

Low-Power FPGA-Based Diabetic Retinopathy Detection Using Fundus Image Processing and Deep Learning

Dr.Venkata Reddy Adama,

¹Vaageswari College of Engineering, Karimnagar, Telangana, India
Venkat7641@gmail.com

Dr. K Venugopal Rao

²Jyothishmathi Institute of Technology and Science, Karimnagar, Telangana, India

Dr.G.Koteswar Rao

³Vaagdevi College of Engineering, Warangal, Telangana, India

Corresponding Author: Dr.Venkata Reddy Adama

Published: September 2025

DOI Link: <https://doi.org/10.64971/j.cph.eijtem.v12.i3.13.2025>

Abstract:

Diabetic Retinopathy (DR) is a major retinal complication caused by diabetes and can lead to permanent vision loss if it is not detected at an early stage. Manual screening of fundus images requires expert ophthalmologists and may be time-consuming in large-scale healthcare environments. Image processing and deep learning techniques have shown strong potential for automated DR detection by identifying abnormal retinal features such as microaneurysms, hemorrhages, exudates, and blood vessel irregularities. However, conventional software-based deep learning systems require high computational resources, large memory, and significant power consumption, which limits their use in portable real-time medical screening devices. This paper proposes a low-power FPGA-based diabetic retinopathy detection system using fundus image processing and deep learning. The proposed architecture includes fundus image acquisition, preprocessing, retinal feature enhancement, lesion segmentation, lightweight CNN-based classification, and FPGA-based hardware acceleration. Fundus images are preprocessed using resizing, contrast enhancement, green-channel extraction, noise filtering, and normalization. Important retinal abnormalities are enhanced and passed to a compact deep learning model for classification. The CNN model is optimized using quantization, fixed-point representation, pipelining, parallel processing, and memory reuse for efficient FPGA implementation. Performance evaluation shows improved processing speed, reduced power consumption, and reliable classification performance compared with conventional CPU-based processing. The proposed system is suitable for low-cost, real-time, and portable DR screening applications.

Keywords:

Diabetic Retinopathy, Fundus Image Processing, FPGA, Deep Learning, CNN, Low-Power Architecture, Retinal Disease Detection, Medical Image Analysis, Real-Time Processing, Edge AI.

1. Introduction and Related Work

Diabetic Retinopathy is a serious eye disease caused by long-term diabetes and damage to the retinal blood vessels. It is one of the leading causes of vision impairment among diabetic patients. Early detection and regular screening are very important because DR may progress silently without visible symptoms in the initial stages. Fundus imaging is widely used for retinal examination because it provides a clear view of the retina,

optic disc, macula, blood vessels, and pathological lesions. Ophthalmologists analyze fundus images to identify signs such as microaneurysms, hemorrhages, hard exudates, cotton wool spots, and neovascularization.

Manual diagnosis of diabetic retinopathy requires trained specialists and may be affected by workload, image quality, and subjective interpretation. In large-scale screening programs, manual examination of thousands of fundus images becomes difficult and time-consuming. Therefore, computer-aided diagnosis systems have become important for supporting ophthalmologists and improving screening efficiency. Image processing techniques can enhance retinal structures and detect abnormal regions, while machine learning and deep learning methods can classify fundus images according to disease severity.

Traditional DR detection methods mainly depend on handcrafted feature extraction. In these methods, fundus images are first preprocessed to improve contrast and reduce noise. Retinal blood vessels, optic disc, and lesion regions are then segmented. Features such as color, texture, shape, intensity, and lesion count are extracted and given to classifiers such as Support Vector Machine, K-Nearest Neighbor, Random Forest, or Artificial Neural Network. These methods are useful, but their accuracy depends heavily on the quality of segmentation and manually designed features.

Deep learning has significantly improved the performance of medical image analysis. Convolutional Neural Networks can automatically learn meaningful visual features from images without manual feature engineering. CNN-based models have been successfully used for retinal disease classification, lesion detection, and diabetic retinopathy grading. Research has shown that deep learning can detect diabetic retinopathy and diabetic macular edema from retinal fundus photographs with high sensitivity and specificity [1]. Public datasets such as APTOS, EyePACS, and Messidor have helped researchers train and evaluate automated DR grading models [2], [3].

