iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.79050

www.ijraset.com

Call: ✆08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Automated Five-Class Diabetic Retinopathy Detection Using Transfer Learning and Web-Based Clinical Deployment

Aravinda Karri[1], M. Radhika Mani[2]
[1]M.Tech Scholar, [2]Professor, Department of Computer Science and Engineering, Pragati Engineering College, Surampalem, India

Abstract: Diabetic Retinopathy (DR) is a serious eye condition caused by diabetes and is a major contributor to vision impairment across the world. Detecting the disease at an early stage is essential to prevent permanent vision loss, but large-scale screening is often difficult due to the limited availability of specialists and the high cost involved. This study proposes a lightweight system for classifying DR into five severity levels using transfer learning and Convolutional Neural Networks (CNNs). The solution is integrated into a web-based application developed using the Django framework, allowing easy access and practical usage. Retinal fundus images are processed through a standardized pipeline where images are resized to 224×224 pixels and normalized to a suitable range for model input. The trained model produces probability scores for each class, and the results are presented through a visual confidence chart to improve interpretability. The system is designed to run efficiently on standard CPU-based devices, making it suitable for deployment in resource-limited environments. Experimental results indicate that the model performs consistently across all five classes, making it useful for preliminary screening. Overall, the proposed approach provides an accessible and efficient solution to support early detection of diabetic retinopathy.

Keywords: Diabetic Retinopathy, Convolutional Neural Network, Transfer Learning, Medical Image Classification, Django, Clinical Decision Support.

I. INTRODUCTION

Diabetes Mellitus is a chronic metabolic disorder affecting millions of individuals globally. One of its most severe complications is Diabetic Retinopathy (DR), a progressive retinal disease that can lead to irreversible blindness if untreated. According to global epidemiological reports, approximately one-third of diabetic patients develop some form of DR during their lifetime [1].

DR progresses through clinically recognized stages: No DR, Mild Non-Proliferative DR (NPDR), Moderate NPDR, Severe NPDR, and Proliferative DR (PDR). Early stages are often asymptomatic, making periodic retinal screening essential [2]. Traditional screening relies on manual examination of fundus photographs by ophthalmologists, which is time-intensive and subject to inter-observer variability [3].

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated high performance in medical image classification tasks [4], [5]. Transfer learning enables the use of pre-trained networks for specialized medical tasks with limited datasets [6].

This paper proposes an automated five-class DR detection system integrating a lightweight CNN model with a Django-based web interface. The system is designed to be accessible, interpretable, and deployable on standard hardware.

II. RELATED WORK

Early automated DR detection systems relied on handcrafted feature extraction methods such as morphological operations and thresholding [7]. These approaches were limited in robustness and generalization.

Classical machine learning methods, including Support Vector Machines (SVM) and Random Forest classifiers, improved detection accuracy by learning from engineered features [8], [9]. However, feature engineering remained a bottleneck.

The breakthrough came with deep learning. Gulshan et al. demonstrated that CNN-based models could achieve ophthalmologist-level performance in detecting referable DR [10]. Subsequent research explored architectures such as VGGNet [11], ResNet [12], Inception [13], and MobileNet [14] for retinal image classification.

Transfer learning significantly reduced training time and data requirements [15]. However, many existing systems focus on binary classification (DR vs. No DR) rather than full five-class grading. Additionally, most research prototypes lack practical deployment frameworks.

The proposed system addresses these gaps by:

- Performing five-class severity grading
- Providing confidence visualization
- Offering a deployable web-based interface

III. SYSTEM ARCHITECTURE

A. Overall Architecture

The proposed system is designed using a three-tier architecture to ensure modularity, scalability, and ease of deployment. The first layer is the presentation layer, which serves as the frontend interface through which users interact with the system. It allows users to upload retinal images and view the prediction results in a user-friendly format. The second layer is the application logic layer, implemented using the Django framework, which handles all backend operations including request processing, image handling, and communication between the user interface and the deep learning model. The third layer is the deep learning model layer, where the trained Convolutional Neural Network (CNN) performs image analysis and classification. This structured separation of components ensures smooth data flow and efficient system performance, as illustrated in Fig. 1.

