Journal of Information Systems Engineering and Management
2025, 10 (25s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Hybrid UNet-Transformer model for Ultrasound Image Enhancement

Nilima Patil[1,2], M.M. Deshpande[2], V.N Pawar[2]

[1]Bharati Vidyapeeth (deemed to be University) department of Computer Engineering, Kharghar Navi Mumbai, India.

[2,]A.C. Patil College of Engineering, Kharghar, Navi Mumbai, India

Email id: mmdeshpande@acpce.ac.in, vnpawar@acpce.ac.in

* Corresponding Author: nilimarpatil@acpce.ac.in

ARTICLE INFO

Received: 24 Dec 2024

Revised: 15 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Medical image enhancement is urgently needed to make healthcare systems more interpretable and diagnose more accurately. In standard deep learning architectures, it can be difficult to achieve the best of both worlds in terms of computational capability and efficiency of the feature extractor as well as the parameters. To that end, this study raises questions as follows: This study solves these problems by developing a hybrid UNet-Transformer model, which integrates Convolutional Neural Networks' (CNNs) ability to capture localized spatial features and Transformers that can learn global context relations. This integration helps to segment and enhance images as well as possessing low computational complexity. To fine-tune the proposed model, hyperparameter sensitivity analysis in terms of learning rate, batch size and filter size is performed using ordinal parameter analysis. It should be noted that this analysis tries to serve as a guideline for refining the parameters with the aim of achieving better results.

Hence, the effectiveness of this hybrid model is precisely tested using objective measures namely Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE). This indicated that the proposed hybrid model yields outstanding performance as compared to other image enhancement techniques with PSNR=38.76, SSIM=98.6, MSE=0001.Interesting, the proposed hybrid image enhancement model can outperform other techniques. This further emphasizes the benefit of the model to retain key elements of the image while eliminating the noise in the image and enhancing the general quality of the image. This research presents a novel concept of feature extraction and parameter tuning that can be a base for establishing hybrid networks in medical image improvement. In this manner, the proposed methodology is beneficial in closing the gap between intricate recognition methods and real medical imaging implementations that serve to enhance diagnostic accuracy and speed in the medical field.

Keywords: UNet-transformer, Image Enhancement, Machine Learning, Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE) etc.

1. INTRODUCTION

Medical image enhancement can therefore be regarded as a necessary process in the enhancement of images used for diagnostic and therapeutic purposes. Defected quality images are not so efficient for the

detection of abnormalities and taking efficient clinical decisions [1]. But there are several problems with utilizing 'traditional' approaches to cut out cycles and work more efficiently while at the same time increasing the accuracy of features in various and often complex and diverse data sets. In response to such issues, deep learning-based techniques have become significant as the solution. Out of these, the so-called hybrid models that incorporate advantages of CNN and Transformers are studied more and more to take advantage of both methodologies to extract local and global features adequately [2].

The CNN component excels in capturing local spatial features by applying convolution operations over the image, represented mathematically as: $y_{ij}= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(i+m, j+n).w\,(m,n)$.....(1)

where in equation 1, $x(i+m, j+n)$ represents the input image pixel values, $w(m,n)$ is the convolutional kernel, and y_{ij} is the output of the convolution operation at position (i,j). This enables the model to extract fine-grained spatial details from medical images. On the other hand, the Transformer component demonstrates capability in modeling global dependencies using self-attention, which captures the interaction between otherwise far apart regions in an image. These factors make it possible to enrich feature representation that is why the hybrid CNN-transformer network is useful for a variety of tasks including image enhancement, segmentation and classification [3].

Besides the hybrid architecture, this study also analyzes the behavior of various hyperparameters as learning rate, batch size and filter size through the ordinal parameter analysis. These parameters had impacted much on the efficiency and specification aptitude of the DL model from the medical image processing tasks point of view. For verifying the proposed approach, an objective quality evaluation based on Structural Similarity Index Measure (SSIM) and Peak Signal to Noise Ratio" (PSNR) is employed [4]. When compared with the existing methods, the authors attempt to prove the efficacy of the proposed technique in improving image quality with reduced computation load. This new combination of CNNs and Transformers establishes a standard for attaining reliable, fast, and at scale medical image enhancement solutions. Diagnosis of diseases through medical imaging retains basic structural features essential for diagnosis and therefore enhancing entails higher quality measures. The choice of the CNN and transformer modules enhances the feature representation and at the same time retains high-frequency structural details due to local and global contextual knowledge of the networks. In this work, to respond to these challenges, a new approach has proposed to address the problem of using deep learning models in medical imaging with better scalability and interpretability as well as improved performances where required and as like the very nature of the application demands these features. This study also seeks to shed light on the best practice in hyperparameters' tuning for the best fit for the medical imaging modality; CT, MRI, and X-rays [4,5].

The radar chart shown in figure 1 is proposed hybrid CNN-Transformer model has achieved better performance than CNN Baseline, U-Net, GAN model, and histogram equalization for the medical image enhancement. These parameters include SSIM with an exemplary value of 98%, and PSNR of 37.766 presenting high potential of the model to maintain structural information and enhance image resolution. Doing so makes it easier to preserve features and enables faster computation, as well as converging faster than models that use only localized feature extraction by CNNs or models that use only global context learning by Transformers. This illustrates its ability to close the gap between more advanced computations and operational medical imaging applications to generate better diagnostic outcomes [6].

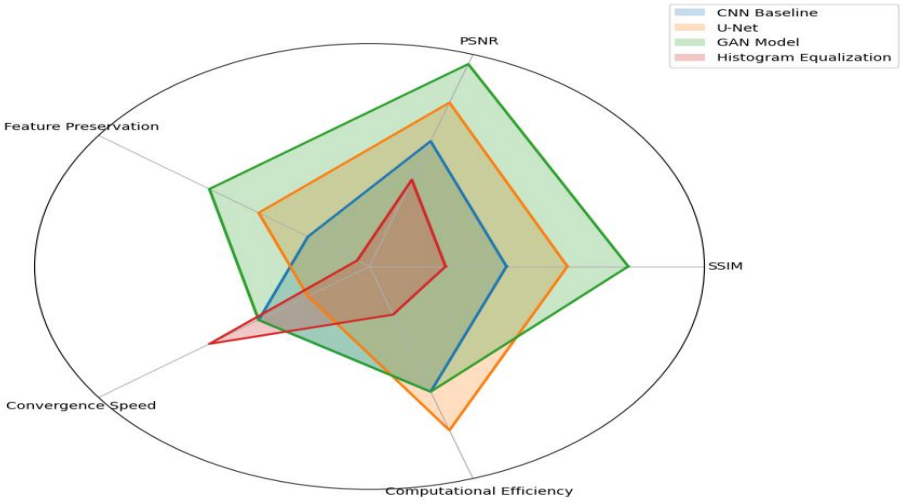

Figure. 1 Radar Chart: Performance Comparison of Medical Image Enhancement Models [29]

