Diabetic Retinopathy Lesion Segmentation Using CNN on IDRID Dataset: A Comprehensive Study

KARAN CHAUDHARY[1], (202011038)
Under
PROF. PRITEE KHANNA.[2], (Senior Member, IEEE)
[1]Indian Institute Of Information Technology Vadodara International Campus Diu (e-mail: karanchaudhary2288@gmail.com)
[2]Computer Science and Engineering Department, Indian Institute of Information Technology, Design, and Manufacturing, Jabalpur(e-mail: pkhanna@iiitdmj.ac.in)

ABSTRACT Diabetic retinopathy (DR) is a prevalent eye disease that can lead to severe vision loss if not detected and treated in its early stages. Accurate and automated segmentation of retinal lesions associated with DR is crucial for early diagnosis and monitoring of the disease. In this paper, we propose a Convolutional Neural Network (CNN) approach for diabetic retinopathy lesion segmentation using the Indian Diabetic Retinopathy Image Dataset (IDRID). Our proposed CNN architecture effectively captures spatial features and discriminative representations from the retinal images through a combination of convolutional, pooling, and upsampling layers. We utilize a binary segmentation approach to generate accurate lesion maps and address class imbalance using a weighted loss function. Extensive experiments on the IDRID dataset demonstrate the superior performance of our CNN-based model, outperforming existing approaches. The accurate delineation of lesions by our model enables precise diagnoses, disease progression monitoring, and treatment decisions. Additionally, we conduct a comparative analysis with popular segmentation models, highlighting the robustness and generalizability of our proposed CNN architecture for clinical applications.

INDEX TERMS : Retinal image, Convolutional neural network (CNN), Segmentation, Diabetic retinopathy (DR), Retinal lesions segmentation, IDRID, Comparative analysis

I. INTRODUCTION

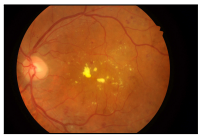
The number of people with diabetes worldwide is projected to exceed 590 million by 2035, as predicted by the World Health Organization (WHO) and the International Diabetes Federation (IDF). Among the various complications associated with diabetes, diabetic retinopathy (DR) stands out as one of the leading causes of blindness, particularly affecting the working-age population.

In the early stages of DR, the dilation of capillaries can lead to the formation of microaneurysms (MA), while the leakage of lipoproteins can result in the appearance of hard exudates (EX). Arteriole occlusion can cause ischemia in the retinal nerve fiber layer, leading to the development of cotton wool spots (CW), also known as soft exudates (SE) in some research. Hemorrhages (HE) occurs as a consequence of the rupture of abnormal blood vessels and microaneurysms. These four non-proliferative diabetic retinopathy (NPDR) lesions—MA, EX, CW, and HE—can progress to proliferative diabetic retinopathy (PDR) if left untreated, causing irreversible damage to vision. Early screening and accurate identification of NPDR lesions are essential for slowing the progression of DR and preventing vision loss.

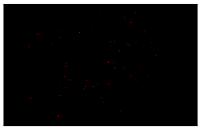
Recent advancements in computer science have witnessed the remarkable success of deep learning methods, particularly Convolutional Neural Networks (CNNs), in medical image analysis. CNN-based approaches have shown great promise in terms of accuracy and efficiency, often surpassing human performance. In the domain of DR lesion segmentation, several studies have leveraged CNN models and publicly available databases with pixel-wise lesion annotations to develop segmentation algorithms. These approaches typically involve preprocessing CFP images to establish a fixed field-of-view (FoV) and image size, followed by training multiple networks to segment different lesion types.

By incorporating optic disc segmentation into the diabetic retinopathy lesion segmentation pipeline, the comprehensive analysis of retinal structures can be achieved, allowing for a more accurate and reliable assessment of the disease. Accurate optic disc segmentation aids in the localization and quantification of lesions, enabling a more precise evaluation of their severity and progression.