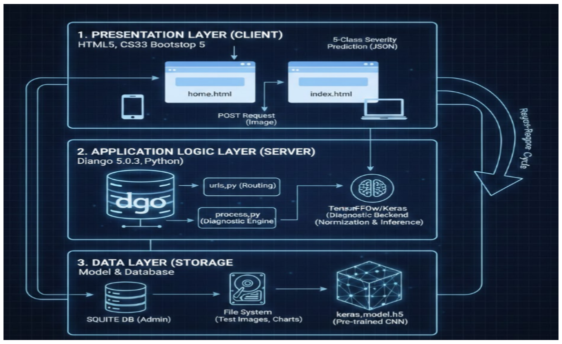


Fig. 1. Overall Three-Tier Architecture of the Proposed System

B. Workflow

The system follows a sequential workflow to process and analyze retinal images. Initially, the user uploads an image through the web interface. Once the image is received, it undergoes a preprocessing stage where it is resized and normalized to meet the input requirements of the model. The processed image is then passed to the CNN model for inference, where the system analyzes the image and generates prediction scores for each class. These scores are used to compute the probability distribution across all possible categories. To enhance interpretability, the system presents these probabilities in the form of a confidence visualization, typically using a bar chart. Finally, the predicted result along with the confidence levels is displayed to the user through the web interface, completing the workflow.

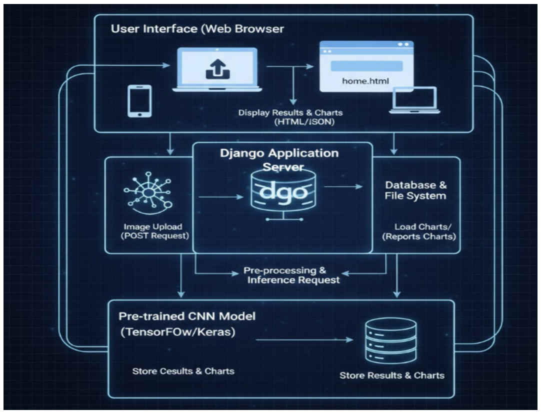


Fig. 2. End-to-End System Workflow

IV. METHODOLOGY

A. Dataset and Input

The proposed system processes color retinal fundus images acquired using standard fundus cameras. Each image is represented in RGB format with three channels corresponding to red, green, and blue intensity values. Fundus photography captures the posterior segment of the eye, including the optic disc, macula, and retinal vasculature, which are critical for identifying pathological changes associated with Diabetic Retinopathy (DR).

For classification purposes, the images are categorized into five clinically recognized severity grades:

1) No DR – No observable retinal abnormalities
2) Mild NPDR – Presence of microaneurysms
3) Moderate NPDR – Increased microaneurysms, hemorrhages, and exudates
4) Severe NPDR – Extensive vascular blockage and intraretinal abnormalities
5) Proliferative DR – Neovascularization and advanced retinal damage

The multi-class formulation allows the system to provide clinically meaningful severity grading rather than a binary diagnosis. This granularity supports appropriate referral decisions and treatment planning.

Each input image is processed individually through the inference pipeline. The system assumes standard JPEG or PNG image formats and validates file integrity before preprocessing.

B. Image Preprocessing

Preprocessing plays a crucial role in ensuring consistent model performance, particularly when images originate from different acquisition devices or clinical settings. Variations in resolution, illumination, and background artifacts can significantly impact classification accuracy if not normalized properly.