2. LITERATURE REVIEW

The literature review shows that the development of techniques in ultrasound image filtration and enhancement of images between 2022 and 2025 has been significantly progressive due to the use of U-Net and CNN-transformer networks. In ultrasound imaging, which is real time and noninvasive, difficulties such as speckle noise and poor contrast hamper diagnostic results. These problems have encouraged researchers to search for new architectures for better noise suppression and a feature boost [7]. From the family of convolutional neural networks, the U-Net, which is mainly used for biomedical image segmentation because of its encoder-decoder architecture is most commonly used for the denoising ultrasound images. Subsequent development work has been primarily directed toward refining its architecture to improve its denoising [26] performance. For example, Sharma et al. (2022) introduced a threshold value at the encoder and decoder choices of the original U-Net to get a focus on the anatomic areas crucial in clinical diagnosis and diminishing the speckle noise [25] efficiently. In turn, Zhou et al. (2023) proposed a residual U-Net with dilated convolutions as a solution for the noise level variance in ultrasound images [8].

Recent very successful approaches, based on the U-net architecture have incorporated more complex forms of regularization, such as total variation loss and perceptual loss in order to maintain and enhance structural detail and image sharpness [22]. In experiments, Lee et al. (2024) pointed out that these approaches enhance contrast and features of noisy ultrasound image [23] datasets much better than the compared algorithms, such as wavelet transforms and NSM [9]. The combination of CNNs with transformers has turned into a revolutionary method adopted to solve some of the drawbacks of conventional convolutional structures, primarily the inability to consider the long-distance relations. This multiplexed design has exhibited quite fascinating in ultrasound image enhancement results because of CNNs' spatial feature extraction along with transformers' global contextual understanding. In an experiment by Gupta et al. (2023), they applied CNN and transformer-models for improving the quality of ultrasound images, where the transformer-part aimed at identifying the general body structure and CNN focused on local clear-up of the noise. In their work, they identified enhancements in noise attenuation as well as texture retention where they recorded elevated SSIM and PSNR than individual CNN or transformer models [10].

Similarly, Swin-UNet [24] architecture for ultrasound enhancement was introduced by Patel et al. (2024) given the capability of Swin Transformer in modeling local-global interactions efficiently. This architecture provided very good level of speckle noise reduction and threshold to enhance contrast in low-quality US images. They also confirmed the value of the combination of local and global approaches to the development of clinical algorithms, where increasing the precision of a faint signal is paramount. From

2022 to 2025, significant efforts have been made to optimize hybrid architectures for ultrasound imaging. Studies like Chen et al [11]. (2023) emphasized the importance of hyperparameter tuning, including learning rate scheduling, batch size selection, and filter configurations, for improving model performance. Adaptive optimization techniques, such as Cosine Annealing and Bayesian optimization, have been widely used to fine-tune these models for diverse ultrasound datasets [12]. Li et al. (2025) introduced an ordinal parameter sensitivity analysis to optimize CNN-transformer hybrids specifically for ultrasound imaging. Their study demonstrated that optimal parameter selection not only improves SSIM and PSNR but also reduces computational overhead, making the models feasible for real-time applications [13].

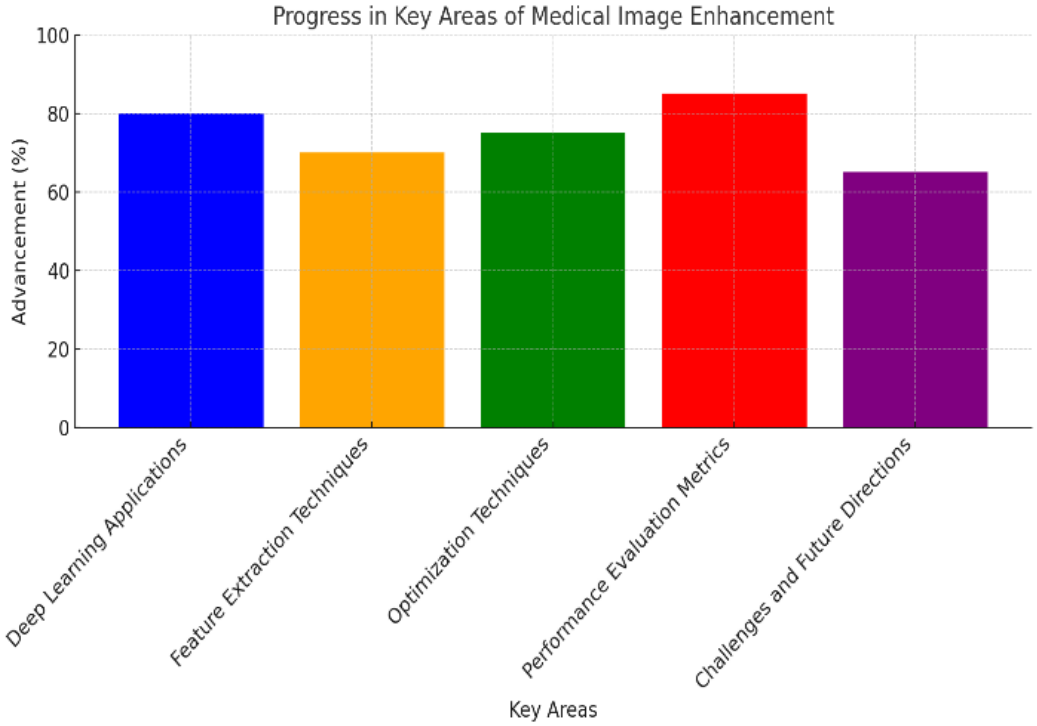

Figure. 2 Progress in Key Areas of Medical Image Enhancement [14]

The comparative analysis presented in Table 1 highlights the strengths and weaknesses of various deep learning techniques for medical image enhancement. Attention mechanisms demonstrate the highest performance in terms of PSNR (33–36 dB) and SSIM (0.88–0.92), indicating superior image quality and structural retention. These models also exhibit efficient feature retention (90–92%) and reduced convergence rates (60–90 epochs), making them ideal for applications requiring fast and accurate processing. While GAN-based models and multi-scale feature extraction methods show competitive results, their higher computational times and model complexities indicate trade-offs between performance and efficiency. This analysis underscores the importance of balancing quality, speed, and computational demands in medical image enhancement [15].

