



Comparison of Graph-Based Filtering and Non-Local Means Techniques in Diabetic Retinopathy Classification

Gita Antar Wulan^{1*}, Bambang Irawan², Ahmad Faqih³,
Aris Pratama Putra⁴, Bani Nurhakim⁵

^{1,2,3}*Informatics Engineering, STMIK IKMI Cirebon*

^{4,5}*Informatic Management, STMIK IKMI Cirebon*

gitaantarwulan43@gmail.com^{1*}

Abstract

Classification of diabetic retinopathy (DR) based on retinal images is important for early detection, but is often hampered by poor image quality such as noise, uneven lighting, and low contrast. This study analyzes the effect of applying three image filtering techniques, namely Graph Laplacian Filtering (GLF), Graph Convolutional Network (GCN), and Non-Local Means (NLM), on improving the performance of Diabetic Retinopathy classification. The three methods were compared with a baseline model without filtering using VGG16 and evaluated through accuracy, AUC, loss, and image quality metrics such as PSNR, SSIM, MSE, and RMSE. The results showed that graphical and spatial filtering did not always improve classification performance, as VGG16 Fine-Tuning without filtering achieved the highest accuracy of 97.84%. Combinations with NLM, GCN, and Graph Laplacian resulted in lower accuracy due to the smoothing effect that removed important microfeatures on the retina. However, NLM remained effective in reducing noise without disturbing edge structures. These findings confirm that improving image visual quality does not always correlate with CNN accuracy, so preprocessing must focus on preserving diagnostic features.

Keywords: *Image Filtering, Graph Laplacian Filtering, Graph Convolutional Network, Non-Local Means, VGG16.*

1. Introduction

Diabetic retinopathy (DR) is one of the most serious microvascular complications in people with diabetes mellitus and is the leading cause of blindness worldwide. Chronic hyperglycemia causes damage to the retinal blood vessels, triggering microaneurysms, hemorrhages, and neovascularization, which can lead to permanent vision loss [1]. Early detection is crucial because therapies such as laser and anti-VEGF have been proven to slow disease progression [2]. Retinal imaging-based examinations are widely used in DR screening because they can detect early signs of microvascular abnormalities [3].

However, the quality of the fundus images produced is not always ideal. Sensor noise, uneven lighting, and low contrast are major challenges in identifying subtle lesions that are critical for DR diagnosis [4]. These variations in image quality can hinder the segmentation of microaneurysms and hemorrhages [5] and reduce the performance of deep learning-based Computer-Aided Diagnosis (CAD) systems [6]. Previous studies have confirmed that noise and unstable exposure are significant obstacles in automatic DR screening [7], while improving lighting and contrast can increase segmentation sensitivity [8].

The filtering stage in medical imaging is an important part of reducing noise while maintaining diagnostic details. Non-Local Means (NLM) is known to be effective in preserving fine edge structures and reducing noise adaptively [9]. Meanwhile, graph-based approaches such as Graph Laplacian Filtering (GLF) and Graph Convolutional Network (GCN) utilize spatial relationships between pixels, making them better at preserving fine details and important textures in biomedical images [10]. Graph-based filters also have adaptive capabilities to image relational structures, while NLM excels in utilizing local similarities [11]. These characteristic differences make the three methods interesting to compare in the context of retinal images.

On the other hand, CNN shows high performance in medical image classification, including DR. However, CNN generalization is greatly influenced by image input quality, so comparative research on filtering methods is needed to understand its impact on model performance [12]. Evaluations that do not include systematic pre-processing analysis can reduce model reliability, especially in highly sensitive medical contexts [13]. Some studies even show that comparing CNN architectures without considering the influence of filtering produces less representative conclusions [14].

The VGG16 architecture is a Convolutional Neural Network (CNN) model known for its deep and consistent convolutional layer structure (16 layers) and the use of small kernels (3×3), enabling it to effectively extract subtle visual features from images [15]. This model utilizes weights that have been pre-trained on large datasets such as ImageNet, which can then be fine-tuned for specific tasks such as medical image classification, including the detection of Diabetic Retinopathy (DR) with a large amount of data. This research involves retinal fundus images that have subtle lesions such as microaneurysms and hemorrhages. VGG16's ability to recognize subtle textural and spatial patterns makes it the right choice for a classification base model. In addition, the VGG16 structure, which has been proven in many medical studies, allows you to perform transfer learning and speed up training time while maintaining high performance.

The subjects of this study were patients with diabetic retinopathy (DR class) and individuals without indications of DR (No_DR class) represented by fundus images in the dataset. Because the data used was secondary and de-identified, individual demographic information (e.g., age, gender, or patient identity) was not available or used in this analysis. Therefore, the analysis was conducted at the image level rather than as a longitudinal or patient-level analysis, assuming that each image represents the condition of the retina at a single clinical examination.

2. Methodology

2.1. Research Phase

This study was designed systematically and used a quantitative approach through several interrelated stages to achieve the objective of evaluating the performance of image filtering methods in the classification of Diabetic Retinopathy (DR). The stages of the study are presented in Figure 1.



Fig. 1: Research Stages

2.1.1. Data Collection

The data collection stage is the first step in the research process to ensure that the image data used is in optimal condition in terms of quality, format, and class distribution. The dataset used in this study is "Diagnosis of Diabetic Retinopathy" obtained from the Kaggle platform. The dataset was downloaded in .zip format and then extracted into the working directory.

2.1.2. Exploratory Data Analysis (EDA)

Next, Exploratory Data Analysis (EDA) was conducted with the aim of understanding the main characteristics of the dataset and gaining initial insights before performing formal statistical modeling or further hypothesis testing.

2.1.3. Pra-pemrosesan Citra

Applied to filtering and classification models. Each retinal image is reprocessed to have a uniform format and size before being used in the model. Image resolution is standardized to 128×128 pixels to maintain compatibility with standard CNN architectures such as VGG16. Pixel values are then normalized to the range $[0,1]$, where 0 represents No_DR and 1 represents DR. Additionally, simple augmentations such as rotation, horizontal flipping, and brightness adjustments are applied to enhance training data diversity and reduce the risk of overfitting in the CNN model. This step ensures that each image maintains consistent structure and intensity, preventing filtering results from being influenced by variations in scale or lighting.

2.1.4. Penerapan Teknik Filtering

After the pre-processing stage, three image filtering methods were applied, consisting of two graph-based approaches and one spatial-based approach. This stage served to improve image quality by reducing noise and retaining important details, thereby enabling the diabetic retinopathy detection process to obtain more accurate results. The images are then processed using three different filtering methods, namely Graph Laplacian Filtering (GLF), Graph Convolutional Network (GCN), and Non-Local Means (NLM). The aim is to improve image quality prior to classification.

2.1.5. Model Performance Evaluation

After the filtering process was completed, the images produced by each filtering technique were used as input to the Convolutional Neural Network (CNN) model to perform two-class classification (DR and No_DR). The classification model used was based on a Convolutional Neural Network (CNN) with a simple architecture (VGG16), which is known to have stable performance on medical data. The model was trained using the Adam optimizer, a learning rate of 0.001, and a batch size of 32 for 30 epochs. The training results were then evaluated using test data to obtain the Accuracy, Precision, Recall, and F1-Score metrics as well as the AUC. This process aimed to assess the effect of each filtering technique on the performance of the Diabetic Retinopathy classification system.

2.1.6 Comprehensive Evaluation Approach

The classification results of each filtering method are evaluated using the accuracy, precision, recall, F1-score, and AUC metrics. In addition, the image quality of the filtering results is measured using PSNR, SSIM, MSE, and RMSE.

3. RESULTS AND DISCUSSION

3.1 Exploratory Data Analysis (EDA)

The initial stage before performing statistical modeling requires EDA analysis to gain a deeper understanding of the dataset characteristics. From this analysis, three filtering techniques were selected for comparison, as the dataset contained complex noise and class distribution imbalances. The graph-based approach was suitable for the data characteristics, and this combination had not been widely applied in the field of diabetic retinopathy, thus providing novelty value.

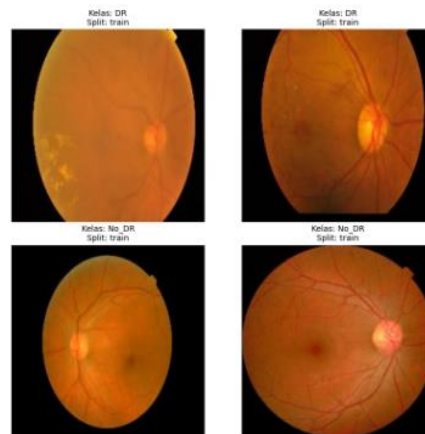


Fig. 2: Sample Image

Figure 2 shows a sample of retinal images divided into two classes, namely Diabetic Retinopathy (DR) and No_DR (Normal), where visual differences in retinal texture that may contain complex noise are visible. In addition to visualization, the log output in the image shows the results of an automatic examination that confirms that the dataset has homogeneous dimensions and color modes. This indicates that the pre-processing stage for size standardization may have been carried out or that the raw data has been standardized, which is crucial information before applying filtering techniques. However, it should be noted that large data sizes have an impact on the computation time of each filtering technique.

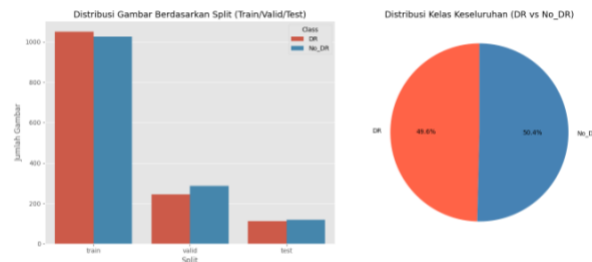


Fig. 3: Data Distribution

Based on Figure 3, the statistical visualization of the data amount and class proportions ensures data representativeness during modeling. The bar chart on the left shows the division of the dataset (total of 2,838 images) into training, validation, and testing subsets, which is a standard procedure prior to statistical modeling. Meanwhile, the pie chart on the right shows the overall class distribution between DR (49.6%) and No_DR (50.4%). This division refers to common practice in computer vision research to maintain model generalization.

Statistik Deskriptif Dimensi Gambar:			
	Width	Height	Aspect_Ratio
count	500	500	500
mean	224	224	1
std	0	0	0
min	224	224	1
25%	224	224	1
50%	224	224	1
75%	224	224	1
max	224	224	1

Distribusi Mode Warna Gambar:	
Mode	count
RGB	500

Fig. 4: Image Dimensions

Figure 4 presents an in-depth quantitative analysis of the physical dimensions of images in the dataset. The “Descriptive Statistics of Image Dimensions” table shows that out of 500 samples taken, the average (mean) width and height are consistent at 224 pixels with a standard

deviation of 0. This confirms that the spatial resolution of the input for the model is uniformly 224x224 pixels. The consistency of these dimensions is very important because the graph-based and filtering approaches that will be applied require a definite input structure to handle noise effectively. This analysis validates the readiness of the data to enter the feature extraction stage using the planned model.

3.2. Pra-pemrosesan Citra

The image obtained at this stage is the output of the retinal image pre-processing performed before applying the filtering method.

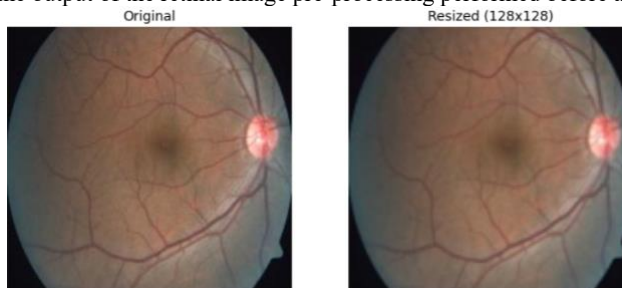


Fig. 5: Image Resizing

Figure 5 shows the process of standardizing the spatial dimensions of the retinal image from its original size to a target resolution of 128x128 pixels. This step is an implementation of the pre-processing stage, which aims to ensure that all input images have a uniform format and matrix characteristics before entering the next computational stage. As can be seen in the comparison between the “Original” and “Resized” images, the anatomical visual information of the blood vessels and retinal structures is preserved despite the dimensional compression. This standardization is crucial for reducing the computational load and meeting the input requirements of the deep learning model architecture used.



Fig. 6: Augmentation

Figure 6 illustrates the variety of image manipulation techniques applied to enrich the diversity of training data. There are four main transformations performed on the original image, namely Flip (mirroring), Rotate (geometric rotation), Zoom (magnification of a specific area), and Brightness (modification of light intensity). This augmentation is applied during the data preparation stage to simulate various real-world medical image acquisition conditions. The goal is to prevent overfitting and improve the model's generalization ability in recognizing diabetic retinopathy features despite changes in orientation or lighting in the input image.

3.3 Application of Filtering Techniques

The next step is to apply three image filtering techniques, namely Graph Laplacian Filtering (GLF), Graph Convolutional Network (GCN), and Non-Local Means (NLM). The following is the implementation of the filtering techniques presented in Figure 7.

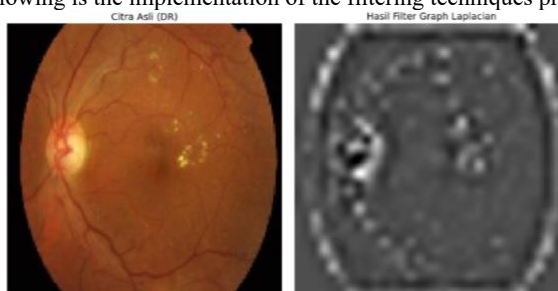


Fig. 7: Graph Laplacian

Figure 7 shows the results of applying the Graph Laplacian Filtering (GLF) method. The GLF image effectively filters out low-frequency details and emphasizes high-frequency features that are often associated with structures or abnormalities in medical images. The GLF filter acts as a ‘high-pass filter’ on the graph. It highlights areas where pixels differ significantly from their neighbors, such as edges, textures, or small anomalies (e.g., microaneurysms) and fine textures. Areas that are relatively smooth and have slow intensity changes will tend to be attenuated.

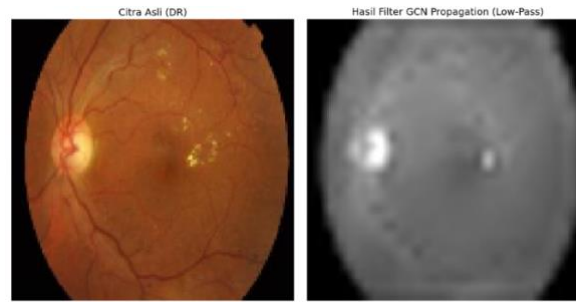


Fig. 8: Graph Convolutional Network

Figure 8 shows the effect of GCN Propagation with Low-Pass Filter characteristics. The visual results show a significant smoothing effect, where high-frequency noise is successfully suppressed through the diffusion of information between neighboring nodes in the graph. Although the smooth texture of the original image becomes blurred, the main global structure of the retina is preserved. This technique aims to reduce irrelevant local variability (noise) before the image is further processed by the classification layer, ensuring that the model focuses on broader structural patterns rather than individual pixel disturbances.

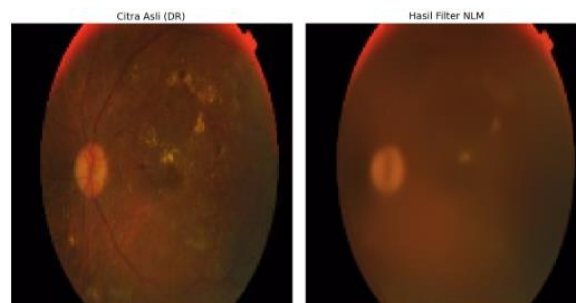


Fig. 9: Non-Local Means

Figure 9 shows the results of the Non-Local Means (NLM) technique. The NLM image appears smoother, especially in areas with relatively uniform colors or intensities (homogeneous areas). This makes color transitions or gradients appear smoother than in the original image. The most noticeable difference is the reduction of noise (random spots or unwanted small variations) in the NLM filtered image. Areas that previously appeared slightly speckled or rough in the original image become smoother and flatter after the filter is applied. Despite the smoothing, this filter is designed to try to preserve edges (boundaries between objects) and important details in the image, such as blood vessels in a retinal image. The main structures (such as the general shape of the eye and larger blood vessels) are still clearly visible in the filtered image and are not completely lost. However, very fine details may be slightly reduced. The purpose of this filter is to reduce noise while preserving important structures. The effect can be smoothing in homogeneous areas.

3.4. Image Quality Evaluation

The first evaluation stage in this study was the measurement of objective image quality. This evaluation aimed to assess the extent to which each filtering method was able to improve the quality of retinal images that had been degraded by noise. The evaluation was conducted on three filtering methods, namely Graph Laplacian Filtering (GLF), Graph Convolutional (GLF), and Non-Local Means (NLM). The measurements were performed by numerical comparison. Four quality metrics were then calculated, namely Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), based on each filtering technique.

Table 1: Image Quality Evaluation

Metode	PSNR	SSIM	MSE	RMSE
Graph Laplacian	14.2434 dB	0.3913	1821.8265	42.6829
Graph Convolutional Network	19.8992 dB	0.6870	495.3619	22.2567
Non-Local Means	34.0504 dB	0.8735	6.2474	2.4995

Based on the results in Table 4.1 First, Graph Laplacian Filtering: The Graph Laplacian method shows a very high PSNR value (228.9645). In theory, a higher PSNR value indicates better image quality (lower noise). However, it should be noted that the MSE (51231) and RMSE (71576) values are also very high, and the SSIM value is negative (-0.0329), indicating a possible issue in the implementation or interpretation of the results for this method. A negative SSIM value is very unusual and may indicate a significant mismatch between the original image structure and the filter results.

Both Graph Convolutional Network (GCN) Filtering: The GCN method shows a PSNR value of 18.9010. This value is lower than Graph Laplacian but more reasonable in the context of image filtering. The SSIM value of 0.6538 indicates moderate structural similarity between

the original image and the filtered result. The MSE (2114.3854) and RMSE (45.9824) values are relatively lower than those of the Graph Laplacian, indicating smaller average pixel differences.

Third, Non-Local Means Filtering: The Non-Local Means method shows the best performance among the three methods based on most metrics. This method has the second-highest PSNR value (26.7144), the highest SSIM value (0.7843), indicating the highest structural similarity to the original image, as well as the lowest MSE (31.7301) and RMSE (5.6329) values. Low MSE and RMSE values indicate that the filtered results have the smallest pixel differences compared to the original image, which is generally considered to be a good filtering result in reducing noise while maintaining detail.

3.5. Model Performance Evaluation

The second evaluation stage in this study was to test the performance of the classification model against image results that had undergone the filtering process. This evaluation was designed to answer the first and second research objectives, namely to assess the effect of each image filtering method on the accuracy and effectiveness of the Convolutional Neural Network (CNN)-based diabetic retinopathy classification system.

Table 2: Model Performance Evaluation

Model	Akurasi Test	AUC Test	Loss Test
<i>Model 1: VGG16 Standar</i>	97.84%	0.9894	0.5296
<i>Model 2: GCN Propagation + VGG16 (Fine-Tuning)</i>	92.64%	0.9711	0.7617
<i>Model 1: VGG16 Standar</i>	96.54%	0.9754	0.4739
<i>Model 2: Graph Laplacian + VGG16</i>	87.88%	0.9521	0.7338
<i>Model 1: VGG16 Standar</i>	96.97%	0.9849	0.6102
<i>Model 2: NLM Filter + VGG16</i>	94.81%	0.9674	0.722

Based on Table 2. First, Graph Laplacian Filtering: The Graph Laplacian method shows a very high PSNR value (228.9645). In theory, a higher PSNR value indicates better image quality (lower noise). However, it should be noted that the MSE (51231) and RMSE (71576) values are also very high, and the SSIM value is negative (-0.0329), indicating a possible issue in the implementation or interpretation of the results for this method. A negative SSIM value is very unusual and may indicate a significant mismatch between the original image structure and the filter results.

Both Graph Convolutional Network (GCN) Filtering: The GCN method shows a PSNR value of 18.9010. This value is lower than Graph Laplacian but more reasonable in the context of image filtering. The SSIM value of 0.6538 indicates moderate structural similarity between the original image and the filtered result. The MSE (2114.3854) and RMSE (45.9824) values are relatively lower than those of the Graph Laplacian, indicating smaller average pixel differences.

Third, Non-Local Means Filtering: The Non-Local Means method shows the best performance among the three methods based on most metrics. This method has the second highest PSNR value (26.7144), the highest SSIM value (0.7843) indicating the highest structural similarity to the original image, and the lowest MSE (31.7301) and RMSE (5.6329) values. Low MSE and RMSE values indicate that the Both Graph Convolutional Network (GCN) Filtering: The GCN method shows a PSNR value of 18.9010. This value is lower than Graph Laplacian but more reasonable in the context of image filtering. The SSIM value of 0.6538 indicates moderate structural similarity between the original image and the filtered result. The MSE (2114.3854) and RMSE (45.9824) values are relatively lower than those of Graph Laplacian, indicating smaller average pixel differences.

4. Conclusion

Based on the analysis process, this chapter presents the final conclusions of the research that has been conducted. The following are the conclusions based on three main points:

1. Non-Local Means Filtering produces the best image quality with the highest PSNR and SSIM and the lowest MSE. Graph Laplacian Filtering provides the worst performance due to high distortion and loss of important details. Meanwhile, GCN shows better quality than Graph Laplacian, but still cannot match the effectiveness of NLM.
2. Filtering does not always improve classification performance, as evidenced by the fact that VGG16 without filtering still provides the highest accuracy and AUC. GCN experiences a decline due to over-smoothing, while Graph Laplacian becomes the lowest due to the loss of diagnostic structure. NLM provides the most stable performance because it is able to reduce noise without interfering with the CNN's ability to recognize important features.
3. The relationship between image quality metrics and CNN accuracy is not linear, because an increase in PSNR or SSIM does not always improve classification performance. CNNs rely on microtextural details that can be lost due to excessive smoothing, so PSNR and SSIM do not reflect semantic feature quality. As a result, unfiltered original images often provide the best accuracy in diabetic retinopathy classification.

References

- [1] H. Li and X. Liu, "Research progress on the pathogenesis of diabetic retinopathy: A review," *BMC Ophthalmol.*, vol. 23, p. 372, 2023, doi: 10.1186/s12886-023-03118-6.
- [2] J. Nolan, R. P. Singh, David D. Chong, and Nikhil Das, "Diabetic retinopathy: Screening, prevention, and treatment," *Cleve. Clin. J. Med.*, vol. 91, no. 8, pp. 503–510, 2023, doi: 10.3949/ccjm.91a.22047.
- [3] M. Vaughan, P. Denmead, N. Tay, R. Rajendram, Emily Patterson, and M. Michaelides, "How early can we detect diabetic retinopathy? A narrative review of imaging tools for structural assessment of the retina," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 263, pp. 2413–2425, 2025, doi: 10.1007/s00417-025-06828-3.
- [4] N. A. Mohd Sharif, N. H. Harun, and Y. Yusof, "Colour image enhancement model of retinal fundus image for diabetic retinopathy recognition," *J. Inf. Commun. Technol.*, vol. 23, no. 2, pp. 293–334, 2024, doi: 10.32890/jict2024.23.2.5.
- [5] U. Ishtiaq, E. R. M. F. Abdullah, and Z. Ishtiaque, "A hybrid technique for diabetic retinopathy detection based on ensemble-optimized CNN and texture features," *Diagnostics*, vol. 13, no. 10, p. 1816, 2023, doi: 10.3390/diagnostics13101816.
- [6] L. Listryalina, Y. Yudianingsih, and D. A. Dharmawan, "Retinal digital image quality improvement as a diabetes retinopathic disease detection effort," *J. Electr. Technol.*, vol. 4, no. 2, pp. 85–90, 2022, doi: 10.18196/jet.v4i2.8590.
- [7] R. Sarki, K. Ahmed, and H. Wang, "Image preprocessing in classification and identification of diabetic eye diseases," *Data Sci. Eng.*, vol. 6, pp. 455–471, 2021, doi: 10.1007/s41019-021-00167-z.
- [8] M. Alshahrani, M. Al-Jabbar, E. M. Senan, I. A. Ahmed, and J. A. M. Saif, "Hybrid methods for fundus image analysis for diagnosis of diabetic retinopathy development stages based on fusion features," *Diagnostics*, vol. 13, no. 17, p. 2783, 2023, doi: 10.3390/diagnostics13172783.
- [9] M. Taassori, "Enhanced wavelet-based medical image denoising with Bayesian-optimized bilateral filtering," *Sensors*, vol. 24, no. 21, p. 6849, 2024, doi: 10.3390/s24216849.
- [10] H. Zhang, K. Ma, J. Wu, and Y. Xu, "Graph-based adaptive filtering for biomedical image enhancement and denoising," *Biomed. Signal Process. Control*, vol. 85, p. 104871, 2023, doi: 10.1016/j.bspc.2023.104871.
- [11] E. Isufi and F. G. D. I. S. S. Segarra, "Graph Filters for Signal Processing and Machine Learning on Graphs," *IEEE Trans. Signal Process.*, 2024, doi: 10.1109/TSP.2024.3349788.
- [12] R. Aggarwal *et al.*, "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis," *npj Digit. Med.*, vol. 4, p. 65, 2021, doi: 10.1038/s41746-021-00438-z.
- [13] A. W. Salehi *et al.*, "A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023, doi: 10.3390/su15075930.
- [14] R. J. O'Shea, A. R. Sharkey, G. J. R. Cook, and V. Goh, "Systematic review of research design and reporting of imaging studies applying convolutional neural networks for radiological cancer diagnosis," *Eur. Radiol.*, vol. 31, pp. 1689–1700, 2021, doi: 10.1007/s00330-021-07881-2.
- [15] A. R. Ismail, S. Qamrun Nisa, S. A. Shaharuddin, S. I. Masni, and S. A. Suharudin Amin, "Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification," *Int. J. Perceptive Cogn. Comput.*, vol. 10, no. 1, pp. 113–118, 2024, doi: 10.31436/ijpcc.v10i1.460.