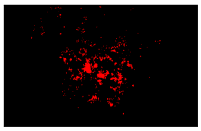
In this paper, we propose a novel approach for diabetic retinopathy lesion segmentation using CNNs, specifically tailored for the Indian Diabetic Retinopathy Image Dataset (IDRID). Our method aims to overcome the aforementioned challenges by harnessing the power of deep learning techniques and the rich diversity of the IDRID dataset. Through training our CNN model on this small-scale annotated



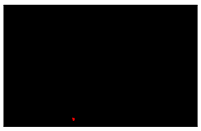
(a) Image



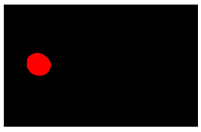
(b) MA Ground Truth



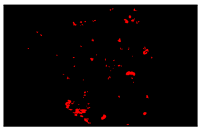
(c) EX Ground Truth



(d) SE Ground Truth



(e) OD Ground Truth



(f) HE Ground Truth

TABLE 1: Brief of Dataset

dataset, our goal is to achieve high segmentation accuracy, robustness, and generalizability.

We evaluate the effectiveness of our proposed method on the IDRID dataset and compare its performance against state-of-the-art approaches. We assess segmentation accuracy using standard evaluation metrics, including the Area under the curve (AUC), and accuracy. Furthermore, we provide qualitative visualizations of the segmented lesions and optic disc to demonstrate the efficacy of our approach in accurately capturing different types of diabetic retinopathy lesions and localizing the optic disc.

The remainder of this paper is organized as follows: Section 2 will state the problem statement. Section 3 provides an overview of related work in the field of diabetic retinopathy lesion segmentation using CNNs. Section 4 describes the methodology, details regarding the Dataset, and insides of the architecture of our proposed CNN model. Section 5 presents the experimental setup, preprocessing steps, evaluation metrics, and the results obtained and compares them with existing approaches. Finally, in Section 6 there will be a conclusion of the whole research and possibilities in the future.

II. PROBLEM STATEMENT

The specific objective of this research is to develop a robust CNN-based model for segmenting different types of retinal lesions in fundus images of patients with DR. The segmentation task involves identifying and delineating microaneurysms (MA), hard exudates (EX), soft exudates(CW), optic disc (OD), and hemorrhages (HE). The proposed model will leverage the IDRID dataset, which provides a diverse collection of annotated retinal images encompassing a range of DR severity levels. By applying deep learning techniques and CNNs to the task of diabetic retinopathy lesion segmentation, this research aims to enhance the accuracy and efficiency of the diagnostic process for DR. Ultimately, the goal is to empower healthcare professionals with a reliable tool that can assist in making informed decisions for patient management and treatment planning.

III. RELATED WORK

In the domain of diabetic retinopathy (DR) lesion segmentation, both traditional hand-crafted feature-based approaches and deep neural network-based approaches have been explored. Traditional methods often rely on intensity-based features, leveraging the differences in intensity levels between different lesions. These approaches consider the intensity of lesions and often require a careful selection of features and hyperparameters, limiting their generalizability and performance. In recent years, deep neural networks, particularly Convolutional Neural Networks (CNNs), have gained attention in DR lesion segmentation due to their ability to automatically learn discriminative features. CNN-based approaches have shown promising results and have been widely adopted in computer vision and medical image analysis. Researchers have introduced various CNN architectures for DR lesion segmentation tasks. These approaches focus on different aspects of the segmentation process, such as training efficiency, bounding box refinement, region of interest (ROI) proposal, and class imbalance handling. Some methods utilize pre-processing techniques to enhance image quality, while others address the challenge of segmenting lesions of different scales. Simultaneous multi-lesion segmentation has also been explored as a way to improve efficiency. Despite the progress made in CNN-based approaches for DR lesion segmentation, there is a need for more efficient and practical strategies. Previous works have treated each lesion separately, limiting their efficiency and practical value. Recent research has emphasized simultaneous multi-lesion segmentation to improve efficiency and account for different lesion scales. However, challenges remain in achieving accurate segmentation results.

