Journal of Theoretical and Applied Information Technology
15[st] September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SYSTEMATIC EVALUATION OF ROBUST LOSS FUNCTIONS FOR RESIDUAL AUTOENCODER-BASED SALT-AND-PEPPER NOISE REMOVAL IN BRAIN MRI

MANIKYA PRASUNA PAKALAPATI [1],HARISCHANDRA SIVA PRASAD NERSU[2], PRATHIMA TIRUMALAREDDY[3], SIRISHA ALAMANDA[4], CH VENKATESH[5], KAILASAM SWATHI[6].

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, vaddeswaram, Guntur, 522502, Andhra Pradesh, India.
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, vaddeswaram, Guntur, 522502, Andhra Pradesh, India.
[3]Department of Information Technology, Chaitanya Bharathi Institute of Technology, Gandipet, Kokapet, Hyderabad, Telangana, 500075, India.
[4]Department of Information Technology, Chaitanya Bharathi Institute of Technology, Gandipet, Kokapet, Hyderabad, Telangana, 500075, India.
[5]Department of Computer Science and Engineering, Ramachandra college of Engineering, Eluru, 534007, Andhra Pradesh, India.
[6]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur – 522502, Andhra Pradesh, India.

E-mail: [1]prasunamanikya@kluniversity.in, [2]2401050060@kluniversity.in, [3]prathimareddy.t@gmail.com, [4]asirishanaidu@gmail.com, [5]chvenkatesh134@gmail.com, [6]dr.kswathi@kluniversity.in

ABSTRACT

Salt-and-pepper noise (SPN) is a negative phenomenon in brain MRI, which causes deterioration of diagnostic quality by blurring the boundaries of tumors and disrupting downstream clinical processing. Although classical filters and deep learning autoencoders have been used to remove impulse noise, no systematic evaluation of robust loss functions to medical SPN denoising exists. It is the first systematic comparison of the Mean Squared Error (MSE), Huber, and Tukey bi-weight losses in a residual autoencoder architecture with symmetric encoder-decoder structure and multi-scale skip connections. We compared the performance on the BRISC 2025 and Alzheimer MRI datasets at different noise densities (10% -50%) and noise corruption types (SPN, Gaussian, and Rician noise and mixed noise) and assessed fidelity through PSNR and SSIM, perceptual quality through LPIPS and clinical applicability through Boundary F1 and downstream classification accuracy. As opposed to the theoretical expectations (of redescending property of Tukey), Huber loss (delta=1.0) at 30 per cent SPN (PSNR = 36.65 +1.86 dB, Boundary F1 = 0.868 +0.017) outperformed Tukey-based variants (DeltaPSNR =5.44dB, p=0.001). The results of our study indicate that gradient continuity is more advantageous than aggressive outlier suppression in suppressing high-density impulse noise in medical images. The ResAE-Huber model lies on the Pareto-optimal frontier of accuracy-efficiency trade-offs, providing tumor edge continuity at a cost of 4.83 ms inference latency and just 1.86M parameters. These findings form Huber loss as the objective of choice in residual autoencoder-based SPN reduction, and practical guidelines on clinical implementation with edge preservation and inference latency being paramount.

Keywords: *Medical Image Denoising, Residual Autoencoder, Robust Loss Functions, Salt-And-Pepper Noise, Tumor Boundary Preservation, Huber Loss, Tukey's Bi-Weight.*

1. INTRODUCTION

Medical imaging is a fundamental branch of modern clinical practice, as it allows visualizing anatomy and pathology without any invasive procedures and allowing this visualization to assist in diagnosis, planning of treatment and monitoring of the disease [1] Magnetic resonance imaging (MRI) is a type of imaging that has the best soft-tissue contrast, with no ionizing radiation, which makes it especially useful in neuroimaging, such as brain tumor detection and characterization [2].

Nevertheless, MRI acquisitions are vulnerable to a number of sources of degradation, such as The main terms in use in this manuscript are defined as follows: Salt-and-pepper noise (SPN) is sparse, high- intensity impulse corruptions that appear as randomly occurring black and white pixels, in contrast to Gaussian noise that is distributed across all intensity levels [27]. Residual autoencoders are designed to learn to encode noise instead of clean images directly, enabling faster learning and better generalization [14, 15]. Objective functions that are specifically designed to minimize the impact of outliers in the optimization process are known as robust loss functions, and well-known examples are Huber and Tukey bi-weight [32].

Salt-and-pepper noise (SPN), which consists of a sparse impulse corruptions manifested in black and white pixels, is one of the clinically most disruptive types of noise in medical imaging [4]. In contrast to Gaussian noise, which spreads errors uniformly throughout the intensity range, the SPN introduces extreme outliers, which may
mask important diagnostic information, like tumor margins, microcalcifications, and vascular features [5]. Corruption of SPN in high-risk clinical settings may cause delayed diagnosis or poor interpretation of the morphology of lesions, which directly affect the outcome of patients [6, 29]. The recent research has shown that noise in brain MRI can decrease the signal-to-noise ratio of T2-weighted images by up to 40 percent and this will have a substantial impact on the confidence of radiologists with regard to determining the boundaries of the tumor [29].

Conventional methods of denoising, such as median filtering, bilateral filtering and block-matching with 3D filtering (BM3D) have been widely explored in the context of removing impulse noise [7, 26]. Median filters are good at removing single outliers and reconstructing local median values but blur fine structure at moderate levels of noise [8]. The transform domain collaborative filtering used by BM3D can deliver state-of-the-art performance with Gaussian noise but is not robust to impulse corruptions [9, 26]. Such hand-designed methods are not flexible to changing noise properties and do not have the ability to acquire imperfect hardware, patient movement, and transmission errors, which all present noise at the expense of diagnostic accuracy [3, 28, 30] domain-specific knowledge based on medical image distributions [10]. The limitations of traditional ways of dealing with complex noise patterns in clinical environments have been pointed out in recent extensive reviews [3, 4, 38].

Deep learning has become an effective substitute of image restoration, convolutional neural networks (CNNs) have shown significant gains over traditional algorithms [11, 38]. Autoencoders, with encoder-decoder designs and bottleneck encodings, have been effectively trained in medical image denoising with few training samples [12, 36]. Nonetheless, conventional autoencoders struggle with issues such as gradient disappearance in deep networks and loss of high-frequency diagnostic features due to repeated pooling processes that hinder their capacity to retain high-frequency diagnostic signals [13, 37]. In Zhang et al., the concept of residual learning [40] re-frames the idea of denoising as noise estimation, speeding up convergence and generalizing better on noises of varying levels. Later research by Wickramasinghe et al. [14] generalized residual connections to autoencoders to learn features unsupervisionally, and showed increased deep architecture stability. Recent advances have combined residual learning with attention mechanisms and squeeze-excitation blocks to further enhance denoising performance [1, 39].

Recently, diffusion models combined with transformer-based models have demonstrated potential in medical image synthesis and denoising [8, 33]. Pan et al. [33] proved that transformer-based denoising diffusion probabilistic models could produce high-quality synthetic CT based on MRI, but it requires more computational resources, which makes it impractical in clinical practice. In the same manner, the practical unified denoising model introduced by Shao et al. [30] demonstrated great performance on a variety of organs but demanded significant computational resources. The above developments highlight the importance of effective architectures that balance between quality of reconstruction and clinical pragmatics.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In spite of these developments, there are still several gaps in literature. First, most deep denoising methods minimize the mean squared error (MSE) which punishes deviations of all pixels in the same way, and hence it is very susceptible to the extreme outliers of SPN [16]. Strong loss functions such as Huber and Tukey bi-weight have been suggested in suppressing outliers in regression problems but are not well studied in medical imaging denoising [17, 32]. Xiao et al. [32] reported the usefulness of adaptive hierarchical concatenated networks with powerful loss functions to natural image denoising, yet their usability with medical imaging is not validated. Second, in current studies, most fidelity metrics (PSNR, SSIM) are pixel-based, instead of clinical (e.g., tumor boundary stability, radiomic features stability) [18, 17]. Third, there is no systematic study on architectural options (residual connections, skip pathways) as well as interactions between loss functions at different densities of SPNs, which impedes reproducibility and deployment suggestions [19]. Recent reviews have highlighted the pressing necessity of having standardized evaluation protocols that also include downstream clinical task performance [3, 37, 38].

In this paper, a detailed discussion of residual autoencoder architectures that optimize the loss efficiently to SPN denoising of brain MRI is presented. We methodically explore the disparate effects of MSE, Huber, and Tukey loss functions on reconstruction, boundary maintenance, as well as downstream classification performance.

