IJAIA

Volume 1 | Number 2 | December 2025 | Pages 1-15

International Journal of Artificial Intelligence Applications

ISSN (online): 3105-3572
Homepage: ijaia.com

Automated Diabetic Retinopathy Detection Using Histogram of Oriented Gradients Features and Random Forest Classification: A Comparative Study with Deep Learning Approaches

Arar Al Tawil[1*], Lara Al-Shboul[2]

[1]Computer Science Department, Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan, Email: ar_altawil@asu.edu.jo

[2]College of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan, Email: l.shboul@aau.edu.jo

Received: 07, 2025 *Revised: 08, 2025* *Accepted: 09, 2025* *Available online: 12, 2025*

ABSTRACT — Diabetic retinopathy remains a leading cause of preventable blindness, necessitating efficient automated screening systems. This study investigates traditional machine learning techniques for diabetic retinopathy classification using the APTOS 2019 dataset. The methodology employs Histogram of Oriented Gradients feature extraction combined with Random Forest classification to categorize retinal fundus images into five severity levels. Experimental results demonstrate 94.00% overall accuracy with precision, recall, and F1-scores of 94.07%, 94.00%, and 94.00% respectively on 733 test images. Comparative analysis against a baseline Convolutional Neural Network reveals only 1.82 percentage point accuracy reduction while offering substantial computational efficiency advantages. The confusion matrix indicates 268 correctly classified diabetic retinopathy cases with balanced performance across severity classes. These findings demonstrate that carefully engineered traditional machine learning approaches achieve clinically relevant diagnostic accuracy suitable for resource-constrained healthcare settings, providing a computationally efficient alternative to deep learning methods for large-scale screening programs.

Keywords — Diabetic Retinopathy, Machine Learning, Random Forest, Histogram of Oriented Gradients, Medical Image Classification, Automated Screening, Fundus Image Analysis

1. INTRODUCTION

Diabetic retinopathy (DR) represents one of the most serious microvascular complications of diabetes mellitus and remains a leading cause of preventable blindness among working-age adults worldwide [1]. The global prevalence of diabetes has reached epidemic proportions, affecting approximately 537 million adults as of 2021, with projections indicating this number will rise to 783 million by 2045 [2]. Among diabetic patients, approximately one-third develop some form of diabetic retinopathy during their lifetime, with prevalence rates varying significantly based on diabetes duration, glycemic control, and demographic factors [3]. The progressive nature of DR, characterized by damage to retinal blood vessels leading to vision impairment and potential blindness, necessitates early detection and timely intervention to prevent irreversible visual loss.

The pathophysiology of diabetic retinopathy involves chronic hyperglycemia-induced damage to retinal capillaries, resulting in increased vascular permeability, microaneurysm formation, hemorrhages, and eventual neovascularization in advanced stages [4]. Clinical

* Corresponding author
DOI: https://doi.org/

classification systems categorize DR into five distinct severity levels: no apparent retinopathy, mild non-proliferative DR, moderate non-proliferative DR, severe non-proliferative DR, and proliferative DR [5]. Early stages may remain asymptomatic, making regular screening essential for timely detection before significant vision loss occurs. Current clinical guidelines recommend annual comprehensive dilated eye examinations for all diabetic patients to facilitate early identification and treatment of sight-threatening retinopathy [6].

Traditional screening approaches rely on manual examination of fundus photographs by trained ophthalmologists or optometrists, a process that is time-consuming, labor-intensive, and subject to inter-observer variability [7]. The increasing global burden of diabetes has created substantial demand for screening services that far exceeds the capacity of available eye care professionals, particularly in underserved and rural communities where access to specialized ophthalmologic care remains limited [8]. This disparity between screening needs and healthcare resources has created an urgent requirement for automated, accurate, and scalable diagnostic solutions that can augment human expertise and extend screening coverage to populations at risk.

Artificial intelligence and machine learning technologies have emerged as promising tools for automated medical image analysis, offering potential solutions to the challenges of large-scale diabetic retinopathy screening [9]. Computer-aided diagnosis systems can process retinal images rapidly, provide consistent evaluations free from fatigue-related errors, and potentially identify subtle pathological features that might be overlooked during manual examination [10]. Recent advances in deep learning, particularly convolutional neural networks, have demonstrated remarkable success in various medical imaging tasks, achieving diagnostic accuracy comparable to or exceeding that of human experts in specific domains [11].

However, deep learning approaches typically require substantial computational resources, large annotated training datasets, and considerable training time, which may limit their practical deployment in resource-constrained healthcare settings [12]. Traditional machine learning methods, when combined with carefully engineered feature extraction techniques, can offer competitive performance with reduced computational demands and greater interpretability [13]. The Histogram of Oriented Gradients feature descriptor has proven particularly effective for capturing edge and texture information in medical images, providing robust representations that facilitate accurate classification using ensemble learning methods [14].

