Journal of Scientific & Industrial Research
Vol. 84, February 2025, pp. 148-161
DOI: 10.56042/jsir.v84i02.8327

GAN-CNN based Structure-Preserving Mixed Noise Removal Model for Enhancing Medical Image

Vishal H Shah* & Prajna Parimita Dash

Department of Electronics & Communication Engineering, Birla Institute of Technology, Mesra, Ranchi 835 215, Jharkhand, India

Received 26 January 2024; revised 28 October 2024; accepted 17 December 2024

The current era of the Internet of Medical Things (IoMT) and Medical Artificial Intelligence (MAI) makes medical imaging a prominent mode of providing effective solutions in diagnosis and prognosis. The main issue with these images is the presence of noise that requires enhancement through effective edge preservation and noise reduction. The proposed work introduces a two-stage Deep Learning (DL) model, utilizing Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNN) for jointly reducing speckle, impulse, and Gaussian noise while preserving edge information in noisy medical images. The work also explores the probabilistic evaluation of generators and discriminators for compensating lossy patches to ensure image quality. The performance of the proposed model is investigated by considering three different performance metrics, namely, PSNR, FSIM, and SSIM. Moreover, non-parametric statistical tests like the Sign test, Wilcoxon Signed rank tests and Friedman tests are also conducted to assess the dominance of the proposed model over other state-of-the-art approaches. Two-stage GAN-based models generate realistic, high-quality images by effectively suppressing inherently present spurious noise in medical images and simultaneously preserving the edge information.

Keywords: Deep learning, Edge preservation, Image enhancement, Non-parametric statistical testing, Patch selection

Introduction

In the era where everyone is connected to everything through the convergence of social, mobility, analytics, and cloud through the Internet of Everything (IoE), the Internet of Medical Things (IoMT) is gaining popularity.[1] IoMT enables the treatment to be readily available and accessible to people by establishing a network of connected devices over a network. One of the applications of IoMT is Medical Imaging, a non-evasive approach to medical diagnosis and prognosis. Accurate diagnosis and efficient prognosis are crucial steps in this domain. The transmission of medical images in these scenarios could result in noises that hamper the quality of the images. Furthermore, during the acquisition and other processing of medical images, the inclusion of noise is the basic phenomenon that may occur due to different factors, such as non-linearity in sensors, electrical disturbances, motion errors, etc. The most probable noises that may be introduced at different stages of medical image processing are Speckle, Impulse, Gaussian, multiplicative, Poisson, Gamma, and uniform noise etc., which are superimposed on a medical image. These noises are inherently present in medical images resulting in the loss of important information at critical places, influencing the decision-making on any medical diagnosis.

The authors conducted an extensive review of various denoising algorithms for digital images.[2] It is highlighted that the classical filters like mean, median, Gaussian, non-local mean, etc., lack adaptability to suppress any noise from medical images. Moreover, these fixed filters are suitable for a specific type of noise present in a specific modality, but their performance degrades for mixed noise suppression from medical images. Kumar & Mishra have extensively surveyed medical image denoising.[3] Several researchers have worked on the removal of different types of noises involved in medical images, like salt and pepper noise removal from bio-X-ray images.[4] Barcelos *et al.*[5] proposed a variational approach to reconstruct images corrupted by non-uniformly distributed mixed noise. Sagar *et al.*[6] proposed a circular adaptive median filter for suppressing salt and pepper noise from MRI images. Machine learning methods are integrated into the basic denoising process to provide adaptability and improve the performance of these filters to eliminate different spurious noises from digital images.[3] There has been a significant achievement in developing image filters for suppressing various noises using one

*Author for Correspondence
E-mail: vishalhshah@bitmesra.ac.in

of the variants of Artificial Neural Networks (ANN), i.e., Convolutional Neural Networks (CNNs). A complex-valued CNN has been implemented for denoising medical images.[7] Though image denoising has been implemented in several works, very few works have been considered to suppress mixed noise.[8,9] However, simultaneous noise elimination and edge preservation are not addressed in most of the literature.

The proposed two-stage GAN-based models significantly advance image generation by dividing complex tasks into manageable stages, enhancing quality, stability, and flexibility. Integrating GANs with CNNs boosts realism by leveraging CNNs' feature extraction, upsampling, and downsampling capabilities. The proposed work's innovations promise improved generation quality, training stability, and diversity, vital for medical image processing.

Related Work

Many researchers have proposed various adaptive models for denoising of medical images by implementing Machine Learning (ML) and Deep Learning (DL) based models. Jin *et al.*[10] integrated Deep Learning (DL) with boosting algorithms where conventional handcrafted boosting units are incorporated with CNN. The boosting with CNN improves the performance in terms of computation speed. However, the main drawback of this approach is its ability to suppress only the Gaussian noise, and its performance degrades for other noises. Lin *et al.*[11] proposed an attentive-based GAN model with two sub-networks for denoising real photographs. The first subnetwork finds the noisy regions and provides an attention map as an output. The second subnetwork takes the attention map and noisy map as input to provide a noise-free image. This approach is also designed to remove Gaussian noise from noisy images. A novel approach known as the grouped residual dense network, which aims to effectively denoise the actual image, has been proposed.[12] The proposed approach utilizes a residual dense network with some minor adjustments. Better denoising performance is achieved by cascading grouped residual dense blocks. However, the limitation of the work forces for real image acquisition from cameras whose calibration parameters are known in advance.

Deep Feed Forward CNN (DFF-CNN) is proposed to evaluate a model-based filter for denoising images affected by a mixture of Gaussian and impulse noise.[13] Here, the CNN model directly estimates noise-free images from noisy images and performs well in preserving edge details. Both the impulse and Gaussian noise can be denoised by the structured dictionary learning models.[14] In this work impulse noise is removed by a median-based filter and LP norm fidelity. L_2 regularization is utilized to remove Gaussian noise. DL is integrated with traditional model-based variational methods to remove the mixture of the Gaussian and impulse noise.[15] Mixed noise removal is considered to be a six-step process: smoothness, synthesis, noise put back, noise parameter estimation, noise classification and convergence checking. CNN is used for the first stage of smoothness to remove noise. CNN is used to remove mixed noise having Gaussian and impulse noise.[16] The study introduced a CNN model consisting of two parts: one component is designated for the elimination of impulse noise, and the other for the removal of Additive White Gaussian Noise (AWGN). However, though the method removes a mixture of these two noises, it affects the structure of the image. The images are pre-processed with rank order filtering and upscaled before passing to convolutional layers. This method requires a large dataset for improved performance. Here, noise is estimated in each layer of CNN, and the final estimated noise is added to the input noisy image to obtain the final noise-free images. This denoiser is trained for each of the noises separately before being used for denoising purposes. A combination of CNN and Particle Swarm Optimization (PSO) effectively eliminates high-density impulsive noise commonly found in images.[17] Here, PSO finds the optimum threshold for detecting pixels corrupted by impulse noise. This method is found to work well only for high-density impulse noises. A blind CNN model is designed for denoising random valued impulse noise.[18] Here, multiple CNN models are trained for different noise ratios parallelly. The noise ratio is estimated patch-wise by using the features of pixel deviation, edge pixel difference and multi-rank ordered pixel difference. Based on patch-wise results, the overall noise ratio for the image is determined. Finally, the trained CNN network for that noise ratio is implemented to denoise the image. Couturier *et al.*[19] proposed CNN with parametric rectified linear units for denoising the image. In their research, the output image from CNN is processed by the adaptive