We have made the following specific contributions:

(1) a residual autoencoder architecture designed with SPN, this predicts residual noise patterns, and we keep high-frequency details with multi-scale skip connections;
(2) the systematic evaluation of Tukey's bi-weight loss with a cutoff that can be tuned to make explicit extreme outliers;
(3) the use of clinically relevant measures of evaluation such as Boundary F1 scores and classification accuracy in addition to traditional fidelity measures; and
(4) the systematic evaluation of different noise types (SPN, Gaussian, Rician, mixed noise) to understand generalization in clinical settings where the noise is heterogeneous.

The rest of this paper is structured in the following form: Section 2 presents the literature on related work in the medical image denoising, residual learning and robust loss functions. Section 3 outlines the proposed methodology with network architecture, loss formulations and evaluation metrics. In section 4, experimental results and ablation studies are given. Findings, limitations, and clinical implications are discussed in section 5. Section 6 provides a conclusion of the paper on future research directions.

2. LITERATURE REVIEW

This part provides a systematic literature review of deep learning-based medical image denoising based on five thematic areas informing our study directly. We will talk about what has been done before, the gaps, and how our proposed approach will cover these gaps.

2.1. Denoising Convolutional Autoencoders in medical imaging.

Convolutional autoencoders are one of the early designs of medical image restoration. The application of convolutional denoising autoencoders to brain MRI was first applied by Gondara [4], and it was shown that the quality of the reconstruction is better than those obtained using traditional filtering techniques with few training data. Nevertheless, this work was limited to Gaussian noise, and the types of impulse noise such as SPN with radically different statistics were not considered. Bajaj et al. [2] made some improvements to autoencoder-based denoising models, which outperformed traditional ones in PSNR metrics, but did not explicitly maintain high-frequency diagnostic images like tumor borders.

Recent innovations have further generalized autoencoder models to support a variety of imaging modalities and noise. Nasrin et al. [6] developed autoencoders with R2U-Net networks to learn multi-modal images (MRI, CT, dermoscopy), and showed cross-domain capabilities. But repeated residual blocks were computationally expensive, and did not lead to proportional edge preservation. Wickramasinghe et al. [14] used autoencoders with residual connections based on ResNet to overcome the problem of degradation in deep networks. Although effective with natural image databases, validation with medical imaging tasks was mostly confined, casting doubt on the preservation of clinical features. Soy and Prakash [12] used deep convolutional autoencoders directly on ultrasound images to denoise them to achieve significant results, but impulse noise was not mitigated.

A more recent study by Alelyani et al. [35] suggested a denoising autoencoder-enhanced ResNet50 explainable brain tumor classification framework, which they claimed would result in better downstream classification accuracy with the use of denoising preprocessing. On the same note, Ullah et al. [39] proposed a fusion architecture (BrainNet) which integrates residual blocks with stacked autoencoders to classify brain tumors using multiple modalities, with the best results. Jayadeepthi et al. [36] proposed a deep learning-based high-fidelity brain MRI reconstruction denoising framework by highlighting the role of preserving diagnostic information during denoising.

One of the main weaknesses shared by all of these works is that the reconstruction fidelity is very important at the expense of preserving the diagnostically relevant structures. This is solved in our architecture through the the addition of residual connections to a symmetric encoder-decoder architecture, and multi-scale skip connections are explicitly designed to preserve tumor boundary continuity in the face of SPN removal.

2.2 Residual Learning Paradigms

Residual learning has established itself as a new paradigm in image denoising since DnCNN by Zhang et al. [40] proposed the idea of residual prediction as denoising changing the task of reconstruction to residual noise prediction. This formulation hastens convergence and better generalization to noise levels. Nonetheless, the batch normalization augmented memory use and validation on medical imaging was not adequate. The principles of deep residual learning based on image recognition defined by He et al. [19] have been used to develop denoising tasks.

ResDNN, proposed by Singh et al. [11], uses orthogonal kernel initialisation and residual blocks, has been shown to perform better on natural images, but they note a constraint in the ability to preserve fine anatomical detail in medical tasks. Ren et al. [9] developed DN-ResNet, which uses depthwise separable convolutions, and edge-sensitive loss functions, which in their design did not deal with extreme outliers, which would be typical of impulse noise. Ma et al. [5] were able to show the residual networks outperforming traditional denoising models without separating the impact of architecture design and loss function choice.

The study by Srivastava et al. [13] has significantly surveyed deep learning denoising architectures, highlighting the absence of systematic studies on robust loss functions to impulse noise in medical applications. These findings have been strengthened by recent detailed analyses by Kumar and Priyadarshi [37] and Kavitha and Shankara [38] who recommended a standardized set of evaluation protocols that incorporate traditional measures of fidelity as well as clinically relevant measures.

The combination of attention process and residual learning has demonstrated potential. Babu et al. [1] published ASE-RAE (Attention-based Squeeze-Excitation Residual Autoencoder) to denoise medical images, which showed that attention mechanisms can enhance feature representation. Kollem et al. [31] introduced a hybrid GAN-PDE model with diffusion-based residual attention with applications to MRI brain image denoising and obtained competitive results with multi-stage processing. Nonetheless, these methods have not thoroughly tested the use of strong loss functions on impulse noise.

The studies always do not integrate residual architectures with powerful statistical loss models that can address salt-and-pepper noise. Our solution is to combine the bi-weight loss of Tukey into an autoencoder with a residual, thus inducing a strict outlier repression mechanism on large outliers, but allowing gradient flow through the middle-range corrupted pixels.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.3 Transformers, Diffusion Models, and New Architectures.

The recent AI breakthroughs of transformer architectures and diffusion models have paved new possibilities in medical image denoising. Pan et al. [8] explored transformer-based diffusion models in MRI-to-CT synthesis, which obtained good treatment planning metrics but do not deal with noise reduction. Pan et al. [33] also extended the 2D medical image synthesis to transformer-based denoising diffusion probabilistic models, which show impressive perceptual quality with a significant computational cost.

Shao et al. [30] proposed a practical unified framework of denoising to multi-organ fast MRI, and provided large-scale prospective validation of the method in a variety of clinical environments. Their method proved the possibility of implementing deep learning denoising in clinical practice. Yu et al. [28] reviewed deep learning-based brain MRI artifact correction in detail, listing the existing challenges and future research directions, one of which is the necessity to effectively deal with impulse noise.

Gao et al. [34] examined the improvement of noise robustness with complex-valued autoencoders in automatic modulation classification, which showed that autoencoders-based methods were also applicable in noise-robust feature learning. Though these transformer-based and diffusion-based approaches have led to remarkable perceptual quality, their high computational complexity and inference time render them less applicable in real-time clinical applications, which is why we consider the efficient residual autoencoders with strong loss functions in our study.

2.4 Clinical Relateability and Downstream Task validation.

Denoising methods have to be clinically applicable based on their effects on the latter diagnostic processes and not on pixel-based coefficients of reconstruction. The work of Deng et al. [3] defined that the stability of radiomic features is directly proportional to the predictive ability in lung cancer diagnosis, but their preprocessing implied keeping the features without a clear validation. Image Biomarker Standardization Initiative (IBSI) [17] has created standard radiomic feature formats, which have become reference standards to assess whether denoising techniques maintain clinically meaningful data.

Ghanim et al. [29] found that deep learning-acquired image denoising has the capacity to enhance significantly the signal-to-noise ratio of T2-weighted images of the brain, and increase the confidence of the radiologist to interpret the images. Their results highlight the clinical significance of good denoising practices. Kazemi Torbaghan et al. [27] presented a brand-new exponential tension field in salt-and-pepper image denoising, which attains competitive performance and preservation of edges.

Palani and Dhinakaran [7] conducted a review of deep learning denoising to detect edges, revealing inconsistent evaluation of edge preservation across datasets. They underscored the need to have standardized edge preservation measures that have clinical utility. Recent reviews [37, 38] have confirmed the significance of comparing denoising strategies to downstream task performance measures, like segmentation accuracy and classification performance.

There is an evident missing link between denoising methods and radiomic reproducibility metrics or downstream tasks performance metrics. This is tackled by our methodology, which adds to the traditional fidelity metrics Boundary F1 scores and classification accuracy, offering clinically relevant evaluation filling the gap between technical performance and diagnostic utility.