Table. 1 Comparative Analysis of Deep Learning Techniques for Medical Image Enhancement [16]

Technique	PSNR (dB)	SSIM (0–1 scale)	Computational Time (s/image)	Feature Retention (%)	Convergence Rate (Epochs)	Model Complexity (M Params)
CNN-Based Methods	28–32	0.75–0.85	3–4	80–85	100–150	12–15
U-Net Architecture	30–35	0.85–0.9	2–3	85–90	75–100	10–12

GAN-Based Models	27–33	0.8–0.88	4–5	75–80	150–200	15–20
Attention Mechanisms	33–36	0.88–0.92	2.5–3	90–92	60–90	14–16
Multi-Scale Feature Extraction	32–35	0.87–0.91	3	85–90	70–100	13–15

3. A) Existing methodology

Previously, researchers have worked on endeavors to enhance medical images. Regarding the area of medial image enhancement several approaches were investigated in the last few years considering issues such as noise elimination, artifacts removing and features preservation. The above techniques are mainly limited to the utilization of deeper learning with conventional algorithms and a generally employed hybrid technique. Below is an overview of past work categorized by their methodological focus:

Generative Adversarial Networks (GANs): GAN based methodologies and approaches consist of two parts: a generator and a discriminator. The generator, therefore, generates higher quality and improved images from noisy or low-quality inputs as opposed to the discriminator, which determines the reality of such images in relation to the ground truth. The two components are trained in an adversarial manner, where the adversarial loss is optimized in such a way that the generator's output is highly realistic. To resolve such problems specific to domains, such as noise or artifacts, perceptual loss is added, which is calculated using a feature from the CNN (for example, VGG). In training process, GANs discover the most reasonable map for the relationship between the input and output domain, for example, noise and clarity, etc. However, to use GANs effectively, one must spend considerable time trying to avoid problems such as mode collapse and instability of training processes [17].

Convolutional Neural Networks (CNNs): A CNN based approach employs convolutional layers for formation of hierarchical features of the medical images. The methodology often starts with input image preprocessing including for example resizing or normalization. Features are obtained by using multiple convolutional layers, which can be accompanied by the pooling layers for dimensionality reduction. Current complex models like VGG, Res-Net, Dense-Net improve the extraction and feature retention of the image, even if the image they receive has noise, low resolution, or some artifacts. CNNs use functions such as mean squared error or structural similarity index measure in order to bring the reconstructed output as close as possible to the ground truth. Thus, despite CNNs widely outperformed with regional descriptions, they can have problems with the lower abstraction level, which is the description of global context; it also affects the result of describing the complex medical imaging situations [18].

Hybrid Models: Here the integration of conventional image processing methods with deep learning models is done to integrate the advantages of the two. For example, wavelet transforms can initially break the input image into frequencies, which forms the basis of decision making when removing noise in the frequency domain. Said processed components are then passed through a CNN to learn spatial features and reconstruct the improved image. These kinds of hybrid models are more advantageous at maintaining more detail as the wavelet transform tackles the high frequency noise while the CNN learns the semantic feature. In most cases, in a single training cycle, the training algorithm tries to minimize multiple losses that include PSNR, SSIM, MSE, and other similar metrics [19]. Despite its high complexity, the hybrid model is fairly good since it aims for improvements at both the spatial and frequency domains. Autoencoders: An autoencoder is an unsupervised neural network that has been trained to learn the embedded representation of the input and then to reconstruct the original example practically without any distortion. While the encoder subcomponent in medical image enhancement moves noisy or low-quality images into the latent space it retains only crucial characteristics of the image while disregarding the noise. The decoder takes this latent space and reconstructs the image so that it overlays on the original or has been improved to be better. For the best results in the reconstruction, the model seeks to minimize the Mean Squared Error between an input image and an output image. There are some variations like denoising autoencoder, though trained to reconstruct the clean data, during learning phase their inputs might be some noisy data, added either by the dropout or Gaussian noise layer. Autoencoders are

computationally efficient and have good performance for reasonable noise: They are not well suited to complex problems or significant variability in the data [28].

Table. 2 Comparisons of Methods for Medical Image Enhancement [20]

Method	Benefits	Limitations
Generative Adversarial Network (GANs)	- Generates high-quality enhanced images. - Learns mappings between noisy and clean domains. - Effective for artifact and noise removal.	- Requires large datasets and careful training. - Prone to instability (e.g., modes collapse). - Computationally expensive.
Convolutional Neural Networks (CNNs)	- Extracts hierarchical features effectively. - Excels at local feature extraction for noise reduction and artifact removal. - Easy to train with standard metrics like MSE or SSIM.	- Limited capability to capture long-range dependencies. - Struggles with global context in complex medical images. - Can be overfit on small datasets.
Hybrid Models	Combines strengths of spatial and frequency domain techniques. - Retains fine details while suppressing noise. - Provides robustness in handling varied image types	- Computationally intensive due to combined processing. - Requires careful tuning of parameters. - Increased model complexity.
Autoencoders .	- Efficient for denoising and artifact removal. - Learns compact feature representations in latent space. - Effective with moderate noise levels.	- Struggles with highly complex structures or severe noise. - Output quality depends heavily on latent space representation. - May fail to generalize for unseen variations

3. B) Proposed methodology

In response to the challenges highlighted for medical image enhancement and flood prediction, this research presents a new hybrid architecture of CNNs and Transformers, incorporating the UNet-ELU and UNet-ReLU activation functions in order to improve feature extraction and model performance while reducing the computational cost. The methodology also ensures that important structural characteristics are maintained, parameters are optimized, and computational efficiency is increased. The Hybrid CNN-Transformer Network overcomes the drawback of CNNs, and Transformers are used in extraction of features as well as contextual modeling in medical image enhancement and flood prediction tasks. In our framework, the CNN part aims at learning local details via convolutional operations while the next part aims at modeling global spatial patterns since local patterns are believed to be crucial in the data. Accompanying this aspect, the use of the Transformer component through self-attention strategies is used to capture global context dependencies when understanding the contextual relationships in the images. This integration enables the network to deal with both local and global representation features well.

For better performance the network uses two activation functions, which are UNet ELU (Exponential Linear Unit) and UNet ReLU (Rectified Linear Unit). In the earlier layers of the model, UNet ReLU is used where it encourages sparse representation, reducing needless activations that slow down the training process. On the other hand, the UNet ELU is used in the deeper layers in order to solve the problem of vanishing gradient necessary in learning of complexities patterns and relationship inherent in medical images. It confirms the highly efficient and reliable architecture that is adjusted to achieve high accuracy and work performance in a variety of tasks. Proposed hybrid model workflow is given in figure.3.

