Journal of Systems Engineering and Electronics (ISSN NO: 1671-1793) Volume 35 ISSUE 5 2025

Wavelet-CNN Based Medical Image Fusion: A Hybrid Approach for Enhanced Diagnosis

[1]P.Saravanakumar, [2]P.KiranKumar, [3]K.uma, [4]T.Gopu,[5] S.Oyyathevan
Assistant Professor[1], Professor[2], Associate Professor [3,4,5]
[1,2,3,4,5]Department of Computer Science & Technology
Sasi Institute of Technology and Engineering, Andhrapradesh-534101

ABSTRACT Precise Medical image fusion is vital for improving diagnostic precision and treatment planning. Combining clinical images has been used to extract valuable information from multimodal clinical imaging data. This sophisticated hybrid fusion technique merges Wavelet Transform (DWT) and Convolutional Neural Networks (CNNs) for multimodal medical image fusion. By utilizing the wavelet transform, this method decomposes both CT and MRI images into multiple frequency bands, enhancing image details. The CNN captures high-level features, further strengthening the fusion of these images. This method retains essential information from both modalities and enhances soft tissue details of MRI while preserving dense tissue characteristics of CT. The performance of the fusion is assessed using Root Mean Square Error (RMSE) and Peak Signal-to-Noise Ratio (PSNR), demonstrating the method's ability to enhance diagnostic accuracy, image quality, and overall clinical usefulness. This hybrid strategy holds great promise for medical image analysis, aiding medical professionals in making better and more precise decisions.

INDEX TERMS Medical Image Fusion; Computed Tomography (CT); Magnetic Resonance Imaging (MRI); Wavelet Transform; Root Mean Square Error (RMSE); Peak Signal-to-Noise Ratio (PSNR), Convolutional Neural Networks (CNNs).

I.INTRODUCTION

Brain medical imaging plays a vital role in today's healthcare, offering essential support in diagnosing conditions, planning treatments, and tracking patient progress. Techniques such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Ultrasound each provide valuable insights into the body's structure and function. However, no single imaging method is flawless—each comes with its own strengths and limitations. For instance, while CT scans are excellent for detailed anatomical views, they struggle to differentiate soft tissues. MRI excels in showing soft tissue contrast but is less effective when it comes to bone clarity. Meanwhile, functional imaging tools like PET and Single-Photon Emission Computed Tomography (SPECT) are great for detecting metabolic activity but often lack spatial resolution[1].

Due to these individual drawbacks, combining multiple imaging modalities has become increasingly important. Medical image fusion helps integrate the strengths of different scans into a single, more informative image. This allows healthcare providers to make better, faster, and more accurate clinical decisions. For example, fusing CT's structural clarity with PET's functional insights can help doctors precisely locate tumors and assess their behavior. Similarly, merging MRI with SPECT enhances the ability to detect and interpret neurological conditions by offering both anatomical and functional context. Without such fusion, clinicians would need to interpret separate images independently, which not only consumes more time but also increases the risk of missing critical details. As a result, image fusion has become an indispensable tool for improving diagnostic precision and efficiency in modern medical imaging systems[2].

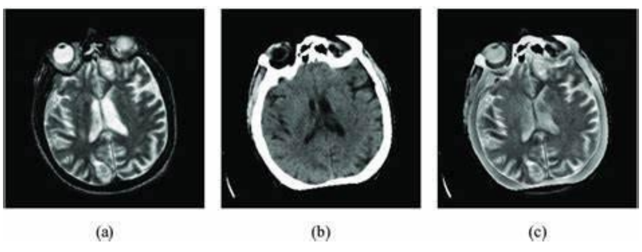

Figure 1: Comparison of MRI (a), CT (b), and Fused Image (c) for Brain Scan

Despite its diagnostic value, medical imaging faces challenges such as varying image quality across modalities. CT offers high spatial detail but limited quality soft tissue contrast, while MRI provides better contrast but slower acquisition. PET shows metabolic activity but lacks spatial precision [5].

Image noise and artifacts from motion, scanner limits, or interference can degrade quality and lead to misdiagnosis. Misalignment between modalities further complicates fusion, demanding advanced registration techniques for accurate alignment [6].