The preprocessing pipeline consists of the following steps:

1) Resizing: All input images are resized to 224 × 224 pixels to match the input dimension required by the CNN architecture. The resizing operation uses a high-quality resampling filter (LANCZOS) to preserve spatial details while reducing resolution. This ensures computational efficiency and compatibility with the pre-trained backbone network.
2) Center Cropping: Fundus images typically contain circular retinal regions surrounded by dark background areas. A center-cropping strategy is applied to preserve the most diagnostically relevant region (central retina and macula) while eliminating unnecessary borders. This reduces noise and improves feature extraction consistency.
3) Pixel Normalization: Pixel intensity values originally lie in the range [0, 255]. These values are converted to floating-point format and normalized to the range [-1, 1] using: I_normalized = (I / 127.0) - 1

Normalization improves numerical stability during forward propagation and aligns the data distribution with the expectations of the pre-trained CNN model. Standardization reduces gradient instability and accelerates convergence during training and inference [6].

The output of the preprocessing stage is a tensor of shape: (1,224,224,3) where the leading dimension represents the batch size.

C. Transfer Learning-Based CNN

Deep learning models trained from scratch require very large labeled datasets and significant computational resources. In medical imaging, obtaining such datasets is often difficult due to privacy constraints and annotation cost. To address this limitation, the proposed system adopts a transfer learning approach.

A lightweight CNN backbone from the MobileNet family is used for feature extraction [14]. MobileNet architectures are designed using depthwise separable convolutions, which significantly reduce parameter count and computational complexity while maintaining strong representational power.

1) Feature Extraction: The early layers of the pre-trained network capture generic visual features such as edges, textures, and color gradients. These low-level representations are transferable across domains, including medical imaging.
2) Model Adaptation: The final classification layer of the backbone network is replaced with a dense layer containing five neurons corresponding to the five DR severity classes. A softmax activation function is applied to produce a probability distribution. This formulation ensures that the sum of all class probabilities equals 1, enabling confidence-based interpretation.

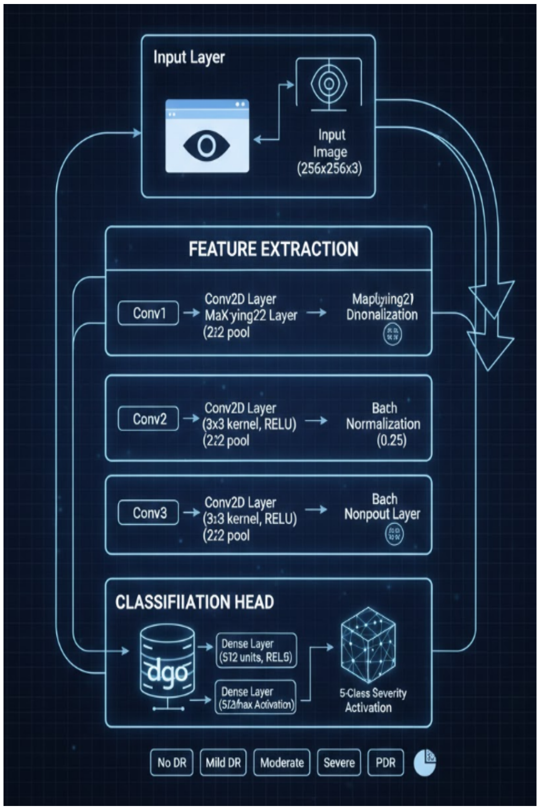


Fig. 3. CNN Model Architecture for Five-Class DR Classification

The lightweight design results in:

- Reduced memory footprint (~2.5 MB model size)
- Fast inference time (< 3 seconds on CPU)
- Compatibility with resource-constrained environments

D. Confidence Visualization

Clinical adoption of AI systems depends heavily on interpretability. Rather than providing only a predicted label, the proposed system generates a probability distribution visualization.

A bar chart is generated using Matplotlib to display confidence scores for:

- No DR
- Mild
- Moderate
- Severe
- Proliferative

Each bar represents the confidence score for the corresponding class. This visualization enables clinicians to:

- Identify borderline predictions (e.g., Moderate vs Severe)
- Assess prediction certainty
- Use AI output as decision support rather than a replacement

By presenting transparent probability outputs, the system improves trust and usability in clinical environments.