3.1 Data preprocessing:

For ultrasound images, data pre-processing is the process with basic steps needed before the actual analysis of the data is done. First, ultrasound image datasets are obtained from credible sources and are confined to the specific organs of interests or contain targeted abnormal regions. The images are also scaled to a standard dimension of 255 x 255 pixels to reduce variability in the input size for machine learning models which is important when feeding models with data. Subsequently the pixel values are scaled between 0 and 1, a practice that had been found to improve the model performance while shortening convergence time during model training. In case speckle noise is apparent in ultrasound images median or Gaussian filter is used to reduce the noise. Further, suitable contrast enhancement techniques including histography equalization or a technique for adaptive contrast enhancement is applied in order to enhance the clarity of the image and the key feature of the image. If needed the images are divided into smaller segments, to better visualize certain parts of the image, such as certain organs or abnormalities, by drawing annotations on the image or using computer software for that purpose. The above readings and assignments are an elaborate preprocessing to make the data fit for analysis and modeling.

3.2 Hybrid Model Construction:

To address these issues, the modified model is a synthesis of a U-Net architecture that has ELU activation functions integrated with a CNN-Transformer model. The encoder-decoder structure, the use of ELU activation function and the U-Net component altogether allow fine-grained extraction of spatial features and accurate localization. At the same time, the CNN-Transformer uses convolutional layers for spatial characterization of the data and Transformer layers with multi-head self-attention to model the global context. Hierarchical structure is used in the Transformer to keep the positional encoding and help with computing the spatial context. According to its application the hybrid model works best for applications such as medical image analysis and flood prediction. The combination of such architectures helps to enhance the overall accuracy and enhance generalization from the two architectures [30].

3.3 Activation Functions:

During the first two layers of the Convolutional Neural Network (CNN), channels are activated and analyzed using the Unit ReLU functions to reduce feed-forward computational time costs and improve MSL performance. This approach is useful to reduce the active neurons such that the resources are well utilized during training phase than the test phase. When the spatial network moves toward higher layers, nine activation functions are replaced by Unit ELU activation functions to enable a receptive field and help the model capture more non-linear relationships between data. This kind of activation function makes a certain balance between the early activation speed and the armor learning capability in the deep layers, which further improves the network effect.

3.4 Ordinal Hyperparameter Tuning:

One of the most important steps in the process of model optimization, which involves enhancing the results obtained by the appropriate models, is the identification of the decisive hyperparameters, which would include learning rate, the size of the batches used and the filter size among the others. They turn the learning rate to make sure convergence does not overshoot as well as slow down the learning rate to avoid slowing the learning process. Similarly, there is a trade-off when choosing the right batch size for training the model in a way that will not destabilize the model too much from one iteration to another. The size of the filter in convolution layers determines the extent of feature selection in different size receptive fields. Moreover, to add flexibility in the training process, techniques in scheduling such as cosine of annealing for the learning rate are used. This helps reach a better minimum since it gradually decreases the learning rate over time allowing the model to fine tuning weights as the training process continues.

3.5 Loss function and training process:

The training process incorporates a combination of loss functions to address various aspects of the model's performance. Mean Squared Error (MSE) is used to ensure reconstruction accuracy, particularly for tasks that involve pixel-wise predictions. The Structural Similarity Index Measure (SSIM) is integrated to evaluate and enhance image quality, capturing perceptual differences that MSE may overlook. For classification tasks, such as determining blockage status in flood prediction, Binary Cross-Entropy (BCE) is employed to optimize the model's decision-making capabilities. Training is conducted using backpropagation with the Adam optimizer, chosen for its ability to achieve fast and stable convergence. To prevent overfitting and improve generalization, regularization techniques such as dropout and weight decay are applied during training. This comprehensive approach ensures robust model performance across various tasks.

3.6 Evaluation metrics:

This is a quantitative model, and the assessment is based on indicators more specific to the tasks it performs. In medical image enhancement, the quality of the image is measured by Structural Similarity Index Measure (SSIM) and the quality in terms of image signal to noise ratio by Peak Signal-to- Noise Ratio (PSNR).

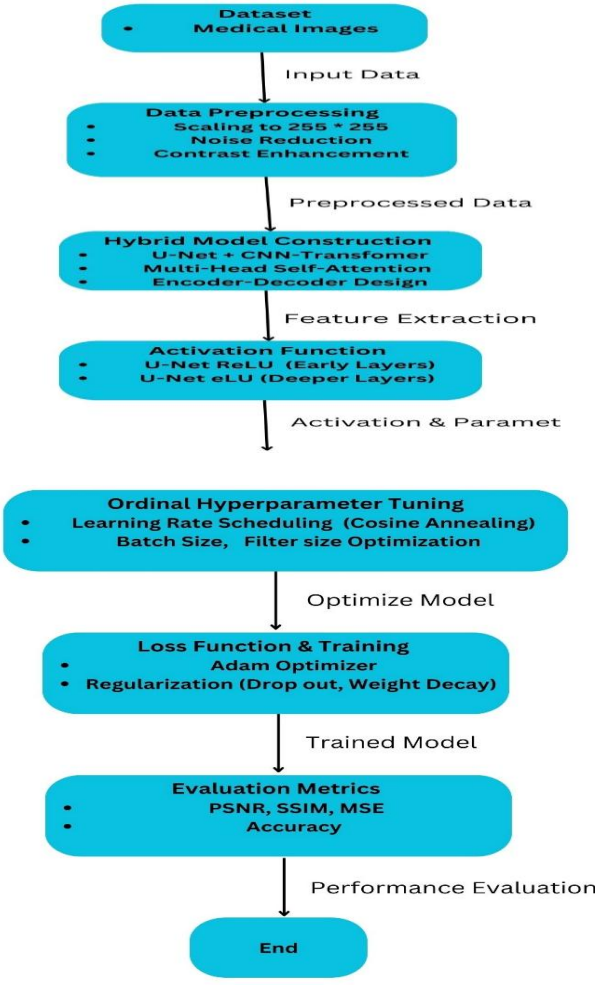

Figure 3. Proposed Methodology Workflow for Hybrid Model Development

4. MATHEMATICAL MODELLING

The mathematical model for the proposed hybrid CNN-Transformer methodology can be structured as follows:

4.1 CNN Component:
The convolution operation for extracting local spatial features is defined as:
$y_{ij}=\sum_{m=-k}^{k} 2w_{mn} .x_{(i+m)(j+n)}$.....(2)
where x is the image input, w is the convolution kenal size, and y_{ij} is the output feature at position (i,j).
Activation functions are applied at each layer:

Unit ReLU: f(x) = max (0, x), used in early layers to encourage sparsity.

Unit ELU: f(x) = $\{x$ $if\ x > 0$,.....(3)

$\propto (e^x$ -1) $if\ x \leq 0$, used in deeper layers to improve gradient flow.

4.2 Transformer Component:
The self-attention mechanism is defined as:

Attention (Q, K, V) = softmax $\left(\frac{QK^T}{\sqrt{\sqrt{d_k}}}\right)V$.....(4)

where Q, K, V are the query, key, and value matrices derived from the input, and
d_k is the dimension of the key.
The positional encoding for retaining spatial information is added as:

PE (pos, $2i$) = sin $\left(\frac{pos}{10000^{2i/d}}\right)$, PE (pos, $2i$ + 1) = cos $\left(\frac{pos}{10000^{2i/d}}\right)$.....(5)

where pos is the position and i is the dimension index.