Treatment planning, especially for radiation therapy, requires high precision in tumor localization to minimize damage to surrounding healthy tissues. Image fusion is widely used in pre-surgical assessments and intraoperative guidance, particularly in neurosurgical and orthopedic procedures, where detailed imaging is crucial for making real-time decisions. Furthermore, fused images support AI-based clinical decision-making systems for automated tumor segmentation, organ classification, and anomaly detection. Deep learning-based fusion methods improve the speed and accuracy of real-time medical assessments, enabling more efficient workflows in healthcare settings [3].

Medical image fusion is critical in healthcare for:

Enhancing Diagnosis: Combined CT-MRI images provide both structural and soft tissue details.

Improving Treatment Planning: Helps in radiation therapy by precisely locating tumors.

Reducing Redundancy: Removes unnecessary information while preserving important details.

Increasing Accuracy: Helps radiologists and AI models detect abnormalities with higher precision.

Better Visualization: Provides clearer images for surgical planning and medical research.

Medical image fusion is advancing rapidly, with key roles in diagnostics, treatment, and AI-driven decisions. This study focuses on CT-MRI fusion but offers methods applicable to other multimodal combinations, improving medical data interpretation across fields [3].

The wavelet-CNN-based fusion method presented in this study is developed to effectively combine CT & MRI images, enhancing diagnostic precision. CT imaging is particularly effective for identifying issues in dense structures like bones and tumors, while MRI excels in capturing detailed soft tissue contrasts. By merging the strengths of both modalities, this approach generates a high-quality image that is rich in information, with improved resolution and reduced noise—ultimately supporting more accurate medical assessments and diagnoses [6].

This research reviews spatial and transform domain fusion methods, integrating modern AI techniques. It emphasizes how fusion reduces diagnostic uncertainty, enhances treatment planning, and supports AI-based decision-making in medical imaging [8].

The main driving force of this study is to address the limitations of individual imaging techniques by developing an advanced fusion methodology that enhances visualization, improves diagnostic accuracy, and facilitates AI-driven medical decision-making. The study specifically focuses on wavelet-based fusion techniques and their potential integration with deep learning models, ensuring efficient and accurate image synthesis for clinical applications [11].

Beyond image quality, this research also explores the real-world clinical impact of medical image fusion. In applications such as oncology, neurology, and cardiology, accurate imaging is critical for early detection and intervention. PET-CT fusion is widely used for tumor localization, while MRI-SPECT fusion helps in diagnosing neurodegenerative disorders like Alzheimer's and Parkinson's. This study focuses on advancing fusion methodologies to support the creation of AI-driven diagnostic tools, which can streamline and enhance clinical decision-making in these areas [9].

Furthermore, this work is motivated by the increasing demand for real-time medical imaging solutions, particularly in telemedicine and remote diagnostics. With the rise of cloud-based healthcare systems and AI-powered imaging platforms, the ability to perform efficient and accurate image fusion on-the-fly has become a necessity. The proposed system, if successfully implemented, could play a crucial role in expanding access to high-quality diagnostic imaging in underserved regions, thereby improving global healthcare outcomes [15].

II. MEDICAL IMAGE FUSION

Multimodal medical image fusion technologies and techniques have greatly advanced, contributing to more precise and informed clinical decision-making. Selecting the most effective imaging solution for a particular diagnostic or investigative purpose demands a comprehensive understanding of the anatomy and functionality of the target organs. Since no single imaging technique can capture all the necessary structural and functional information, combining data from multiple modalities becomes vital for ensuring thorough analysis and dependable diagnostic outcomes.

As illustrated in Figure 2, medical image fusion research focuses on three main goals: (a) recognizing, enhancing, and developing imaging technologies that are well-suited for fusion applications, (b) formulating and refining various fusion algorithms and strategies to integrate diverse medical imaging data effectively, and (c) utilizing these fusion methods to examine specific human organs more accurately, especially when assessing diseases or abnormalities. Together, these efforts help enhance diagnostic accuracy by combining the advantages of various imaging tools while reducing their individual shortcomings.