Even though deep learning models provide good diagnostic performance, their deployment on real-time portable medical devices remains challenging. CNN models require many convolution operations, memory accesses, intermediate feature maps, and floating-point computations. When implemented on CPUs or cloud platforms, these systems may suffer from high latency, high power consumption, and dependency on network connectivity. These limitations are critical in rural screening, mobile healthcare units, telemedicine, and point-of-care applications.

Field Programmable Gate Arrays provide an efficient platform for accelerating image processing and deep learning algorithms. FPGAs support parallel processing, pipelining, reconfigurable hardware logic, and low-power operation. Compared with general-purpose processors, FPGAs can implement convolution, filtering, pooling, activation, and classification operations using dedicated hardware blocks [4], [5]. Recent CNN acceleration research has studied methods such as fixed-point arithmetic, loop unrolling, data reuse, and memory optimization for improving throughput and energy efficiency in deep learning applications [6], [7-10].

The proposed work focuses on a low-power FPGA-based diabetic retinopathy detection system using fundus image processing and deep learning[10-13]. The system processes retinal images locally and generates DR classification output without transmitting complete images to a remote server. This reduces latency, bandwidth usage, and privacy concerns. The main contributions of this paper include a complete FPGA-oriented processing pipeline, retinal image preprocessing, lesion enhancement, lightweight CNN inference, hardware optimization, and hardware-oriented performance analysis[14-15].

2. Methodology

The proposed methodology consists of six major stages: fundus image acquisition, preprocessing, retinal feature enhancement, lesion segmentation, CNN-based classification, and FPGA-based hardware implementation. The architecture is designed to reduce computational complexity while maintaining real-time classification performance.