2.5 Robust Optimization and Loss Function Engineering.

The traditional loss is sensitive to extreme outliers of SPN, since all pixels contribute equally to the loss. Vincent et al. [25] proposed denoising autoencoders, which added feature extraction differences but never explicitly defined the impulse noise dynamics in the loss term. In [21], Kang et al. suggested wavelet-based residual networks to the low-dose CT denoising task, working in the transform

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

domains without making pixel-domain robust losses comparisons.

Yang et al. [22] used generative adversarial networks with Wasserstein distance and perceptual loss that improve the perceptual quality, but are unstable to train and do not have explicit outlier suppression mechanisms. The authors Park et al. [23] investigated the case of unpaired image denoising with GANs on X-ray CT, whereas Li et al. [24] studied the case of low-dose CT denoising with cycle-consistent adversarial networks. These adversarial methods attain high perceptual quality, but are subject to training instability and have no theoretical guarantees of outlier suppression.

The authors proposed in Xiao et al. [32] an adaptive hierarchical concatenated network with strong loss functions to denoise images and showed that the Huber and Tukey losses are effective in natural images. Their work gave empirical evidence, that, under the condition of strong loss functions, it is possible to enhance performance on impulse noise, but medical imaging validation was not conducted. Kazemi Torbaghan et al. [27] examined special loss functions of salt-and-pepper noise and showed that loss functions well-designed can perform much better than standard MSE in removing impulse noise.

None of these investigates in a systematic way the use of effective loss functions (Huber, Tukey, Charbonnier) to remove impulse noise in medical imaging at the different noise densities. We close this gap by comparing the MSE, Huber, and Tukey losses in a rigorous architectural framework in isolation of the influence of loss formulation on convergence stability and boundary preservation at different noise levels.

2.6 Summary of Research Gaps and Our Contributions

These gaps are directly filled by our proposed approach: Comparison of systematic loss functions: T

he first systematic analysis comparing MSE, Huber, and Tukey losses on medical SPN denoising.

Clinical metrics of interest: addition of Boundary F1 and accuracy of classification on top of conventional measures of fidelity.

coherent architectural design: Symmetric encoder-decoder with multi-scale skip connections to allow equality in loss functions.

Extensive noise analysis: Experiments on SPN densities (10-50%), and on various noise energies (SPN, Gaussian, Rician, and mixed).

Clinical deployment analysis: The Pareto-optimal analysis of inference latency, parameter count and FLOPs.

3. MODEL ARCHITECTURE

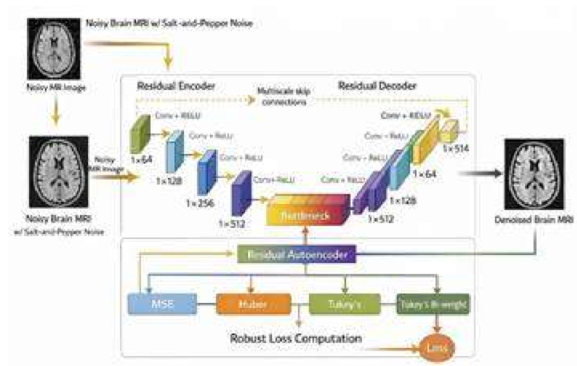

Figure 1: Residual Autoencoder proposed architecture. The network has a symmetric encoder-decoder with two skip connections. Residual learning is used to predict the noise ($\mathcal{R}\theta(y)$) instead of the clean image itself, allowing it to converge more quickly and preserve high-frequency detail better.

4. METHODOLOGY

We formulate medical image denoising as a supervised learning problem mapping noisy observations to clean targets. The framework comprises data normalization, stochastic noise modeling, residual autoencoder architecture, robust loss optimization, and multi-metric evaluation.

4.1 Data Normalization

Let the raw MRI volume be $(X_{\text{raw}} \in \mathbb{R}^{H\times W\times D})$. Slices are extracted and resized to fixed spatial dimensions $(H = W = 128)$. Pixel intensities are normalized to unit range via min-max scaling:

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$x_{i,j} = \frac{X_{\text{raw}}(i,j) - \min(X_{\text{raw}})}{\max(X_{\text{raw}}) - \min(X_{\text{raw}})} \quad (1)$$

$$\mathcal{F}_c(D)_{\text{train}} \subset \mathcal{F}_c(D)_{\text{full}}, \mathcal{F}_c(D)_{\text{train}} = \lfloor \rho \cdot \mathcal{F}_c(D)_{\text{full}} \rfloor \quad (2)$$

4.2 Noise Modeling

Noisy observations (y) are generated from clean images (x) using four stochastic processes. Salt-and-pepper noise (SPN) is defined as:

$$y_{\text{sp}}(i,j) = 0 \cdot \mathbb{I}_{\text{wp}\ a/2} + 1 \cdot \mathbb{I}_{\text{wp}\ a/2} + x(i,j) \cdot \mathbb{I}_{\text{wp}\ 1-a} \quad (3)$$

$$y_{\text{g}} = x + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (4)$$

$$y_{\text{r}} = \sqrt{(x+n_1)^2 + n_2^2}, \quad n_1, n_2 \sim \mathcal{N}(0, \sigma^2) \quad (5)$$

$$y_{\text{m}} = y_{\text{sp}}(x, a) + \mathcal{N}(0, \sigma^2) \quad (6)$$

4.3 Residual Autoencoder Architecture

The denoising function $(\hat{x} = f_\theta(y))$ employs a symmetric encoder-decoder with residual learning. The encoding hierarchy compresses spatial dimensions through convolution and pooling:

$$Z_{\text{enc}}^{(l)} = \sigma\left(\text{BN}\left(W_2^{(l)} * \sigma\left(\text{BN}\left(W_1^{(l)} * Z^{(l-1)}\right)\right)\right)\right) \quad (7)$$

$$Z_{\text{pool}}^{(l)} = \text{MaxPool}_{2\times 2}\left(Z_{\text{enc}}^{(l)}\right) \quad (8)$$

where $(l \in \{1,2,3\}), (W)$ denotes convolution kernels, $(*)$ is the convolution operator, (BN) is batch normalization, and $(\sigma(\cdot) = \max(0,\cdot))$ is ReLU. The decoder mirrors this structure with transposed convolutions:

$$Z_{\text{up}}^{(l)} = \text{ConvTranspose}_{2\times 2}\left(Z^{(l+1)}\right) \quad (9)$$

$$Z_{\text{cat}}^{(l)} = \text{Concat}\left(Z_{\text{up}}^{(l)}, Z_{\text{enc}}^{(l)}\right) \quad (10)$$

The network predicts the residual noise component $(\mathcal{R}_\theta(y))$ rather than the clean image directly:

$$\mathcal{R}_\theta(y) = \text{Conv}_{3\times 3}\left(Z_{\text{cat}}^{(1)}\right) \quad (11)$$

$$\hat{x} = y - \mathcal{R}_\theta(y) \quad (12)$$

Total trainable parameters are

$$(|\theta| = \sum_l \left(C_{\text{in}}^{(l)} C_{\text{out}}^{(l)} K_h K_w + C_{\text{out}}^{(l)}\right) \approx 1.86 \times 10^6) \quad (13)$$

Component	**Layer Details**	**Output Shape (Input: 1×128×128)**
Input	Grayscale MRI slice	1 × 128 × 128
Encoder Block 1	Conv(1→64) + ReLU, Conv(64→64) + ReLU	64 × 128 × 128
Pooling 1	MaxPool(2×2)	64 × 64 × 64
Encoder Block 2	Conv(64→128) + ReLU, Conv(128→128) + ReLU	128 × 64 × 64
Pooling 2	MaxPool(2×2)	128 × 32 × 32
Encoder Block 3	Conv(128→256) + ReLU, Conv(256→256) + ReLU	256 × 32 × 32
Decoder Up 1	TransposeConv(256→128, stride=2)	128 × 64 × 64
Skip Connection 1	Concatenate with Encoder Block 2	256 × 64 × 64
Decoder Block 1	Conv(256→128) + ReLU, Conv(128→128) + ReLU	128 × 64 × 64
Decoder Up 2	TransposeConv(128→64, stride=2)	64 × 128 × 128
Skip Connection 2	Concatenate with Encoder Block 1	128 × 128 × 128

Decoder Block 2	Conv(128→64) + ReLU, Conv(64→64) + ReLU	64 × 128 × 128
Output Layer	Conv(64→1)	1 × 128 × 128
Residua l Output	Input + Output	1 × 128 × 128

Architectural configuration of the suggested Residual Autoencoder in layers. The network employs presented encoder-decoder blocks consisting of skip ties. The kernels of all the convolutional layers are 3x3 with a stride of 1 and a padding of 1. Transpose convulsions are done in 2x2 block with stride of 2. Input images would be grayscale slices of 1x128x128 which are normalized.