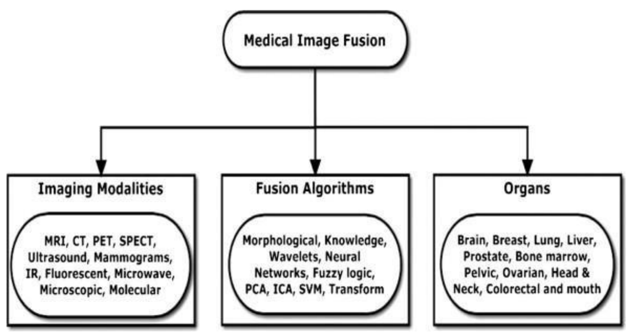

Figure 2: Overview of Medical Image Fusion – Modalities, Algorithms, and Application Areas

III.LITERATURE REVIEW

This section provides an overview of existing research on medical image fusion utilizing advanced wavelet-CNN transform methods. The goal of medical image fusion is to combine images from different modalities—such as CT, MRI, and PET—to improve the accuracy of diagnoses. While conventional approaches have mainly used wavelet-based techniques, newer developments have increasingly adopted deep learning strategies to achieve more effective and precise fusion results.

Y. Jie et al. [1] proposed a Integrated system for combining multiple medical imaging modalities leveraging artificial intelligence techniques to integrate structural and textural information. Their study demonstrated significant improvements in image quality and diagnostic interpretability.

R. Zhu et al. [2] proposed a medical image fusion method for MRI and CT that utilizes a synchronized anisotropic diffusion model. This approach effectively balances the fusion process while improving spatial consistency and preserving important edge details.

Similarly, X. Liang et al. [3] developed MCFNet, a deep learning-based multi-layer concatenation fusion network that optimally retains structural details and contrast.

S. Singh et al. [4] presented a hybrid layer decomposition method incorporating CNN-based feature mapping alongside structural clustering, significantly improving texture retention and edge definition.

L. Wang et al. [5] proposed a wavelet-based fusion technique using Gabor representation, combining multiple CNNs and fuzzy neural networks to enhance image clarity and texture.

Wavelet Transform and Machine Learning Approaches Recent research has incorporated machine learning (ML) and deep learning (DL) techniques with wavelet-based fusion for enhanced accuracy.

Y. Ling et al. [6] introduced MTANet, a multi-task attention-based network for segmenting medical images and enhancing feature focus classification, significantly improving segmentation accuracy and robustness.

X. Li et al. [7] developed a Laplacian redecomposition framework, ensuring enhanced contrast and

structural preservation in MRI and CT fusion.

G. Wang et al. [8] proposed a gradient-enhanced decomposition model, effectively integrating functional (e.g., PET) and anatomical (e.g., MRI) imaging modalities while retaining high gradient fidelity.

Z. Guo et al. [9] explored CNN-based multimodal fusion, improving segmentation accuracy by leveraging hierarchical feature extraction.Deep Learning and Image Fusion Techniques Deep learning Architectural innovations have revolutionized medical image fusion, leading to improved clinical outcomes.

F. Zhao et al. [10] introduced adaptive structure decomposition for CT and MR image fusion, ensuring edge preservation and texture clarity.

Q. Zuo et al. [11] developed DMC-Fusion, a deep multi-cascade framework that progressively refines fusion results, significantly improving structural retention.

Y. Zhao et al. [12] proposed TUFusion, a transformer-based universal fusion algorithm, demonstrating superior spatial consistency and structural retention.

K. Shi et al. [13] introduced a multi-level bidirectional feature interaction network, improving spatial consistency and contrast preservation in multimodal fusion.

S. Iqbal et al. [14] explored deep learning-based feature engineering for multimodal medical image retrieval, integrating CNNs with hybrid feature extraction mechanisms for enhanced precision.

W. Li et al. [15] introduced multi-modal sensor-based fusion using guided image filtering, achieving superior contrast enhancement and noise reduction.

R. Shen et al. [16] proposed a cross-scale coefficient selection framework, optimizing fusion across multiple scales to ensure minimal information loss.

R. R. Nair et al. [17] developed a multi-resolution approach for color medical image fusion, improving visualization in dermatology and histopathology applications.

W. Kong et al. [18] introduced a local difference analysis-based multimodal fusion technique, effectively preserving local structural variations while enhancing spatial coherence.

M. T. Irshad et al. [19] proposed gradient compass-based adaptive fusion, dynamically adjusting the fusion process for enhanced edge preservation and contrast.

Y. Tong et al. [20] developed a multi-focus image fusion algorithm, integrating images captured at different focal depths to improve clarity in ophthalmology and endoscopy applications.