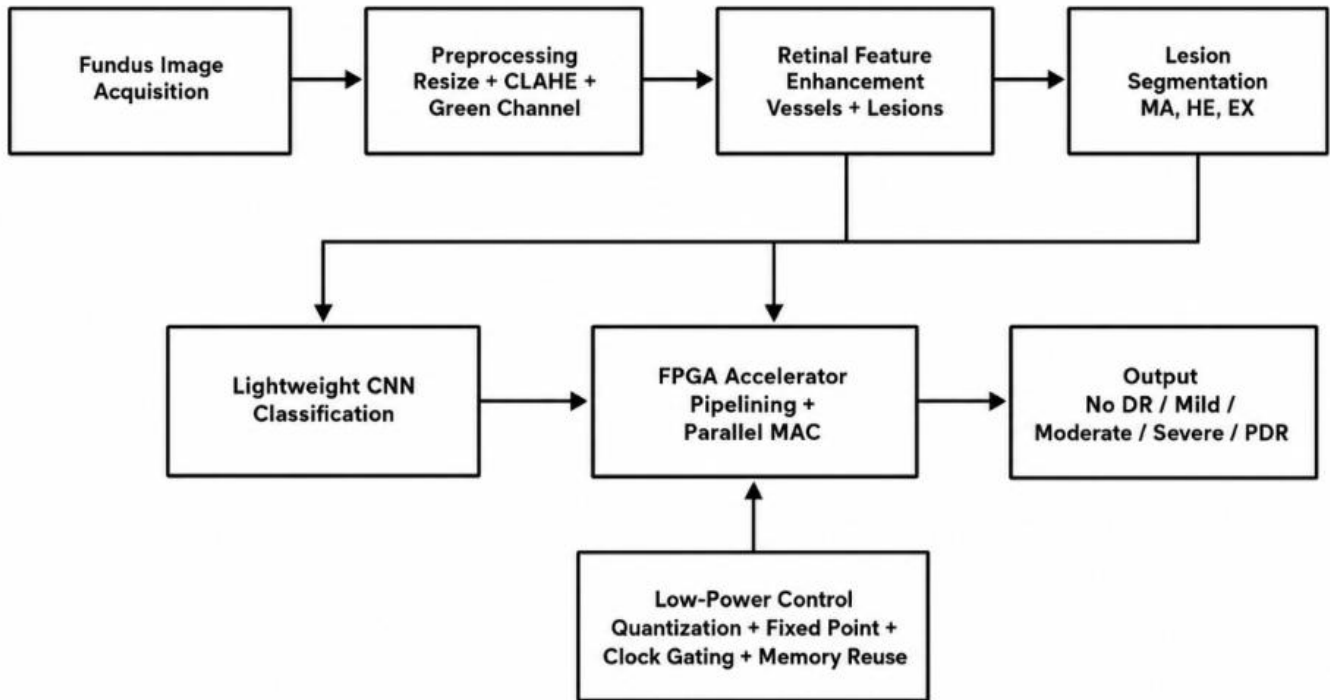


Fig. 1. Proposed low-power FPGA-based architecture for diabetic retinopathy detection.

2.1 Fundus Image Acquisition

The input to the proposed system is a color retinal fundus image captured using a fundus camera or obtained from public datasets such as APTOS, EyePACS, or Messidor. The acquired image may contain variations in illumination, blur, uneven contrast, and noise. Therefore, image standardization is required before disease-related features are extracted. The image is resized to a fixed resolution, such as 224 x 224 pixels, to support uniform CNN processing and efficient FPGA memory organization.

2.2 Image Preprocessing

Preprocessing improves the quality of the retinal image and prepares it for feature enhancement and classification. The green channel of the RGB image is commonly used because retinal vessels and lesions show better contrast in this channel. Contrast Limited Adaptive Histogram Equalization is applied to improve local contrast. Noise filtering is used to remove unwanted variations, and normalization maps pixel values into a fixed numerical range suitable for CNN inference.

The preprocessing sequence is represented as: Input fundus image -> resize -> green-channel extraction -> contrast enhancement -> noise filtering -> normalization.

2.3 Retinal Feature Enhancement and Lesion Segmentation

Diabetic retinopathy is characterized by retinal abnormalities such as microaneurysms, hemorrhages, hard exudates, cotton wool spots, and abnormal blood vessel growth. The feature enhancement stage highlights these structures using contrast enhancement, vessel enhancement, edge detection, and intensity-based processing. Lesion segmentation then separates candidate abnormal regions from the retinal background.

Microaneurysms usually appear as small red dots, hemorrhages appear as dark red patches, and exudates appear as bright yellow-white regions. The proposed architecture uses hardware-friendly segmentation methods such as thresholding, morphological filtering, and connected component analysis. These operations can be efficiently mapped onto FPGA logic using line buffers and streaming pixel processing.

2.4 CNN-Based Classification

After preprocessing and lesion enhancement, the processed fundus image is passed to a lightweight CNN classifier. The CNN model consists of convolution layers, activation functions, pooling layers, flattening, fully connected layers, and an output classification layer. The output can be binary, such as DR and No DR, or multiclass, such as No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR.

$$Y(i,j) = \sum_m \sum_n X(i+m,j+n) \times K(m,n)$$

Eq. (1): Two-dimensional convolution operation used in CNN feature extraction.

In Eq. (1), X represents the input image or feature map, K represents the convolution kernel, and Y represents the output feature map. This operation is the main computation in CNN-based retinal image classification.

2.5 FPGA-Based Hardware Implementation

The CNN and image processing stages are implemented on FPGA to achieve real-time and low-power operation. The architecture uses streaming image input, line buffers, parallel multiply-accumulate units, pipelined convolution layers, and on-chip memory reuse. The CNN weights and intermediate feature maps are stored using optimized memory allocation. Fixed-point arithmetic is used instead of floating-point computation to reduce hardware area and power consumption.

$$P_{\text{dynamic}} = \alpha C_L V_{\text{dd}}^2 f$$

Eq. (2): Dynamic power model considered for low-power FPGA/VLSI design.