bilateral filter to provide image smoothening and sharpness. The model is designed as a blind Gaussian denoising filter**.** An attention-guided denoising CNN for noise suppression from camera images has been proposed.[20] The salient feature of this model is the integration of feature enhancement modules to reduce the training complexity and improve efficiency. Valsesia *et al.*[21] proposed a CNN structure that utilises a combination of local and non-local variational methods. In their work, the neural network has been trained to handle various noise levels by utilising a unified set of parameters. However, this approach has been found to be effective in suppressing AWGN from digital images. The problem of local self-similarity-based convolutions is solved in typical CNN denoising with graph convolutional layers. Graph convolutions exploit local and non-local self-similarities and can construct neighbourhoods that detect correlations in the feature maps. The method is investigated for camera images corrupted with Gaussian noise.

Cruz *et al.*[22] have proposed an integrated denoising method by combining a local multiscale denoising-based CNN model and a non-local denoising-based non-local filter. Compared to other complicated CNN models, this proposed model is modular and has less computational complexity, making it more suitable for Gaussian noise suppression. Zhang *et al.*[23] introduced a versatile CNN model that exhibits the capability to effectively address a broad spectrum of noise levels. CNN receives the tunable noise level map as input and generates the noise filter map. This filter map is then subtracted from the noisy image, resulting in the filtered digital image that is corrupted by AWGN. A deep iterative down-up CNN is employed to efficiently remove noise from images.[24] The model being suggested is an amended version of the U-net model, which is usually used for segmentation tasks. In their research work, the proposed model demonstrates the capability of dealing with a variety of Gaussian noise levels inside a single network architecture. A deep CNN is proposed by Chen *et al.*[25] that is adjustable to different noise levels in the input image, where soft shrinkage is used as an activation function. The use of soft shrinkage makes the network more adaptable for eliminating various levels of noise present in the digital image. Thakur *et al.*[26] used CNN with parametric rectified linear units for denoising the image. The output image from CNN is further processed by adaptive bilateral filter to provide image smoothening and sharpness. The solution is designed for blind Gaussian denoising. However, the model provides satisfactory results only for a wide range of Gaussian noise levels.

Zhong *et al.*[27] proposed a Generative Adversarial Network (GAN) to perform denoising of images. In their approach, the generator component of the GAN is implemented using Densenet, while the loss function employed is Wasserstein-GAN. The Wasserstein distance is employed to regulate the fidelity of generated images. Again, the scope of the methodology is limited to captured images by real cameras.

Further, GAN for noise modelling is carried out, where the noise distribution in the input mask is estimated to generate noise samples. The massive number of noise samples generated by GAN are used for subsequent training of the CNN model by which the problems of limited training set for blind denoising have been solved. Liu *et al.*[28] proposed a cycle-consistent adversarial network (Cycle GAN) to denoise natural and medical images. However, this method lacks ground truth alignment, mode collapse with limited diversity, unusual artefact generation and computational cost. Chen *et al.*[29] proposed a Dual-domain Attention-enhanced Encoder-decoder GAN (DAEGAN) to assist the medical practitioner in obtaining PET images with low-dose radiotracers injections. Tripathi *et al.*[30] used the modified conditional generative adversarial network to remove motion artifacts from MRI scans.

Though there are many works as specified in the literature, we have found many limitations that have given rise to the following research gaps.

(i) Deep learning-based denoising is superior to other neural network-based adaptive filters, but they lack adaptivity to a wide range of noise levels. Moreover, their performance is low when noise levels in noisy images differ from trained images.

(ii) Most CNN-based denoising models are developed to remove Gaussian and impulse noises separately. Therefore, mixed noise removal from medical images remains an open area for researchers. Very limited work has been done using GAN-based CNN models for denoising medical images.

(iii) Most of the CNN denoising models used L_1 or L_2 regularization-based loss function without consideration for visual quality metrics. Due to this, structure preservation in images is compromised.

The research gaps have enabled us to propose the current work, the contributions of which are highlighted below:

(i) A two-stage GAN-CNN model is suggested to effectively eliminate the presence of mixed noise, specifically Gaussian, Impulse, and Speckle noise, while simultaneously preserving the integrity of image edges.

(ii) Patch selection criteria are considered to improve the noise modelling of GAN. A novel loss function is considered for the effective preservation of the image structure information.

(iii) The proposed approach is implemented on a set of medical images collected from a benchmark data set available athttps://openi.nlm.nih.gov.

(iv) The performance of the proposed model is evaluated by comparing it with other established methodologies, considering three performance measures: Peak Signal Noise Ratio (PSNR), Feature Similarity Indexing Method (FSIM), and Structural Similarity Indexing Method (SSIM). Non-parametric statistical testing is also conducted to examine the dominance of the proposed GAN-CNN model over other state-of-the-art methodologies.

Dataset & Preprocessing

The proposed approach is implemented on medical images collected from a benchmark data set at https://openi.nlm.nih.gov. accessed on 11 January 2022. This repository provides a wide range of publicly available medical images, including radiographs, CT scans, and other imaging modalities, accompanied by metadata and diagnostic reports. For this work, we focused on MRI images due to their prevalence and clinical importance.

The dataset comprises around 10000 images with varying resolutions, contrast, and noise levels. The data was selected to include diverse cases representing common challenges encountered in medical imaging, such as Gaussian noise, Salt and Pepper noise, and Speckle noise.

Preprocessing

To ensure the data was suitable for training the GAN-CNN-based noise removal model, the following preprocessing steps were performed:

Noise Augmentation

Synthetic noise patterns were added to clean images to simulate mixed noise conditions commonly found in medical imaging. Specifically, Gaussian noise (with a mean of 0 and varying standard deviations) and salt-and-pepper noise (with noise densities ranging from 0.1 to 0.5) were superimposed on the images. These noisy images were paired with their clean counterparts to form the training set.

Normalization

All pixel intensities were normalized to a range of [0, 1] to standardize the input for the GAN-CNN model and ensure consistent performance across different image modalities.

Image Resizing:

Images were resized to a fixed dimension of 256 × 256 pixels to maintain computational efficiency and ensure compatibility with the model architecture.

Data Splitting

The dataset was split into training, validation, and test sets using an 80:10:10 ratio. Care was taken to maintain a balanced distribution of noise levels and image types across the subsets.

Data Augmentation

Additional transformations, such as rotation, flipping, and contrast adjustment, were applied to augment the dataset and improve model robustness.

Metadata Exclusion

No textual metadata or diagnostic information was included in the model training to focus solely on image enhancement. The resulting dataset comprised 8000 training images, 1000 validation images, and 1000 test images.

By ensuring a rigorous preprocessing pipeline, we aimed to create a robust benchmark for evaluating the performance of the proposed GAN-CNN model in mitigating mixed noise while preserving the structural details critical for medical diagnosis.