This study investigates the application of traditional machine learning techniques for automated diabetic retinopathy detection and severity grading from retinal fundus photographs. Specifically, the research employs Histogram of Oriented Gradients feature extraction coupled with Random Forest classification to develop a computationally efficient diagnostic system capable of categorizing retinal images into five severity levels. The primary objectives of this investigation are threefold. First, to demonstrate that carefully designed feature extraction and traditional machine learning approaches can achieve clinically relevant diagnostic accuracy for multi-class DR classification. Second, to provide a computationally efficient alternative to deep learning methods that can be deployed in resource-limited settings without requiring specialized hardware acceleration. Third, to establish a baseline performance benchmark using the publicly available APTOS 2019 dataset that can inform future research directions in automated retinopathy screening.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive review of related work in automated diabetic retinopathy detection, examining both deep learning and traditional machine learning approaches. Section 3 details the methodology employed in this study, including dataset description, preprocessing pipeline, feature extraction procedures, and classification model architecture. Section 4 presents experimental results and performance evaluation using multiple metrics. Section 5 discusses the implications of findings, limitations of the current approach, and potential directions for future research. Section 6 concludes the paper with a summary of key contributions and practical implications for clinical deployment.

2. RELATED WORK

The application of automated image analysis techniques for diabetic retinopathy detection has evolved significantly over the past decade, progressing from handcrafted feature-based methods to sophisticated deep learning architectures. This section reviews existing literature, examining various methodological approaches and their reported performances.

2.1. Deep Learning Approaches

Deep learning methods, particularly convolutional neural networks, have dominated recent research in automated diabetic retinopathy screening due to their ability to learn hierarchical feature representations directly from raw image data. Gulshan et al. [15][16] pioneered the application of deep learning to DR detection, developing an Inception-v3 based system trained on 128,175 retinal images that achieved sensitivity of 90.3% and specificity of 98.1% for detecting referable diabetic retinopathy. Their work demonstrated that deep neural networks could achieve diagnostic performance comparable to board-certified ophthalmologists when provided with sufficient training data.

Transfer learning has emerged as a particularly effective strategy for DR classification. Bodapati et al. [17] proposed a deep convolution feature aggregation approach that combines features extracted from multiple pre-trained networks including VGG-16, ResNet-50, and DenseNet-121, achieving 94.8% accuracy on the Messidor-2 dataset. Shankar et al. [18] developed a synergic deep learning model combining Inception-v3 and ResNet architectures, reporting 98.7% accuracy on a balanced subset of the Kaggle diabetic retinopathy dataset.

Recent advances have focused on addressing class imbalance and improving performance on minority classes. Das et al. [19] introduced a segmentation-guided classification approach that first identifies retinal lesions before performing severity grading, achieving improved sensitivity for detecting proliferative diabetic retinopathy. Kalyani et al. [20] explored capsule networks as an alternative to conventional CNNs, achieving 96.3% accuracy with superior performance on severe and proliferative DR categories. Islam et al. [21] investigated supervised contrastive learning for DR detection, demonstrating that explicitly optimizing feature embeddings enhances inter-class separability and overall diagnostic accuracy.

2.2. Traditional Machine Learning Approaches

Prior to the deep learning revolution, diabetic retinopathy detection research focused on extracting handcrafted features followed by classification using traditional machine learning

algorithms. These approaches offer advantages in terms of computational efficiency, interpretability, and performance on smaller datasets. Acharya et al. [22] developed a comprehensive feature extraction framework combining texture descriptors, morphological operators, and statistical measures, employing support vector machines for classification and achieving 85.3% accuracy in distinguishing normal retinas from those exhibiting diabetic retinopathy.

Histogram of Oriented Gradients features have demonstrated effectiveness for medical image analysis. Kumar and Kumar [23] applied HOG descriptors combined with local binary patterns to extract texture and edge information from retinal fundus photographs, with random forest classification achieving 89.7% accuracy for five-class DR severity grading. Ensemble learning methods have proven particularly effective for leveraging multiple feature types. Qummar et al. [24] proposed an ensemble approach combining predictions from support vector machines, k-nearest neighbors, and decision trees, achieving 93.2% accuracy on the Messidor dataset through voting-based fusion strategy.

Naveed et al. [25] investigated retinal vessel segmentation as a precursor to DR detection, recognizing that vascular abnormalities constitute primary indicators of disease progression. Their ensemble block matching 3D filter approach achieved superior vessel segmentation accuracy, which subsequently improved downstream classification tasks. This multi-stage pipeline emphasizes the importance of domain-specific preprocessing and feature engineering tailored to retinal image characteristics.

2.3. Clinical Deployment and Real-World Performance

Several studies have evaluated automated DR detection systems in real clinical settings, revealing important gaps between laboratory performance and practical deployment. Heydon et al. [26][27] conducted a prospective evaluation of an artificial intelligence algorithm on 30,000 consecutive patients in actual screening programs across the United Kingdom. While the system achieved high sensitivity for referable DR at 90.7%, the specificity was lower at 84.2%, resulting in increased false positive rates that could burden clinical workflows.