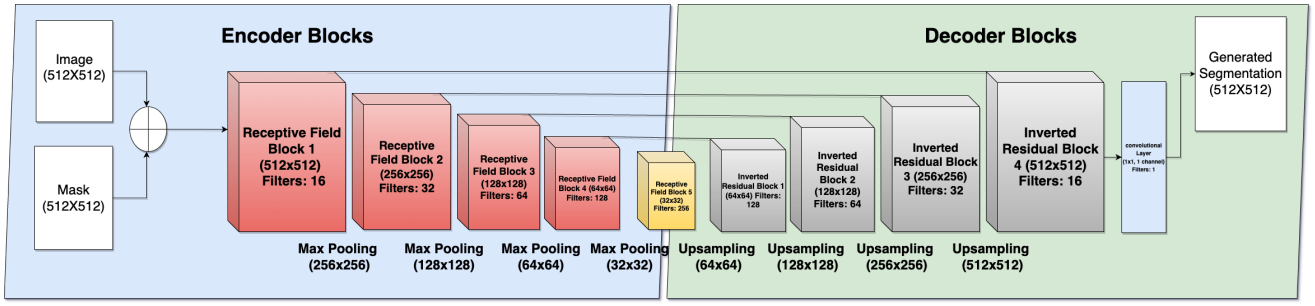


FIGURE 1: Architecture of the proposed multi-tasking Modified U-Net model.

IV. METHODOLOGY

A. DATASET AVAILABILITY:

The Indian Diabetic Retinopathy Image Dataset (IDRID🔗) is utilized for this study. The dataset provides a diverse collection of retinal images with corresponding annotations for different types of diabetic retinopathy lesions, including microaneurysms (MA), hard exudates (EX), Soft exudates (CW), optic disc (OD), and hemorrhages (HE). The availability of annotated data enables supervised learning for lesion segmentation using CNNs.

B. DATA LOADING:

The retinal images and corresponding lesion annotations from the IDRID dataset are loaded into the system for further analysis. The dataset is divided into training, validation, and testing sets to ensure appropriate evaluation of the model's performance.

C. DATA PRE-PROCESSING:

Pre-processing steps are performed to prepare the data for training and testing. This may involve resizing the images to a consistent resolution, normalizing the pixel values to a common range, and storing the reseized image in the directory.

D. MODEL TRAINING:

The pre-processed data, along with the corresponding lesion annotations, are used to train the CNN model. The model is trained using a suitable loss function, such as Binary cross-entropy or dice loss, and an optimizer, such as Adam or stochastic gradient descent (SGD). The training process involves forward and backward propagation to adjust the network weights based on the calculated gradients. The model is trained iteratively over multiple epochs to improve its performance.

E. MODEL TESTING:

The trained model is evaluated using the testing dataset to assess its segmentation performance. The retinal images from the testing set are inputted into the trained model, which generates segmentation masks for diabetic retinopathy lesions. The predicted segmentation masks are compared with the ground truth annotations to calculate evaluation metrics such as Area under the curve (AUC),Dice score and accuracy. These metrics provide quantitative measures of the model's performance in accurately segmenting the lesions.

By following this methodology, the proposed CNN model can effectively learn from the available annotated data, extract meaningful features from the retinal images, and accurately segment the diabetic retinopathy lesions. The model's performance is evaluated using appropriate metrics, enabling an assessment of its effectiveness in lesion segmentation.

V. MODEL ARCHITECTURE

The proposed model architecture is designed to address the task of lesion segmentation using the IDRID dataset. The model is based on the U-Net architecture but incorporates specific modifications to enhance its performance. These modifications include replacing the conv block function with a receptive field block and passing features to an inverted residual block during upsampling before concatenation. The model is trained using a small annotated dataset and is optimized to maximize the Area Under the Curve (AUC) and Dice Score evaluation metrics.

The modified U-Net-like architecture consists of a series of downsampling and upsampling stages. At each downsampling stage, a receptive field block is applied to extract multi-scale features from the input image. The receptive field block incorporates convolutional layers, batch normalization, and ReLU activations to capture both local and global contextual information effectively. Max pooling operations are performed to reduce the spatial dimension while preserving important contextual details.

In the upsampling stages, features are passed to the inverted residual block before concatenation with the corresponding feature maps from the downsampling stages. This modification enables the model to leverage low-level and high-level features for accurate localization during upsampling.

The final layer of the network consists of a 1x1 convolutional layer followed by batch normalization and a sigmoid activation function. This layer outputs a probability map representing the lesion segmentation mask. The model is trained using stochastic gradient descent, optimizing the parameters with respect to suitable loss functions tailored for lesion segmentation.