Summary of Methodologies Used

1. **Wavelet-Based Approaches:** DWT, SWT, and NSCT remain widely used for medical image fusion.
2. **Hybrid Wavelet + ML Models:** Integration of CNNs, Autoencoders, and PCA with Wavelet Transform enhances feature extraction.
3. **Deep Learning Models:** GANs, CapsNet, and CNN-based architectures improve accuracy and fusion quality.
4. **Transformer-Based Fusion:** TUFusion leverages long-range dependencies for improved structural retention.
5. **Multi-Stage Feature Refinement:** DMC-Fusion progressively enhances fusion results through a multi-cascade process.
6. **Ensemble Learning:** Combining multiple CNN models, such as EfficientNet and VGG16, further enhances classification performance.
7. **Multi-Focus and Adaptive Fusion:** Techniques like gradient compass-based adaptive fusion and multi-focus fusion improve visualization and contrast.

This literature review highlights the evolution from traditional wavelet-based techniques to hybrid and deep learning-enhanced fusion models. Although wavelet-based approaches remain valuable, incorporating deep learning techniques into medical image fusion has opened new avenues for enhancing diagnostic precision and supporting more informed clinical decisions.Future research could explore more efficient multi-scale fusion techniques and self-supervised learning models to further enhance Medical image analysis

IV. METHODS FOR IMAGE FUSION

Image fusion techniques span a broad spectrum, yet they are mainly classified into two core groups: spatial domain approaches and transform domain approaches. A brief summary of each category is outlined below. These are explained below:

i. Spatial Domain Fusion Techniques

In spatial domain fusion approaches, fusion is carried out by operating directly on the pixel values of the input images. These methods manipulate pixel intensities to achieve the desired fused output. The strategy relies on grey-level transformations, and the enhancement outcome depends on the specific mapping technique applied. While these methods are relatively straightforward, a major drawback is the potential for spatial distortions in the resulting fused image. Often, these techniques sacrifice spectral integrity and introduce unwanted artifacts.
A variety of spatial domain fusion methods exist, including the Averaging Method, Maximum Selection, Minimum Selection, and IHS (Intensity-Hue-Saturation) based methods. These Techniques are typically applied directly to the source images without any transformation, which may result in lower signal-to-noise ratios, particularly when simple averaging is used.

Average Method

This method involves computing the average of the corresponding pixel values from the input images to produce the fused image. It works effectively when the images come from the same type of sensor and have consistent brightness and contrast levels. The mathematical expression is:
It’s described as:

$$F(x, y) = \frac{P(x,y) + Q(x,y)}{2} \qquad (1)$$

Where, F(x, y) is the final fused image, P(x, y) and Q(x, y) are two input images

Select Maximum Method

This approach is Opposite to the maximum method, this technique selects the minimum pixel value between the two input images at each location:. The fused image is formed by retaining the highest intensity value at each pixel position:

$$F(i, j) = \Sigma^{M}{}_{i=1} \Sigma^{N}{}_{j=2} \max A(i, j)\, B(I, j) \qquad (2)$$

In this case, A & B were also used as source images, and F represents the resulting fused image.

Select Minimum Method

In this approach, the fused image is generated by selecting the minimum pixel intensity value from the corresponding positions in the pair of input images.

$$F(i, j) = \Sigma^{M}{}_{i=1} \Sigma^{N}{}_{j=2} \max A(i, j)\, B(i, j) \qquad (3)$$

In this case, A & B were also input images, and F is the fused image.

ii. Transform domain fusion technique

Transform domain techniques involve converting the image into its frequency components before fusion. Instead of working with raw pixel values, these methods operate on transformed coefficients, making them ideal for preserving image data obtained through frequency content. This class of methods typically yields improved outcomes regarding feature preservation and image quality.
Some of the most widely used transform-based fusion methods in medical imaging include:
Discrete Fourier Transform
Discrete Cosine Transform
Discrete Wavelet Transform

Discrete Fourier Transform

Image processing tools play a vital role in decomposing an image into its sine and cosine frequency components through the Fourier Transform. While the input remains in the spatial domain, the output of the transformation provides a frequency domain representation of the image. Each coordinate in this domain corresponds to a specific frequency component derived from the spatial structure of the base image. The Fourier Transform is widely applied in various fields, including image reconstruction, filtering, analysis, and compression. When applied to a 2D image of dimensions N × N, the Discrete Fourier Transform (DFT) is mathematically represented as:

$$F(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j)\, e^{-2\pi\left(\frac{ki}{N} + \frac{lj}{N}\right)} \qquad (4)$$