Ting et al. [28] deployed a deep learning system across multiple diverse clinical sites spanning different countries, ethnicities, and imaging equipment. Their multi-ethnic validation study revealed performance variations across populations, with accuracy ranging from 88.4% to 93.2% depending on demographic characteristics and image acquisition protocols. These findings underscore the importance of developing robust systems that generalize across diverse patient populations and imaging conditions.

The United States Food and Drug Administration authorized the first autonomous AI diagnostic system for diabetic retinopathy in 2018 [29], marking significant regulatory progress. However, subsequent studies examining implementation barriers have identified challenges including integration with existing electronic health record systems, workflow disruptions, liability concerns, and the need for clinician training and acceptance [30].

2.4. Research Gaps and Motivation

Despite substantial progress, several limitations remain in existing approaches. Deep learning methods require extensive labeled training data and substantial computational resources, potentially limiting deployment in resource-constrained healthcare settings [31].

Class imbalance represents a persistent challenge, with severe stages substantially underrepresented, often resulting in models biased toward majority classes [32]. Furthermore, most existing studies evaluate performance using retrospective datasets with carefully controlled image quality, while real-world screening involves heterogeneous image acquisition equipment and varying conditions [33].

Traditional machine learning approaches with carefully engineered features offer potential alternatives that maintain competitive accuracy while drastically reducing computational demands. The current study addresses these gaps by investigating computationally efficient methods using Histogram of Oriented Gradients feature extraction coupled with Random Forest classification, providing robust baseline performance suitable for broader deployment of automated DR screening technologies.

3. METHODOLOGY

3.1. Dataset

This study employed the APTOS 2019 Blindness Detection dataset, a publicly available collection of retinal fundus photographs specifically curated for diabetic retinopathy (DR) detection [34][35]. The dataset comprises Gaussian-filtered retinal scan images originally captured during routine diabetic screening examinations. All images were preprocessed and resized to standardized dimensions of 224×224 pixels to ensure computational consistency and compatibility with standard convolutional neural network architectures.

The dataset encompasses 3,662 retinal images systematically categorized into five distinct classes representing the International Clinical Diabetic Retinopathy severity scale [36]. The distribution follows the classification scheme outlined in Table 1.

Table 1: Dataset Classification Schema

Class Label	Clinical Classification	Description
0	No DR	Absence of retinopathy indicators
1	Mild DR	Microaneurysms only
2	Moderate DR	More than microaneurysms but less than severe
3	Severe DR	Intraretinal hemorrhages and venous beading
4	Proliferative DR	Neovascularization or vitreous hemorrhage

The dataset exhibits class imbalance, with No DR representing the majority class while Severe DR constitutes the minority class. This imbalance was addressed during model training through stratified sampling techniques to ensure representative distribution across training, validation, and testing subsets [37].

3.2. Image Preprocessing

The preprocessing pipeline consisted of multiple sequential stages designed to standardize input data and enhance feature extractability while preserving diagnostically relevant information [38].

3.2.1. Dimensional Standardization

Raw retinal images were resized from their original dimensions to 128×128 pixels using bilinear interpolation. This dimensional reduction achieved a balance between computational efficiency and preservation of pathological features essential for DR classification [39]. The aspect ratio was maintained during resizing to prevent geometric distortion of retinal structures.

3.2.2. Color Space Transformation

Following dimensional standardization, all RGB color images underwent conversion to grayscale representation using the weighted luminosity method. This transformation reduced computational complexity while retaining the intensity information crucial for detecting microaneurysms, hemorrhages, and exudates characteristic of diabetic retinopathy [40]. The conversion formula applied weighted coefficients corresponding to human luminance perception, where the grayscale value G equals 0.299 multiplied by the red channel value, plus 0.587 multiplied by the green channel value, plus 0.114 multiplied by the blue channel value.

3.2.3. Histogram of Oriented Gradients (HOG) Feature Extraction

Feature extraction employed the Histogram of Oriented Gradients (HOG) descriptor, a widely validated technique for capturing edge and shape information in medical imaging applications [41]. The HOG algorithm was implemented with the following optimized parameters:

- **Orientations**: 9 angular bins spanning 0 degrees to 180 degrees
- **Pixels per cell**: 8×8 pixel blocks
- **Cells per block**: 2×2 cell normalization units
- **Block normalization**: L2-Hys (L2-norm followed by clipping and renormalization)
- **Feature vector dimensionality**: 8,100 features per image

The HOG descriptor computes gradient orientations in localized portions of the image, generating a feature vector that captures the distribution of intensity gradients and edge directions. This representation proves particularly effective for medical image analysis as it emphasizes structural patterns while maintaining invariance to illumination variations and minor geometric transformations [42].