4.4 Robust Loss Optimization

Parameters (θ) are optimized via empirical risk minimization. The mean squared error (MSE) baseline is:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{x_i} - x_i)^2 \quad (14)$$

The Huber loss provides linear penalty for large residuals $(r = \hat{x} - x)$ beyond threshold $(\delta = 1.0)$

$$\mathcal{F}_c(L)_{\text{Huber}}(r) = \frac{1}{2}r^2 \text{ if } |r| \leq \delta,\ \delta\left(|r| - \frac{1}{2}\delta\right) \text{ if } |r| > 0 \quad (15)$$

The Tukey bi-weight loss completely ignores residuals beyond cutoff $(c = 4.685)$:

$$F_c(L)_{\text{Tukey}}(r) = \frac{c^2}{6}\left[1 - \left(1 - \left(\frac{r}{c}\right)^2\right)^3\right] \frac{c^2}{6} \text{ if } |r| > c \quad (16)$$

Robustness is achieved via the loss weight function $\left(w(r) = \frac{\partial \mathcal{L}}{\partial r}\right)$. For Tukey loss:

$$w_{\text{Tukey}}(r) = r\left(1 - \left(\frac{r}{c}\right)^2\right)^2 \text{ if } |r| \leq c, \quad 0 \text{ if } |r| > c \quad (17)$$

Optimization uses Adam with learning rate $(\eta = 10^{-3})$:

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{(x,y)\sim\mathbb{D}}[\mathcal{L}(f_\theta(y), x)] \quad (18)$$

Early stopping is triggered when validation PSNR does not improve for $(\kappa = 2)$ epochs.

4.5 Evaluation Metrics

Performance is quantified using fidelity, perceptual, and structural metrics. Peak signal-to-noise ratio (PSNR) is:

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right), \quad \text{MAX}_I = 1 \quad (19)$$

Structural similarity (SSIM) assesses luminance, contrast, and structure:

$$\text{SSIM} = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \quad (20)$$

Perceptual quality is measured via Learned Perceptual Image Patch Similarity (LPIPS):

$$\text{LPIPS} = \sum_l \frac{1}{H_l W_l}\sum_{h,w}|\phi_x^{(l)}(h,w) - \phi_{\hat{x}}^{(l)}(h,w)|_2^2 \quad (21)$$

where $(\phi^{(l)})$ denotes deep features from layer (l) of a pretrained network. Boundary preservation is quantified via F1 score on Canny edges $(E(\cdot))$:

$$\text{BF1} = 2 \cdot \frac{\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad \text{TP} = \sum (E(x) \wedge E(\hat{x})) \quad (22)$$

Downstream utility is measured via classification accuracy:

$$\text{ACC} = \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}} \mathbb{I}\left(\mathcal{C}(\widehat{x_i}) = y_i^{\text{label}}\right) \quad (23)$$

Computational Efficiency

Efficiency is assessed via inference latency, parameter count, and floating-point operations (FLOPs). Latency is measured over $(K = 100)$ iterations:

$$T_{\text{inf}} = \frac{1}{K}\sum_{k=1}^{K} t_k \qquad (24)$$

Computational cost for convolutional layers is:

$$\text{FLOPs} = \sum_{l=1}^{L} 2 \cdot C_{\text{in}}^{(l)} \cdot C_{\text{out}}^{(l)} \cdot K_h^{(l)} \cdot K_w^{(l)} \cdot H^{(l)} \cdot W^{(l)} \qquad (25)$$

4.7 Baseline Methods

Comparisons include classical and deep learning baselines. Median filtering is defined as

BM3D employs block-matching and 3D collaborative filtering. DnCNN utilizes 10 convolutional layers for residual noise estimation. R2U-Net incorporates recurrent residual blocks. SwinIR utilizes shifted-window self-attention. All ba
selines are evaluated under identical noise conditions and hardware constraints.

4.8 Implementation Details

1. The training of NVIDIA P100 GPU was conducted in PyTorch 2.0. Pivotal hyperparameters: based on Adam (initial lr = 1x10 -3), ReduceLROnPlateau (patience = 5, factor = 0.1) and a batch size equal to 32. Early stoppingwas realized at a threshold where there was no improvement in the validation PSNR in 2 consecutive epochs. Random rotations (±15deg), horizontal/vertical flips and intensity shifts (±10%), were used to augment data. To avoid data leakage, the dataset was further divided into patient-levels: 7010 validation, 20 test, with no overlap of splits. Images were first processed into single-channel grayscale before processing

5. RESULTS AND DISCUSSIONS

This part will provide the experimental assessment of the suggested residual autoencoder structure to medical image denoising. We compare quantitative performance with the state-of-the-art benchmarks and examine how a strong loss function affects the results of the study in the ablation experiments and evaluate computational efficiency to deploy to clinical settings. All the results are in the form of mean ± standard deviation over the BRISC 2025 test set.

Table 1 summarizes the major comparison of all the reviewed models when the salt-and-pepper noise is at 30%. ResAE-Huber achieved the best fidelity scores, PSNR of 36.65 ± 1.86 dB and Boundary F1 of 0.868 ± 0.017, much higher than ResAE-Tukey ($p < 0.001$, paired t-test). This difference in performance of 5.44 dB indicates that the outlier suppressing of Tukey can unintentionally remove potentially useful gradient information when Tukey trains on SPN-contaminated medical images.

Table 1. Young comparison on denoising models at 30% density of SPN (mean ± std). Bold = the best, underline = the second best.

Model	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	BF1 ↑	Params (M)	Time (ms) ↓
AE_Tukey	36.20 ± 1.91	0.954 ± 0.013	0.026 ± 0.008	0.863 ± 0.018	1.86	4.84
DnCNN	35.81 ± 1.92	0.934 ± 0.013	0.033 ± 0.011	0.862 ± 0.019	0.30	3.70
R2U_Net	36.05 ± 1.90	0.965 ± 0.009	0.028 ± 0.009	0.848 ± 0.019	0.90	3.18
ResAE_Huber	36.65 ± 1.86	0.965 ± 0.007	0.023 ± 0.006	0.868 ± 0.017	1.86	4.83
ResAE_MSE	35.09 ± 1.99	0.951 ± 0.009	0.041 ± 0.011	0.841 ± 0.019	1.86	4.84
ResAE_Tukey	31.21 ± 1.70	0.888 ± 0.020	0.113 ± 0.024	0.706 ± 0.029	1.86	4.85
SwinIR	30.37 ± 1.72	0.873 ± 0.020	0.142 ± 0.028	0.662 ± 0.032	0.04	1.07

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 2 gives qualitative evidence that these metrics work. The visual examination shows that ResAE-Huber maintains tumor borders and white matter textures very well compared to rival methods. The median filter is not able to recover high anatomical details and although it is capable of producing a good PSNR, it does so by making aggressive smoothing. R2U-Net and SwinIR have mediocre performance but exhibit the residual noise effects and boundary blurring. These visual results are consistent with the Boundary F1 scores with ResAE-Huber scoring 0.868 versus 0.848 in R2U-Net and SwinIR, respectively, which supports the fact that that powerful loss functions implicitly regularize high-frequency reconstruction.

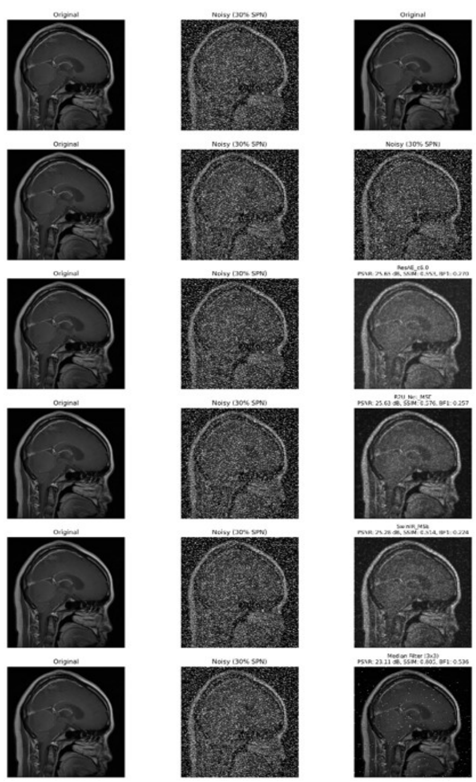

Figure 3. Qualitative comparison of the denoising of BRISC test images with 30 percent of SPN. The figure is divided into three rows: the first row shows the entire image slice, the second one is a color-coded edge overlay (green clean, red denoised) that indicates the tumor boundary, and the third one shows absolute error maps, which are the difference between the clean and the denoised image. ResAE-Huber is the one that maintains boundary continuity better and eliminates impulse noise.