GAN Model

GAN has been implemented by different researchers to estimate generative models having less complexity which are difficult to approximate using other conventional CNN models. GAN has two network components, i.e., generative and discriminative. Discriminative networks are trained to predict whether the sample is real or generated. The generative network is trained to provide the given sample close to the real one to deceive the discriminative network. The two networks compete to outsmart each other during the training stage, and as a

result, a synthetic sample is generated that is as close to the real sample. The configuration of generative and discriminative components of GAN used in this work is presented in Fig. 1 and Fig. 2.

Table 1 — Filter configuration for GAN

	Tensor Flow Layers	No. of Nodes
Generative network	B.1	256
	B.2	128
	B.3	64
Discriminative network	A.1	64
	B.1	128
	B.2	256
	B.3	512

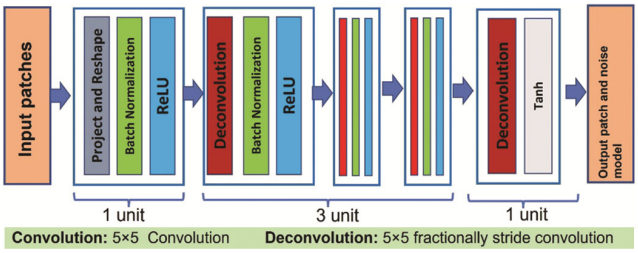

Fig. 1 — Configuration of generative network

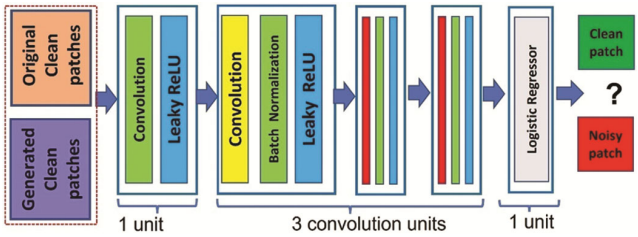

Fig. 2 — Configuration of discriminative network

As depicted in Fig. 1, the generative network generates new data that is similar to the data used in the training process. The generative network uses an input patch to produce an output patch. The discriminative network, as shown in Fig. 2 is responsible for distinguishing between real data and generated data produced by the generator. The filter configuration for GAN with tensor flow layers with the number of nodes used in generative and discriminative networks in our application is presented in Table 1. For the generator, the Dense layer: 8 × 8 × 256 = 16384, 8 × 8 × 256 = 16384 nodes and Filters in transposed convolution layers: 256, 128, 64, 32, 3 (decreasing as the spatial dimensions increase) were utilized for the processing. For Discriminator, Convolution layers with increasing filters: 64, 128, and 256 (increasing as the spatial dimensions decrease) and Final Dense layer: 1 node (for binary classification) were utilized for the processing.

Proposed Approach

The proposed denoising framework using the two-stage GAN model has been delineated in Fig. 3. To train the model, a training data set has been developed by artificially adding different noises with different intensities to the images. As depicted in Table 2, three types of noise, viz, Gaussian noise, Impulse noise and Speckle noise with varying intensities, have been considered to corrupt the original images for testing and validation. The corrupted images with various noise levels are represented as M1, M2…. M8 in this table. The pixels of the grey-scale medical image I_k,

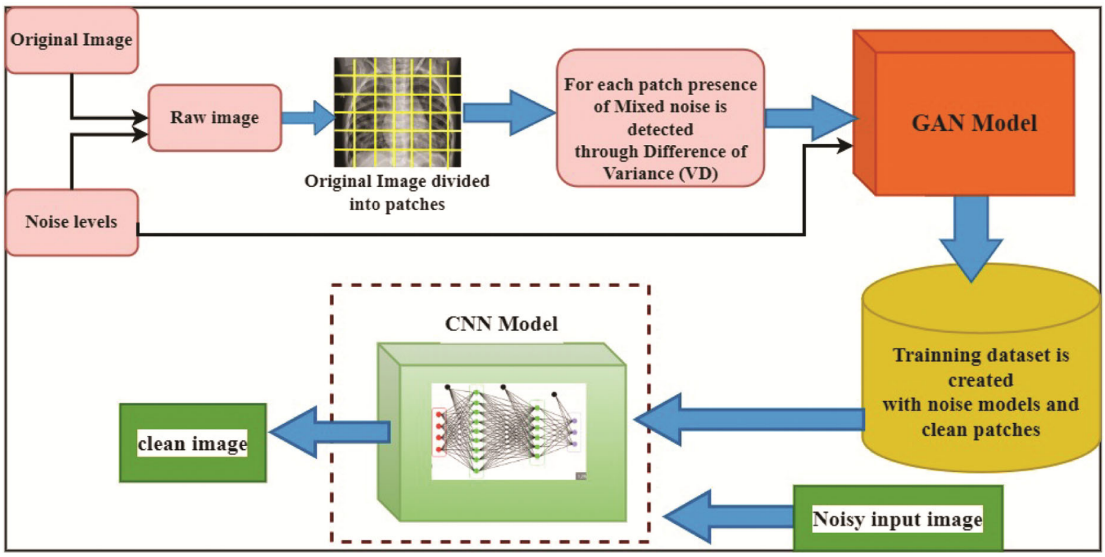

Fig. 3 — Proposed two-stage GAN model

Table 2 — Mixed noise combination

ID	Gaussian noise		Impulse noise		Speckle noise	
	5	10	10%	50%	10%	50%
M1	✓	—	✓	—	✓	—
M2	✓	—	✓	—	—	✓
M3	✓	—	—	✓	✓	`—
M4	✓	—	—	✓	—	✓
M5	—	✓	✓	—	✓	—
M6	—	✓	✓	—	—	✓
M7	—	✓	—	✓	—	✓
M8	—	✓	—	✓	✓	—

where k ranges from 1 to n, are influenced by mixed noise and can be represented as

$$(I_k{}^N)_{i,j} = \begin{bmatrix} R_{i,j} & \forall Pf(p_1) \\ r_{min} & \forall Pf(p_2) \\ r_{max} & \forall Pf(p_3) \\ (I)_{k_{i,j}} + a_{i,j} & \forall Pf(p_4) \end{bmatrix} \quad \ldots (1)$$

where, $pf(p_i)$ is the probability function for the input probability of p_i, $R_{i,j} \in [rmax_{minj}$ is the random value impulse noise with densityp_1. When the pixel value is $r_{\min}$ or $r_{\max}$, noise is speckle. $(I_k)_{i,j}$ is the original pixel value, and $a_{i,j}$ is the additive white Gaussian noise with zero mean and variance is σ^2. The distribution is done in such a way that $p_1 + p_2 + p_3 + p_4 = 1$

Thus, the proposed approach considers the pixel corrupted with a mixture of AWGN, random valued impulse noise (RVIN) and speckle noise.