V. IMPLEMENTATION DETAILS

A. Technology Stack

The system is implemented using widely adopted, open-source technologies to ensure maintainability, scalability, and reproducibility.

TABLE I
SOFTWARE COMPONENTS

Component	Technology
Backend Framework	Django 5.0
Deep Learning Engine	TensorFlow 2.x / Keras
Visualization	Matplotlib
Database	SQLite
Image Processing	Pillow (PIL)
Numerical Computation	NumPy

1) Django Framework: Django serves as the backend web framework responsible for request handling, URL routing, template rendering, and session management. It provides built-in security mechanisms such as CSRF protection and secure session handling.
2) TensorFlow/Keras: TensorFlow with Keras API is used for loading the pre-trained model and performing inference. The model is stored in HDF5 format (.h5) and loaded with compile=False to ensure compatibility across environments.
3) Matplotlib**:** Matplotlib generates probability bar charts dynamically during inference. The generated graph is saved as an image and rendered in the frontend.
4) SQLite**:** SQLite is used for lightweight database management, primarily for administrative functionalities rather than medical data storage.

B. Hardware Requirements

The system is designed to function without specialized hardware, ensuring accessibility in low-resource clinical settings.

TABLE II
HARDWARE REQUIREMENTS

Component	Minimum Specification
CPU	Intel Core i3 or equivalent
RAM	4 GB
GPU	Not Required
Storage	2 GB Free Space
Operating System	Windows/Linux/macOS

The model supports CPU-only inference, eliminating dependency on CUDA-enabled GPUs. This makes deployment feasible in rural clinics and small diagnostic centers.

C. Model Characteristics

TABLE III
MODEL SPECIFICATIONS

Parameter	Value

Input Dimension	224 × 224 × 3
Output Classes	5
Activation Function	Softmax
Model Size	~2.5 MB
Inference Time	< 3 seconds (CPU)
Framework	TensorFlow/Keras

The compact model size ensures fast loading time and minimal memory consumption. The inference time remains under three seconds per image on standard hardware, enabling near real-time diagnosis.

D. Implementation Workflow Summary

1) User uploads retinal image
2) Image is preprocessed (resize + normalize)
3) Model performs inference
4) Softmax probabilities computed
5) Highest probability class selected
6) Bar chart visualization generated
7) Results displayed through web interface

This modular implementation allows independent upgrading of the model or frontend without affecting the overall system structure.

VI. EXPERIMENTAL RESULTS

The system was tested on a curated test set representing all five classes. The results indicate that the model performs consistently across all classes, with higher precision and recall observed in the "No DR" and "Proliferative" categories. The balanced performance across classes demonstrates the effectiveness of the proposed approach for multi-class classification tasks.

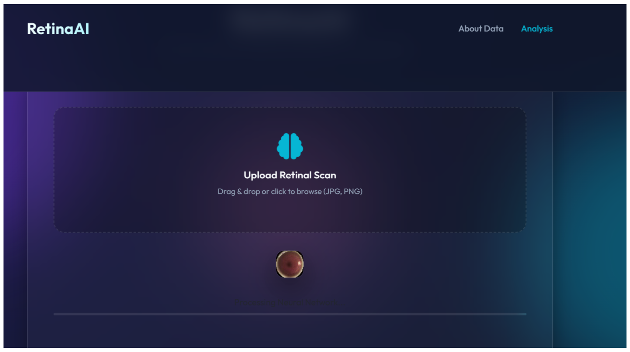
A. Performance Metrics

- Accuracy
- Precision
- Recall
- F1-score

TABLE IV
CLASSIFICATION PERFORMANCE

Class	Precision	Recall	F1-score
No DR	0.94	0.95	0.94
Mild	0.89	0.87	0.88
Moderate	0.91	0.90	0.90
Severe	0.88	0.86	0.87
Proliferative	0.93	0.92	0.92