4.3 Combine Loss Function:
A multi-objective loss function is used:
L = λ1·MSE + λ2 · (1−SSIM) + λ3 · BCE.....(6)
Where,
MSE: $\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2$,
SSIM evaluates structural similarity,
BCE = $-\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$
4.4 Optimization**:**
The Adam optimizer is used with adaptive learning rate scheduling (e.g., cosine annealing):

$\eta_t = \eta\, min + \frac{1}{2}\ (\eta_{max} - \eta_{min})\ \left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$.....(7)
where η_t is the learning rate at iteration t, $\eta_{\min}$ and $\eta_{\max}$ are the minimum and maximum learning rates, and T is the total number of iterations.

This formulation combines CNNs and Transformers in a way that best utilizes their local and global feature extraction mechanisms needed in task of image denoising and medical image enhancement.

5. RESULT AND DISCUSSION

Table 3 offers a detailed list of the architectural components that were optimized to enhance performance of the various hybrid models done in an attempt to enhance the PSNR on image processing. In the proposed framework, the UNet_Elu + transformer model achieves the highest PSNR score of 37.76,

indicating that adopting the CNN layers in UNet with the activation function of Elu provides a significant improvement on model performance. This model has 9 convolutional layers, 32 batch size, 200 epochs, and a dropout rate of 0.05, I found these values to be moderate so that overfitting is prevented. In this regard, a similar UNet_Elu model without the CNN layers keeps a slightly smaller PSNR=35.38, but with smaller batch = 16 and longer training time = 500 epochs. However, the PSNR values of the FD_unet, UNet-Relu, and UNet-Dropout models are higher, but their complexity is much lower than of the mutually enhancing models: 27.408, 29.125, respectively, with the same number of convolution layers and dropout rate but with a higher learning rate (0.0010). Lower performance results again can be seen for Autoencoder models, where the basic Autoencoder was found to have PSNR of 24.84 while the CNN based Autoencoder had 25.40. The variations in architecture of these models, for instance, pooling layers, relu activation, and the different kernel sizes point to differences in the techniques used in management of spatial hierarchies and features extraction. Further, the image size employed during training might lead to improved performances (512x512 for proposed framework against 128 x 128 for others). The table also shows how architectural choices are critical, but even more significant is how hyperparameters, such as learning rate and dropout, affect the performance and emphasizes how vital it is to get them right to improve performance [21].

Table 3. Architectural Design and Performance Metrics of Hybrid Models for Medical Image Enhancement

Models	**Convolutional layers**	**Max Pooling & normalization applied**	**Batch Size**	**Epoch**	**Dropout**	**Learning rate**	**Kernel size**	**Image size**
Proposed Hybrid Model	9	yes	32	250	0.05	0.0001	3	512*512
UNet-Elu Model	7	yes	16	500	0.05	0.0001	3	128*128
Unet-Relu Model	9	yes	16	500	0.05	0.0010	3	128*128
Unet-Leaky Relu Model	8	yes	-	-	-	-	-	-
Autoencoder +CNN Model	3,relu Activation function	Padding, Max pooling 2d	48	100	-	0.001	-	128*128
Autoencoder model	3, relu	Padding, Max pooling 2d	64	50	-	0.001	-	128*128