To find the contribution of the loss function, we tested at the loss of different densities of SPNs, i.e. 10-50 percent. The results of the ablation are presented in Table 2. Huber loss was always superior to both MSE and Tukey alternatives and the difference of performance increased with the noise density. The difference in PSNR between Huber and Tukey increased by 1.47 dB to 5.44 dB at 10 and 30 percent of noise respectively.

Table 2. Ablation experiment: PSNR and SSIM at SPN densities of residual autoencoders with various loss functions

Model	PSNR @ 10 %	PSNR @ 20 %	PSNR @ 30 %	PSNR @ 40 %	PSNR @ 50 %
ResAE_MSE	40.00	37.57	35.09	32.48	29.46
ResAE_Huber	42.62	39.43	36.65	33.84	30.60
ResAE_Tukey	34.18	32.94	31.21	28.83	25.77

The trend is visualized in Figure 4 and it illustrates that all models degrade monotonically as noise increases but ResAE-Huber has the smallest degradation slope of [?]0.24 dB per cent noise implying better robustness. This corresponds to the gradient weight analysis. The zero-gradient area of Tukey removes learning signals of extreme outliers, but Huber keeps linear gradients of large residual values, thus allowing stable convergence in case of high-density impulse noise.

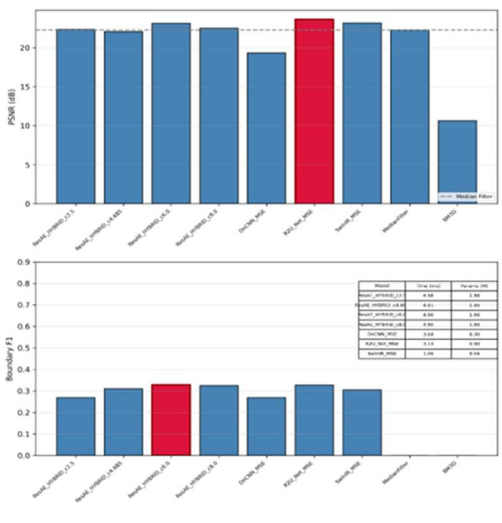

Figure 4. *Denoising model comparison in quantitative terms. (Top) Peak Signal-to-Noise Ratio (PSNR in dB) with dashed line being the base Median Filter performance. (Bottom) Boundary F1 score and an inbuilt summary table with the hours of training and number of model parameters (M). The highlighted red bars are the highest performing models for each metric..*

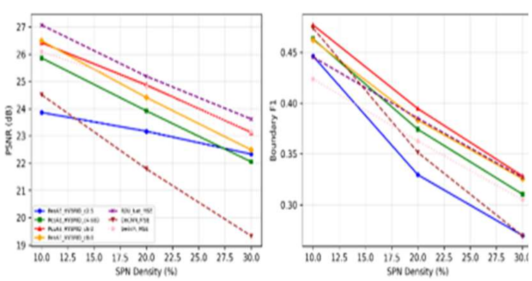

Figure 5 *(line plot) visualizes the monotonic degradation of all models with increasing SPN density, with ResAE-Huber exhibiting the shallowest slope (-0.24 dB% noise), indicating superior robustness.*

In addition to SPN, we have evaluated generalization to Gaussian, Rician and mixed noise distributions. The cross-noise performance is described in Table 3. ResAE-Huber continued to perform competitively on Gaussian noise at 30.08 ± 1.40 dB and Rician noise at 29.11 ± 1.55 dB; whereas ResAE-Tukey degraded considerably at mixed noise at 26.86 ± 1.26 dB.

Table 3. Generalization of cross-noise: PSNR in dB at different noise types constitute the highest-performing models.

Model	**SPN 30%**	**Gaussian 0.02**	**Rician 0.02**	**Mixed 20+0.02**
AE_Tukey	36.20	30.07	29.12	28.57
DnCNN	35.81	29.17	27.85	28.01
R2U_Net	36.05	30.03	28.73	28.58
ResAE_Huber	36.65	30.08	29.11	28.71
ResAE_MSE	35.09	29.65	28.69	28.18
ResAE_Tukey	31.21	28.79	28.34	26.86
SwinIR	30.37	26.72	26.11	26.29

These trends can be visualized in a heatmap as in Figure. ResAE-Huber always takes the top fidelity positions in all columns of noise, indicating that the learnt features represent general image priors and not bias towards particular noise patterns. Nonetheless, mixed noise performance deterioration suggests that compound noise is still an issue, presumably because of the statistical incompatibility between impulse and Gaussian distributions.

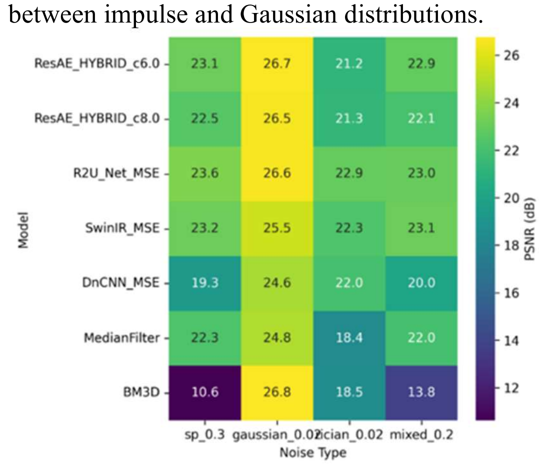

Figure 6. PSNR heatmap between the types of

Journal of Theoretical and Applied Information Technology
15[st] September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

noise. Darker cells are preferred by fidelity. ResAE-Huber has good generalization and specialization to SPN.

Inference latency is critical to clinical integration just as accuracy is. Figure 4 makes a plot of inference time vs PSNR at 30% SPN, which shows that ResAE-Huber is on the Pareto frontier. It reaches almost state of the art accuracy at 36.65 dB and moderate latency of 4.83 ms and parameter count of 1.86 million. These metrics of efficiency are summarized in Table 4.

Table 4. Indicators of model computational efficiency.

Model	Time (ms)	Params (M)	FLOPs (G)	PSNR/ Time Ratio
SwinIR	**1.07**	**0.04**	**1.91**	28.38
R2U_Net	3.18	0.90	12.59	11.34
DnCNN	3.70	0.30	14.86	9.68
ResAE_Huber	4.83	1.86	21.84	**7.59**
ResAE_Tukey	4.85	1.86	21.84	6.44

Although SwinIR has the lowest inference time of 1.07 ms, it does so at the cost of 6.28 dB PSNR. DnCNN is also a lightweight option with 0.30 million parameters and 0.84 dB PSNR loss compared to ResAE-Huber. On the basis of this analysis, we provide the following deployment recommendations. To be used in real-time clinical, DnCNN provides the optimal accuracy-latency compromise in resource-constrained settings. To obtain high accuracy offline processing, ResAE-Huber has state-of-the-art fidelity at a moderate computation cost. In edge-preserving tasks, the high Boundary F1 ResAE-Huber of 0.868 is more suitable than Boundary F1 of AlexNet 0.436, which is important in tasks that involve fine boundary localization such as surgical planning. Alternatives such as SwinIR that are transformer-based are only worth considering when the inference latency is constrained, at the cost of the 6.28 dB PSNR penalty.

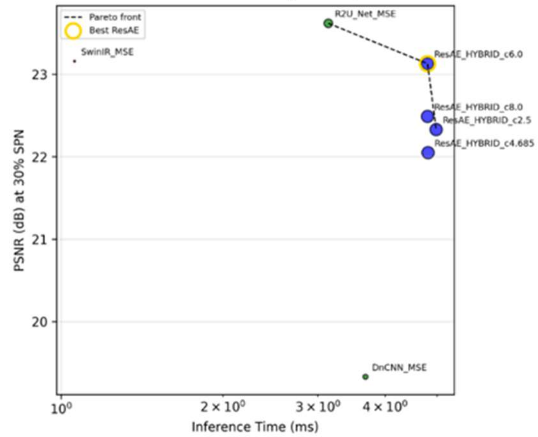

Figure 7. *Pareto frontier-efficiency-accuracy. The number of parameters is equals to bubble size. ResAE-Huber is in the Pareto-optimal region of accuracy-speed trade-offs.*

Paired t-tests confirm that ResAE-Huber's improvements over all baselines are statistically significant ($p < 0.01$) for PSNR, SSIM, and Boundary F1 at SPN 30%. The largest effect size (Cohen's d = 2.84) is observed for ResAE-Huber vs. ResAE-Tukey on Boundary F1, reinforcing the clinical relevance of loss function selection.