The proposed approach has two major functionalities, i.e., training and testing/validation. In the training stage, a paired set of noisy image patches and corresponding clean image patches are selected using GAN. This dataset is used to train the CNN model to predict separate patches from noisy patches. In the testing stage, the input noisy image is split into patches and passed to the trained CNN model to predict the clean patches. The clean patches are then joined together to obtain the denoised image. Here, the loss function of CNN is modified as a structure-preserving metric so that the resulting noise-free image has reduced mixed noises, and at the same time, edge structures are preserved.

Training

Given a set of 'n′ number of clean training images $\{I_1, I_2, ..., I_n\}$, mixed noise images are generated based on the mixed noise model as mentioned in Eq. (1). The input and the corresponding mixed noise images are split into patches of size $m \times m$. Each of the noisy patches is evaluated for its suitability to be used for the next stage. The generation for mixed noise patch is done in terms of the difference of variance given as

$$\mathrm{VD} = |\mathrm{AV_{i,j}} - \mathrm{NV_{i,j}}| \quad \ldots (2)$$

where, $AV_{i,j}$ is the grey level variance of all pixels in the patch given as

$$\mathrm{AV_{i,j}} = \frac{\sum_{k=1}^{m^2}(g_{s_k r_k} - \mathrm{M_1})^2}{m^2} \quad \ldots (3)$$

$g_{s_k r_k}$ is the grey level value of the pixels of the patches. $\mathrm{M_1}$ is the mean grey level of all pixels in the local window. $\mathrm{NV_{i,j}}$ is the grey level variance of the center pixel calculated as

$$\mathrm{NV_{i,j}} = \frac{\sum_{k=1}^{m^2}(g_{s_k r_k} - \mathrm{M_2})^2}{m^2 - 1} \quad \ldots (4)$$

Here, $\mathrm{M_2}$ is the mean grey level of the neighbourhood pixel of the center pixel. The noise patches whose VD is greater than a threshold T are qualified. These qualified noise patches are passed to GAN for noise modelling. The set of qualified noise patches is given as

$$X = \{NP_1, NP_2, NP_3, \ldots\ldots NP_m\} \quad \ldots (5)$$

The CNN model needs a large volume of images for improved performance. In this work, we applied GAN to provide a greater number of noisy data to boost the performance of CNN.

The salient feature of GAN is that it can learn any complex distributions effectively. The back propagation algorithm is used to train the GAN to produce samples close to real input. In the proposed method, a GAN is adopted to estimate the noise distribution over a set of qualified noisy blocks X. Gradient penalty-based Wasserstein GAN (WGAN-GP) is used in this work to learn the noise distribution.[11] The gradient norm penalty encourages the discriminator model to be piecewise linear around the data. Here, the GAN's objective function is given as:

$$\mathrm{L_{GAN}} = \mathrm{E_{\bar{x}\sim P_g}}[\mathrm{D}(\bar{x})] - \mathrm{E_{\bar{x}\sim P_r}}[\mathrm{D}(x)] + \lambda \mathrm{E_{\bar{x}\sim P_x}}[(||\nabla \mathrm{D}\ (\alpha \mathrm{X} + (1-\alpha)\hat{\mathrm{X}})||_2 - 1)^2] \quad \ldots (6)$$

Here, the distribution of the input X is given as P_D, and the generative network distribution is given as P_g.

By providing X as input to the GAN, the result is X'. This X' is a large set of noise samples augmented using GAN. The training images $\{I_1, I_2, ..., I_n\}$ are split into blocks of size $m \times m$ and the set of patches is given as $X = \{x_1, x_2, ..., x_e\}$. The noise blocks in X' are added to the patches to obtain $Y = \{y_1, y_2, ..., y_e\}$ where,

$$y_i = x_i + X_k' \quad \ldots (7)$$

A training dataset is created with noisy image Y as the input, and the difference between noisy image and original image R is called a noise mask is the output of the GAN model and is considered to be the input to the CNN model. The CNN model utilizes these noise mask data as a training dataset and is trained by minimizing loss function. The loss function to be minimized in the CNN is defined as

$$L_{CNN}(\theta) = \frac{1}{2N}\sum_{i=1}^{N}|R(y_i;\theta) - (y_i - x_i)|^2 + SF(y_i, x_i) \quad \ldots (8)$$

where, θ is the CNN parameters, N is the number of training patches, y_i is the noisy patch, x_i is the corresponding clean image. SF is calculated in terms of the difference of Eigen transformation values obtained from the structure tensor of y_i and x_i.[31]

The loss function to be minimized in the CNN is defined as

$$L_{CNN}(\theta) = \frac{1}{2N}\sum_{i=1}^{N}|R(y_i;\theta) - (y_i - x_i)|^2 + SF(y_i, x_i) \quad \ldots (9)$$

where, θ is the CNN parameters, N is the number of training patches, y_i is the noisy patch, x_i is the corresponding clean image. SF is calculated in terms of the difference of Eigen transformation values obtained from the structure tensor of y_i and x_i

$$SF(y_i, x_i) = (L_{y_i} - L_{x_i})^2 \quad \ldots (10)$$

where, L_{y_i} is the Eigen transformation on y_i, and it is formulated as

$$L_{y_i} = \frac{1}{2}(J_{11} + J_{22}) + \sqrt{(J_{11} - J_{22})^2 + 4{J_{12}}^2} \quad \ldots (11)$$

where, J_{11}, J_{12}, J_{21}, and J_{22} are the tensor structure values in the direction of m and n.

$$S = \begin{bmatrix} {i_m}^2 & i_m i_n \\ i_m i_n & {i_n}^2 \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \quad \ldots (12)$$

This proposed loss function, in combination with both the pixel differences and tensor differences, ensures that the structure of the image is persevered while generating the noisy mask by the CNN model.

Testing

Once CNN is trained, it is used to denoise noisy medical images. Given an input image I_K, the image is split into non-overlapping patches of size $m \times m$. Each patch is passed through the trained CNN, and the noisy mask is obtained. The noise mask is subtracted from the original patch to get the clear patch. This process is repeated for all the patches to get the clean patches. All the clean patches are then properly joined to get the clean image.

Results & Discussion

The effectiveness of the proposed study is investigated by considering MRI images collected from the standard repository using the link https://openi.nlm.nih.gov. Some of these medical images are corrupted with noise and most of them are free from noise. For obtaining more noisy images, the noise-free medical images are artificially contaminated with three different noises, i.e., Gaussian, impulse and speckle, with varying distribution to obtain the mixed noise. Gaussian noise levels like 5 and 10 represent the standard deviation of noise in a dataset. The higher the standard deviation, the more intense the noise. Using different noise levels (such as 5 and 10) helps simulate varying real-world conditions, allowing researchers to evaluate and improve the robustness of models, particularly in image processing, signal processing, and machine learning tasks.

By considering these three different noises, eight combinations of noise, i.e., M1, M2, M3, M4, M5, M6, M7 and M8, as shown in Table 2, are generated. The collected noise-free images are intentionally corrupted with each of these combinations of mixed noise. These noisy images are utilized to generate patterns and to train the proposed GAN-CNN model for investigation of the performance in the testing phase.

The performance of the proposed model is compared in terms of three performance indices, such as peak signal-to-noise ratio (PSNR), Feature

Similarity Index Measure (FSIM) and structural similarity index measure (SSIM).