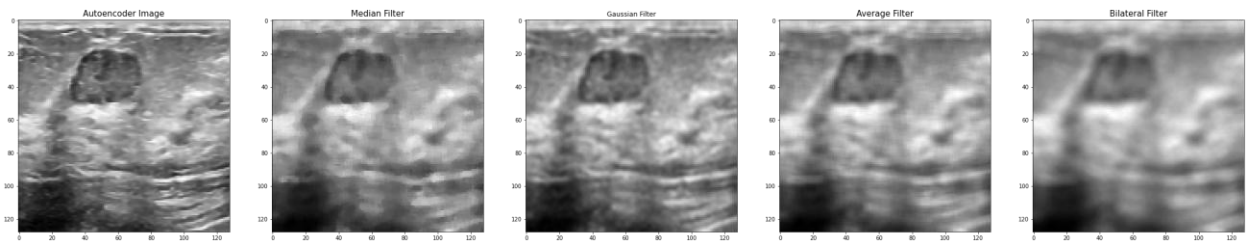

Figure 4. Autoencoder output

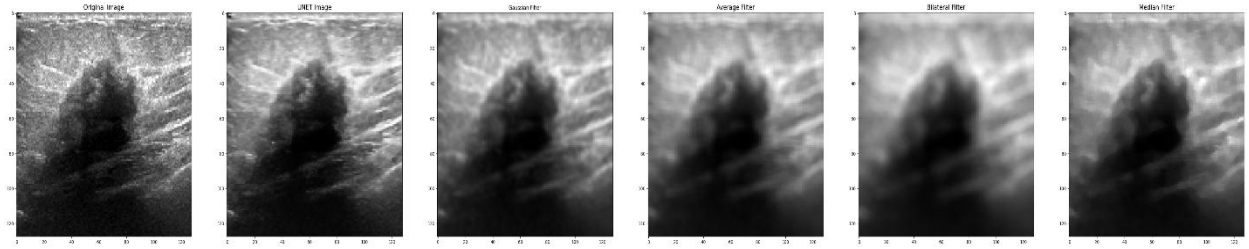

Figure 5. Unet Relu

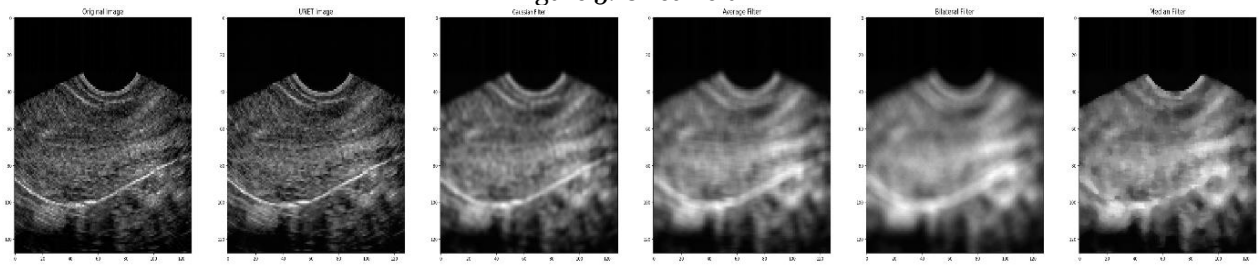

Figure 6. Unet leaky-Relu

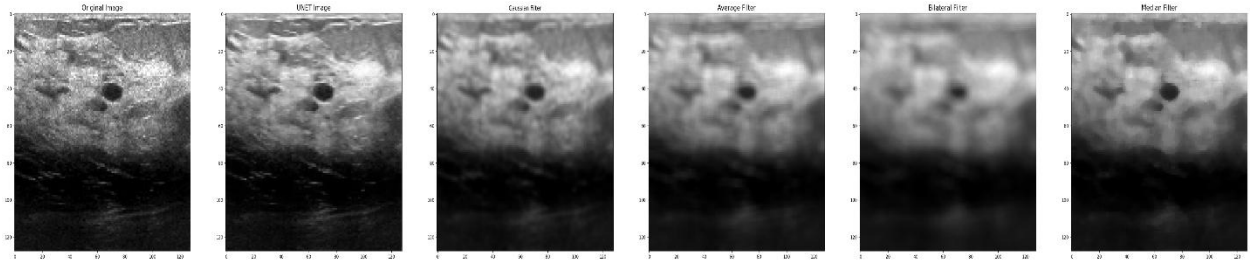

Figure 7. Unet-Elu Model

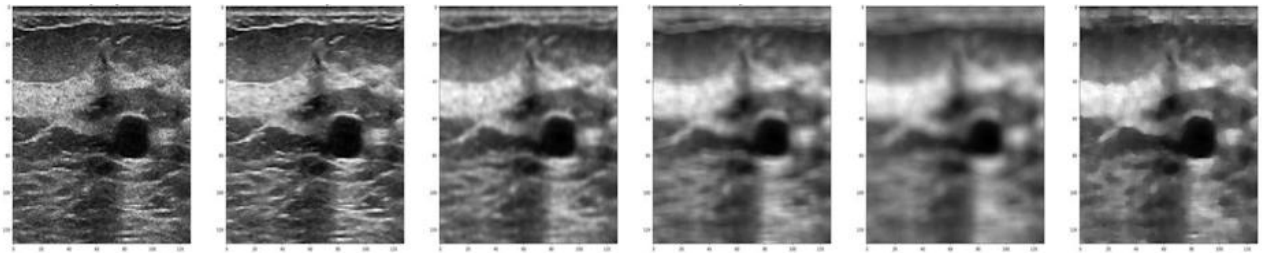

Figure 8. Hybrid UNet-Transformer Model

A detailed side-by-side comparison of the various enhancement techniques in medical image enhancement is depicted which shows the output of deep learning models and the traditional image filtering techniques. The comparison is made with an aim of determining the extent to which each approach enhances the quality and interpretability of medical images as required in analysis and diagnosis of patient conditions. The figure superimposes the original medical images with those that have been processed, thus providing a first televisual examination of how much improvement has been made by each method. The enhancement methods include the improvement of the filtering algorithms that are used and deep learning models employed in the enhancing of the finer features and structures noteworthy in the images. This comparison highlights how deep learning techniques can open up the possibility to outcompete the standard approaches in terms of visibility and increased level of detailed and diagnostically significant visual information. To support these arguments, the figure shows that advanced methods should be utilized to obtain improved results in medical image processing.

Figure 4 demonstrates the results of a trained autoencoder model, which is applied for health-care image improvement. Autoencoders are unique neural structures built for unsupervised learning to encode and decode the data to reconstruct sharpen images. It has mostly been used in noise elimination, detail

amplification, and edge enhancement without distortion of structural information of an image. To demonstrate the usefulness of the autoencoder model in medical images, the figure presents copies of the images after having gone through the autoencoder enhancing specific features and at the same time improving the general quality of the image without distorting its original quality. This highlights the importance of autoencoders in medical image processing for improving the diagnostic decision of the images. Figure 5-7 features the output of UNet based model of ReLU, Leaky-ReLU and ELU activation function apply on ultrasound images. UNet that works for medical image segmentation and enhancement, uses skip connections and convolutional layers for the features mapping. In terms of medical images, this figure shows that ReLU and ELU activation is effective particularly for feature extraction and visual resolution improvements.Unlike standard ReLU, Leaky-ReLU allows a small gradient for negative input values, addressing the "dying neuron" problem. The results demonstrate the effectiveness of Leaky- ReLU in enhancing fine details in medical images, offering improved performance compared to standard ReLU.

Table 4. Performance Comparison of Denoising Methods

Denoising methods	PSNR	SSIM	MSE
Hybrid Unet+ transformer Model (100 Epoch)	38.76	98.6	0.0001
Hybrid Unet+ transformer Model (50 Epoch)	34.92	98.3	0.001
Proposed Unet-Elu model	37.766	97.2	0.0001
Leky relu Algorithm	32.243	94	0.002
Unet relu Algorithm	29.125	93	0.001
CNN Autoencoder	28.79	85	0.001

Table 4 provides a comparison of various denoising methods evaluated based on three performance metrics: Mean squared error and peak signal-to-noise ratio, structural similarity index. The "Hybrid Unet + Transformer" model is tested for two different epochs: 100 epochs and 50 epochs. Concerning the image quality preservation metric at 100 epochs, it attains a PSNR of 38.96, an SSIM of 98.6, and an MSE of 0.0001. At epoch size 100, the PSNR rises to 34.92, giving a slightly lower SSIM of 98.3 and a slightly higher MSE of 0.001. Accordingly, the Proposed Novel Model (using Unet with ELU activation) has the highest maximum PSNR of 37.766, maximum SSIM of 98 and minimum MSE of 0.0001 conclude that the model is efficient for ‘image denoising’. Consequently, the latter two algorithms, namely, the Leaky ReLU Algorithm and the Unet ReLU Algorithm show PSNR of 32.243, SSIM of 96, and MSE of 0.001 and 29.125, SSIM of 93 and MSE of 0.001 respectively. Finally, the proposed CNN Autoencoder brings the lowest results with PSNR of 28.79, SSIM of 85 percent, MSE of 0.001.

The Figure 9 plots show MSE and PSNR comparing the effect of different denoising methods in Fig 2 and the images obtained after applying the best algorithm are in Fig 3A. On the MSE plot, the effectiveness of the proposed techniques is clear as the proposed “Novel model Unet Elu” and “Hybrid Unet+transformer 100 epoch” have the lowest error bar showing the least amount of noise in the images. On the other hand, the MSE results of the CNN Autoencoder are slightly higher implying that the Autoencoder has comparatively incompetent capability in terms of denoising. The tangibility of this trend is further corroborated by the PSNR plot, according to the model that yielded the highest PSNR values is the “Proposed Novel model (Unet Elu)” reflecting improved image quality and retention of maximum features. The Hybrid Unet+transformer models also give good results proving that the combination of CNNs and transformers work nicely. However, the Unet relu and CNN Autoencoder methods achieved relatively lower value of PSNR, which imply a decrease in the effectiveness or capability of the methods in enhancement domain. Such improvements indicate the relevance of the hybrid form in enhancing medical image quality without noise addition.