The equation defines the value at each point F(k,l) as a combination of weighted contributions from the spatial domain, using sine and cosine functions. These functions correspond to different frequency components. Specifically, F(0,0) represents the DC component, which is related to the overall brightness of the image, while F(N−1,N−1) corresponds to the highest frequency in the image. The frequency-domain representation obtained through the Fourier Transform can also be converted back to the spatial domain using the Inverse Fourier Transform, which is mathematically expressed as:

$$f(a, b) = \frac{1}{N2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F(k, l)\, e^{-2\pi\left(\frac{ka}{N} + \frac{lb}{N}\right)} \qquad (5)$$

Notice the factor $1/N^2$, which serves as a normalization constant in the inverse transformation process. To compute the inverse transform, a double summation is required for each spatial point. However, due to the separable nature of the Fourier Transform, it can be computed independently along rows and columns, making it more efficient and easier to implement. This allows the transformation to be expressed in a simplified, step-wise manner as follows:

$$F(k, l) = \frac{1}{N}\sum_{b=0}^{N-1} P(k, b)\, e^{-i2\pi \frac{lb}{N}} \qquad (6)$$

As,

$$P(k, b) = \frac{1}{N}\sum_{a=0}^{N-1} f(a, b)\, e^{-i2\pi \frac{ka}{N}} \qquad (7)$$

To apply the Fourier Transform in two dimensions, the process begins by performing NNN one-dimensional transforms along one axis of the image, producing an intermediate result. This intermediate image is then transformed again using NNN one-dimensional transforms along the other axis, completing the full 2D transformation. This method of using 2N one-dimensional transforms significantly simplifies computation while maintaining accuracy.

The output of the Fourier Transform is typically complex-valued, consisting of both magnitude and phase components. These can be visualized either as separate real and imaginary parts or as magnitude and angle representations. In most image processing applications, only the magnitude component is used, as it contains the essential information about the image's spatial structure and frequency characteristics.

Discrete Cosine Transform

The Discrete Cosine Transform operates through a four-stage coding process. Initially, the visual representation is segmented into smaller blocks, typically of size 8×8 pixels. Each of these blocks is then transformed into the frequency domain using a two-dimensional DCT function. After transformation, the frequency coefficients are quantized, followed by encoding using a lossless entropy algorithm. DCT proves to be highly effective for image compression due to its ability to reduce pixel correlation.

The formula for 1-D DCT is as follows:

$$x_k = \frac{1}{2}(x_0 + (-1)^k x_{N-1}) + \sum_{n=1}^{N-2} x_n \cos\left(\frac{\pi}{N-1} n_K\right) \qquad (8)$$

Where, k = 0,1.................N-1

The equation for 2-DCT:

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] \qquad (9)$$

Here is Figure 3, which provides the basic idea of DCT.

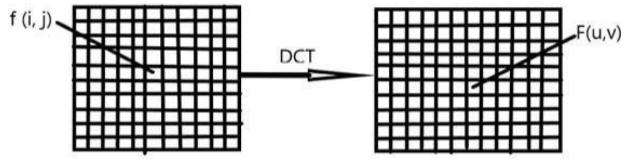

Figure 3: Conversion from Spatial to Frequency Domain

Discrete Wavelet Transform

It is a time domain analysis approach with a fixed window size and convertible forms. In the high frequency section of discrete wavelet transform converted signals, the time differentiated rate is excellent. In addition, the frequency difference rate in the low frequency section is good. It is capable of successfully extracting information from a signal.

Figure 4 depicts a two-channel, one-dimensional filter bank for perfect reconstruction. Filters for low and high analysis are utilized to convolve the input discrete sequence x.

Each one of the pair down sampled signals, xH and sL is transformed according to the algorithm. For well-created filters, x is a signal precisely reconstructed (y=x) .