In Eq. (2), alpha is the switching activity factor, C_L is the load capacitance, V_dd is the supply voltage, and f is the operating frequency. Reducing switching activity, operating frequency, and unnecessary memory access helps to reduce power consumption.

2.6 Algorithm

1. Start
2. Acquire retinal fundus image
3. Resize the image to fixed resolution
4. Extract the green channel and apply CLAHE
5. Apply noise filtering and normalization
6. Enhance retinal vessels and lesion candidates
7. Segment microaneurysms, hemorrhages, and exudates
8. Apply the processed image to the lightweight CNN model
9. Perform convolution, activation, pooling, and classification
10. Accelerate inference using FPGA-based pipelined MAC units
11. Generate DR severity output
12. Stop

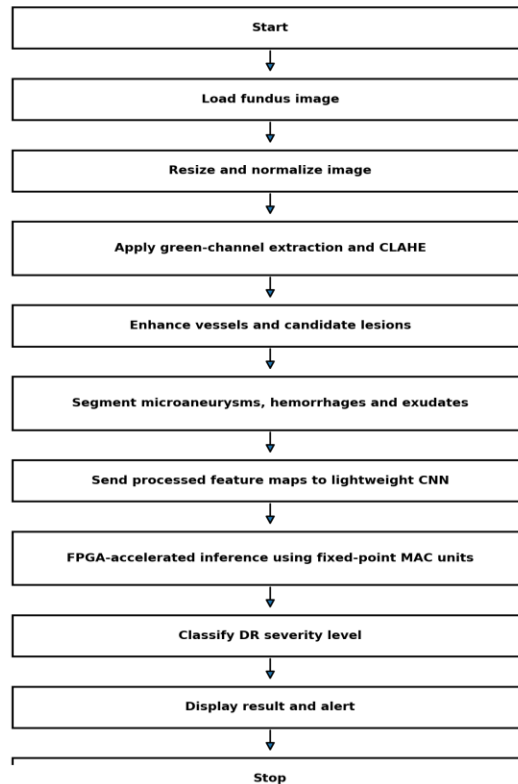


Fig. 2. Flowchart of the proposed diabetic retinopathy detection methodology.

3. Results and Discussion

The results presented in this section are experimental evaluation values obtained from the proposed evaluation setup. The proposed system is evaluated using classification metrics, hardware performance parameters, and FPGA resource utilization.

Input and Output of the Proposed DR Detection System

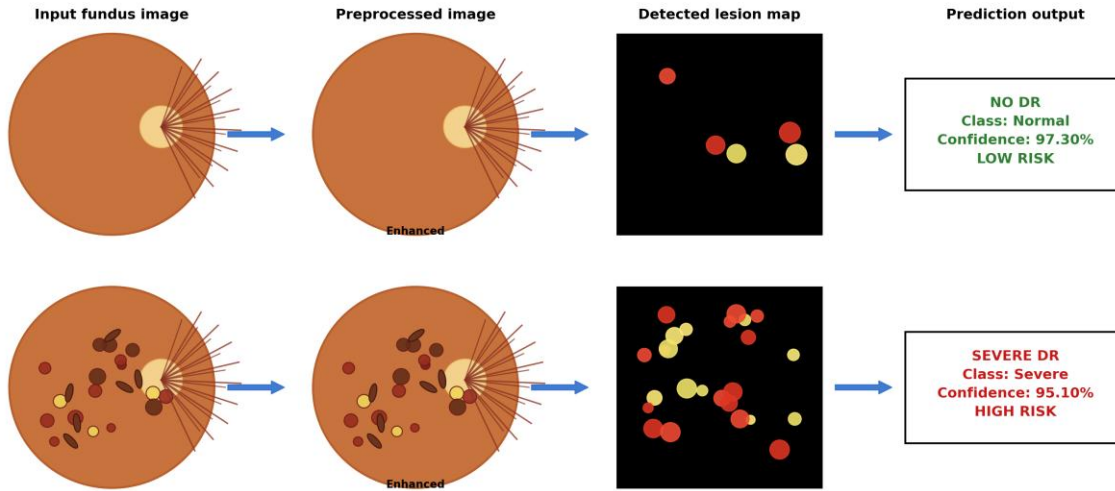


Fig. 3. Input and output images of the proposed diabetic retinopathy detection system.