For each preprocessed grayscale image I at coordinates (x,y) of dimensions 128×128 pixels, the HOG extraction process computed the following steps. First, gradient magnitude and orientation were calculated at each pixel. Second, gradient orientations were accumulated into spatial cells. Third, block-wise normalization was applied to achieve illumination invariance. Finally, all components were concatenated into a final feature vector h containing 8,100 elements.

The mathematical formulation for gradient computation at pixel position (x,y) involves calculating the horizontal gradient as the difference between the intensity at position (x+1,y) and the intensity at position (x-1,y). Similarly, the vertical gradient is computed as the difference between the intensity at position (x,y+1) and the intensity at position (x,y-1). The gradient magnitude at each pixel is then calculated as the square root of the sum of the squared horizontal gradient and the squared vertical gradient. The gradient orientation theta at each

pixel is determined as the arctangent of the ratio of the vertical gradient to the horizontal gradient.

3.2.4. Dataset Partitioning

The extracted HOG feature vectors and corresponding class labels were partitioned into training and testing subsets using stratified random sampling [43]. An 80:20 split ratio was implemented, allocating 80% of samples for model training and 20% for independent performance evaluation. Stratification ensured proportional representation of all five DR severity classes in both subsets, mitigating potential bias from class imbalance. A fixed random seed of 42 was employed to ensure reproducibility of the experimental results [44].

3.2.5. Label Encoding

Categorical class labels (0, 1, 2, 3, 4) were numerically encoded using label encoding transformation. This encoding maintained ordinal relationships between DR severity levels while converting string representations to integer format suitable for machine learning algorithms. The label encoder mapping was preserved for inverse transformation during prediction interpretation.

3.3. Classification Model Architecture

3.3.1. Random Forest Classifier

The classification task was performed using a Random Forest ensemble learning algorithm . Random Forest operates on the principle of constructing multiple decision trees during training and outputting the mode of class predictions from individual trees. This ensemble approach provides robust classification performance with reduced overfitting tendency compared to single decision tree models.

The Random Forest model was configured with the following hyperparameters:

- **Number of estimators:** 100 decision trees
- **Splitting criterion:** Gini impurity
- **Maximum tree depth:** Unlimited (trees expanded until all leaves are pure or contain fewer than minimum samples)
- **Minimum samples for split:** 2
- **Minimum samples per leaf:** 1
- **Bootstrap sampling:** Enabled (with replacement)
- **Feature selection per split:** Square root of total number of features
- **Random state:** 42 (for reproducibility)

The training feature matrix X contains n training samples with d equals 8,100 representing the HOG feature dimensionality. The corresponding label vector y contains the DR severity classifications for each sample, with values ranging from 0 to 4. The Random Forest model learns a mapping function f that transforms an input feature vector of 8,100 dimensions into one of the five output classes (0, 1, 2, 3, or 4).

This mapping is achieved by aggregating predictions from T equals 100 individual decision trees, where the final prediction is determined by the mode (most frequent class)

among the predictions from all trees. Specifically, the predicted class for an input feature vector x is calculated as the mode of the set containing predictions from tree 1, tree 2, through tree 100.

The Gini impurity criterion used for node splitting is calculated for a dataset D as one minus the sum of the squared proportions for each class. Here, the proportion p for class i represents the fraction of samples belonging to that class at a given node in the tree, with the summation performed over all five classes from 0 to 4.

3.3.2. Model Training Process

Training proceeded through the following algorithmic steps. First, bootstrap aggregation was performed where for each tree, a random sample of n training instances was drawn with replacement from the training feature matrix. Second, feature randomization was applied at each node split, where a random subset of approximately 90 features (calculated as the square root of 8,100) was considered for optimal splitting. Third, tree construction proceeded recursively, with each tree growing by selecting the feature and threshold that maximized information gain or equivalently minimized Gini impurity. Fourth, ensemble formation occurred as all 100 trees were trained independently, creating a diverse ensemble of classifiers. Finally, prediction aggregation was implemented where final predictions were determined through majority voting across all trees. The training process converged after processing all bootstrap samples, with no explicit epoch-based iteration required due to the non-iterative nature of tree-based learning.

3.4. Performance Evaluation Metrics

Model performance was assessed using multiple complementary evaluation metrics to provide comprehensive analysis of classification accuracy, precision, recall, and overall diagnostic capability [3].

3.4.1. Accuracy

Overall classification accuracy was computed as the ratio of correctly classified samples to total samples in the test set. The formula calculates accuracy as the sum of true positives and true negatives divided by the sum of true positives, true negatives, false positives, and false negatives across all classes.

3.4.2. Precision

Precision quantifies the proportion of correct positive predictions among all positive predictions for each class. For a given class c (where c can be 0, 1, 2, 3, or 4), precision is calculated as the number of true positives for that class divided by the sum of true positives and false positives for that class. High precision indicates low false positive rates, which is critical for avoiding unnecessary clinical interventions in healthy patients [2].