The definition and formula for these three indices are as follows:

- PSNR (Peak Signal to Noise Ratio): PSNR is the ratio of maximum possible signal power to the power of distorting noise corrupting the image, thereby affecting the quality of the image. It is measured in decibel form to cover a wide dynamic range.

$$PSNR = \frac{10 \log_{10}(peakval)^2}{MSE} \quad \ldots (13)$$

where, $peakval$ is the maximum signal value of the original noise-free image. MSE is the mean square error.

- FSIM (Features Similarity Index Matrix): The Feature Similarity Index Method is a quality index method that maps the features in an image and measures the similarity that exists in the image based on Phase Congruency (PC) and Gradient Magnitude (GM). PC is invariant to light variation in images and is also invariant to contrast. Horizontal and vertical gradients G_x and G_y for an image give the gradient magnitude as

$$\sqrt{G_x^2 + G_y^2}\ .$$

For test image IT and reference image IR, the phase congruencies are PCT and PCR, respectively, and the magnitude gradient maps corresponding to these test and reference images are G_T and G_R, respectively. The similarity of these two images will be given by

$$S_{PC} = \frac{2PC_T PC_R + T_1}{PC_T^2 + PC_R^2 + T_1} \quad \ldots (14)$$

Here, T_1 is a positive constant value for increasing S_{PC} stability, calculated based on PC values. Similarity-based on G_1 and G_2 can be calculated as

$$S_G = \frac{2G_1G_2 + T_2}{G_1^2 + G_2^2 + T_2} \quad \ldots (15)$$

T_2 is a positive constant that depends on the dynamic range of gradient magnitude S_L. The similarity is then calculated as

$$S_L(X) = [S_{PC}(X)]^{\alpha} \times [S_G(X)]^{\beta} \quad \ldots (16)$$

α and β are the weighted parameters to enhance the importance of S_{PC} or S_G.

- SSIM (Structural Similarity Index Measure): SSIM is a method in which the perceived quality of the image is predicted for its quality. The image quality is predicted based on a distortion-free reference image. SSIM is evaluated by taking various windows on an image. SSIM measures two windows 'a' and 'b' with common size $N \times N$, and is mathematically expressed as,

$$SSIM(a,b) = \frac{(2m_a m_b + c_1)(2\sigma_{ab} + c_2)}{(m_a^2 + m_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)} \quad \ldots (17)$$

where, m_a is the mean of window a, m_b is the mean of window b, σ_a^2 is the variance of window a, σ_b^2 is the variance of window b, σ_{ab} is the covariance of the window a and b; c_1 and c_2 are the variables for stabilizing the division with a weak denominator.

The average values of these three performance indices are considered for the performance evaluation of the proposed approach against the four models: (1) the Deep CNN model for denoising[13], (2) the Variational CNN model for denoising[15], (3) CNN with pReLU activation[26] and (4) Attention guided CNN.[20]

The PSNR measures the effectiveness of denoising, and the higher the value of PSNR, the better the denoising performance. The results for the average value of PSNR obtained after denoising the noisy medical image of MRI by considering all the eight combinations of the mixed noise are given in Table 3. The statistical values of PSNR after implementing all five models in eight mixed noise combinations are given in Table 4.

The average value of PSNR is higher for the proposed two-stage GAN-CNN network, as depicted in Table 3. The statistical value shown in Table 4 reveals that the proposed method is symmetric, and the spread of the data around PSNR is highest compared to previous methodologies. It is also visible from the observations that the previous methodologies are skew-symmetric, and the spread is also lower than the proposed method. The average PSNR value of filtered image obtained by the proposed GAN-CNN approach is 16% higher compared to Deep CNN, 19.7% higher compared to variational CNN, 17.4% higher compared to pReLU and 6.92% higher compared to Attention CNN. The variance in PSNR value of the filtered image for all the eight combinations of noise and by implementing the proposed approach is 1.2 which is very low compared

Table 3 — PSNR, FSIM & SSIM Results for mixed noise combinations

Combinations	Deep CNN[28]			Variational CNN[21]			pReLU[30]			Attention CNN[39]			Proposed GAN-CNN		
	PSNR	FSIM	SSIM	PSNR	FSIM	SSIM	PSNR	FSIM	SSIM	PSNR	FSIM	SSIM	PSNR	FSIM	SSIM
M1	31.59	0.901	0.912	32.56	0.914	0.924	32.21	0.928	0.922	35.12	0.932	0.922	37.24	0.9670	0.9570
M2	32.29	0.910	0.914	31.65	0.928	0.931	31.45	0.932	0.918	36.11	0.944	0.931	38.12	0.9677	0.9617
M3	32.56	0.915	0.923	30.76	0.927	0.923	32.56	0.931	0.932	36.21	0.932	0.933	38.56	0.9710	0.9620
M4	33.14	0.921	0.925	30.12	0.926	0.914	33.11	0.939	0.923	36.67	0.942	0.934	39.12	0.9612	0.9562
M5	33.21	0.927	0.926	30.44	0.933	0.924	32.86	0.940	0.923	37.15	0.941	0.931	40.10	0.9810	0.9720
M6	33.56	0.929	0.928	31.91	0.930	0.926	33.21	0.931	0.921	37.33	0.933	0.929	39.36	0.9700	0.9731
M7	34.12	0.931	0.923	32.12	0.921	0.931	31.78	0.924	0.911	37.12	0.925	0.938	41.22	0.9656	0.9816
M8	33.26	0.927	0.911	32.32	0.911	0.922	32.21	0.927	0.931	36.59	0.931	0.929	40.32	0.9712	0.9822
Noise average	32.96	0.817	0.818	31.48	0.821	0.821	32.42	0.828	0.820	36.53	0.831	0.827	39.25	0.861	0.860

Table 4 — Data Summary for PSNR, FSIM & SSIM with N = 8

N = 8	Groups	Deep CNN[13]	Variational CNN[15]	pReLU[26]	Attention CNN[20]	GAN-CNN
PSNR data summary	Min	31.59	30.12	31.45	0	37.24
	Q1	32.4925	30.68	32.1025	36.11	38.45
	Median	33.175	31.78	32.385	36.59	39.24
	Q3	33.335	32.17	32.9225	37.12	40.155
	Max	34.12	32.56	33.21	37.33	41.22
	Mean	32.9662	31.485	32.4238	32.4778	39.255
	SD	0.7907	0.9216	0.627	12.1979	1.2856
FSIM data summary	Min	0	0	0	0	0
	Q1	0.91	0.914	0.927	0.931	0.9656
	Median	0.921	0.926	0.931	0.932	0.9677
	Q3	0.927	0.928	0.932	0.941	0.971
	Max	0.931	0.933	0.94	0.944	0.981
	Mean	0.8179	0.8211	0.8280	0.8311	0.8611
	SD	0.3069	0.3080	0.3105	0.3117	0.3232
SSIM data summary	Min	0	0	0	0	0
	Q1	0.912	0.922	0.918	0.929	0.957
	Median	0.923	0.924	0.922	0.931	0.9672
	Q3	0.925	0.926	0.923	0.933	0.9731
	Max	0.928	0.931	0.932	0.938	0.9822
	Mean	0.8180	0.8217	0.8201	0.8274	0.8606
	SD	0.3068	0.3082	0.3076	0.3103	0.3229

to other approaches. From this, it can be inferred that the proposed approach performs consistently across all the mixture combinations of noises. Also, the PSNR value of the filtered image by applying the proposed model is 41.22 dB which is at least 3 dB higher than other models. Hence, the use of GAN-CNN has added noise discrimination and provided better-filtered images as evidenced by higher PSNR value compared to others.

The FSIM measures the structural similarity between the input and denoised image, where the value is from 0 to 1. The higher the value of FSIM, the better the filtration. The average values of FSIM obtained for each of the eight combinations of noises after implementing five models are given in Table 3. The data statistics of the FSIM across the mixture combinations of noise are given in Table 4.

It is observed that the average value of FSIM of the filtered image obtained from the proposed two-stage GAN-CNN model is higher as compared to others. Also, it is found that the proposed approach symmetrically spreads the data around FSIM more widely as compared to previous methodologies. It is further observed that the previous methodologies are skew-symmetric, and the spread is lower than the proposed method. The average value of FSIM of the filtered image obtained from the proposed model is 3.48% higher compared to Attention CNN, 3.83% higher compared to pReLU, 4.64% higher compared to variational CNN and 5.11% higher compared to

Deep CNN. The maximum variation in FSIM of filtered image for all eight mixtures of noise by applying the proposed GAN-CNN model is 0.32 indicating its consistent performance. Also, the maximum value of FSIM filtered image is 0.98 which is at least 0.04 higher than other models. It can be concluded that due to the use of a tensor-based metric as a loss function, instead, an L_1 or L_2 regularization-based loss function, the structure-preserving property is improved in our proposed GAN-CNN model.

Similarly, the SSIM data summary, as depicted in Table 4, also affirms that the proposed method is superior to other state-of-the-art methodologies. The average value of SSIM from our proposed GAN-CNN model is 3.83% higher compared to Attention CNN, 4.65% higher compared to pReLU, 4.53% higher compared to variational CNN and 4.81% higher compared to Deep CNN. The variation in obtained SSIM value for different mixture combinations of noise, and by applying the proposed model is 0.32 indicating its consistent performance. Moreover, the higher value of SSIM is 0.9822, which is 0.05 higher than the second-best one i.e., the Attention CNN model. Hence, the use of GAN-based training set expansion and structure tensor-based metric as loss functions in CNN ensure higher structured preserving capability. This edge-preserved structure is important for further medical image analysis like segmentation.

Three different masks of patch sizes, i.e., 21 × 21, 34 × 34 and 41 × 41, are utilized by the proposed model for the investigation of its performance. The values of all three performance measures for these three sizes of masks are presented in Table 5. It can be concluded that the PSNR value of the filtered image is higher for a patch size of 34 × 34 as compared to the other two patches while considering the M1 combination of mixed noise on the CT image. Similarly, the value of FSIM of the filtered CT image is higher for the patch size of 41 × 41. The SSIM values for 34 × 34 and 41 × 41 patches are found to be comparable.

The simulation time requirement is also investigated to compare the performance of all the models mentioned in Table 6. The time requirement for our proposed model is 11.51 seconds, which is the minimum compared to others.

It can be observed that the performance of our proposed GAN-CNN feed-forward model having only 17 layers is superior compared to other models having more layers. The original images, noisy images and the corresponding filtered images are shown in Fig. 4 for executing the subjective evaluation of the proposed approaches.

The results show that the proposed model can provide consistent results for all eight combinations of noise, yielding almost the same quality having the

Table 5 — Performance measure for different patches

Patch size	M1	M2	M3	M4	M5	M6	M7	M8	Average
					PSNR				
21 × 21	38.21	38.26	37.81	38.35	38.12	36.33	37.12	37.24	37.68
34 × 34	40.12	41.21	41.10	40.26	41.12	40.32	39.26	39.81	40.4
41 × 41	39.42	38.67	39.12	39.56	38.12	38.34	39.11	38.89	38.9
					FSIM				
21 × 21	0.951	0.962	0.957	0.954	0.951	0.949	0.952	0.952	0.953
34 × 34	0.961	0.973	0.965	0.951	0.966	0.961	0.959	0.962	0.962
41 × 41	0.972	0.984	0.981	0.979	0.980	0.982	0.985	0.986	0.982
					SSIM				
21 × 21	0.961	0.962	0.971	0.964	0.966	0.964	0.962	0.961	0.963
34 × 34	0.981	0.983	0.985	0.981	0.986	0.987	0.983	0.982	0.983
41 × 41	0.982	0.984	0.979	0.982	0.982	0.985	0.981	0.983	0.982

Table 6 — Comparison of time taken, Critical value for the two-tailed sign tests attained for α = 0.05 and α = 0.01 using FSIM as the winning parameter, & Ranking table for the Friedman test

Methods	Time taken (sec)	Critical value for the two-tailed sign tests attained for α = 0.05 and α = 0.01 using FSIM as the winning parameter			Friedman mean ranks
GAN-CNN	**11.51**	**Wins (+)**	**Loss (-)**	**Detected difference**	19.8
Attention CNN	16.72	20	0	α = 0.05	19.2
pRELU	15.12	19	1	α= 0.05	18.4
Variational CNN	18.52	18	2	α = 0.05	18.2
Deep CNN	16.41	17	3	α= 0.05	18.1