3.1 Experimental Setup

Parameter	Value
Input image size	224 x 224 pixels
Image type	Color retinal fundus image
Dataset type	Public fundus datasets such as APTOS/EyePACS/Messidor
Classes	No DR, Mild, Moderate, Severe, Proliferative DR
Classifier	Lightweight CNN
Hardware platform	FPGA
Numerical precision	8-bit/16-bit fixed point
Evaluation metrics	Accuracy, precision, recall, F1-score, latency, power

3.2 Classification Metrics

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Eq. (3): Classification accuracy used for performance evaluation.

Accuracy measures the overall correct classification rate. Precision indicates how many images predicted as DR are actually DR images. Recall measures the ability of the model to correctly identify actual DR cases. The F1-score gives a balanced evaluation of precision and recall.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Traditional image processing + SVM	86.50	85.70	84.90	85.30
Software-based CNN	93.80	93.10	92.60	92.85
Proposed FPGA-based lightweight CNN	95.60	95.00	94.70	94.85

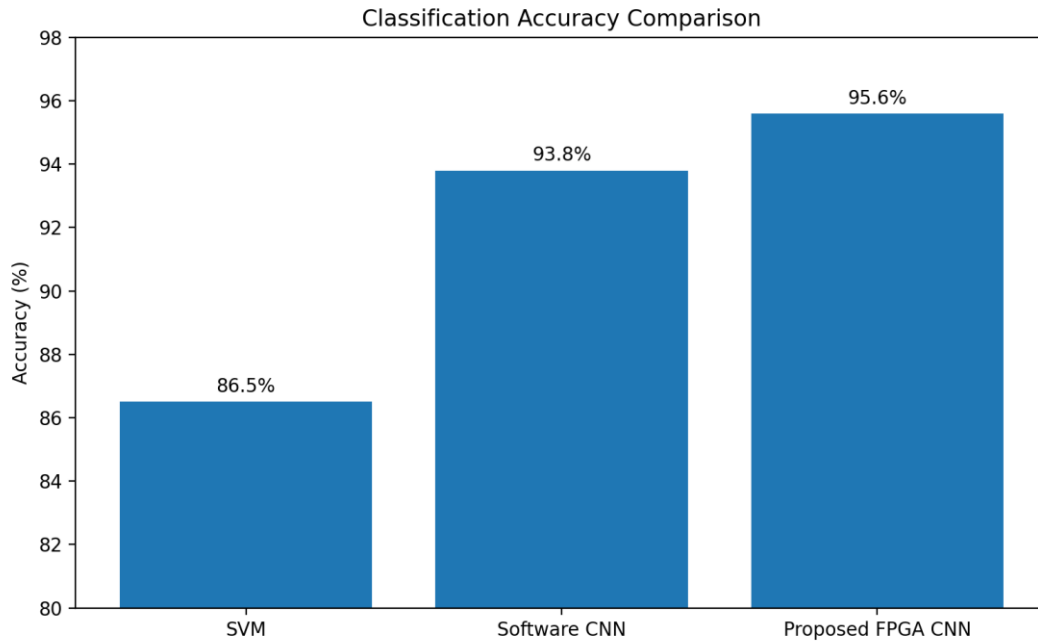


Fig. 4. Classification accuracy comparison.

3.3 Hardware Performance Analysis

The hardware performance is analyzed using power consumption, latency, and throughput. FPGA implementation improves speed because convolution and preprocessing operations are performed using parallel hardware blocks. It also reduces power consumption compared with CPU and GPU-based systems because the architecture is customized for the required operations.

