2017.243.

24. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. “Focal loss for dense object detection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020; 42 (2): 318–327. DOI: https://doi.org/10.1109/TPAMI.2018.2858826.

25. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. “CBAM: Convolutional block attention module”. In *Proc. European Conf. on Computer Vision (ECCV)*. 2018. p. 3–19. DOI: https://doi.org/10.48550/arXiv.1807.06521.

26. Gal, Y. & Ghahramani, Z. “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning”. *arXiv*. 2015. DOI: https://doi.org/10.48550/arXiv.1506.02157.

27. Grandini, M., Bagli, E. & Visani, G. “Metrics for multi-class classification: An overvie”. *arXiv*. 2020. DOI: https://doi.org/10.48550/arXiv.2008.05756.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 12.01.2026
Received after revision 11.03.2026
Accepted 18.03.2026

DOI: https://doi.org/10.15276/aait.09.2026.16
УДК 004.8:616.379-008.64-073.756.8

Діагностика діабетичної ретинопатії за допомогою глибоких нейронних мереж

Кабанова Єлізавета Ігорівна[1)]
ORCID: https://orcid.org/0009-0001-2692-5066; yelizavetakabanova@gmail.com
Кузнєцова Наталія Володимирівна[1)]
ORCID: https://orcid.org/0000-0002-1662-1974; natalia-kpi@ukr.net. Scopus Author ID: 56412465200
Іванько Катерина Олегівна[1)],
ORCID: http://orcid.org/0000-0002-3842-2423; ivanko-ee@lll.kpi.ua. Scopus Author ID: 55819298100
Кулкарні Вішвеш[2) 3)]
ORCID: https://orcid.org/0000-0002-22858652; vishwesh.kulkarni@kcl.ac.uk. Scopus Author ID: 7201425741
[1)] Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, пр. Берестейський, 37. Київ, Україна
[2)] Королівський коледж Лондона, Стренд. Лондон, WC2R 2LS, Велика Британія
[3)] SUSTech – Медична школа Королівства. Шеньчжень, 518 055, Китай

АНОТАЦІЯ

Актуальність: дослідження зумовлена високою поширеністю діабетичної ретинопатії як однієї з провідних причин сліпоти у світі та обмеженою доступністю офтальмологічної діагностики. У медичних застосуваннях оцінка невизначеності прогнозів є критично важливою, оскільки помилкові, але впевнені рішення можуть призвести до серйозних клінічних наслідків. Незважаючи на значний прогрес у розробленні систем діагностики ДР на основі глибокого навчання, недостатню увагу приділено системному аналізу взаємодії стратегій попередньої обробки даних, методів подолання дисбалансу класів та оцінки прогностичної невизначеності, що безпосередньо впливають на надійність моделі. **Мета і завдання:** метою дослідження є розроблення та експериментальна валідація автоматизованої методології класифікації стадій діабетичної ретинопатії, яка забезпечує високу діагностичну точність та водночас системно оцінює надійність і невизначеність прогнозів. **Методи:** запропонований підхід інтегрує методи комп’ютерного зору та глибокого навчання на основі

згорткових нейронних мереж із використанням трансферного навчання та запропонованого адаптивного модуля уваги, що забезпечує динамічне регулювання сили канальної та просторової уваги відповідно до дисперсії ознак. Методологія включає декілька стратегій попередньої обробки зображень, методи балансування класів і оцінювання невизначеності шляхом стохастичного виведення без зміни процедури навчання. Проведено систематичне експериментальне дослідження різних комбінацій попередньої обробки та балансування за фіксованих архітектурних умов. Ефективність моделей оцінювалася за стандартними метриками класифікації та запропонованим інтегральним індикатором надійності, яка одночасно враховує якість класифікації, впевненість прогнозів та прогностичну невизначеність. **Результати:** поєднання комплексної попередньої обробки, стратегії балансування класів та адаптивного механізму уваги суттєво підвищує як кількісні показники якості, так і надійність прогнозів. Запропонована конфігурація досягла найвищі показники, а моделі з урахуванням невизначеності продемонстрували нижчу дисперсію прогнозів і вищу впевненість, особливо у складних або неоднозначних випадках. **Висновки:** запропонований фреймворк формує структуровану методологію класифікації ДР з урахуванням надійності та забезпечує інтегрований підхід до оцінювання, що підвищує практичну придатність систем глибокого навчання для клінічної підтримки прийняття рішень.

Ключові слова: глибоке навчання; розпізнавання зображень; оцінювання невизначеності; діабетична ретинопатія; зображення сітківки ока; адаптивна увага

ABOUT THE AUTHORS

Yelizaveta I. Kabanova - Bachelor in System Analysis, Institute for Applied System Analysis. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37, Beresteiskyi Ave. Kyiv, 03056, Ukraine
ORCID: https://orcid.org/0009-0001-2692-5066; yelizavetakabanova@gmail.com
Research interests: Data Science, computer vision, neural networks, medical image analysis

Кабанова Єлізавета Ігорівна - бакалавр Навчально-наукового інституту прикладного системного аналізу. Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, пр. Берестейський, 37. Київ, 03056, Україна

Nataliia V. Kuznietsova - Doctor of Engineering Sciences, Professor, Department of Mathematical Methods of System Analysis. Institute for Applied System Analysis, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37, Beresteiskyi Ave. Kyiv, 03056, Ukraine
ORCID https://orcid.org/0000-0002-1662-1974; natalia-kpi@ukr.net**.** Scopus Author ID: 56412465200
Research interests: Data science, risks analysis, neural networks, computer science

Кузнєцова Наталія Володимирівна - доктор технічних наук, доцент, професор кафедри Математичних методів системного аналізу Навчально-наукового Інституту прикладного системного аналізу. Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, пр. Берестейський 37. Київ, 03056, Україна

Kateryna O. Ivanko - Candidate of Engineering Sciences, Associate Professor, Acting Head of Department of Electronic Engineering. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37 Beresteysky Ave. Kyiv, 03056, Ukraine
ORCID: https://orcid.org/0000-0002-3842-2423; ivanko-ee@lll.kpi.ua. Scopus Author ID: 55819298100
Research interests: Biomedical Engineering, machine learning, biomedical signal and image processing

Іванько Катерина Олегівна - кандидат технічних наук, доцент, в.о. завідуючої кафедри Електронної інженерії. Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, пр. Берестейський ,37. Київ, 03056, Україна

Vishwesh Kulkarni - PhD in Electrical Engineering, Senior Lecturer, School of Biomedical Engineering & Imaging Sciences, King’s College London; Programme Lead, Biomedical Engineering, SUSTech-King’s School of Medicine (SKMed); King’s College London, Strand, London WC2R 2LS, UK; SUSTech-King’s School of Medicine, Shenzhen 518 055, China
ORCID: https://orcid.org/0000-0002-22858652; vishwesh.kulkarni@kcl.ac.uk. Scopus Author ID: 7201425741
Research interests: Biomedical Engineering, computer science, machine learning, biomedical signals

Кулкарні Вішвеш - доктор філософії з електротехніки, старший викладач. Школа біомедичної інженерії та наук візуалізації. Королівський коледж Лондона; керівник програми з біомедичної інженерії, Медична школа SUSTech-King’s (SKMed); Королівський коледж Лондона, Стренд, Лондон WC2R 2LS, Велика Британія; Медична школа SUSTech-King’s, Шеньчжень 518 055, Китай