The schematic diagram for wavelet based fusion techniques is shown in figure 4:

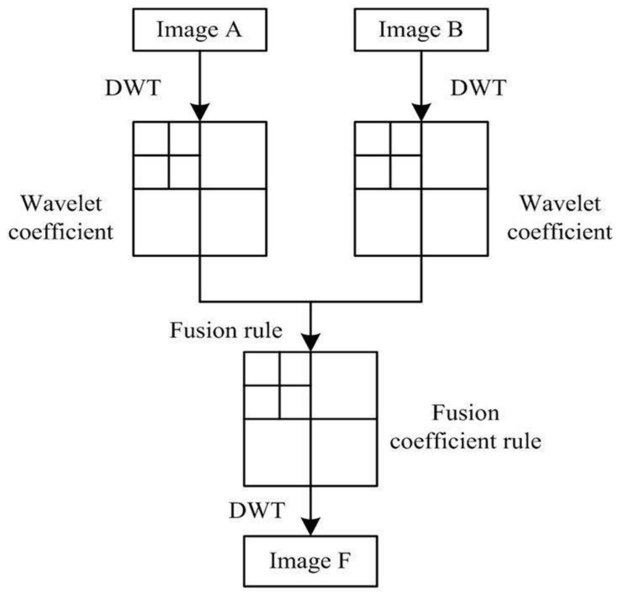

Figure 4: Fusing Multimodal Images with DWT Approach

This diagram represents the Wavelet-CNN Based Image Fusion Process, which integrates two medical images (Image A & Image B) to generate a fused image (Image F). The process follows these steps:

1. **Apply Discrete Wavelet Transform (DWT) to Both Images**
   Image A and Image B undergo DWT, which decomposes them into different frequency sub-bands:

- **Low-Low (LL)**: Approximate details
- **Low-High (LH)**: Horizontal details
- **High-Low (HL)**: Vertical details
- **High-High (HH)**: Diagonal and edge details

2. **Extract Wavelet Coefficients**
   - The wavelet transform produces coefficients that represent different parts of the image.
   - These coefficients contain important structural and texture information from both images.

iii. VGG19 Model

The pretrained VGG19 model is utilized as a key attribute extractor. VGG19 is a Convolutional neural network (CNN) that has were trained on large datasets like ImageNet. It is widely used in image analysis tasks because of its deep layers and ability to capture fine details and advanced pattern structures in images.
The model extracts important features from the high-frequency components of the images (like edges), which are crucial for identifying details in tumor regions.
This architecture represents a typical Convolutional Neural Network (CNN) used for image classification. It begins with an input image of size 32×32×1, representing a grayscale image. The first layer applies convolution using six filters, generating feature maps of size 28×28×6 that capture low-level multi-scale characteristics.These maps are passed through a pooling layer, reducing the spatial dimensions to 14×14×6 to minimize computational complexity and retain important information. A second convolution layer applies 16 filters to extract deeper features, resulting in 10×10×16 feature maps. This is then followed with another pooling layer that it downsamples the data to 5×5×16. The feature maps are then flattened and passed through a series of fully connected layers - first with 120 neurons, then 84, and finally an output layer with 10 neurons. This output layer typically uses soft max activation to classify the input into one of ten possible categories, such as digits or labels.