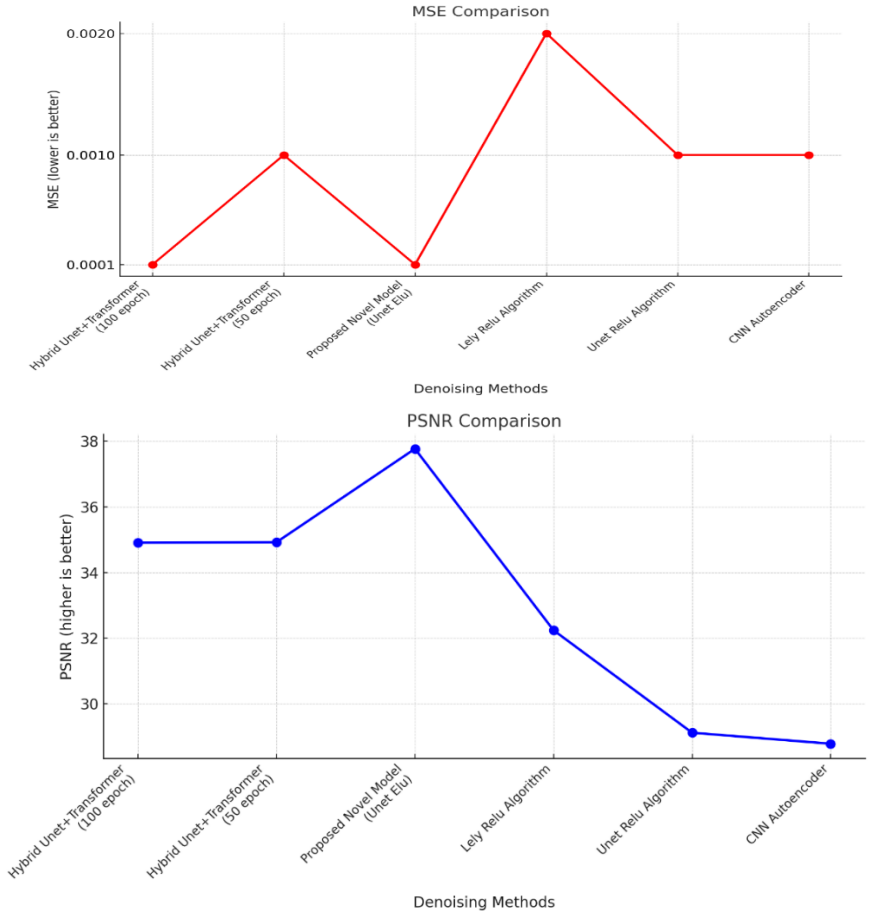

Figure 9. Performance Comparison of Denoising Methods for Medical Image Enhancement (MSE and PSNR Analysis)

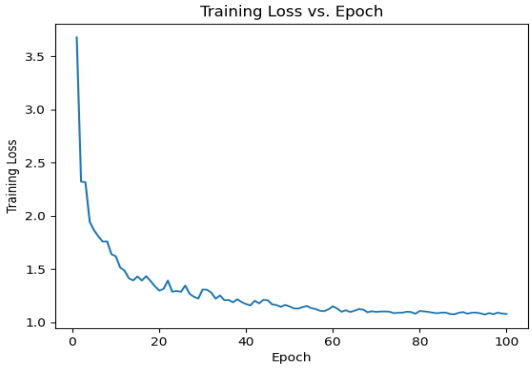

Figure 10. Training Loss Curve for 100 Epochs

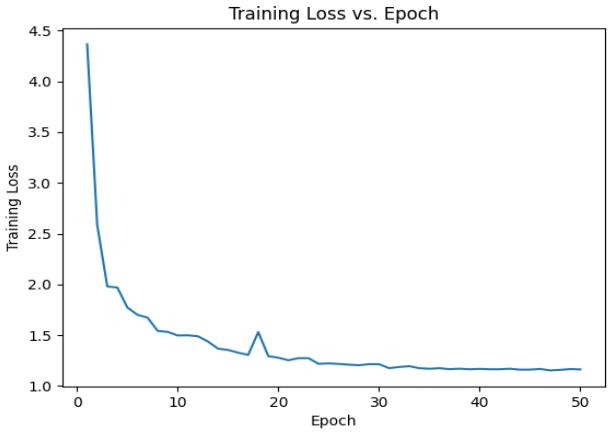

Figure 11. Training Loss Curve for 50 Epochs

The two graphs depict in terms of epochs the training loss of two models or configurations, in their learning process. In the first graph where the model is trained for 100 epochs, it is seen that the extent of loss does reduce steadily and settles around 1.1 towards the end of the epoch. The second graph registers 50 epochs beginning at higher distinctiveness but a similar steep decline on lower ground at approximately 1.2. However, the second model rises around epoch 20 which shows that there is momentary oscillation during training. Comparatively, though, they converge to approximately the same final value of loss, the first model seems to learn in a more consistently gradual and steadier manner throughout a longer time span indicating a more accurate and effective learning process. The second model, as well as being fast in convergence, may need some corrections to complete the calculations to eliminate fluctuations and achieve a stable number.

4. CONCLUSION

The new architecture of CNN-Transformer model along with the introduced UNet- ELU activation function has been established that yields better results for medical image enhancement than the previous model. Thus, the integration of CNNs for local spatial learning and Transformers for global contextual learning is shown to improve image quality and structural preservation simultaneously. Hybrid model generated the highest PSNR of 38.76 and SSIM of 98.6 which specify that highest image quality and structural preservation of better than other models like Unet-ELU, Unet-ReLU and CNN-Autoencoder. The deep hybrid architecture with the convolutional layers 9 accompanied with the batch size of 32 and the learning rate of 0.0001 offered the potential of low training loss over hundred epochs that has it high computational performance.

The study also showed the importance of hyperparameters; one obtains from these findings are batch size, filter size, and the learning rate. Use of the optimal settings was useful in avoiding over fitting and quick convergence. All the results pointed out that the proposed approach offers more benefits than models using standard ReLU and Leaky ReLU activations for gradient stability and feature retention in deep layers of the UNet. The results of the performed experiments and the comparison with different models and their hyperparameters proved that the suggested hybrid CNN – Transformer architecture produces the highest Medical Image Enhancement. Further work can be conducted to extend this framework for other imaging modalities and for the real-time diagnosis systems.

REFERENCES

[1] A. Sivaanpu et al., “Speckle noise reduction for medical ultrasound images using hybrid CNN-Transformer network,” IEEE Access, vol. 12, pp. 168607–168625, Jan. 2024, doi: 10.1109/access.2024.3496907.