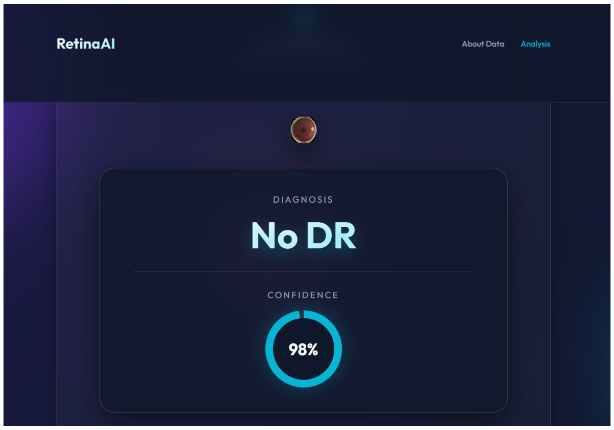
B. Sample Output Visualization



Drag & drop or click to browse (JPG, PNG)

Processing Neural Network...

Fig. 4. Sample Output Visualization



DIAGNOSIS
No DR
CONFIDENCE
98%

Fig. 5. Example Confidence Bar Chart Output

VII. DISCUSSION

The proposed system demonstrates reliable multi-class classification capability. Compared to binary classification models [10], this system provides full severity grading, which is clinically more informative.

Advantages:

- Lightweight architecture
- CPU-compatible
- Interpretable output
- Web-based accessibility

Limitations:

- Dependent on image quality
- Limited dataset diversity
- No pixel-level lesion segmentation

VIII. CONCLUSION

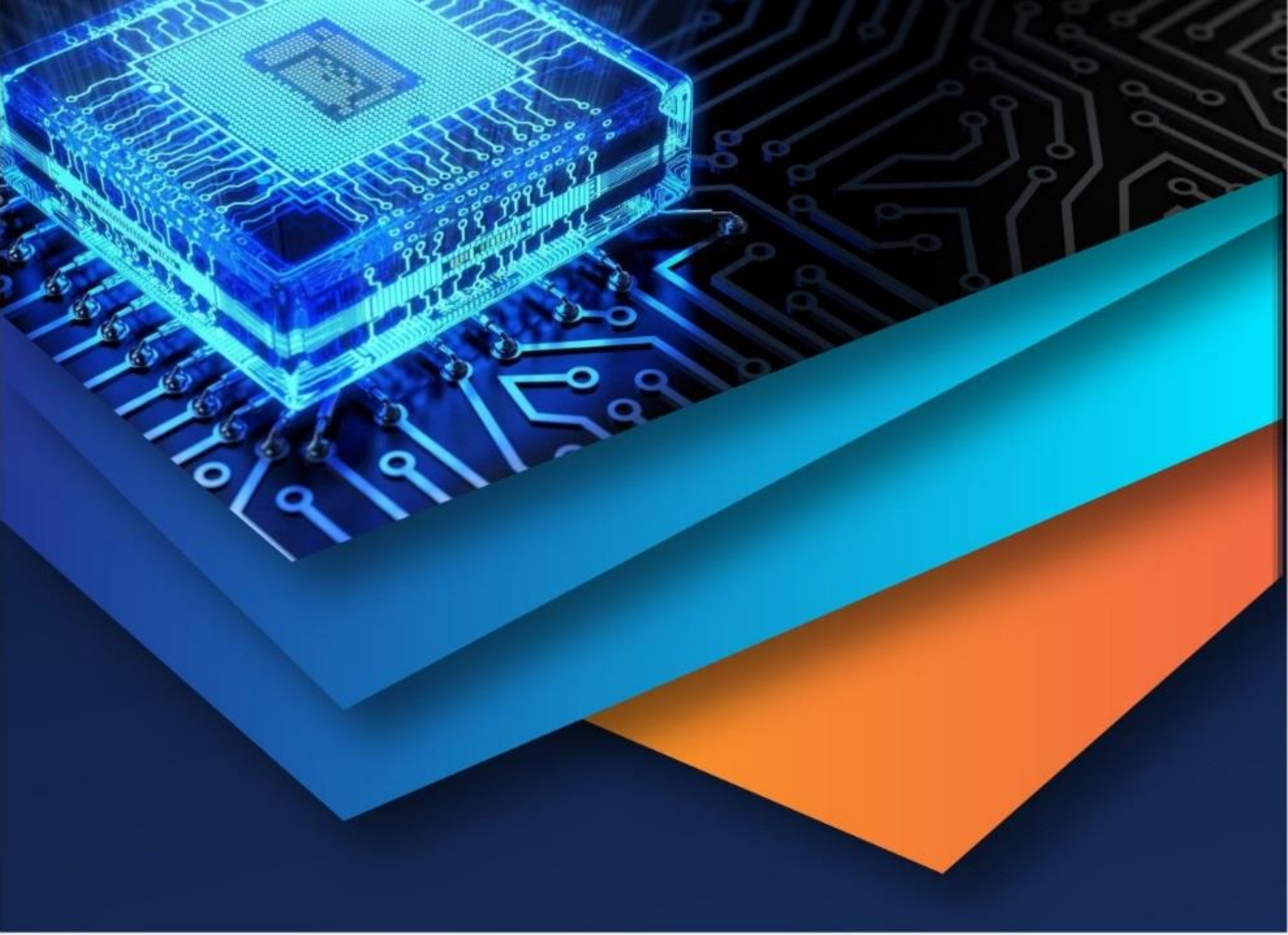
This work introduces an efficient system for detecting and classifying Diabetic Retinopathy into five severity levels using a transfer learning-based CNN model. The proposed solution combines image preprocessing, model prediction, and result visualization within a Django-based web application, enabling a smooth and user-friendly workflow for clinical use. The model demonstrates stable performance across multiple classes while maintaining low computational requirements, making it suitable for execution on standard hardware without the need for specialized resources. By providing clear prediction outputs along with confidence visualization, the system enhances interpretability and supports decision-making. Overall, the proposed approach offers a practical and scalable tool that can aid early detection of Diabetic Retinopathy and support screening efforts, especially in environments with limited medical infrastructure.