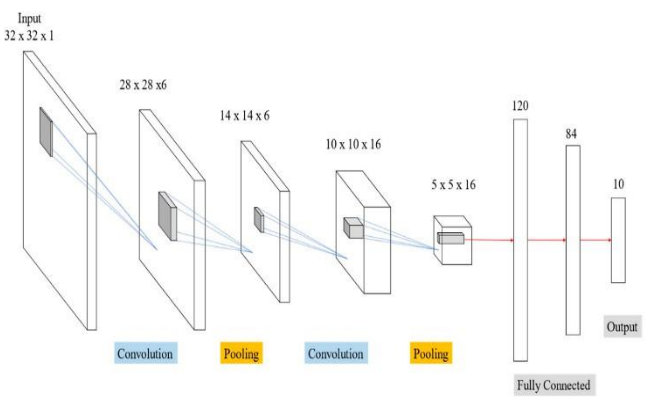

Figure 5: VGG19 Model

V. IMAGE FUSION BASED ON WAVELET-CNN BASED APPROACH

The system architecture is designed to integrate Discrete Wavelet Transform (DWT) and VGG-19-based transfer learning for medical image fusion, combining the strengths of CT and MRI images to enhance diagnostic accuracy. DWT is employed for frequency-based decomposition, while VGG-19, a pre-trained deep learning model, is used for feature extraction and fusion. This combination ensures that the fused image retains both structural details from CT and soft-tissue contrast from MRI, making it more informative for clinical analysis.
The process begins with data acquisition, where MRI and CT images are obtained from medical databases or direct user uploads. These images serve as input for the fusion pipeline. Since medical images often have variations in resolution, noise levels, and intensity distributions, the system applies preprocessing steps such as resizing, noise reduction, and color space conversion to standardize them. Resizing ensures uniform image dimensions, noise reduction eliminates unwanted artifacts, and conversion to the YCbCr color space helps in preserving luminance information, which is crucial for medical imaging.
Next, wavelet decomposition (DWT) is applied separately to both MRI and CT images. This process breaks each image down into four frequency sub-bands:

- LL (Approximation coefficient – Low-frequency details)
- LH (Horizontal detail – High-frequency component)
- LV (Vertical detail – High-frequency component)
- LD (Diagonal detail – High-frequency component)

By decomposing the images into these sub-bands, DWT effectively isolates structural and textural features, making it easier to fuse relevant information from both sources.
For the fusion process, the VGG-19 model—an advanced convolutional neural network pre-trained on the ImageNet dataset—is utilized to extract and integrate features effectively. Instead of traditional pixel or rule-based fusion methods, VGG-19 extracts deep features from each wavelet sub-band and combines corresponding sub-bands from MRI and CT images.

The fusion process ensures that the LL, LH, LV, and LD components capture the most relevant diagnostic details while discarding redundant or less useful information.
Once the fusion process is complete, the system applies Inverse Discrete Wavelet Transform to reconstruct the final fused image. This step merges the fused frequency sub-bands back into a single spatial image, preserving both structural integrity and contrast information. The output is a high-quality fused medical image that provides enhanced visualization of anatomical structures.

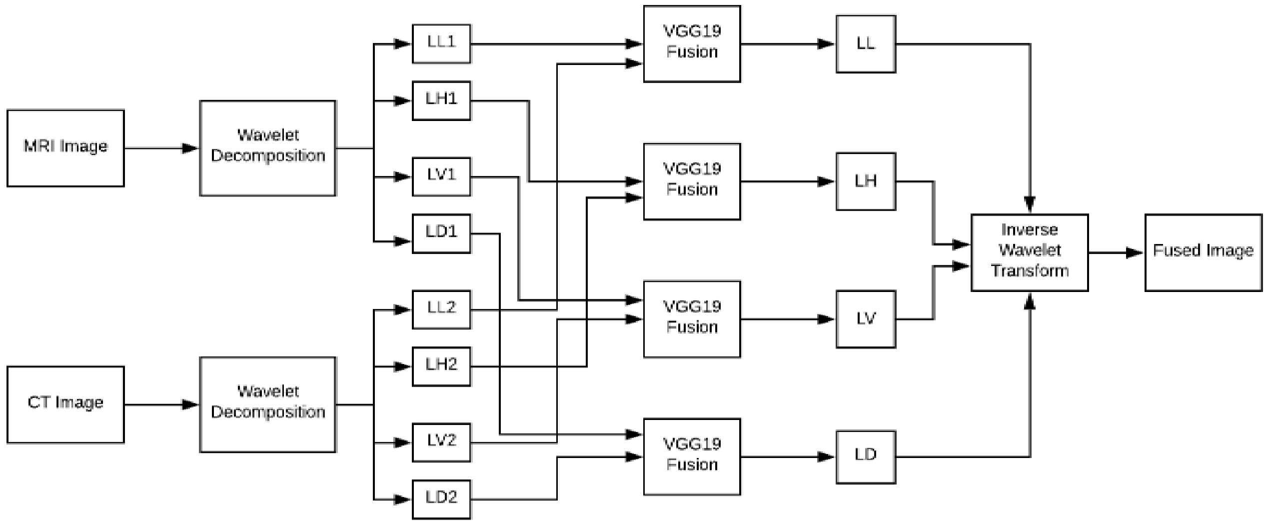

Figure 6: VGG19-Based Hybrid Image Fusion Framework Using Wavelet Decomposition

The algorithm outlined below was designed and applied in Python Programming with several libraries.