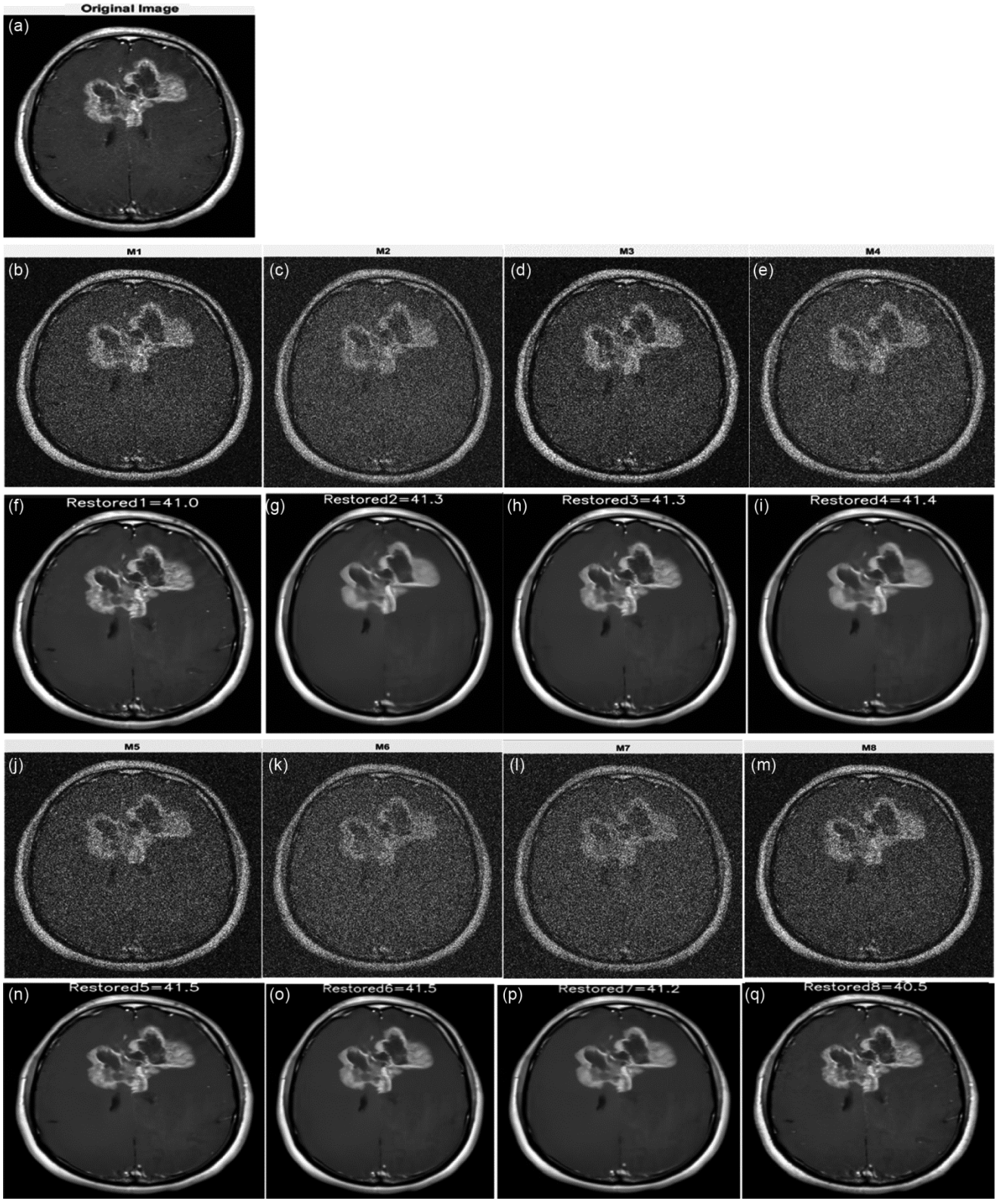

Fig. 4 — Results of the proposed method for an MRI image with different noise levels (a) Original noise-free image, (b, c, d, e, j, k, l, m): Noisy images and (f, g, h, i, n, o, p, q): resulted noise-free images

Table 7 — Minimum number of wins needed to attain significance levels of α = 0.05 and α = 0.01

No. of cases	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
α = 0.05	7	7	8	9	9	10	10	11	12	12	13	13	14	15	15	16	17	18	18
α = 0.01	6	7	7	8	9	9	10	10	11	12	12	13	13	14	14	15	16	16	17

Table 8 — Sign test & Wilcoxon signed test using FSIM score as a winning parameter by applying all five models

	Sign test		Wilcoxon signed test	
Comparison	p-Value	h-value	p-Value	h-value
GAN-CNN to Attention Guided	0.0002	1	0.0002	1
GAN-CNN to pReLU	0.0006	1	0.0025	1
GAN-CNN to Variational CNN	0.0004	1	0.0001	1
GAN-CNN to Deep CNN	0.0003	1	0.0005	1

Table 9 — Ranking table for the Friedman test

Methods	GAN-CNN	Attention Guided	pReLU	Variational CNN	Deep CNN
Mean Ranks	19.8	19.2	18.4	18.2	18.1

Table 10 — Parameter for the Friedman test

Source	Sum of Squares	Degree of Freedom	Mean Square	Chi-Square	Critical value (p)
Column	142.2	3	34.163	55.44	1.782E-11
Error	56.9	74	0.785		
Total	197	100			

PSNR bound around 41 dB. Hence, it can be concluded that our approach provides satisfactory results in terms of both subjective and objective evaluation.

All four state-of-the-art and proposed GAN-CNN models have considered various randomly initialized parameters, and hence, the simulation output also varies in each run. Two well-recognized statistical tests, the Sign test and the Wilcoxon Signed rank test, are also performed for pairwise comparison of the dominance of our proposed GAN-CNN model over others. After that, the Friedman test is also performed to investigate the performance. We have run all five models 25 times each, and one of the performance matrices, i.e., Feature Similarity Index Measure (FSIM), has been noted every time. The minimum number of wins required for achieving victories significance levels of α = 0.05 and α = 0.01 by running the algorithms 5 to 25 times is shown in Table 6.

The comparison among the models by considering the FSIM value as a victory parameter by applying all five models is also indicated in Table 6. The GAN-CNN model has considerable dominance over the others with a magnitude of α = 0.05, as shown in Table 7. By considering the same FSIM value, the p-value and h-value as the triumphant parameter are depicted in Table 8. Additionally, the Friedman test which detects dominant activity between the two models is depicted in Tables 9 and Table 10. Similarly, PSNR and SSIM can be considered and analyzed to investigate the dominance of the proposed GAN-CNN model over others.

Conclusions

In this paper, a two-stage Deep Learning (DL) learning model combining Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) has been successfully developed to suppress mixed noise in medical images. The proposed model also effectively preserves edge information in medical images, enhancing diagnostic accuracy. The probabilistic evaluation of generators and discriminators has successfully compensated for the lossy patches and ensured that the quality of the images is maintained. Experimental results demonstrate that the proposed method outperforms four other state-of-the-art techniques, such as the Deep CNN model, Variational CNN model, CNN with pReLU activation, and attention-guided CNN. The subjective and objective evaluation is successfully conducted by investigating

the obtained filtered images and three performance indices, i.e., PSNR, FSIM and SSIM, respectively. Statistical tests, such as the Sign test, Wilcoxon Signed Rank tests, and Friedman tests, confirm the dominance of the proposed GAN-CNN model over the other four models. Future performance enhancements could be achieved by incorporating additional features from raw image pixels and exploring alternative loss functions.

References

1 Dwivedi R, Mehrotra D & Chandra S, Potential of internet of medical things (IoMT) applications in building a smart healthcare system: A systematic review, *J Oral Biol Craniofac Res*, **12(2)** (2022) 302–318, https://doi.org/10.1016/j.jobcr.2021.11.010.

2 Fan L, Zhang F, Fan H & Zhang C, Brief review of image denoising techniques, *Vis Comput Ind Biomed Art*, **2** (2019) 1–12, https://doi.org/10.1186/s42492-019-0016-7.

3 Kumar M & Mishra S K, A comprehensive review on nature-inspired neural network-based adaptive filter for eliminating noise in medical images, *Curr Med Imaging*, **16(4)** (2020) 278–287, https://doi.org/10.2174/1573405614666180801113345.

4 Lu C T, Chen M Y, Shen J H, Wang L L & Hsu C C, Removal of salt-and-pepper noise for X-ray bio-images using pixel-variation gain factors, *Comput Electr Eng*, **71** (2018) 862–876, https://doi.org/10.1016/j.compeleceng.2017.08. 012.