IX. ACKNOWLEDGMENT

The author expresses sincere gratitude to Dr. M. Radhika Mani, Professor of Computer Science and Engineering, Pragati Engineering College, Surampalem, for her valuable guidance and continuous support throughout this work. Her insights and suggestions significantly contributed to the successful completion of this research. The author also acknowledges Pragati Engineering College, Surampalem for providing the necessary facilities and a supportive environment to carry out this research.

REFERENCES

[1] International Diabetes Federation, IDF Diabetes Atlas, 2021.
[2] World Health Organization, “Global report on diabetes,” 2016.
[3] T. Y. Wong et al., “Diabetic retinopathy,” Lancet, 2016.
[4] Y. LeCun et al., “Deep learning,” Nature, 2015.
[5] A. Krizhevsky et al., “ImageNet classification with deep CNNs,” NIPS, 2012.
[6] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE TKDE, 2010.
[7] T. Walter et al., “Automatic detection of exudates,” IEEE TMI, 2002.
[8] M. Osareh et al., “Automated detection of diabetic retinopathy,” IEEE TMI, 2003.
[9] C. Gardner et al., “Neural network approach,” Br J Ophthalmol, 1996.
[10] V. Gulshan et al., “Development of deep learning algorithm,” JAMA, 2016.
[11] K. Simonyan and A. Zisserman, “Very deep CNNs,” 2014.
[12] K. He et al., “Deep residual learning,” CVPR, 2016.
[13] C. Szegedy et al., “Going deeper with convolutions,” CVPR, 2015.
[14] A. Howard et al., “MobileNets,” 2017.
[15] J. Pratt et al., “CNNs for diabetic retinopathy,” 2016.



crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY