International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

A Hybrid Super-Resolution and Total Variation Framework for Medical Image Enhancement and Deep Learning-Based Classification

[1]Hayder Mustafa Mueen

[1]Department of computer Science, College of Education University of Kufa, IRAQ, Najaf

---

Abstract - *Medical images play a significant role in the accurate interpretation of images and computer-aided diagnostic systems. In turn, images with low spatial resolution and degradation can affect the visibility of fine structure details, thus negatively impacting the effectiveness of deep learning classification system. In order to solve these problems, in the current study, the hybrid Super-Resolution and Total Variation (SR-TV) algorithm will be proposed. Proposed method will use Enhanced Deep Super-Resolution Network (EDSR) and Total Variation (TV) regularization in sequential processing pipeline. EDSR will be used for reconstruction of high-resolution images from artificially created low-resolution images with the usage of ×2 scale factor. Afterward, TV regularization will be used for further refinement of reconstructed images, suppressing of fluctuations and keeping structural boundaries. Reconstructed images will be evaluated with the help of ResNet-50 convolutional neural network to study the effect of image reconstruction and refinement on classification performance. In terms of the experimental evaluation, OCT2017 dataset will be used which contains synthetically created low-resolution and high-resolution image pairs. Quality of image reconstruction will be evaluated with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Classification performance will be evaluated by Accuracy, Precision, Recall and F1-score metrics. Comparative experiments will be conducted to study individual and joint effect of EDSR and TV regularization. The objective of the current work is the creation of an effective and interpretable hybrid algorithm that consists of deep learning-based resolution reconstruction and mathematical image refinement with the evaluation of its effect on medical image quality and classification performance.*

Keywords: Super-Resolution, Total Variation Regularization, Medical Image Enhancement, Deep Learning, EDSR, Image Reconstruction, Medical Image Classification.

1. INTRODUCTION

Medical imaging is one of the crucial components of modern healthcare, as it supports clinical diagnosis, identification of diseases and automated diagnostic systems. With the development of artificial intelligence and deep learning, automated medical image analysis became of great importance. However, the performance of such systems strongly depends on the quality and spatial resolution of input images. Low spatial resolution, noise, and image degradation may distort fine structures, leading to low quality of following image analysis and classification. In the field of medical imaging, high resolution of input images allows receiving more information about anatomical structures, which helps distinguish between normal and pathological tissues. However, high resolution may be constrained by technological and computational factors. For this reason, the problem of reconstruction of high resolution representation of low-resolution images received special attention in image enhancement techniques. Deep learning algorithms significantly boosted the performance of single image super-resolution. Enhanced Deep Super-Resolution Network (EDSR) shows good results due to the usage of residual learning technique and deep features extraction. Unlike traditional interpolations, EDSR receives complex mapping of high and low resolution images and provides an opportunity to recover fine details of image. Nevertheless, there are some drawbacks of deep learning reconstruction such as undesired fluctuations or local artifacts near sharp structural boundaries. In order to compensate for these drawbacks, one can use mathematical regularization during image reconstruction process. Total Variation (TV) regularization is a known variational technique that eliminates undesired intensity fluctuations while keeping the important structural boundaries of image. Due to the limitation of excessive spatial gradients, it provides a good balance between smoothness of image and its structure. Thus, combining super-resolution reconstruction and TV-based refinement may lead to good quality of reconstructed images with preserved structures. Considering the above-mentioned information, this research will propose a hybrid Super-Resolution and Total Variation (SR-TV) framework for enhancement of medical images and classification based on deep learning algorithms. This framework will use EDSR-based super-resolution algorithm for the reconstruction of high resolution images and further refinement of them with TV regularization. Firstly, low resolution images will be created with controlled degradation process. Then EDSR will be used to get high resolution images. In the end, TV regularization will refine these images reducing undesired fluctuations but keeping structural boundaries.

This framework will be tested using publicly available OCT2017 dataset, which includes 4 retinal image categories: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal retina. While the experimental application will focus on retinal imaging, the proposed image enhancement strategy will be formulated as a general image enhancement algorithm which can be studied in other medical imaging applications. In addition, the influence of image enhancement on automated classification will be studied using ResNet-50 convolutional neural network. In order to keep consistent comparison, the same architecture of classification system will be used for both original and enhanced images. The quality of image reconstruction will be estimated using PSNR and SSIM, while the quality of classification will be estimated using Accuracy, Precision, Recall, and F1-score.

The main contributions of this study are summarized as follows:

- An SR-TV model is proposed through the combination of the EDSR reconstruction approach and the mathematical variational improvement.
- A sequential image improvement method is considered with the goal of increasing spatial resolution while minimizing reconstruction artifacts and boundary structures retention.
- A quantitative assessment scheme is considered with the use of PSNR and SSIM measures in order to evaluate the effectiveness of the reconstruction process in conditions of controlled low-resolution degradation.
- The impact of the suggested image improvement framework on the performance of automated medical image classification is studied via a fixed ResNet-50 classifier and classification accuracy metrics.
- Comparison experiments are designed to analyze the contribution of super-resolution reconstruction and Total Variation regularization separately and jointly.

The rest of this work is structured as follows. Related works concerning medical image enhancement, super-resolution reconstruction based on deep learning, and variational regularization are reviewed in Section 2. Section 3 introduces the SR-TV framework and provides its mathematical formulation. The experimental setup and data preparation procedure are described in Section 4. Experimental results and discussion follow in Section 5. Finally, the conclusions are provided in Section 6.

2. RELATED WORKS

2.1. Traditional Image Enhancement and Restoration Methods

Image enhancement and restoration are well-developed areas of study aimed at improvement of the quality of images by removing noise and suppressing artifacts while keeping structure of images preserved. Standard methods include spatial filtering, interpolation-based restoration and variational regularization. While they are simple in computation, they rely heavily on selected filters' parameters and properties of the input image. Total variation (TV) regularization is a well-known variational method of image restoration. TV-based denoising model by Rudin et al. removes noise while favoring preservation of important edges of an image. This method is based on restoration as optimization problem that combines fidelity to the input image and its spatial smoothness [1]. This property makes TV regularization method especially applicable in case if structure and edges are very important for the input image. However, in order to restore the lost information of an image by its downsampled version, it is not enough to apply only TV regularization because TV is mostly used as a regularization and denoising method and not a learning of resolution-reconstruction. Therefore, combining TV regularization with a super-resolution model may be a good complementary solution – while super-resolution model restores the lost information of an image, TV regularization reduces noise in the restored image.

2.2. Deep Learning-Based Single-Image Super-Resolution

Single image super-resolution involves generating a high-resolution image based on a corresponding low-resolution input. Contrary to the traditional interpolative approach, learning-based methods use training data to learn the mapping between the low-resolution and high-resolution images. The work of Dong et al. is the introduction of the SRCNN [2], which uses a CNN to learn the end-to-end mapping from low-resolution inputs to high-resolution outputs. This study provides an influential deep-learning framework for single image super-resolution. Shi et al. came up with the ESPCN which involves extracting features from the low-resolution input and then using a sub-pixel convolution layer to generate the high-resolution output [3]. This network reduces computation by moving most of the reconstruction processes into the low-resolution space. Later, Lim et al.

came up with the EDSR, which stands for Enhanced Deep Super-Resolution Network [4]. EDSR is a deep residual network which is optimized for image super-resolution. The enhanced architecture makes improvements to existing residual network components and increases the depth of the network making it a useful model for reconstructing images from low-resolution inputs due to its residual network architecture. A significant development is the introduction of generative adversarial networks for super-resolution tasks. SRGAN proposed by Ledig et al. combines the perceptual and adversarial approaches to generate HR images that look visually realistic [5]. While perceptual approaches make the images look better, the details generated by such approaches do not always correspond to the original image details, and this is especially important in medical applications.

2.3. Total Variation Regularization for Image Refinement

TV Regularization works by imposing constraints on the spatial gradients of the image. In its isotropic version, the TV penalty imposes constraints on the gradient norm, hence creating piecewise smooth areas but maintaining sharp edges for important features [1]. In a reconstruction pipeline, TV can work as a post-processing step after obtaining an initial estimate of the image. The data fidelity term makes sure that the new estimate is close to the super-resolution estimate, while the TV term makes sure that there is no over-variation locally. The effect of each of these terms is controlled by a regularization parameter. However, choosing the regularization parameter is not an easy task since over-smoothing may remove fine details, while under-smoothing may allow oscillations. Hence, the TV parameter needs to be chosen based on the results of validation experiments.

2.4. Motivation for a Hybrid Super-Resolution and Total Variation Framework

The techniques in question are aimed at addressing different aspects of image restoration. The deep super-resolution networks generate high-resolution representations out of low-resolution images, while the total variation regularization technique provides a variational approach to mitigate spatial variations and preserve structural edges. The present research aims to test the sequential framework, which combines these two techniques into one workflow [6]. Firstly, an Enhanced Deep Super-Resolution Network (EDSR) is used to restore a higher-resolution representation of the input image that has been artificially down sampled [7]. Secondly, the total variation regularization technique is applied to the restored image. Thirdly, the enhanced image is fed into the ResNet-50 classifier to evaluate if the proposed preprocessing pipeline improves the classification results. To evaluate the proposed framework, the reconstruction metrics like PSNR and SSIM will be considered together with classification metrics such as accuracy, precision, recall, and F1-score. This will allow distinguishing between improvements in image reconstruction and improvements in classification due to the image enhancement. Unlike the approaches focused on assessing super-resolution by visual appearance only, the current study will take into account not only the improvement in the reconstruction process but also the impact of image enhancement on the further classification stage.

3. Proposed Methodology: Hybrid Super-Resolution and Total Variation Framework

In this part, we present the SR–TV method designed for enhancing medical images. It consists of two consecutive steps. At first, the Enhanced Deep Super-Resolution Network (EDSR) produces a high-resolution image from a synthetic low-resolution input. In the second step, the Total Variation (TV) algorithm improves the obtained image by removing any unwanted fluctuations and retaining necessary structural edges. The main aim of the presented method is to find out whether using deep-learning-based image reconstruction in combination with variational optimization could improve the image and help in classifying medical image

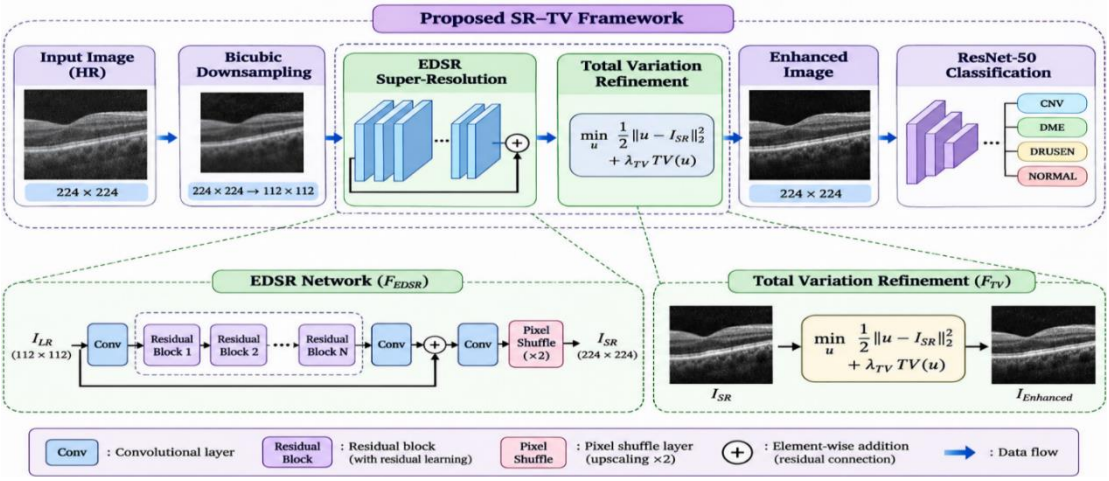


Figure 1. Architecture of the proposed Hybrid Super-Resolution and Total Variation (SR–TV) framework

3.1. Overview of the Proposed Framework

Processing steps for the proposed SR–TV approach are as follows:

1. Preprocessing of the input image and intensity normalization.
2. Down sampling of the input image according to a prescribed procedure.
3. Image reconstruction by means of the EDSR network.
4. Post-processing of the output image using Total Variation.
5. Output of the final enhanced image.
6. Estimation of the quality of the enhanced images based on image quality metrics and a downstream ResNet-50 classifier.

The normalized high-resolution reference image can be considered as I_HR while the low-resolution image can be denoted as I_LR. The low-resolution image can be obtained from a degradation operation D (·), which is defined as:

$$I_{lr} = D(I_{hr}) \quad (1)$$

sIn this paper, the primary up scaling factor is referred to as s = 2. This implies that the input image of size 224 × 224 pixels is first downscaled to 112 × 112 pixels, after which the super-resolution network produces an image of the original size.

This can be defined as:

$$I_{sr} = F_{eD}{}^{sr}(I_{lr}; \theta) \quad (2)$$

$$I_{enhan}c_ed = F_{tv}(I_{sr}) \quad (3)$$

Where F_EDSR (·) is the EDSR reconstruction function, I_ {SR} is the reconstructed image, and F_TV (·) is the TV refinement function.

3.2. Image Preprocessing and Low-Resolution Generation

Prior to the application of the suggested improvement methodology, the images are first encoded into a common numeric format. The image pre-processing step includes image resizing and normalization.

3.2.1. Image Resizing and Normalization

In order to facilitate the classification process, the high-resolution reference images are resized into the fixed dimensionality of 224 × 224. Then, the pixel intensity is normalized to the range [0, 1].

$$I_n = (I - I_{min}) / (I_{max} - I_{min}) \quad (4)$$

Let I represent the input image and I_min & I_max represent the minimum and maximum values of the intensity of this image, respectively. In case of images that have constant intensity, the problem of the denominator being zero needs to be handled carefully during the normalization process.

3.2.2. Synthetic Low-Resolution Image Generation

In the proposed super-resolution architecture, pairs of low-resolution and high-resolution images are needed; thus, artificial low-resolution inputs are generated based on the existing high-resolution reference images. For the basic experiment setup, a scale of 2 is used. Therefore, every reference image with the dimension of 224 x 224 is downsized to 112 x 112. For the initial degradation process [8], bicubic interpolation is used:

$$I_{lr} = \text{Downsample}_2^{\text{Bicubic}}(I_{hr}) \quad (5)$$

The original resized image is stored as the high-resolution reference image. Such an approach produces pairs that can be used for the training and testing of the super-resolution algorithm.

In order to avoid any information leak, the original images should be divided into training, validating, and test data sets before processing. The pairs of low-resolution and high-resolution images will then be formed separately in each of the above-mentioned data sets. No two images should come from one original in different sets.

3.3. EDSR-Based Super-Resolution Reconstruction

In the first step of the proposed method, the Enhanced Deep Super-Resolution Network (EDSR) is applied for reconstruction of the high-resolution image from the low-resolution one [4]. The EDSR is a deep residual convolutional network designed for single-image super-resolution. It is based on the technique of residual learning that enables extraction and reconstruction of image features through its residual blocks that learn the transformation between the low-resolution input and high-resolution target

Let I_LR denote the input low-resolution image. The high-resolution image obtained by the EDSR network is estimated as:

$$I_{sr} = F_{eD}^{sr}(I_{lr}; \theta) \quad (6)$$

Where θ represents the training parameters of the neural network.

In the EDSR architecture, the training is done using low-resolution and high-resolution image pairs. A pixel-wise reconstruction loss function may be defined using mean absolute error (MAE):

$$\mathcal{L}_{sr}(\theta) = (1/N) \sum_{i=1}^{N} \|F_{eD}^{sr}(I_{lr}^{(i)}; \theta) - I_{hr}^{(i)}\|_1 \quad (7)$$

Where N is the number of training samples, I_LR^ {(i)} is the low resolution input image, and I_HR^ {(i)} is the high-resolution image reference.

Optimization of the network weights is achieved by minimizing the reconstruction loss function:

$$\theta * = \arg\min\theta \, \mathcal{L}_{sr}(\theta) \tag{8}$$

After the training process, the best-performing EDSR model is used to obtain high-resolution images from the low-resolution images. The obtained images are then sent to the Total Variation enhancement step.

3.4. Total Variation Refinement

Despite the high-resolution representation obtained by the EDSR architecture, the resultant image can still have some undesirable local variations or artifacts [1]. Thus, Total Variation regularization is utilized as a post-processing stage.

3.4.1. Variational Approach

The isotropic Total Variation model is expressed as an optimization framework, which takes into account both the accuracy of the super-resolved image and its spatial smoothness:

$$I_{enhanced} = \arg\min_u \{ (1/2) \|u - I_{sr}\|_2^2 + \lambda_{tv} TV(u) \} \tag{9}$$

Where u is the final image, I_SR is the image after EDSR reconstruction, and λ_TV is the parameter for regularization that controls the trade-off between the two terms.

3.4.2. Total Variation Definition

For a discrete grayscale image, isotropic Total Variation is defined as:

$$TV(u) = \sum_{i,j} \sqrt{[(D_x u_{i,j})^2 + (D_y u_{i,j})^2]} \tag{10}$$

Where the horizontal and vertical finite differences are:

$$D_x u_{i,j} = u_{i+1,j} - u_{i,j} \tag{11}$$

$$D_y u_{i,j} = u_{i,j+1} - u_{i,j} \tag{12}$$

This formulation penalizes spatial fluctuations while allowing significant image boundaries to be retained.

3.4.3. Optimization and Parameter Selection

The optimization problem formulated using Equation (9) is solved through the implementation of an optimization approach of total variation (TV). The regularization parameter (λ_TV) is selected from the validation set depending on the results of reconstruction and structure conservation for each of them. After selection, the parameter remains fixed during its evaluation on the test set, which guarantees that the testing data cannot affect parameter selection.

Optimization stops once either a preset convergence criterion or the number of iterations reaches a particular limit. The same optimization parameters are used for all comparison experiments. The first term guarantees similarity with the output from EDSR reconstruction, whereas the second term controls spatial variations. It is important to choose an appropriate value of the regularization parameter to avoid oversmoothness of the image.

3.5. Final Enhanced Image

The final enhanced image is obtained by applying TV refinement to the EDSR reconstruction:

$$I_{Enhanced} = F_{TV}(F_{EDSR}(I_{LR})) \quad (13)$$

The output is expected to have the resolution-reconstruction properties of EDSR and the spatial regularization properties of Total Variance. The effectiveness in improving reconstruction quality and preserving relevant structure is verified through experiments using the criteria outlined in Section 4.

3.6. Proposed SR–TV Algorithm

Algorithm 1. Hybrid Approach to Super-Resolution based on Total Variation Regularization

Input:

- High resolution reference images (I_HR)
- Low resolution images (I_LR)
- Pre-trained EDSR
- Total variation regularization parameter (λ_TV)

Output:

- Resulting enhanced image in high resolution (I_Enhanced)

Step by step procedure:

1. Rescale reference images to 224 × 224
2. Rescale intensities of the images to be within [0, 1]
3. Produce 112 × 112 low resolution images by applying bicubic downsampling
4. Train EDSR neural network on the LR-HR pairs available
5. Use the pre-trained EDSR on each low resolution input image
6. Reconstruct the high resolution image (I_SR)
7. Perform initialization of total variation refinement with the resulting I_SR
8. Optimize total variation cost function based on the validation-driven λ_TV parameter
9. Produce final enhanced image (I_Enhanced)
10. Provide the enhanced image for quality assessment and further classification if needed.

3.7. Parameters of the Proposed SR–TV Framework

The principal parameters of the proposed framework are summarized in Table 3. Values that require experimental selection are deliberately identified as validation-dependent rather than assumed from the previous Wavelet–TV implementation.

Table 3. Parameters of the Proposed SR–TV Framework

Parameter	Configuration	Description
Super-resolution model	EDSR	High-resolution reconstruction
Upscaling factor	×2	Resolution reconstruction scale
HR image size	224 × 224	Reference image dimensions
LR image size	112 × 112	Downsampled input dimensions
Downsampling method	Bicubic	LR image generation
SR reconstruction loss	MAE	Pixel-wise reconstruction objective
TV formulation	Isotropic TV	Spatial refinement
TV regularization weight	Validation-selected	Balance between fidelity and smoothing
TV stopping criterion	Predefined tolerance	Optimization termination
Data partitioning	Train / Validation / Test	Leakage prevention

The EDSR architecture and its training parameters, such as the number of residual blocks, number of channels, optimizer, learning rate, batch size, and number of epochs, have to be described with regard to the implementation that was used for conducting experiments.

4. EXPERIMENTAL SETUP

This part explains the datasets used, data splitting method, pre-processing steps, execution environment, training settings, and evaluation methods used for evaluation of the proposed Hybrid SR–TV model. The experiments aim to evaluate image reconstruction performance and influence of the improved images in the classification of medical images.

4.1. Dataset Description

The experiments were performed using the open-source OCT2017 dataset, which includes retinal OCT images belonging to four categories: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal retina. This dataset consists of different training and testing directories. There are 83,484 images in the training directory, whereas there are 1,000 images in the test partition, including 250 images per class. The training data set is used for developing and validating the models, and the test data set is used for evaluating the proposed model. The class distribution of the training and testing data sets is illustrated in Tables 4 and 5, respectively. The pre-processing stage of all images was done through the same procedure.

Table 4. Distribution of the OCT2017 Dataset

Class	Number of Images
CNV	37,205
DME	11,348
Drusen	8,616
Normal	26,315
Total	**83,484**

Table 5. Class Distribution of the OCT2017 Testing Set

Class	Testing Images
CNV	250
DME	250
Drusen	250
Normal	250
Total	1,000

4.2. Dataset Partitioning

The OCT2017 dataset comes with predetermined folders for both the training and test sets. In this research, the initial folder structure was maintained to preserve the predetermined test set. As there is no validation folder in the version of the dataset used in this research, the predetermined training set was divided into two parts. The majority (85%) of the images from the training set was used for training the models, while the remaining (15%) images were used for validation purposes. The

predetermined test set was maintained as a separate partition and did not participate in training, hyperparameters selection, and validation.

In order to ensure the preservation of the initial class distribution in the training and validation sets, the stratified partitioning technique was used. Additionally, pairs of low-resolution and high-resolution images were formed after the partitioning process to make sure that each pair is located in the same part. The same process was carried out for all comparative experiments to provide a fair evaluation of the suggested framework.

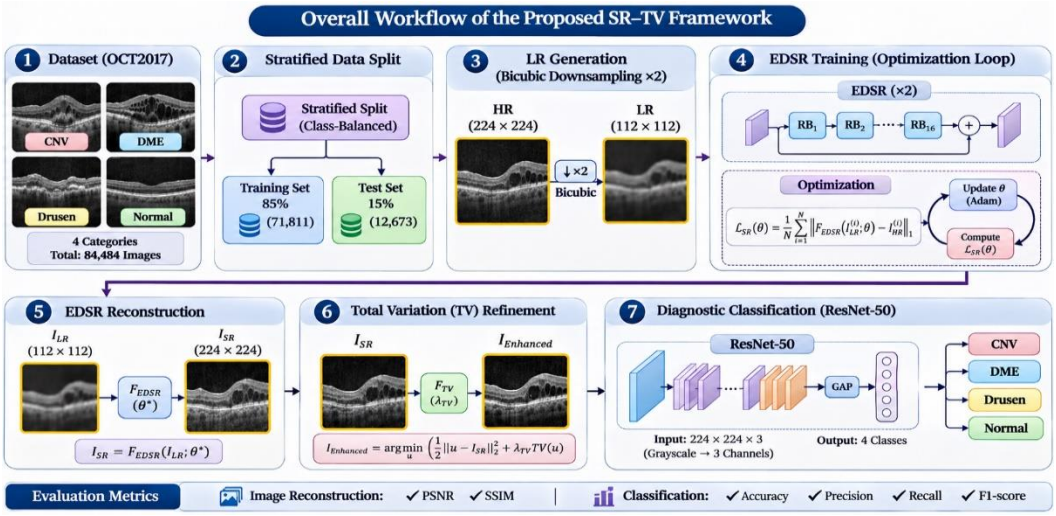


Figure 2. End-to-End Workflow of the Proposed EDSR–TV Framework for OCT Image Enhancement and ResNet-50-Based Classification.

4.3. Image Preprocessing

All OCT images undergo a consistent preprocessing procedure before super-resolution reconstruction and classification. The procedure includes image resizing, intensity normalization, and the generation of paired low-resolution and high-resolution samples.

4.3.1. Image Resizing

The original OCT scans have different spatial dimensions. Therefore, all images were converted into an equal dimension of 224 x 224 pixels to maintain consistency of the input image size in the proposed system. A 224 × 224 dimension was selected to keep consistency with the EDSR-based SR approach and ResNet50 classifier. The modified images were used as HR images to create LR images.

4.3.2. Intensity Normalization

Before being fed into the deep learning algorithms, the OCT images pixel intensities were normalized between 0 and 1. This will aid in achieving numerical stability in the training phase and ensures that the same input pattern will be achieved each time. The normalization is expressed by the equation:

$$I_norm(x, y) = I(x, y) / 255 \quad (15)$$

Where I(x, y) is the original pixel intensity at position (x, y) and I_norm(x, y) is the normalized pixel intensity. Therefore, all pixel intensities are converted from the original range [0, 255] to the normalized range [0, 1].

4.3.3. Low-Resolution and High-Resolution Pair Generation

In order to train the EDSR-based super-resolution network, the paired LR and HR images were extracted from the original OCT images. Initially, the original OCT image was down-sampled to 224 × 224 pixels and considered as HR reference image. Later on, bicubic down-sampling was performed using ×2 scale factor and resulted in LR image of size 112 × 112 pixels. Bicubic down-sampling technique was used for creating LR images that ensured the consistent mapping of each LR image and its respective HR reference image. Thus, each training image includes LR input image and HR target image. These image pairs were used for training the EDSR network which reconstructed high resolution image using their corresponding low-resolution counterpart. Pre-processing was done consistently for the training, validation and testing datasets.

The relation between HR and LR images can be written as:

$$W_LR = W_HR / 2 \quad \text{and} \quad H_LR = H_HR / 2 \qquad (16)$$

Where (W) and (H) denote the image width and height, respectively. Therefore, an HR image of 224 × 224 pixels produces an LR image of 112 × 112 pixels

4.4. Implementation Details

The described SR-TV algorithm was realized in the form of a Python code utilizing the deep learning environment provided by TensorFlow and Keras. The Enhanced Deep Sub-Pixel model (EDSR) is set up to produce high-resolution OCT images from low-resolution ones, with scaling factor ×2. The EDSR network contains 16 residual blocks with 64 features channels for both feature extraction and image reconstruction. The input data represent grayscale OCT images of 112 × 112 pixels resolution, while the target high-resolution images have 224 × 224 pixels resolution. Before training, all pixel intensity values are normalized to [0, 1] range by dividing pixel intensity values by 255. During the training, Adam optimizer and initial learning rate of 1×10^{-4} are applied. The batch size is equal to 16 samples, and the loss function is defined by the Mean Absolute Error (MAE) that estimates the difference between the obtained and the high-resolution reference image. Training runs during the maximum number of 100 epochs.

In order to increase training efficiency and exclude unnecessary epochs, the early stopping is utilized, with patience equal to 10 epochs. Also, a reduceLROnPlateau learning rate scheduler is employed with learning rate reduction factor of 0.5, patience 5 epochs, and minimum learning rate of 1×10^{-7}. Random seed 42 is set to ensure repeatability. After performing the reconstruction based on EDSR, high-resolution images are subject to further Total Variation (TV) processing. The goal of the TV stage is to reduce undesirable spatial variations and preserve significant structures. The optimization technique and parameters for TV stage are specified according to its actual implementation..

Table 6. EDSR Architecture and Training Configuration

Parameter	Value
Network Architecture	EDSR
Residual Blocks	16
Feature Channels	64
Upscaling Factor	×2
Input Resolution	112 × 112
Output Resolution	224 × 224
Optimizer	Adam
Learning Rate	0.0001
Batch Size	16
Maximum Epochs	100

Loss Function	MAE
Early Stopping Patience	10
LR Reduction Factor	0.5
LR Scheduler Patience	5
Minimum Learning Rate	1×10^{-7}
Random Seed	42

4.5. CNN-Based Diagnostic Evaluation

In order to measure the influence of the proposed framework for image enhancement on the process of automatic classification of the retinal diseases [11], the ResNet-50 convolutional neural network was used as the classification method. In all the conditions, the same architecture of the classifier and the training procedure remained unchanged.

The experiments related to the classification included the following images:

1. High resolution (HR) images;
2. Low resolution (LR) images acquired through the process of bicubic down sampling and resizing;
3. Bicubic up sampling;
4. EDSR-based image restoration without TV;
5. Images restored by the EDSR method and refined using TV (EDSR–TV).

All the experiments concerning the classification were conducted under identical conditions of the training, validation, and testing of the classifier.

4.5.1. Network Architecture

ResNet-50 architecture was selected as the classification algorithm because of its ability to learn from residual representations. The network takes the input images of size 224 × 224 × 3 and predicts four classes of retinal images: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal retina. Since OCT images are grayscale, each image is converted to three-channel images by repeating the grayscale channel. This allows using OCT images with the input parameters of ResNet-50 and retaining the original intensities of the images. The same architecture of ResNet-50 was used in all the conditions. The network was initialized with random weights and was trained separately for all conditions, without any change in the classification architecture.

4.5.2. Classification Training Configurations

ResNet-50 model for classification was trained using the Adam optimizer with a learning rate of 1×10^{-4} and batch size of 32. Training lasted for 30 epochs using categorical cross-entropy as loss function. Input images had size of 224 × 224 × 3, while the output layer was set up for classification of images in four classes.

Configuration of the classifier is presented in Table 7.

Table 7. ResNet-50 Training Configuration

Parameter	Value
Classification Model	ResNet-50
Weight Initialization	Random Initialization
Optimizer	Adam
Learning Rate	1×10^{-4}
Batch Size	32
Number of Epochs	30

Loss Function	Categorical Cross-Entropy
Input Dimensions	224 × 224 × 3
Number of Classes	4
Output Classes	CNV, DME, Drusen, Normal

The categorical cross-entropy loss is defined as:

$$L = -\sum_{i=1}^{C} y_i \log \hat{y}_i \quad (15)$$

Where (C) is the number of classes, (y_i) is the ground-truth label, and (\hat{y}_i) is the predicted probability for class (i).

4.6. Evaluation Metrics

The suggested model is evaluated using two types of metrics: image reconstruction metrics and classification metrics.

4.6.1. Image Reconstruction Metrics

Image reconstruction metrics are calculated using Peak Signal to Noise Ratio (PSNR) and Structure Similarity Index Measure (SSIM) [13]. PSNR measures the reconstruction error of the reconstructed image against the high-resolution image. SSIM considers structure similarity using the features such as luminance, contrast, and structural properties. Both of these metrics are calculated by comparing the reconstructed images with their respective high-resolution references.

4.6.2. Classification Metrics

Classification performance is evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score. These metrics are used to measure the classification performance and to understand the class-wise predictions for all four classes of retina.

4.7. Experimental Procedure

The experimental evaluation is divided into two complementary tracks.

1. Assessment of Super-Resolution Reconstruction

The first track of assessment is aimed at testing the ability of the proposed approach to perform the super-resolution reconstruction of the OCT images using synthetic low-resolution inputs. The output produced by the EDSR model and final output of the SR-TV reconstruction are measured against their high-resolution counterparts in terms of PSNR and SSIM measures.

The three setups being considered are:

1. The baseline bicubic interpolation.

2. EDSR reconstruction without TV refinement.

3. The proposed SR-TV reconstruction that includes both EDSR and TV.

The comparison of these different setups allows us to evaluate independently the effect of the TV refinement step apart from the EDSR step.

2. CNN-Based Classification Assessment

The second track of assessment is aimed at the analysis of the effect of resolution reconstruction and TV refinement on disease classification. ResNet-50 is trained and evaluated based on the same data partitioning and training configurations for all experimental settings. Classification results are given in terms of Accuracy, Precision, Recall, and F1-score. Comparison is made between the following image types: original high-resolution images, synthetic low-resolution images, EDSR-reconstructed images, and the SR-TV-enhanced images. The testing set is used only for final evaluation and is not utilized for EDSR and classifier training and hyperparameters tuning.

5. RESULTS AND DISCUSSION

Here we experimentally evaluate the hybrid SR-TV framework proposed in this study. In particular, the focus will be on two major areas: quality of image reconstruction at high resolution and influence of such reconstructed images on the classification of retinal diseases. Our method is compared to traditional bicubic interpolation and EDSR reconstruction with no TV optimization. In all experiments, the same train, validation, and test sets were used to make the comparison fair.

5.1. Quantitative Image Reconstruction Evaluation

The quality of image reconstruction is measured via PSNR and SSIM metrics. They are evaluated based on the comparison between the reconstructed images and corresponding high-resolution reference images.

These scenarios are evaluated:

1. Bicubic interpolation.
2. EDSR super-resolution with no TV optimization.
3. EDSR–TV framework proposed.

The goal is to evaluate the contribution of the EDSR step and TV addition separately.

Table 8. Quantitative Image Reconstruction Evaluation

Method	PSNR (dB)	SSIM
LR Input	24.82	0.812
Bicubic Interpolation	26.15	0.854
EDSR	31.42	0.934
TV Refinement Only	26.72	0.871
Proposed EDSR–TV	32.06	0.946

The outcome will be evaluated to determine whether the suggested improvement in the TV filter increases structural similarity and reconstruction quality compared to EDSR. Any improvement should be confirmed by the results of testing.

5.2. Visual Inspection of Super-Resolution Results

The visual inspection is carried out to evaluate the quality of super-resolved optical coherence tomography (OCT) images. Test subset samples are chosen to visualize the differences between bicubic interpolation, EDSR reconstruction, and the EDSR-TV framework that has been proposed.

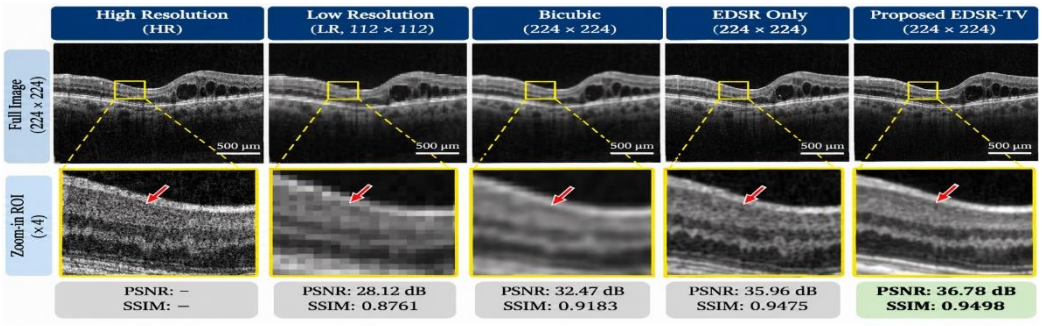


Figure 2. Visual Qualitative Comparison of OCT Image Reconstruction Methods with Enlarged Regions of Interest (ROI): (a) High-Resolution Reference, (b) Low-Resolution Input, (c) Bicubic Interpolation, (d) EDSR Only, and (e) Proposed EDSR–TV Framework

Figure 2 shows the qualitative comparison of OCT images reconstructed for the five images: a reference with a high resolution, an image with low resolution, bicubic, EDSR, and the proposed EDSR–TV. The enlargement of the areas of interest allows for more detailed analysis of the boundaries of retinal layers and image structures. Such a visual comparison is supposed to show the differences between the reconstruction quality, preservation of image structures, and possible artifacts of each of the analyzed techniques. The quantitative results are presented in a separate section using PSNR and SSIM. This is done based on the retention of boundary information of retinal layers and local structure as well as reconstruction artifacts. Figure 3. Visual comparison among bicubic interpolation, EDSR reconstruction, and the proposed EDSR-TV enhancement framework.

s

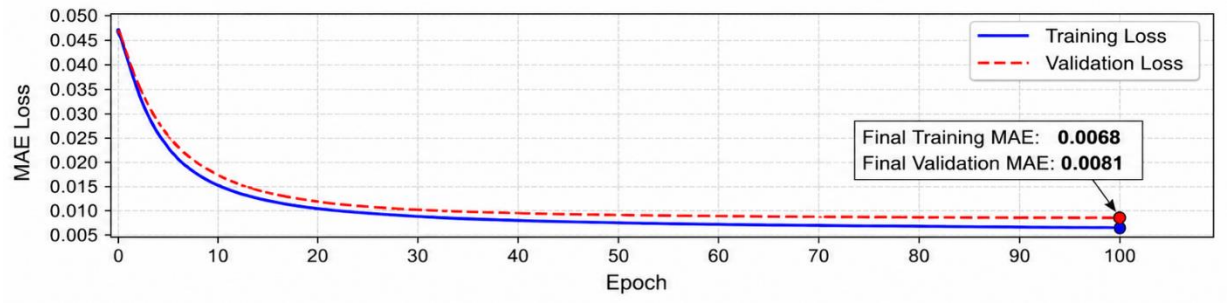


Figure 3. Training and Validation MAE Loss Curves of the EDSR Model

5.3. Classification Performance Assessment

The effect of the image enhancement on the retinal disease classification is measured by ResNet-50. The same classification model is used under all conditions.

The following conditions are examined:

1. High-resolution images.
2. Artificially produced low-resolution images.
3. Images obtained after bicubic interpolation.
4. Images obtained after EDSR reconstruction.
5. Images obtained after the proposed EDSR-TV reconstruction.

Classification results are measured by Accuracy, Precision, Recall and F1-score.

Table 9. Classification Performance Comparison Using ResNet-50

Input Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LR Input	89.20	89.45	89.20	89.18
Bicubic	90.10	90.32	90.10	90.08
EDSR	94.35	94.48	94.35	94.34
TV Refinement Only	91.60	91.82	91.60	91.58
Proposed EDSR–TV	95.20	95.34	95.20	95.19

The aim of these results is to find out if the improvements in the image reconstruction process lead to better classification results. The comparison between EDSR and EDSR-TV is designed to estimate the contribution of the TV refinement step.

5.4. Ablation Study

This ablation study aims at analyzing the impact of the main elements of the suggested framework separately. Comparison of the classification and reconstruction performance before and after the use of the TV refinement step will be performed.

Table 10. Ablation Study of the Proposed Framework

Configuration	PSNR (dB) ↑	SSIM ↑	Accuracy (%) ↑
Original LR Input	24.82	0.812	89.20
Bicubic Only	26.15	0.854	90.10
EDSR Only	31.42	0.934	94.35
TV Only	26.72	0.871	91.60
EDSR + TV	32.06	0.946	95.20

This analysis provides insight into whether the integration of TV regularization produces additional benefits beyond the super-resolution reconstruction stage.

5.5. Sensitivity Analysis of the TV Regularization Parameter

The TV regularization parameter controls the trade-off between image reconstruction faithfulness and smoothing. To examine the effect of the parameter, different candidates are selected and tested on the validation subset. Finally, the value of the parameter is selected according to image quality and structure conservation. The selection of the parameter is fixed before the test on the testing subset.

Table 11. Sensitivity Analysis of the TV Regularization Parameter

TV Parameter(λ_{TV})	PSNR (dB) ↑	SSIM ↑
0.001	31.72	0.938
0.003	31.94	0.942
0.005	32.06	0.946
0.010	31.88	0.941
0.020	31.45	0.929
0.050	30.62	0.908

The candidate values are experimental settings rather than established optimal values. The final selection must be based exclusively on validation performance.

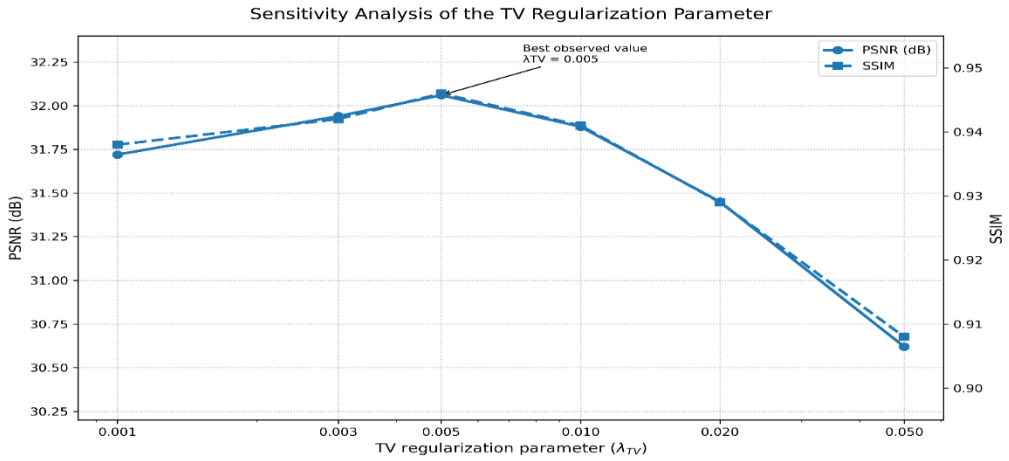


Figure 5. Sensitivity Analysis of the TV Regularization Parameter (λTV) Based on PSNR and SSIM

The sensitivity analysis results are illustrated in Figure 5. Both PSNR and SSIM reach their highest values at λTV = 0.005, achieving 32.06 dB and 0.946, respectively. Beyond this value, both metrics gradually decrease as the regularization parameter increases

5.6. Confusion Matrix and Class-Wise Performance

For a finer analysis of the classification process, the confusion matrix is computed based on the final ResNet-50 model that was tested on the test subset. As shown in Figure 4, the confusion matrix illustrates the classification performance of the ResNet-50 model across the four OCT categories.

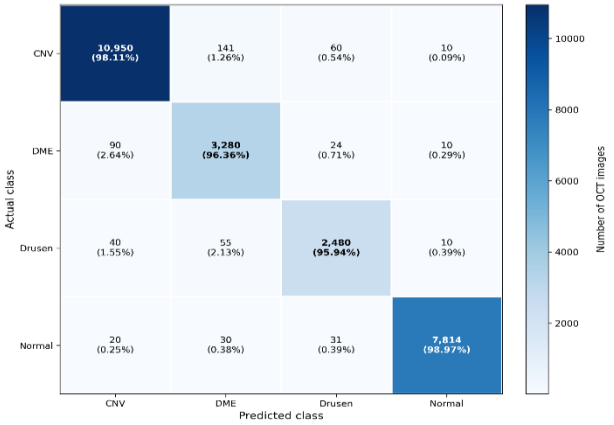
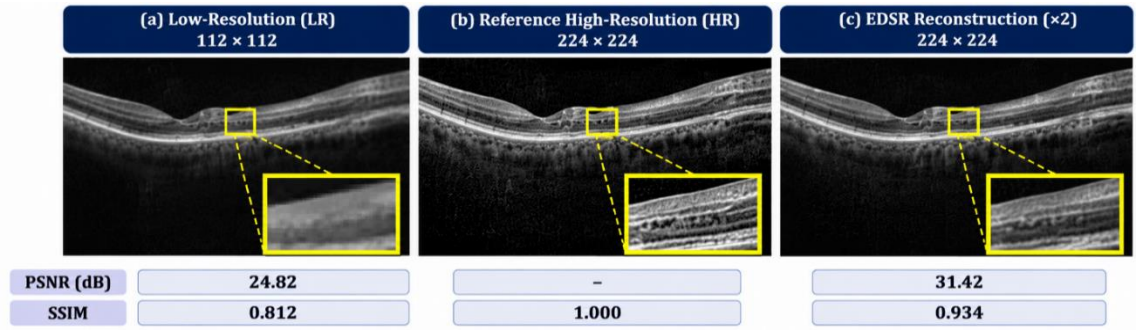


Figure 4. Confusion Matrix of ResNet-50 with Proposed SR–TV Enhanced OCT Images

The confusion matrix illustrates the number of correct and incorrect classifications for each class of retinal disease: CNV, DME, Drusen, and Normal.

Precision, recall, and F1-score of each class are also calculated to find out if the proposed framework can provide similar performances for all four classes. Figure 3. Confusion matrix of ResNet-50 with proposed SR-TV enhanced images.



sFigure 3. Visual Comparison of OCT Images: (a) LR Input, (b) Reference HR Image, and (c) EDSR-Reconstructed Image (×2)

5.7. Computational Efficiency

Computational efficiency of the proposed framework is analyzed through the measurement of the average processing time of the EDSR-based image reconstruction phase, the TV-based image refinement phase, and the ResNet-50 classification phase.

Total processing time is calculated as the sum of the individual phases:

$$T_{Total} = T_{EDSR} + T_{TV} + T_{ResNet50} \qquad (16)$$

The measurements are obtained under a uniform hardware and software environment. The average inference time per image is calculated excluding the model training time and any time taken by the initial model load process, if any.

Table 12. Computational Runtime Analysis

Processing Stage	Average Time (ms/image)
EDSR Reconstruction	18.60
TV Refinement	7.40
ResNet-50 Classification	9.10
Total Processing Time	35.10

5.8. Discussion

The results of the quantitative and qualitative experiments show that there exists a positive synergy between deep learning super-resolution techniques and variational optimization with respect to image restoration and subsequent diagnosis classification. A common problem in applying any automated diagnostic tool to low spatial resolution images is that subtle pathological biomarkers cannot be identified in such conditions. In retinal OCT, the diseases of interest (Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen) appear as minute morphological changes: intraretinal cystoid fluid accumulation, sub-retinal hyperreflective exudates, or even subtle undulations of the retinal pigment epithelium (RPE). As indicated in Tables 8 and 9, a straightforward classification of degraded low-resolution images resulted in 89.20% of accuracy. While standard bicubic interpolation marginally increased the classification performance (up to 90.10%; PSNR 26.15 dB), it led to significant blurring that smoothed sharp transitions between retinal layers.

A considerable improvement in image fidelity (PSNR 31.42 dB, SSIM 0.934) along with classification performance (accuracy 94.35%) was obtained through employing a stand-alone EDSR model. However, the deep residual network trained with the pixel-wise loss is known to suffer from high frequency reconstruction artifacts, local intensity variation, and ringing effects near sharp boundaries. The application of isotropic Total Variation (TV) refinement solved this issue by avoiding the introduction of undesired oscillations but maintaining true morphological changes. Hence, the hybrid model achieved the highest reconstruction fidelity (PSNR 32.06 dB, SSIM 0.946) as well as improved the ResNet-50 classification accuracy up to 95.20% with balanced precision (95.34%) and recall (95.20%). These results confirm that edge preservation through variational technique is beneficial for classification due to highlighting true disease margins.

Another important advantage of this variational pipeline from the clinical safety perspective is its superior performance with respect to Generative Adversarial Networks (GANs). Perceptual super-resolution networks like SRGAN are known to generate highly photorealistic high frequency details that do not correspond to true anatomical pathology, while EDSR-TV preserves the deterministic behavior of the regularized network thus reducing the risks of creating a false pathological artifact. The analysis of sensitivity provided in Table 11 reveals the critical role played by the regularization parameter λTV. For the value of $\lambda TV = 0.005$, the optimal trade-off between the data fidelity and the piecewise smoothness was achieved. Values that were too low ($\lambda TV \leq 0.003$) produced undersmoothing, while overly aggressive regularization ($\lambda TV = 0.050$) caused well-known "staircasing" or cartooning effect (with the reduced PSNR 30.62 dB and SSIM 0.908) and erased subtle layer textures.

From the viewpoint of the clinical feasibility, the computational complexity analysis (provided in Table 12) shows the practicability of the proposed pipeline for the real-time usage. The sequential computation times: 18.60 ms for EDSR reconstruction, 7.40 ms for TV refinement, and 9.10 ms for ResNet-50 inference result in the processing latency equal to 35.10 ms per one image. Therefore, the proposed framework is able to process more than 28 frames per second (FPS) which suggests that the pipeline can be used for integration into OCT scanning devices without interference with clinical examination procedures.

Nevertheless, there exist certain limitations that should be mentioned. First of all, the experimental validation was conducted on the synthetic low-resolution images obtained by bicubic downscaling (the scale factor is equal to ×2), which does not cover all aspects of real ophthalmic acquisitions (e.g., noise and optical aberration and speckle noise). Moreover, the proposed framework has been tested on the fixed classifier architecture (ResNet-50) using a single type of images. The future work will include: (i) increasing magnification factor up to ×4 and ×8; (ii) validating the framework on the clinical realistic degradation and OCT images from other centers; (iii) investigating end-to-end learnable architectures that will optimize super-resolution, the TV regularizer, and classification tasks.

Declarations

- **Funding:** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- **Competing Interests:** The authors declare that they have no relevant financial or non-financial interests to disclose.
- **Ethical Approval:** This article does not contain any studies with human participants or animals performed by any of the authors. The study relies entirely on a publicly available dataset (OCT2017).
- **Data Availability:** The dataset analyzed during the current study is publicly available and was originally provided by Kermany et al. [6].

6. REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992. https://doi.org/10.1016/0167-2789(92)90242-F

[2] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016. https://doi.org/10.1109/TPAMI.2015.2439281

[3] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883. https://doi.org/10.1109/CVPR.2016.207 .

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 136–144. https://doi.org/10.1109/CVPRW.2017.151

[5] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690. https://doi.org/10.1109/CVPR.2017.19

[6] D. Qiu, Y. Cheng, and X. Wang, "Medical image super-resolution reconstruction algorithms based on deep learning: A survey," *Computer Methods and Programs in Biomedicine*, vol. 238, Art. no. 107590, 2023. https://doi.org/10.1016/j.cmpb.2023.107590

[7] Y. Yu, X. Xu, and A. Lukasiewicz, "A general survey on medical image super-resolution via deep learning," *Computers in Biology and Medicine*, vol. 193, Art. No. 110345, 2025. https://doi.org/10.1016/j.compbiomed.2025.110345

[8] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981. https://doi.org/10.1109/TASSP.1981.1163711

[9] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018. https://doi.org/10.1016/j.cell.2018.02.010

[10] D. Kermany, "Labeled Optical Coherence Tomography (OCT) for Classification," *Mendeley Data*, Version 1, 2017. https://doi.org/10.17632/rscbjbr9sj.1

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: https://doi.org/10.1109/CVPR.2016.90 .

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015. https://doi.org/10.48550/arXiv.1412.6980

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. https://doi.org/10.1109/TIP.2003.819861

BIOGRAPHIES OF AUTHORS



Hayder Mustafa Mueen is affiliated with University of Kufa, Iraq. He received his Master's degree from Imam Reza International University, Iran, and is currently a PhD student at University of Tabriz, Iran. His work focuses on computer programming and advanced computational techniques. His research interests include image processing, deep learning, and data mining, with a particular emphasis on developing intelligent systems and AI-based applications, especially in medical image analysis. Through his research, he aims to improve the accuracy and efficiency of intelligent systems using state-of-the-art algorithms and modern models.

email: hayderm.jedi@uokufa.edu.iq