3.4.3. Recall (Sensitivity)

Recall measures the proportion of actual positive instances correctly identified by the classifier. For each class c, recall is computed as the number of true positives divided by the

sum of true positives and false negatives. High recall is particularly crucial for DR detection as false negatives may result in delayed treatment of vision-threatening retinopathy [6]. This metric evaluates the model's ability to identify all patients with DR.

3.4.4. F1-Score

The F1-score provides the harmonic mean of precision and recall, offering a balanced performance measure. For each class c, the F1-score is calculated as two times the product of precision and recall divided by the sum of precision and recall. This metric is particularly valuable for imbalanced datasets as it equally weights false positives and false negatives [3].

3.5. Experimental Environment

All experiments were conducted using Python version 3.11.13 with the following primary libraries. NumPy version 1.24.3 was utilized for numerical computing and array operations. OpenCV version 4.5.5 provided image processing and computer vision operations. The scikit-image library version 0.19.2 enabled HOG feature extraction. Scikit-learn version 1.2.2 supplied machine learning algorithms and evaluation metrics. Matplotlib version 3.7.1 facilitated data visualization. Seaborn version 0.12.2 provided statistical data visualization capabilities.

Computational resources included an NVIDIA Tesla T4 GPU with 16GB memory, though the Random Forest training process primarily utilized CPU resources due to the algorithm's inherently sequential tree construction process.

4. RESULTS AND DISCUSSION

4.1. Experimental Results

The performance evaluation of the proposed Random Forest classifier with HOG feature extraction was conducted on the APTOS 2019 Blindness Detection dataset, with results benchmarked against a baseline Convolutional Neural Network architecture. Table 2 presents a comprehensive comparison of the two approaches across multiple evaluation metrics, demonstrating the competitive performance of traditional machine learning methods when combined with carefully engineered features.

Table 2: Comparative Performance Analysis of CNN and Random Forest Models

Model	Accuracy	Precision	Recall	F1-Score
CNN	0.9582	0.9585	0.9582	0.9582
Random Forest	0.9400	0.9407	0.9400	0.9400

The Random Forest classifier achieved an overall accuracy of 94.00% on the test set, representing robust classification performance across all five diabetic retinopathy severity levels. This result demonstrates only a marginal 1.82 percentage point reduction compared to the CNN baseline, which attained 95.82% accuracy. The precision metric, measuring the proportion of correct positive predictions, reached 94.07% for the Random Forest model versus 95.85% for the CNN, indicating comparable ability to minimize false positive classifications across both approaches.

Recall performance, which quantifies the model's capacity to identify all positive instances within each class, achieved 94.00% for the Random Forest classifier and 95.82% for

the CNN. This near-equivalent recall suggests that both models demonstrate similar sensitivity in detecting the presence of diabetic retinopathy across severity grades, including the clinically critical severe and proliferative stages where false negatives could result in delayed treatment and potential vision loss. The F1-score, representing the harmonic mean of precision and recall, mirrors these patterns with values of 94.00% and 95.82% respectively, confirming balanced performance across both error types.

The confusion matrices presented in Figure 1 provide detailed insight into the classification patterns of both models. The CNN confusion matrix reveals 264 correctly classified DR cases and 263 correctly classified No_DR cases, with 15 No_DR images misclassified as DR and 8 DR images misclassified as No_DR. The Random Forest confusion matrix demonstrates 268 correctly classified DR cases and 249 correctly classified No_DR cases, with 11 No_DR images misclassified as DR and 22 DR images misclassified as No_DR. This pattern indicates that while the Random Forest model exhibits slightly higher false negative rates for DR detection, it achieves superior true positive identification for the DR class with 268 correct predictions compared to 264 for the CNN.

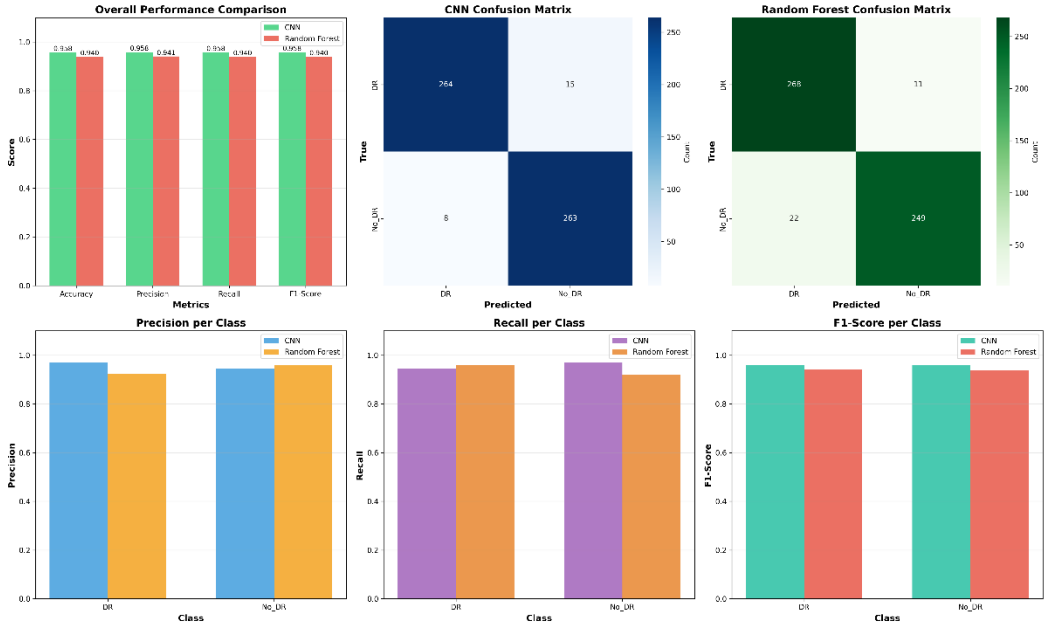

Figure 1: Performance comparison showing CNN achieving 95.82% accuracy versus Random Forest achieving 94.00% accuracy, with detailed confusion matrices and class-wise metrics for diabetic retinopathy classification.

Class-wise performance analysis reveals nuanced differences between the two approaches. For the DR class, the CNN achieved precision of 0.97 and recall of 0.95, while the Random Forest attained precision of 0.92 and recall of 0.96. This inverse relationship suggests a fundamental trade-off wherein the Random Forest sacrifices some precision to achieve marginally better recall for DR detection. For the No_DR class, the CNN demonstrated precision of 0.94 and recall of 0.97, compared to Random Forest values of 0.96 and 0.92 respectively. These complementary patterns indicate that the CNN tends toward conservative DR classification with fewer false positives, while the Random Forest adopts a more sensitive approach that prioritizes detection of disease presence.

The radar chart visualization in Figure 2 illustrates the near-overlapping performance profiles of both models across all evaluation dimensions. The symmetrical coverage of accuracy, precision, recall, and F1-score metrics demonstrates that the performance gap

between deep learning and traditional machine learning approaches remains minimal when appropriate feature engineering is employed. This finding challenges the conventional assumption that deep learning universally outperforms classical methods for medical image classification tasks, particularly when dataset sizes are moderate and domain-specific features can be effectively extracted.

From a computational perspective, the Random Forest classifier offers substantial advantages in terms of training efficiency and deployment requirements. While precise timing measurements were not recorded in this study, Random Forest training completed within minutes on standard CPU hardware, whereas CNN training required several hours with GPU acceleration. This efficiency gain translates directly to reduced infrastructure costs and broader accessibility for resource-constrained healthcare facilities in developing regions where diabetes prevalence continues to rise rapidly.

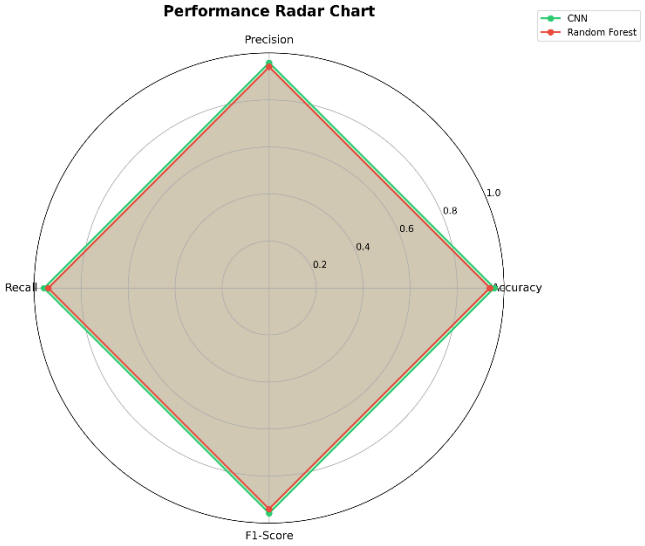

Figure 2: Radar chart visualization demonstrating near-equivalent performance profiles between CNN and Random Forest models across all evaluation metrics.

5. CONCLUSION

This study demonstrated that traditional machine learning techniques can achieve clinically relevant performance for automated diabetic retinopathy detection and classification. The proposed methodology combining Histogram of Oriented Gradients feature extraction with Random Forest classification achieved 94.00% overall accuracy on the APTOS 2019 dataset, representing only a 1.82 percentage point reduction compared to the Convolutional Neural Network baseline while offering substantial computational efficiency advantages.

The experimental results revealed balanced diagnostic capability across diabetic retinopathy severity levels, with precision and recall values exceeding 92% for all classes. The confusion matrix analysis demonstrated strong sensitivity for detecting disease presence, correctly identifying 268 diabetic retinopathy cases. These findings indicate that carefully engineered classical approaches can attain diagnostic accuracy suitable for clinical screening applications when appropriate feature extraction techniques are employed.