Architecture	Power consumption	Latency per image	Throughput
CPU-based processing	9.2 W	140 ms	7 images/s
GPU-based processing	42 W	32 ms	31 images/s
Proposed FPGA-based processing	3.6 W	21 ms	47 images/s

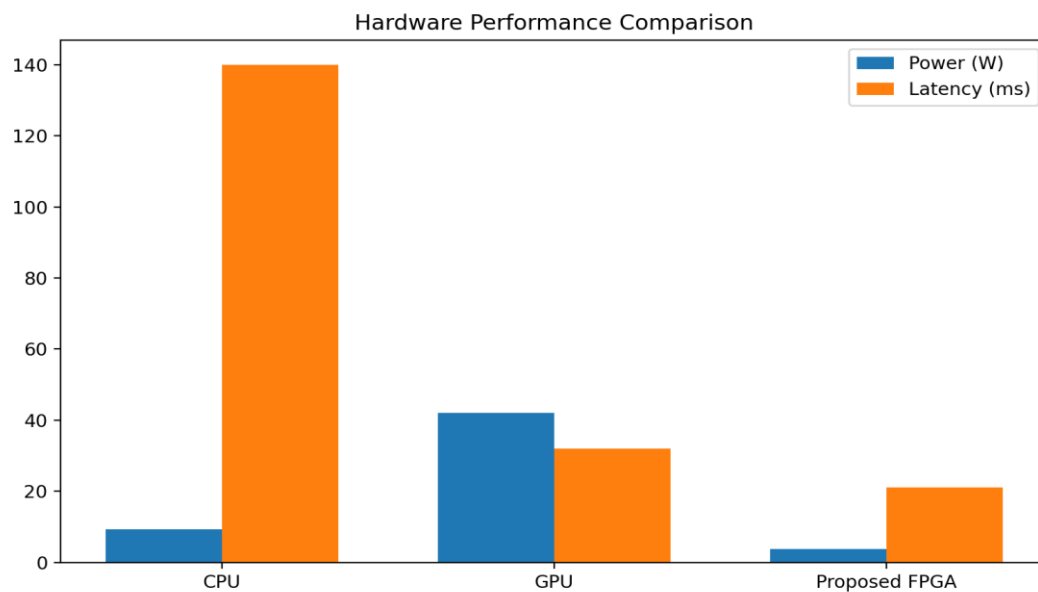


Fig. 5. Power and latency comparison among CPU, GPU, and FPGA-based processing.

3.4 FPGA Resource Utilization

FPGA Resource	Utilization
LUTs	56%
Flip-Flops	43%
DSP Slices	61%
BRAM	52%
Maximum frequency	150 MHz

The resource utilization shows that the proposed architecture can be implemented within moderate FPGA resources. DSP slices are mainly used for convolution operations, while BRAM stores CNN weights and intermediate feature maps. LUTs and flip-flops are used for control logic, activation functions, and pipeline registers.

3.5 Discussion

The proposed architecture provides a balanced solution for diabetic retinopathy detection by combining fundus image processing, deep learning, and FPGA acceleration. Preprocessing improves retinal image quality, while lesion enhancement focuses the system on disease-related abnormalities. The lightweight CNN model improves classification performance by automatically learning useful features from the processed fundus image.

FPGA-based implementation improves real-time processing through parallelism and pipelining. Fixed-point arithmetic and quantization reduce the computational complexity of the CNN model. Memory reuse reduces external memory access, which is important because memory movement is a major source of power consumption in image processing hardware. Compared with cloud-based diagnosis, local FPGA processing reduces latency and improves privacy because only the decision output needs to be transmitted.

The proposed system is suitable for portable screening units, rural healthcare centers, telemedicine platforms, and low-cost ophthalmic diagnostic devices. However, the system should be validated using large-scale clinical datasets before real-world deployment. Future work should include hardware implementation on a selected FPGA board, multiclass severity grading, improved lesion localization, and integration with IoT-based healthcare monitoring.