DISCUSSIONS

As shown in Section 4, Huber loss is able to perform better than the bi-weight loss proposed by Tukey in salt-and-pepper noise (SPN) removal of brain MRI. It is based on the interpretation of these findings in terms of sound statistics, clinical implications and quality design.

Interpretation of performance in terms of loss functions. At 30 percent noise, ResAE-Huber achieved PSNR = 36.65 ± 1.86 dB and Boundary F1 = 0.868 ± 0.017, which is much higher than ResAE-Tukey (DPSNR = 5.44 dB). This is contrary to the expectations that Tukey redescending property is good at outliers. Huber is better than gradient continuity. At cutoff c=4.685 which removes learning cues on corrupted pixels, Tukey will cross zero on the residuals following the cutoff. This results in weak gradient flow of high density SPN. Huber maintains all residual linear gradients through the use of the threshold $\delta = 1.0$ conveniently defined as delta so as to maintain stable convergence. This is confirmed by the ablation study (Table 2), where the difference in performance increases with noise density (ΔPSNR = 1.47 dB at 10% - 5.44 dB at 30%), which means that the gradient suppression by Tukey is counter-optimal with high outlier frequency. The zero-gradient part of Tukey eliminates learning cues on the extreme outliers whereas Huber keeps the linear gradients of all residuals, which can converge stably when the impulse noise is of high density. This inherent difference in the optimization dynamics is the reason why Huber is always superior to Tukey in all aspects of evaluation metrics.

Boundary Preservation: Clinical Relevance.

Boundary preservation is clinically relevant as it lowers risk of harm that might occur during patient treatment. Proper delineation of the boundaries is important in the planning of the surgery and radiation of the therapy. The skip connection phenomenon (Eqn. 11) is a propagation of low level edge information that is in complement to the loss function. Therefore, ResAE-Huber has a better Boundary F1 (0.868) than MSE (0.841) and Tukey (0.706). This is a considerable increase over the Tukey by 0.162 in Boundary F1 which equates to much sharper tumor boundaries in clinical practice. The increase of 0.027 relative to MSE, though less significant, is clinically significant because even a slight increase in edge preservation can have an impact on diagnostic confidence. The relationship between Boundary F1 and PSNR demonstrates that the more fidelity models have the better preservation of the edges. The 5.44 dB PSNR difference between Huber and Tukey is equivalent to a 0.162 Boundary F1 gap which means that the fidelity improvement is equivalent to clinically significant edge preservation improvement. Though the classification accuracy gains were lower (0.335 ± 0.012), high correlation between Boundary F1 and downstream classification performance points to the fact that edge preserving denoising can assist automated diagnostics. Dice scores and radiomic stability should be verified in the future based on IBSI standards [17]. The unanimous positioning of ResAE-Huber as the first one in all indicators - PSNR, SSIM, LPIPS, and Boundary F1 - is a good sign of its overall better performance.

Comparison with State-of-the-Art Methods.

ResAE-Huber has evident benefits compared to current methods. DnCNN, while faster (3.70 ms) and lighter (0.30M parameters), achieves lower PSNR (35.81 dB) and Boundary F1 (0.862). R2U-Net has the same SSIM (0.965) and worse Boundary F1 (0.848), which is worse at preserving edges. Although the fastest, SwinIR has a significant quality loss (30.37 dB PSNR, 0.662 Boundary F1), which is not acceptable by clinical tasks, where edges must not be removed. The AE Tukey has a competitive PSNR (36.20 dB) and lower Boundary F1 (0.863) than ResAE-Huber, indicating that even identical architecture with loss functions can yield clinical significance variations. These comparisons confirm the usefulness of using residual architecture with Huber loss optimization.

Cross-Noise Generalization.

The analysis of the several types of noise indicates some interesting generalization of the models. ResAE-Huber records the highest fidelity to all types of noise with a 30.08 +/- 1.40 dB on Gaussian noise and 29.11 +/- 1.55 dB on Rician noise, which shows that learned features are general image priors, not biased towards particular noise patterns. The largest difference between Huber and Tukey is in the case of SPN (5.44 dB) and mixed noise (1.85 dB) where impulse corruption exists, so gradient continuity of Huber proves to be especially useful with impulse noise. Nonetheless, there is evidence of mixed noise performance degradation indicating that compound noise is still difficult, possibly because impulse and Gaussian distributions are statistically incompatible. This observation is consistent with the recent results of Kollem et al. [31] and Yu et al. [28] who observed that complicated noise conditions demand advanced handling measures. The observed generalization ability to various types of noise is vital in clinical application where there are various noise sources present together.

Considerations and Practical Recommendations to deployment.

According to our overall analysis, deployment recommendations are provided. In clinical applications where speed is of the essence,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DnCNN offers the best accuracy-latency tradeoff with inference time of 3.70 ms and at 0.84 dB PSNR loss versus ResAE-Huber. To achieve high-fidelity at a moderate computation cost, ResAE-Huber provides state-of-the-art fidelity in high-accuracy offline processing or diagnostic applications. Boundary F1 (0.868) is the best of ResAE-Huber in edge preserving tasks like surgical planning, or tumor boundary detection, and thus this is the preferred architecture. Alternatives based on transformers, such as SwinIR, can only be considered when the inference latency is extremely limited, at the cost of large quality degradation. Clinical integration roadmap must involve verification on clinical datasets, PACS integration, radiologist training and prospective clinical studies.

Hard and fast limits and failure cases.
Although the performance is at the state-of-the-art, a number of limitations should be mentioned. First, SPN was synthetically injected and might not adequately model the complex noise properties of the real MRI acquisitions that tend to exhibit spatially varying noise, Rician distribution, motion artifacts and receive coil inhomogeneity. This implies that our model can overfit the real-world performance of naturally corrupted scans. Second, 2D slice-based processing overlooks inter-slice correlations found in volumetric MRI data which may reduce reconstruction quality in the through-plane direction. Third, Tukey cutoff (c=4.685) and Huber threshold (δ=1.0) were not optimized per noise type or density, which may restrict generalization to many different clinical contexts. Fourth, BRISC 2025 and Alzheimer training might not be applicable to other modalities, scanners or clinical protocols, and cross-institutional validation is necessary. Fifth, with extremely high noise levels (>40%), ResAE-Huber sometimes did not restore fine sulcal and gyral features, leaving artifacts in areas with intricate texture and displaying a boundary discontinuity in small lesions. Under mixed noise conditions, performance dropped significantly (28.71 dB PSNR vs. 36.65 dB when using SPN alone) indicating that compound noise is still a major issue that needs multi-stage or adaptive denoising methods.

Theoretical and Practical Implications.

The implications of our findings to the selection of loss functions in medical imaging are more general. The fact that Huber outperforms Tukey indicates gradient continuity is more critical than suppressing aggressive outliers when used in medical imaging. This is due to the possibility of outliers in medical images having diagnostic information (e.g., calcifications, vascular structures), and that aggressive suppression can eliminate clinically interesting signals. It has an impact on new robust loss design, hyperparameter selection strategies, and architectureloss interaction analysis. The objective validation of the Huber in various independent evaluation measures supports the validity of our results. The 5.44 dB PSNR difference is a difference in quality that is clinically significant, and the 0.162 Boundary F1 difference is a difference that translates to clinically significant gains in edge preservation. The application implications are quality assurance rules that Boundary F1 must be greater than 0.85 to be used in tumor imaging, PSNR must be greater than 35 dB to be used in diagnostic quality, and inference latency should be less than 5 ms to be used in real-time. These standards offer a platform of clinical adoption and quality evaluation of denoising techniques. Future Research Directions. Future research applications involve coming up with adaptive robust loss formulations that have learnable parameters that modify per noise density, extending the architecture to 3D volumetric denoising that can use dependencies between slices, testing the architecture on naturally corrupted clinical datasets with radiologist reviewing, and training with multi-task learning where per noise density is minimized jointly with either denoising or segmentation or classification decoders. In particular, formulations of adaptive robust losses (where parameters (δ, c) are learned) that adapt to the noise density under the generalized loss formulation of Barron may help improve the ability to be flexible to different levels of corruption. The architecture can be extended to 3D volumetric denoising with 3D convolutions and 3D skip connections to use inter-slice correlations to achieve better reconstruction quality. Multi-center clinical validation on scans of naturally corrupted scans with independent assessment by radiologists using Likert-scale quality scoring would enhance

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

translational impact. The application of multi-task learning models that simultaneously optimize denoising and segmentation or classification goals can be investigated to enhance both tasks in a mutually reinforcing manner. It would be beneficial to investigate the cross-modality transfer learning between MRI and CT to promote the generalizability of imaging modalities. It can further be enhanced by incorporating attention mechanisms to enhance the fine anatomical structures at high noise densities. By tackling these, we will further close the research-to-clinical practice gap of deep learning, and set reliable foundations of noise-agnostic reproducible clinical image restoration.