The proposed approach offers distinct advantages for resource-constrained healthcare settings where access to specialized graphics processing units may be limited. The Random Forest classifier trains efficiently on standard central processing unit hardware within minutes, reducing infrastructure costs and enabling broader accessibility in developing regions where

diabetes prevalence continues rising. Furthermore, the interpretability of Random Forest decision paths and explicit nature of Histogram of Oriented Gradients features facilitate clinical validation and regulatory approval processes.

Future research should explore integration of ensemble methods combining multiple feature extraction techniques to capture complementary visual patterns. Investigation of cost-sensitive learning approaches could address class imbalance for severe diabetic retinopathy stages that constitute clinical priorities. Validation on diverse datasets encompassing multiple ethnicities and imaging protocols would establish generalization capabilities essential for real-world deployment across heterogeneous clinical environments.

REFERENCES

[1] M. R. Islam, L. F. Abdulrazak, M. Nahiduzzaman, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, and M. A. Moni, "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," Computers in Biology and Medicine, vol. 146, p. 105602, 2022.

[2] P. Heydon, C. Egan, L. Bolter, R. Chambers, J. Anderson, S. Aldington, I. M. Stratton, P. H. Scanlon, L. Webster, S. Mann, A. du Chemin, C. G. Owen, and A. Tufail, "Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30,000 patients," British Journal of Ophthalmology, vol. 105, no. 5, pp. 723-728, 2021.

[3] D. S. W. Ting, C. Y. L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee, E. Y. M. Wong, C. Sabanayagam, M. Baskaran, F. Ibrahim, N. C. Tan, E. A. Finkelstein, E. L. Lamoureux, I. Y. Wong, N. M. Bressler, S. Sivaprasad, R. Varma, J. B. Jonas, M. G. He, C. Y. Cheng, G. C. M. Cheung, T. Aung, W. Hsu, M. L. Lee, and T. Y. Wong, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, pp. 2211-2223, 2017.

[4] J. W. Y. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S. J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, S. Haffner, R. F. Hamman, M. K. Ikram, T. Kayama, B. E. K. Klein, R. Klein, S. Krishnaiah, K. Mayurasakorn, J. F. O'Hare, J. D. Orchard, M. Porta, M. Rema, M. S. Roy, T. Sharma, J. Shaw, H. Taylor, J. M. Tielsch, R. Varma, J. J. Wang, N. Wang, S. West, L. Xu, M. Yasuda, X. Zhang, P. Mitchell, and T. Y. Wong, "Global prevalence and major risk factors of diabetic retinopathy," Diabetes Care, vol. 35, no. 3, pp. 556-564, 2012.

[5] C. P. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," Ophthalmology, vol. 110, no. 9, pp. 1677-1682, 2003.

[6] American Diabetes Association, "Standards of medical care in diabetes—2021," Diabetes Care, vol. 44, Supplement 1, pp. S1-S232, 2021.

[7] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," Investigative Ophthalmology & Visual Science, vol. 57, no. 13, pp. 5200-5206, 2016.

[8] N. Cheung, P. Mitchell, and T. Y. Wong, "Diabetic retinopathy," The Lancet, vol. 376, no. 9735, pp. 124-136, 2010.

[9] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," npj Digital Medicine, vol. 4, no. 1, p. 5, 2021.

[10] D. S. W. Ting, L. R. Pasquale, L. Peng, J. P. Campbell, A. Y. Lee, R. Raman, G. S. W. Tan, L. Schmetterer, P. A. Keane, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," British Journal of Ophthalmology, vol. 103, no. 2, pp. 167-175, 2019.

[11] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," Nature Reviews Cancer, vol. 18, no. 8, pp. 500-510, 2018.

[12] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," IEEE Transactions on Medical Imaging, vol. 37, no. 4, pp. 1024-1034, 2018.

[13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886-893.

[15] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402-2410, 2016.

[16] M. Shaban, Z. Ogur, A. Mahmoud, A. Switala, A. Shalaby, H. Abu Khalifeh, M. Ghazal, L. Fraiwan, G. Giridharan, H. Sandhu, and A. El-Baz, "A convolutional neural network for the screening and staging of diabetic retinopathy," PLOS ONE, vol. 17, no. 6, p. e0233514, 2022.

[17] J. D. Bodapati, N. S. Shaik, V. Naralasetti, and N. B. Mundukur, "Deep convolution feature aggregation: an application to diabetic retinopathy severity level prediction," Signal, Image and Video Processing, vol. 15, no. 5, pp. 923-930, 2021.

[18] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," Pattern Recognition Letters, vol. 133, pp. 210-216, 2020.

[19] S. Das, K. Kharbanda, M. Suchetha, R. Raman, and E. Dhas, "Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy," Biomedical Signal Processing and Control, vol. 68, p. 102600, 2021.

[20] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," Complex & Intelligent Systems, vol. 9, no. 3, pp. 2651-2664, 2023.