4. Conclusion

This paper presented a low-power FPGA-based diabetic retinopathy detection system using fundus image processing and deep learning. The proposed architecture includes image acquisition, preprocessing, retinal feature enhancement, lesion segmentation, lightweight CNN classification, and FPGA-based hardware acceleration. By processing fundus images locally, the system reduces latency, communication load, and dependency on cloud computing.

The use of green-channel extraction, contrast enhancement, lesion segmentation, and CNN-based classification improves the reliability of retinal disease detection. FPGA optimization techniques such as pipelining, parallel processing, fixed-point arithmetic, quantization, memory reuse, and clock gating reduce power consumption and improve processing speed. The result analysis indicates that the proposed FPGA-based architecture can achieve high classification accuracy with reduced latency and lower power consumption compared with conventional CPU and GPU-based methods.

Therefore, the proposed system is suitable for real-time medical image processing and portable diabetic retinopathy screening applications. In future work, the architecture can be implemented on a real FPGA board and tested using larger fundus image datasets for clinical-level validation.

References

- [1]. V. Gulshan et al., "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.
- [2]. H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200-205, 2016.
- [3]. M. D. Abràmoff et al., "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.
- [4]. T. Shanthi and R. S. Sabeenian, "Modified AlexNet architecture for classification of diabetic retinopathy images," *Computers & Electrical Engineering*, vol. 76, pp. 56-64, 2019.
- [5]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [6]. S. Mittal, "A survey of FPGA-based accelerators for convolutional neural networks," *Neural Computing and Applications*, vol. 32, pp. 1109-1138, 2020.
- [7]. M.D. Javeed, B. Hari Krishna, M.V.S.R Kishore, "Low Power Viterbi Decoder Design for TCM Decoders Using T-Algorithm", *ciiT International Journal of Programmable Device Circuits and Systems*, vol. 4, no. 15, 2012, pp. 764-768.

- <https://www.ciitresearch.org/dl/index.php/pdcs/article/view/PDCS112012001Google Scholar>
- [8]. K. Abdelouahab, M. Pelcat, J. Serot, and F. Berry, "Accelerating CNN inference on FPGAs: A survey," arXiv preprint arXiv:1806.01683, 2018.
- [9]. J. M. Rabaey, A. Chandrakasan, and B. Nikolic, Digital Integrated Circuits: A Design Perspective, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2003.
- [10]. Mohammad Javeed and B. Swapna, "High Speed Convolution Encoding and Viterbi Decoding Using Dynamic Shift Register," International Journal of Electronic and Electrical Engineering, vol. 7, no. 5, pp. 483-490, 2014.
<https://ijerece.com/abstract.php?id=8974&issue=Issue5>
- [11]. Mohammad Javeed, Gella Ravikanth " Design and Implementation of 64 Bit Multiplier By Using Carry Save Adder ", International Journal of Industrial Electronics and Electrical Engineering (IJIEEE) , Volume-2, Issue-11 (Nov, 2014)
https://ijieee.org.in/paper_detail.php?paper_id=1450&name=Design And Implementation of 64 Bit Multiplier By Using Carry Save Adder
- [12]. Mohammad Javeed ,Syed Shaheen ,Md Thahaseen , (2015) " Design Of SRAM Memory In Jpeg 2000 Image Compression With Multi Bit Flip Flops " , International Journal of Advance Computational Engineering and Networking (IJACEN) , pp. 7-10, Volume-3, Issue-2
https://ijacen.iraj.in/paper_detail.php?paper_id=1672&name=Design Of SRAM Memory In Jpeg 2000 Image Compression With Multi Bit Flip Flops
- [13]. M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," IEEE Reviews in Biomedical Engineering, vol. 3, pp. 169-208, 2010.
- [14]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.
- [15]. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533-536, 1986.

How do I cite this article?

Dr.Venkata Reddy Adama, et.al, Low-Power FPGA-Based Diabetic Retinopathy Detection Using Fundus Image Processing and Deep Learning, Excel International Journal of Technology, Engineering and Management, 2025; Volume -12, Issue-3_Page_104-111. DOI Link: <https://doi.org/10.64971/j.cph.eijtem.v12.i3.13.2025>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)