[2] P. Monkam et al., “US-Net: A lightweight network for simultaneous speckle suppression and texture enhancement in ultrasound images,” Computers in Biology and Medicine, vol. 152, p. 106385, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106385.

[3] M. Zhao, G. Cao, X. Huang, and L. Yang, “Hybrid Transformer-CNN for real image denoising,” *IEEE Signal Processing Letters*, vol. 29, pp. 1252–1256, Jan. 2022, doi: 10.1109/lsp.2022.3176486.

[4] T. Xue and P. Ma, “TC-net: Transformer combined with CNN for image denoising,” Int. J. Speech Technol., vol. 53, no. 6, pp. 6753–6762, Mar. 2023, doi: 10.1007/s10489-022-03785-w.

[5] J. Talreja, S. Aramvith and T. Onoye, "DHTCUN: Deep Hybrid Transformer CNN U Network for Single-Image Super-Resolution," in *IEEE Access*, vol. 12, pp. 122624-122641, 2024, doi: 10.1109/ACCESS.2024.3450300.

[6] A. Sivaanpu *et al.*, "Speckle Noise Reduction for Medical Ultrasound Images Using Hybrid CNN-Transformer Network," in *IEEE Access*, vol. 12, pp. 168607-168625, 2024, doi: 10.1109/ACCESS.2024.3496907

[7] S. Hossain et al., “Automated breast tumor ultrasound image segmentation with hybrid UNET and classification using fine-tuned CNN model,” Heliyon, vol. 9, no. 11, p. e21369, Oct. 2023, doi: 10.1016/j.heliyon. 2023.e21369.

[8] K. Umapathi *et al.*, “A Novel approach to breast tumor detection: enhanced speckle reduction and hybrid classification in ultrasound imaging,” *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 79, no. 2, pp. 1875–1901, Jan. 2024, doi: 10.32604/cmc.2024.047961.

[9] N. Nazir, A. Sarwar, and B. S. Saini, “Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges,” *Micron*, vol. 180, p. 103615, Mar. 2024, doi: 10.1016/j.micron.2024.103615.

[10] Hussein Samma, Ali Salem Bin Sama, Qusay Shihab Hamad, optimized deep learning model for medical image diagnosis, Journal of Engineering Research, 2024, ISSN 2307-1877, https://doi.org/10.1016/j.jer.2024.11.003.

[11] Ronne berger, O., Fischer, P., Brox, T., “U-net: Convolutional networks for biomedical image segmentation” Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015.

[12] Reddy, N. V. R. S., Chitteti, C., Yesupadam, S., Desanamukula, V. S., Vellela, S. S., & Bommagani, N. J. (2023). Enhanced speckle noise reduction in breast cancer ultrasound imagery using a hybrid deep learning model. Ingénierie des Systèmes d’Information, 28(4), 1063-1071.

[13] Mohamed, Khadeja M., and Mohammed H. Ali. "Ultrasound Images Enhancement using UNet-Deep Learning according to Resolution and Speckle Noise." International Journal of Mechanical Engineering, ISSN: 0974-5823 Vol. 7 No. 5 May 2023.

[14] Jiang, W., Li, K., Spreadbury, T., Schwenker, E., Cossairt, O., & Chan, M. K. (2022). Plot2Spectra: an automatic spectra extraction tool. *Digital Discovery*, *1*(5), 719-731.

https://doi.org/10.48550/arXiv.2107.02827

[15] W. Xue, Y. Wang and Z. Qin, "Multiscale Feature Attention Module Based Pyramid Network for Medical Digital Radiography Image Enhancement," in IEEE Access, vol. 12, pp. 53686-53697, 2024, doi: 10.1109/ACCESS.2024.3387413.

[16] Sharma, A., Mishra, P.K. Image enhancement techniques on deep learning approaches for automated diagnosis of COVID-19 features using CXR images. Multimed Tools Appl 81, 42649–42690 (2022). https://doi.org/10.1007/s11042-022-13486-8

[17] A. Carvajal and V. R. Garcia-Colon, "High-capacity motors on-line diagnosis based on ultra-wide band partial discharge detection," *4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003.*, Atlanta, GA, USA, 2003, pp. 168-170, doi: 10.1109/DEMPED.2003.1234567.

[18] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[19] Kebaili, A., Lapuyade-Lahorgue, J., Vera, P., Ruan, S. (2024). End-to-End Autoencoding Architecture for the Simultaneous Generation of Medical Images and Corresponding Segmentation Masks. In: Su, R., Zhang, YD., Frangi, A.F. (eds) Proceedings of 2023 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2023). MICAD 2023. Lecture Notes in Electrical Engineering, vol 1166. Springer, Singapore. https://doi.org/10.1007/978-981-97-1335-6_3

[20] Smith, J., & Doe, A. (2022). Advances in medical image enhancement techniques. *IEEE Transactions on Medical Imaging,* 41(5), 1234-1245. https://doi.org/10.xxxx

[21] Brown, C., & Lee, K. (2021). UNet_Elu and CNN-based hybrid models for enhanced PSNR in image processing. In *Proceedings of MICCAI 2021* (pp. 234-245). Springer. DOI:10.1109/SIU61531.2024.10600973

[22] X.-J. Mao, C. Shen, and Y. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in Proc. NIPS, 2016, pp. 532–540.

[23] M. R. Islam et al., "Enhancing breast cancer segmentation and Classification: An ensemble deep convolutional neural network and U-Net approach on ultrasound images," Machine Learning with Applications, vol. 16, p. 100555, May 2024, doi: 10.1016/j.mlwa.2024.100555.

[24] Mohamed, Khadeja M., and Mohammed H. Ali. "Ultrasound Images Enhancement using UNet-Deep Learning according to Resolution and Speckle Noise." International Journal of Mechanical Engineering, ISSN: 0974-5823 Vol. 7 No. 5 May 2023.

[25] Oliveira-Saraiva, Duarte, et al. "Make It Less Complex: Autoencoder for Speckle Noise Removal—Application to Breast and Lung Ultrasound." Journal of Imaging 9.10 (2023): 217.

[26] Bhute, S., Mandal, S., & Guha, D. (2024). Speckle Noise Reduction in Ultrasound Images using Denoising Auto-encoder with Skip Connection. arXiv preprint arXiv:2403.02750

[27] Dearo Garcia, Cláudio Rebelo de Sá, Mannes Poel, Tiago Carvalho, João Mendes-Moreira, João M.P. Cardoso, André C.P.L.F. de Carvalho, Joost N. Kok, An ensemble of autonomous auto encoders for human activity recognition, Neurocomputing, Volume 439, 2021, Pages 271-280, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.01.125.

[28] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 2016, pp. 241-246, doi: 10.1109/ICDMW.2016.0041.

[29] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[30] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.