5 Barcelos C A Z & Barcelos E Z, A well-balanced and adaptive variational model for the removal of mixed noise, *Comput Electr Eng*, **40(7)** (2014) 2027–2037, https://doi.org/10.1016/j.compeleceng.2014.06.005.

6 Sagar P, Upadhyaya A, Mishra S K, Pandey R N, Sahu S S & Panda G,A circular adaptive median filter for salt and pepper noise suppression from MRI images,*J Sci Ind Res*, **79** (2020) 1–11, https://doi.org/10.56042/jsir.v79i10.43588.

7 Rawat S, Rana K P S & Kumar V, A novel complex-valued convolutional neural network for medical image denoising, *Biomed Signal Process Control*, **69** (2021) 102859, https://doi.org/10.1016/j.bspc.2021.102859.

8 Mafi M, Izquierdo W, Cabrerizo M, Barreto A, Andrian J, Rishe N D & Adjouadi M, Survey on mixed impulse and Gaussian denoising filters, *IET Image Process*, **14**(16) (2020) 4027–4038, https://doi.org/10.1049/iet-ipr.2018.6335.

9 Shah V & Dash P, Two-stage self-adaptive cognitive neural network for mixed noise removal from medical images, *Multimed Tools Appl*, **83** (2023) 1–23, https://doi.org/10.1007/s11042-023-15423-9.

10 Jin L, Zhang W, Ma G & Song E, Learning deep CNNs for impulse noise removal in images, *J Vis Commun Image Represent*, **62** (2019) 193–205, https://doi.org/10.1016/j.jvcir.2019.05.005.

11 Lin K, Li T H, Liu S & Li G, Real photographs denoising with noise domain adaptation and attentive generative adversarial network, *Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops*, (2019), https://doi.org/10.1109/cvprw.2019.00221.

12 Kim D W, Chung J R & Jung S W, GRDN: grouped residual dense network for real image denoising and GAN-based real-world noise modeling, *Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops*, (2019), https://doi.org/10.1109/cvprw.2019.00261.

13 Mafi M, Izquierdo W, Martin H, Cabrerizo M & Adjouadi M, Deep convolutional neural network for mixed random impulse and Gaussian noise reduction in digital images, *IET Image Process*, **14(15)** (2020) 3791–3801, https://doi.org/10.1049/iet-ipr.2019.0931.

14 Zhu H & Ng M K, Structured dictionary learning for image denoising under mixed Gaussian and impulse noise, *IEEE Trans Image Process*, **29** (2020) 6680–6693, https://doi.org/10.1109/tip.2020.2992895.

15 Wang F, Huang H & Liu J, Variational-based mixed noise removal with CNN deep learning regularization, *IEEE Trans Image Process*, **29** (2019) 1246–1258, https://doi.org/10.1109/tip.2019.2940496.

16 Islam M T, Rahman S M M, Ahmad M O & Swamy M N S, Mixed Gaussian-impulse noise reduction from images using convolutional neural network, *Signal Process Image Commun*, **68** (2018) 26–41, https://doi.org/10.1016/j.image.2018.06.016.

17 Khaw H Y, Soon F C, Chuah J H & Chow C O, High-density impulse noise detection and removal using deep convolutional neural network with particle swarm optimization, *IET Image Process*, **13(2)** (2019) 365–374, https://doi.org/10.1049/iet-ipr.2018.5776.

18 Chen J, Zhang G, Xu S & Yu H, A blind CNN denoising model for random-valued impulse noise, *IEEE Access*, **7** (2019) 124647–124661, https://doi.org/10.1109/access.2019.2938799.

19 Couturier R, Perrot G & Salomon M, Image denoising using a deep encoder-decoder network with skip connections,in *Neural Information Processing: 25th Int Conf ICONIP 2018 Siem Reap Cambodia Dec 13–16 2018 Proc Part VI 25*, Springer Int Publ, (2018) 554–565, https://doi.org/10.1007/978-3-030-04224-0_48.

20 Tian C, Xu Y, Li Z, Zuo W, Fei L & Liu H, Attention-guided CNN for image denoising, *Neural Netw*, **124** (2020) 117–129, https://doi.org/10.1016/j.neunet.2019.12.024.

21 Valsesia D, Fracastoro G & Magli E, Image denoising with graph-convolutional neural networks, *Proc IEEE Int Conf Image Process (ICIP)*, (2019) 2399–2403, https://doi.org/10.1109/icip.2019.8803367.

22 Cruz C, Foi A, Katkovnik V & Egiazarian K, Nonlocality-reinforced convolutional neural networks for image denoising, *IEEE Signal Process Lett*, **25(8)** (2018) 1216–1220, https://doi.org/10.1109/lsp.2018.2850222.

23 Zhang K, Zuo W & Zhang L, FFDNet: toward a fast and flexible solution for CNN-based image denoising, *IEEE Trans Image Process*, **27(9)** (2018) 4608–4622, https://doi.org/10.1109/tip.2018.2839891.

24 Yu S, Park B & Jeong J, Deep iterative down-up CNN for image denoising, *Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops*, (2019), https://doi.org/10.1109/cvprw.2019.00262.

25 Isogawa K, Ida T, Shiodera T & Takeguchi T, Deep shrinkage convolutional neural network for adaptive noise reduction, *IEEE Signal Process Lett*, **25(2)** (2017) 224–228, https://doi.org/10.1109/lsp.2017.2782270.

26 Thakur R S, Yadav R N & Gupta L, PReLU and edge-aware filter-based image denoiser using convolutional neural

network, *IET Image Process*, **14(15)** (2020) 3869–3879, https://doi.org/10.1049/iet-ipr.2020.0717.

27 Zhong Y, Liu L, Zhao D & Li H, A generative adversarial network for image denoising, *Multimed Tools Appl*, **79** (2020) 16517–16529, https://doi.org/10.1007/s11042-019-7556-x.

28 Liu J, Li C, Liu L, Chen H, Han H & Zhang B, Speckle noise reduction for medical ultrasound images based on cycle-consistent generative adversarial network, *Biomed Signal Process Control*, **86** (2023) 105150, https://doi.org/10.1016/j.bspc.2023.105150.

29 Chen S, Tian X, Wang Y, Song Y, Zhang Y, Zhao J & Chen J, DAEGAN: generative adversarial network based on dual-domain attention-enhanced encoder-decoder for low-dose PET imaging, *Biomed Signal Process Control*, **86** (2023) 105197, https://doi.org/10.1016/j.bspc.2023.105197.

30 Tripathi V R, Tibdewal M N & Mishra R, Denoising of motion artifacted MRI scans using conditional generative adversarial network, *Multimed Tools Appl*, **83** (2024) 11923–11941, https://doi.org/10.1007/s11042-023-15705-2.

31 Åström F & Felsberg M, On the choice of tensor estimation for corner detection, optical flow and denoising, in *Computer Vision – ACCV 2014 Workshops, ACCV 2014 Lecture Notes Comput Sci*, vol 9009, Springer, Cham, (2015) 554–565, https://doi.org/10.1007/978-3-319-16631-5_2.