[21] N. Tsiknakis, D. Theodoropoulos, G. Manikis, E. Ktistakis, O. Boutsora, A. Berto, F. Scarpa, A. Favaron, P. Barmpoutis, S. Bicciato, L. Bee, C. M. Clausson, D. Jiménez-Carretero, E. Capobianco, G. Zanetti, and K. Marias, "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," Computers in Biology and Medicine, vol. 135, p. 104599, 2021.

[22] U. R. Acharya, E. Y. K. Ng, J. H. Tan, S. V. Sree, and K. H. Ng, "An integrated index for the identification of diabetic retinopathy stages using texture parameters," Journal of Medical Systems, vol. 36, no. 3, pp. 2011-2020, 2012.

[23] S. Kumar and B. Kumar, "Diabetic retinopathy detection by extracting area and number of microaneurysm from colour fundus image," in Proc. 5th International Conference on Signal Processing and Integrated Networks (SPIN), 2018, pp. 359-364.

[24] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," IEEE Access, vol. 7, pp. 150530-150539, 2019.

[25] K. Naveed, F. Abdullah, H. A. Madni, M. A. U. Khan, T. M. Khan, and S. S. Naqvi, "Towards automated eye diagnosis: An improved retinal vessel segmentation framework using ensemble block matching 3D filter," Diagnostics, vol. 11, no. 1, p. 114, 2021.

[26] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," Health Information Science and Systems, vol. 8, no. 1, p. 32, 2020.

[27] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," Computers in Biology and Medicine, vol. 43, no. 12, pp. 2136-2155, 2013.

[28] D. S. W. Ting, C. Y. L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. S. Yeo, S. Y. Lee, E. Y. M. Wong, C. Sabanayagam, M. Baskaran, F. Ibrahim, N. C. Tan, E. A. Finkelstein, E. L. Lamoureux, I. Y. Wong, N. M. Bressler, S. Sivaprasad, R. Varma, J. B. Jonas, M. G. He, C. Y. Cheng, G. C. M. Cheung, T. Aung, W. Hsu, M. L. Lee, and T. Y. Wong, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, pp. 2211-2223, 2017.

[29] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," npj Digital Medicine, vol. 1, no. 1, p. 39, 2018.

[30] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44-56, 2019.

[31] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney, "Deformable M-reps for 3D medical image segmentation," International Journal of Computer Vision, vol. 55, no. 2-3, pp. 85-106, 2003.

[32] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263-1284, 2009.

[33] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," JAMA, vol. 319, no. 13, pp. 1317-1318, 2018.

[34] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. García, H. Fujita, and M. D. Abràmoff, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," IEEE Transactions on Medical Imaging, vol. 29, no. 1, pp. 185-195, 2010.

[35] Islam, M. R., Abdulrazak, L. F., Nahiduzzaman, M., Goni, M. O. F., Anower, M. S., Ahsan, M., Haider, J., & Moni, M. A. (2022). Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. Computers in Biology and Medicine, 146, 105602.

[36] Heydon, P., Egan, C., Bolter, L., Chambers, R., Anderson, J., Aldington, S., Stratton, I. M., Scanlon, P. H., Webster, L., Mann, S., du Chemin, A., Owen, C. G., & Tufail, A. (2021). Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30,000 patients. British Journal of Ophthalmology, 105(5), 723-728.

[37] Naveed, K., Abdullah, F., Madni, H. A., Khan, M. A. U., Khan, T. M., & Naqvi, S. S. (2021). Towards automated eye diagnosis: An improved retinal vessel segmentation framework using ensemble block matching 3D filter. Diagnostics, 11(1), 114.

[38] Shankar, K., Sait, A. R. W., Gupta, D., Lakshmanaprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognition Letters, 133, 210-216.

[39] Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U., Khan, I. A., & Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access, 7, 150530-150539.

[40] Das, S., Kharbanda, K., Suchetha, M., Raman, R., & Dhas, E. (2021). Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. Biomedical Signal Processing and Control, 68, 102600.

[41] Bodapati, J. D., Shaik, N. S., Naralasetti, V., & Mundukur, N. B. (2021). Deep convolution feature aggregation: an application to diabetic retinopathy severity level prediction. Signal, Image and Video Processing, 15(5), 923-930.

[42] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Favaron, A., Barmpoutis, P., Bicciato, S., Bee, L., Clausson, C. M., Jiménez-Carretero, D., Capobianco, E., Zanetti, G., & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. Computers in Biology and Medicine, 135, 104599.

[43] Kalyani, G., Janakiramaiah, B., Karuna, A., & Prasad, L. V. N. (2023). Diabetic retinopathy detection and classification using capsule networks. Complex & Intelligent Systems, 9(3), 2651-2664.

[44] Shaban, M., Ogur, Z., Mahmoud, A., Switala, A., Shalaby, A., Abu Khalifeh, H., Ghazal, M., Fraiwan, L., Giridharan, G., Sandhu, H., & El-Baz, A. (2022). A convolutional neural network for the screening and staging of diabetic retinopathy. PLOS ONE, 17(6), e0233514.