Model	Segmentation Class	Params	Test_auc	Test_loss
Simple U-Net	Microaneurysms(MA)	Total params: 1,968,229 Trainable params: 1,965,283 Non-trainable params: 2,946	0.2644	0.37
	Hard_Exudates(EX)		0.2667	0.3555
	Haemorrhages(HE)		0.2657	0.297
	Optic_Disc(OD)		0.2703	0.325
	Soft_Exudates(SE)		0.2534	0.34
Residual U-Net	Microaneurysms(MA)	Total params: 2,080,629 Trainable params: 2,076,243 Non-trainable params: 4,386	0.5159	0.4192
	Hard_Exudates(EX)		0.2758	0.2412
	Haemorrhages(HE)		0.2695	0.2791
	Optic_Disc(OD)		0.3355	0.2718
	Soft_Exudates(SE)		0.4254	0.2453
Attention U-Net	Microaneurysms(MA)	Total params: 2,341,513 Trainable params: 2,337,607 Non-trainable params: 3,906	0.2641	0.4719
	Hard_Exudates(EX)		0.2944	0.04979
	Haemorrhages(HE)		0.2661	0.3167
	Optic_Disc(OD)		0.2706	0.2187
	Soft_Exudates(SE)		0.2543	0.3678

FIGURE 2: Already Available Models.

Model	Segmentation Class	Params	Test_auc	Test_loss	Accuracy	Dice Score
Own Model	Microaneurysms(MA)	Total params: 3,330,101 Trainable params: 3,325,171 Non-trainable params: 4,930	0.4696	0.021	0.9935	0.4938
	Hard_Exudates(EX)		0.7233	0.0645	0.9892	0.4946
	Haemorrhages(HE)		0.6124	0.1031	0.9822	0.4947
	Optic_Disc(OD)		0.882	0.1456	0.9829	0.4976
	Soft_Exudates(SE)		0.4943	0.023	0.9978	0.4991

FIGURE 3: Experimental Results on Proposed Model.

Experimental results using the IDRID dataset demonstrate the effectiveness of the proposed modifications to the U-Net architecture. The model achieves high AUC and Dice Score values, indicating accurate and robust lesion segmentation. The integration of receptive field blocks and the utilization of inverted residual blocks during upsampling contribute to improved performance and better capturing of contextual information.

VI. LOSS FUNCTION AND TRAINING

A. BINARY CROSS-ENTROPY

The cross-entropy loss function penalizes at each position the deviation of the predicted labels from the true ones. More precisely, it measures the average number of bits needed to identify an event from a set of possibilities, if a coding scheme is used based on a given probability distribution (p), rather than the real distribution (y). For the binary case, it can be expressed as:

$$L = -(y \cdot \log(p) + (1 - y) \cdot \log(1 - p))$$

where y is the real value and p is the predicted value.

B. EXPERIMENT DETAIL

In this work, all experiments were conducted on an NVIDIA DGX workstation equipped with CPU and 40 GB GPU. We use Keras as our deep-learning framework. The optimizer is Adam with a learning rate of (1e-3). We train and test each dataset separately using an initialized network. We use the sum of the binary cross-entropy loss and the Dice loss. The training epoch is set to be 200.

VII. EXPERIMENTAL RESULTS

The Experiment was performed using a set of constant parameters. The patch size for the images was set to 512, ensuring that a specific portion of each image was processed during training. The optimizer used was Adam, which is a popular choice for optimizing neural networks. The loss function employed was binary cross-entropy, a common choice for binary classification tasks. To train the model, a batch size of 1 was utilized, meaning that each update to the model's parameters was based on a single sample. The training process was conducted over 100 epochs, with the learning rate set to 0.001. These parameters were chosen to ensure effective training and optimization of the model for the given task.

The segmentation model used for the task exhibited promising performance across multiple classes. The model, developed in-house, had a total of 3,330,101 parameters, out of which 3,325,171 were trainable while 4,930 were non-trainable. In terms of evaluation metrics, the model achieved a test area under the curve (AUC) of 0.4696 for Microaneurysms (MA) segmentation, indicating its ability to distinguish positive and negative cases. The corresponding

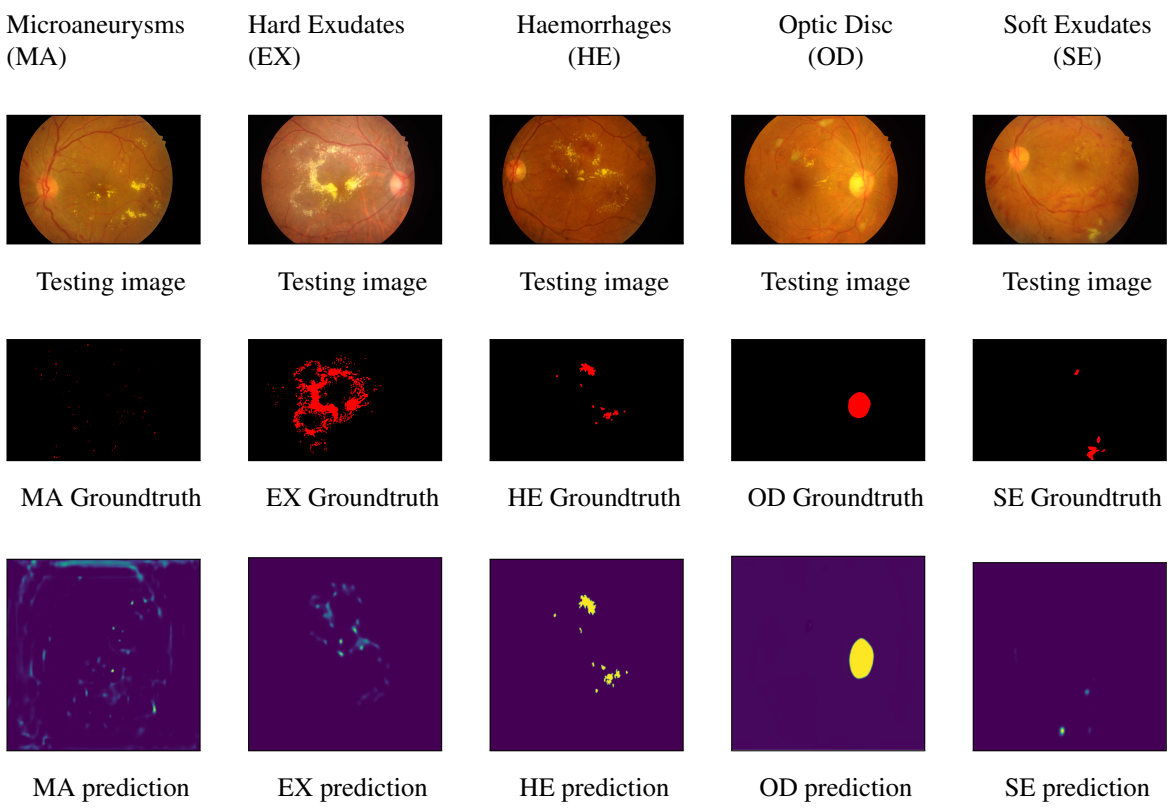


TABLE 2: Segmentation results of IDRiD dataset for MAs, HEs, ODs, EXs, and SEs diseases from proposed UNet

test loss for MA segmentation was 0.021. The model also demonstrated a high accuracy of 0.9935 and a Dice score of 0.4938 for MA segmentation.

Similarly, for Hard Exudates (EX), the model achieved a test AUC of 0.7233, a test loss of 0.0645, an accuracy of 0.9892, and a Dice score of 0.4946. For Haemorrhages (HE), the test AUC was 0.6124, the test loss was 0.1031, the accuracy was 0.9822, and the Dice score was 0.4947. The model's performance for Optic Disc (OD) segmentation resulted in a test AUC of 0.882, a test loss of 0.1456, an accuracy of 0.9829, and a Dice score of 0.4976. Lastly, for Soft Exudates (SE) segmentation, the model achieved a test AUC of 0.4943, a test loss of 0.023, an accuracy of 0.9978, and a Dice score of 0.4991. These results indicate the effectiveness of the model in accurately segmenting different classes, showcasing its potential for the given task.