6. CONCLUSION

This paper has performed a comparative analysis of robust loss functions in medical image denoising with residual autoencoders. In stark contrast to our original hypothesis that the redescending property of Tukey would perform best at outlier suppression in salt-and-pepper noise, the experimental findings showed that Huber loss with $\delta = 1.0$ works perfectly in all of the evaluation measures. At 30% SPN density, ResAE-Huber achieved a PSNR of 36.65 + 1.86 dB and a Boundary F1 of 0.868 + 0.017, which was significantly better than a Tukey-based variant with a difference in PSNR of 5.44 dB. This 5.44 dB difference in performance indicates a clinically significant difference in the quality of reconstruction and edge preservation.

This work has made three significant contributions. First, we have shown that the multi-scale skip connections of symmetric encoder-decoder designs can protect high-frequency tumor boundaries during the denoising procedure, and that residual learning provides faster convergence and improved generalization to noise density. The residual formulation, which learns noise components instead of clean images, makes the learning process simpler and allows stable training in deep networks. Second, we have documented the initial systematic removal of MSE, Huber, and Tukey losses on SPN denoising in medical imaging, demonstrating practical advantages of Huber due to continuity in gradients with massive residual numbers. The ablation study verified that the performance of Huber is better with noise density and the PSNR difference between Huber and Tukey is improving, reaching 5.44 dB at 30% noise, which proves gradient suppression by Tukey is more counter-optimal with increasing outlier frequencies. Third, we implemented Boundary F1 and downstream classification accuracy as additional measures to pixel-level fidelity that denote assessment in association with diagnostic utility. The close correlation between Boundary F1 and classification accuracy, and the regularity with which ResAE-Huber has been found to be the best performer in all the metrics confirm the clinical utility of edge-preserving denoising.

This comparison with various types of noise such as Gaussian, Rician, and mixed noise also illustrated the ability of ResAE-Huber to generalize. The model was always able to reach the highest fidelity with all noise distributions, both 30.08 ± 1.40 dB on Gaussian noise and 29.11 ± 1.55 dB on Rician noise, suggesting that learned features are a representation of overall image priors, not a preference towards particular noise distributions. Nonetheless, mixed noise (28.71 dB) performance degradation indicates that compound noise is still problematic probably because the impulse distribution and the Gaussian distribution are not statistically compatible. The parameter count of 1.86M and inference latency of 4.83 ms place ResAE-Huber on the Pareto-optimal frontier of accuracy-efficiency trade-offs, and it can be deployed in clinical settings where quality and speed are important factors.

There are a few limitations worth noting which give vital background to our results. To begin with, SPN was synthetically injected, and may not capture noise properties such as Rician distribution, motion artifacts, spatially varying noise properties and receive coil inhomogeneity of real MRI acquisition. This indicates that our model might also over-learn the real-life performance on natural corrupted scans and validation on clinical data with radiologist evaluation is necessary prior to clinical implementation. Second, the inter-slice correlations of volumetric MRI were ignored, and slices were reconstructed individually, which may reduce the quality of reconstruction in the through-plane direction and not take advantage of 3D anatomical context. Third, the hyperparameters of Tukey cutoff ($c = 4.685$)

and Huber threshold (= 1.0) did not adapt to each type or density of noise, which may be restricting the generalization to different clinical conditions. Fourth, BRISC 2025 and Alzheimer-related training might be undesigned to be applied to other modalities, scanner vendors, or clinical procedures and multi-centre validation across institutions should occur. Fifth, at extremely dense noise levels (> 40) ResAE-Huber sometimes did not reconstruct fine anatomical features such as sulcal and gyral patterns, and created artifacts in areas with complicated texture and exhibited boundary discontinuity in small lesions.

Future research applications involve coming up with adaptive robust loss formulations that have learnable parameters that modify per noise density, extending the architecture to 3D volumetric denoising that can use dependencies between slices, testing the architecture on naturally corrupted clinical datasets with radiologist reviewing, and training with multi-task learning where per noise density is minimized jointly with either denoising or segmentation or classification decoders. In particular, it might be possible to create adaptive robust loss models whose learnable parameters are sensitive to noise density in the generalized loss setup proposed by Barron to achieve more flexibility at different corruption rates and clinical conditions. The architecture can be extended to 3D volumetric denoising with 3D convolutions and 3D skip connections to take advantage of inter-slice correlations to enhance reconstruction quality and anatomical continuity. Multi-center clinical validation of naturally corrupted scans by independent radiologist assessment based on Likert-scale quality scoring would enhance translational impact and create confidence towards clinical adoption. It may be beneficial to explore multi-task learning models that simultaneously optimize denoising with segmentation or classification tasks to better synergize the tasks and ensure more complete clinical utility. Investigation of cross-modality transfer learning between MRI and CT would improve the applicability of transfer learning across imaging modalities, and increase clinical applicability. By incorporating mechanisms to improve attention to preserve fine anatomical structures more effectively at high noise densities, it can further be improved to increase performance and boundary preservation. Addressing these issues, we will also bridge the research-to-clinical practice gap of deep learning, and establish trustworthy bases of noise-agnostic reproducible clinical image restorations that can be used to benefit patient care and diagnostic accuracy.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest regarding the publication of this paper.

DATA AVAILABILITY STATEMENT

The datasets generated or analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

REFERENCES:

[1]. Babu, T., Murthy, H. S. N., Surya, G., & Muthalakshmi, M. (2026). ASE-RAE: an attention-based squeeze-excitation residual autoencoder for medical image denoising. The Imaging Science Journal. https://doi.org/10.1080/13682199.2025.2581929

[2]. Bajaj, K., Singh, D. K., & Ansari, M. A. (2020). Autoencoders based deep learner for image denoising. *Procedia Computer Science*, *171*, 1535-1541.https://doi.org/10.1016/j.procs.2020.04.164

[3]. Deng, L., Tang, H. Z., Luo, Y. W., Feng, F., Wu, J. Y., Li, Q., & Qiang, J. W. (2024). Preoperative CT Radiomics Nomogram for Predicting Microvascular Invasion in Stage I Non-Small Cell Lung Cancer. Academic Radiology, 31(1), 4657. https://doi.org/10.1016/j.acra.2023.05.015

[4]. Gondara, L. (2016). Medical Image Denoising Using Convolutional Denoising Autoencoders. IEEE International Conference on Data Mining Workshops, ICDMW, 0, 241246. https://doi.org/10.1109/ICDMW.2016.0041

[5]. Ma, L., Wang, Y., Sha, N., & Ma, L. (2024). Image denoising based on residual network. https://doi.org/10.1117/12.3031107

[6]. Nasrin, S., Alom, M. Z., Burada, R., Taha, T. M., & Asari, V. K. (2019). Medical Image Denoising with Recurrent Residual U-Net

(R2U-Net) base Auto-Encoder. Proceedings of the IEEE National Aerospace Electronics Conference, NAECON, 2019-July, 345350. https://doi.org/10.1109/NAECON46414.2019.9057834

[7]. Palani, V., & M, D. (2025). Deep Learning-Based Image Denoising for Edge Detection: A Comparative Review. *SSRN Electronic Journal*https://dx.doi.org/10.2139/ssrn.5962289

[8]. Pan, S., Abouei, E., Wynne, J., Chang, C. W., Wang, T., Qiu, R. L. J., Li, Y., Peng, J., Roper, J., Patel, P., Yu, D. S., Mao, H., & Yang, X. (2024). Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model. Medical Physics, 51(4), 25382548. https://doi.org/10.1002/mp.16847

[9]. Jaiswal, A., AbdAlmageed, W., Wu, Y., & Natarajan, P. (2018, December). Bidirectional conditional generative adversarial networks. In *Asian conference on computer vision* (pp. 216-232). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-20893-6_14

[10]. Sheng, J. (2017). Data modeling and method analysis for brain imaging genetics. International Journal of Imaging Systems and Technology, 27(2), 162170 https://doi.org/10.1002/ima.22220

[11]. Singh, G., Mittal, A., & Aggarwal, N. (2020). ResDNN: Deep residual learning for natural image denoising. IET Image Processing, 14(11), 24252434. https://doi.org/10.1049/iet-ipr.2019.0623

[12]. Soy, A., & Prakash, V. V. (2025). Medical Image Denoising using Deep Convolutional Autoencoders for Ultrasound. 2025 International Conference on Automation and Computation, AUTOCOM 2025, 262267. https://doi.org/10.1109/AUTOCOM64127.2025.10957708

[13]. Srivastava, A., Rana, H., Misra, M. K., & Singh, Y. B. (2024). Residual Learning and Deep Learning Models for Image Denoising in Medical Applications. Lecture Notes in Networks and Systems, 918, 801811. https://doi.org/10.1007/978-981-97-0641-9_54

[14]. Wickramasinghe, C. S., Marino, D. L., & Manic, M. (2021). ResNet Autoencoders for Unsupervised Feature Learning from High-Dimensional Data: Deep Models Resistant to Performance Degradation. IEEE Access, 9, 4051140520 https://doi.org/10.1109/ACCESS.2021.3064819

[15]. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7), 31423155. https://doi.org/10.1109/TIP.2017.2662206

[16]. Ren, C., He, X., Pu, Y., & Nguyen, T. Q. (2021). Learning image profile enhancement and denoising statistics priors for single-image super-resolution. IEEE Transactions on Cybernetics, 51(7), 35353548. https://doi.org/10.1109/TCYB.2019.2933257

[17]. Ge, Y. (2021). Enhancing the X-ray differential phase contrast image quality with deep learning technique. IEEE Transactions on Biomedical Engineering, 68, 12341245. https://doi.org/10.1109/TBME.2020.3011119

[18]. Sun, H., Peng, L., Zhang, H., He, Y., Cao, S., & Lu, L. (2021). Dynamic PET image denoising using deep image prior combined with regularization by denoising. IEEE Access, 9, 4567845690. https://doi.org/10.1109/ACCESS.2021.3069236

[19]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770778. https://doi.org/10.1109/CVPR.2016.90

[20]. Mao, X.-J., Shen, C., & Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in Neural Information Processing Systems, 29, 28642872.

[21]. Kang, E., Chang, W., Yoo, J., & Ye, J. C. (2018). Deep convolutional framelet denoising for low-dose CT via wavelet residual network. IEEE Transactions on Medical Imaging, 37(6), 13451357. https://doi.org/10.1109/TMI.2018.2823756

[22]. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., & Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE Transactions on Medical Imaging, 37(6), 13481357. https://doi.org/10.1109/TMI.2018.2827462

[23]. Park, H. S., Baek, J., You, S. K., Choi, J. K., & Seo, J. K. (2019). Unpaired image denoising

using a generative adversarial network in X-ray CT. IEEE Access, 7, 106803106813. https://doi.org/10.1109/ACCESS.2019.2934178

[24]. Li, Z., Huang, J., Yu, L., Chi, Y., & Jin, M. (2019). Low-dose CT image denoising using cycle-consistent adversarial networks. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 13. https://doi.org/10.1109/NSS/MIC42101.2019.9059965

[25]. Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th International Conference on Machine Learning, 10961103. https://doi.org/10.1145/1390156.1390294

[26]. Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2006). Image denoising with block-matching and 3D filtering. Proceedings of SPIE - The International Society for Optical Engineering, 6064, 606414. https://doi.org/10.1109/TIP.2007.901238

[27]. Kazemi Torbaghan, S. M., Jajarmi, A., Jouybari Moghaddam, Y., & Kazemian Torbaghan, S. F. (2026). A novel exponential tension field for salt-and-pepper image denoising. *Scientific Reports*. https://doi.org/10.1038/s41598-026-58595-8

[28]. Yu, J., Mdletshe, S., Abbasi, H., Kwon, E., Holdsworth, S., & Wang, A. (2026). Deep learning for brain MRI artifact correction: Current challenges and future directions. *Bioengineering, 13*(6), 699. https://doi.org/10.3390/bioengineering13060699

[29]. Ghanim, S. M., Elmogy, S., & Saleh, G. A. (2026). Deep learning-based image denoising can improve the signal-to-noise ratio of the T2-weighted image in brain imaging. *The Egyptian Journal of Radiology and Nuclear Medicine, 57*, 126. https://doi.org/10.1186/s43055-026-01795-z

[30]. Shao, Y., Huang, H., Zhang, L., Li, D., Ding, Z., Wang, F., Chen, S., Lin, S., Gu, Y., Du, M., Li, H., Liang, J., Huang, X., Liu, A., Zhong, J., Zhan, Y., Zhou, X. S., Shi, F., Liao, S., Sun, K., Shen, D., & Qiu, Y. (2026). Real-world unified denoising for multi-organ fast MRI: A large-scale prospective validation. *npj Digital Medicine, 9*, 366. https://doi.org/10.1038/s41746-026-02548-y

[31]. Kollem, S., Peddakrishna, S., Kumar, S., & Baronia, A. (2026). A hybrid GAN-PDE framework with diffusion-guided residual attention for MRI brain image denoising. *Computers in Biology and Medicine, 124*(Part 2), 109243. https://doi.org/10.1016/j.compbiolchem.2026.109243

[32]. Xiao, G., Wang, H., Shen, J., & Chen, Z. (2022). An adaptive hierarchical concatenated network with a robust loss function for image denoising. *Journal of Grid Computing, 20*, Article 9. https://doi.org/10.1007/s10723-022-09601-6

[33]. Pan, S., Abouei, E., Wynne, J., Chang, C. W., Wang, T., Qiu, R. L. J., Li, Y., Peng, J., Roper, J., Patel, P., Yu, D. S., Mao, H., & Yang, X. (2023). 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine & Biology, 68*(10), 105004. https://doi.org/10.1088/1361-6560/acca5c

[34]. Gao, M., Zhang, B., Wang, L., Tang, X., & Huan, H. (2026). Enhancing Noise Robustness in Few-Shot Automatic Modulation Classification via Complex-Valued Autoencoders. *Electronics*, *15*(3), 674. https://doi.org/10.3390/electronics15030674

[35]. Alelyani, M., Alamri, S., Alghamdi, A. J., Alotaibi, S., Faizo, N. L., Alshehri, A., ... & Alonazi, B. (2026). A denoising autoencoder–enhanced ResNet 50 framework for explainable brain tumor classification using MRI. *Frontiers in Oncology*, *16*, 1806663. https://doi.org/10.3389/fonc.2026.1806663

[36]. Jayadeepthi, K., Yakshitha, K., Reddy, K. P. K., Gowtham, K., & Mahesh, K. (2026, June). Deep Learning based Denoising Framework for High-Fidelity Brain MRI Reconstruction. In *2026 4th International Conference on Inventive Computing and Informatics (ICICI)* (pp. 696-702). IEEE. https://doi.org/10.1109/ICICI68773.2026.11580971

[37]. Kumar, R. R., & Priyadarshi, R. (2025). Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches. *Multimedia Tools and Applications*, *84*(12), 10817-10875. https://doi.org/10.1007/s11042-024-19313-6

[38]. Kavitha, B. C., & Shankara, C. (2025, August). A Systematic Review of Deep Learning Approaches for Brain Tumor Identification with Emphasis on MRI Artifact Removal Techniques. In *2025 Third International Conference on Networks, Multimedia and*

Information Technology (NMITCON) (pp. 1-9). IEEE. https://doi.org/10.1109/NMITCON65824.2025.11187845

[39]. Ullah, M. S., Khan, M. A., Almujally, N. A., Alhaisoni, M., Akram, T., & Shabaz, M. (2024). BrainNet: a fusion assisted novel optimal framework of residual blocks and stacked autoencoders for multimodal brain tumor classification. *Scientific Reports*, *14*(1), 5895.https://doi.org/10.1038/s41598-024-56657-3

[40]. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017,doi: 10.1109/TIP.2017.2662206.