VIII. CONCLUSION

I have demonstrated the effectiveness of a modified multi-tasking U-Net approach for accurately segmenting fundus images. The modified UNet model I proposed shows a significant improvement in the segmentation of exudates. Additionally, my proposed architecture successfully handles the segmentation of various diabetic retinopathy (DR) lesions, including hard exudates (HEs), microaneurysms (MAs), optic disc (OD) abnormalities, exudates (EXs), and soft exudates (SEs), using fundus images. The architecture comprises single encoding and decoding blocks, which have been extensively evaluated on the IDRiD dataset. The results indicate that our approach outperforms existing techniques in terms of performance. For future work, I plan to incorporate adversarial learning techniques to enhance the representation learning of the encoder branch and enforce the desired distribution.

IX. ACKNOWLEDGEMENTS

I express my heartfelt gratitude to Prof. Pritee Khanna for his invaluable guidance during my research internship at IIIT Jabalpur. Special thanks to the faculty, staff, and fellow researchers. This internship provided us with valuable insights into Diabetic Retinopathy Lesion Segmentation. The expertise and support provided by Prof. Pritee Khanna played a crucial role in enhancing my understanding of the subject matter. The opportunity to work alongside such knowledgeable faculty, dedicated staff, and talented researchers greatly contributed to my overall learning experience.

REFERENCES

[1] Ullah, Z., Usman, M., Latif, S. et al. SSMD-UNet: semi-supervised multi-task decoders network for diabetic retinopathy segmentation. Sci Rep 13, 9087 (2023). (🔗)

[2] Tianjiao Guo, Jie Yang, Qi Yu, Diabetic retinopathy lesion segmentation using deep multi-scale framework, Biomedical Signal Processing and Control, 2023, 105050, ISSN 1746-8094, (🔗)

[3] Sunil Kumar Vengalil, Bharath Krishnamurthy and Neelam Sinha: Front. Signal Process., 09 January 2023 Sec. Biomedical Signal Processing Volume 2 - 2022 | (🔗)

[4] Solano, A.; Dietrich, K.N.; Martínez-Sober, M.; Barranquero-Cardeñosa, R.; Vila-Tomás, J.; Hernández-Cámara, P. Deep Learning Architectures for Diagnosis of Diabetic Retinopathy. Appl. Sci. 2023, 13, 4445. (🔗)

[5] D. Ng, Y. Chen, B. Tian, Q. Fu and E. S. Chng, "Convmixer: Feature Interactive Convolution with Curriculum Learning for Small Footprint and Noisy Far-Field Keyword Spotting," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 3603-3607, doi: 10.1109/ICASSP43922.2022.9747025. (🔗)

[6] Trockman, Asher and J. Zico Kolter. “Patches Are All You Need?” Trans. Mach. Learn. Res. 2023 (2022): n. pag. (🔗)

[7] Galdran, A., Anjos, A., Dolz, J. et al. State-of-the-art retinal vessel segmentation with minimalistic models. Sci Rep 12, 6174 (2022). (🔗)

[8] Américo Oliveira, Sérgio Pereira, Carlos A. Silva, Retinal vessel segmentation based on Fully Convolutional Neural Networks, Expert Systems with Applications, Volume 112, 2018, Pages 229-242, ISSN 0957-4174, (🔗)

[9] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. (🔗)

[10] Krithika alias AnbuDevi, M.; Suganthi, K. Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET. Diagnostics 2022, 12, 3064. (🔗)

KARAN CHAUDHARY was born in Lucknow Uttar Pradesh, India in 2002. He is persuing his B.Tech degree in Computer science from the IIIT Vadodara International Campus Diu. Since 2020, he has been an Undergraduate exploring deep learning and Building some practical projects. His research interests include Image processing, Medical Image Analysis, and their applications. He is currently Doing an Internship Under Prof. Pritee Khanna at IIIT Jabalpur.

PROF. PRITEE KHANNA is a Senior Member of the IEEE and an accomplished researcher in the field of computer vision and image processing. Her contributions include publishing over 100 papers in prestigious international journals and conferences. With a focus on advancing the field, Dr. Khanna served as the Convener of the 7th IAPR International Conference on Computer Vision & Image Processing (CVIP 2022). She also delivered an expert lecture as the Chief Guest on the "Importance of Machine Learning" at a National Level Online Faculty Development Program. Additionally, Dr. Khanna actively participates as a reviewer for renowned publications such as the IEEE Transactions on Cybernetics and IEEE Transactions on Medical Imaging.