STEPS:

i. Begin by importing the CT and MRI scans into the system.
ii. Examine and verify the dimensions of both image inputs.
iii. If a mismatch in image dimensions is detected, generate a prompt to notify the user and stop further execution.
iv. If both images are of equal size, allow the fusion workflow to continue.
v. Inspect the images for inconsistencies or distortions and apply necessary adjustments to ensure accurate spatial alignment between the CT and MRI data.
vi. Perform a Discrete Wavelet Transform (DWT) on each image to decompose them into frequency subbands and obtain their corresponding wavelet coefficients.
vii. Combine the extracted coefficients from both images using a suitable fusion method to integrate information.
viii. Execute the Inverse Discrete Wavelet Transform (IDWT) on the merged coefficients to reconstruct a single composite image.
ix. Input the resulting fused image into a Convolutional Neural Network (CNN) for further analysis and classification.
x. Analyze the CNN's prediction outcome:
xi. If the network categorizes the image as free of abnormalities, label the scan as "normal" and conclude the analysis.
xii. If the network flags anomalies, move on to advanced analysis involving lesion or tumor localization.
xiii. Conduct object detection procedures to isolate and highlight suspicious regions potentially indicating tumor presence
xiv. Quantify the overall performance of the fused output by calculating metrics such as Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE).
xv. Evaluate and contrast the time efficiency of this method against other image fusion techniques to determine computational performance.
xvi. Finally, display a visual representation of the original input images, their decomposed wavelet components, and the resulting fused image for review.

A .Block Diagram for Image Fusion

The proposed system architecture is designed for the fusion and analysis of CT and MRI images, primarily aimed at enhancing medical diagnostics such as tumor detection. The process begins with the loading of CT and MRI images, which serve as complementary modalities—CT images provide high-resolution structural information, while MRI images offer superior soft tissue contrast. After loading, the system performs a check to ensure that both images are of the same size. This is a critical step, as mismatched image sizes can lead to alignment errors during the fusion process. If the images are not the same size, an error message is displayed, prompting the user to correct the inputs. If the images are compatible, the system proceeds to the next stage where it identifies any alignment errors and matches the corresponding anatomical structures using image registration techniques.
Once the images are aligned, the Discrete Wavelet Transform (DWT) is applied to both CT and MRI scans. This mathematical technique decomposes each image into different frequency components, capturing both spatial and frequency information. The system then extracts wavelet coefficients from both

modalities, which represent various image features at different resolutions.

The structural flow of the image fusion technique is demonstrated in Figure 7.

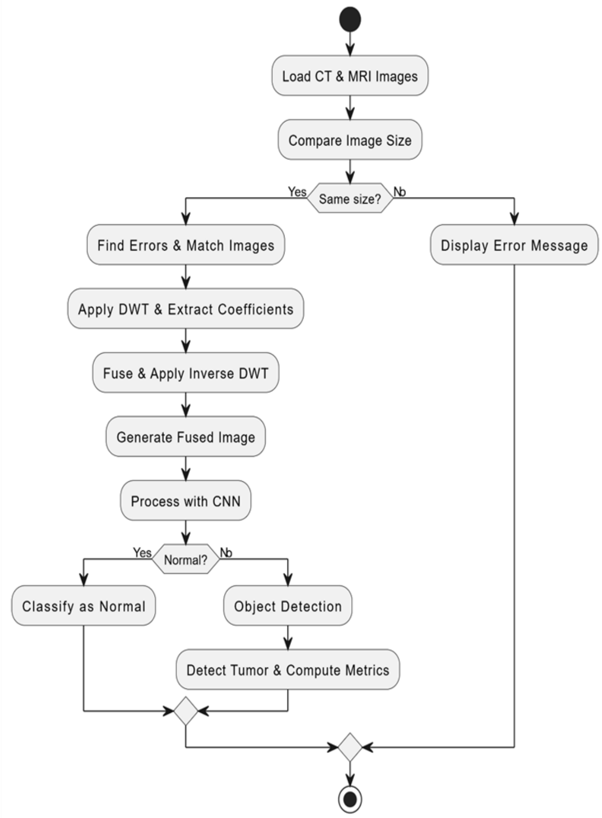

Figure 7: Block Diagram of medical image fusion

These coefficients are fuses using a predefined strategy—often through techniques such as averaging or maximum selection—to retain the most relevant features from both modalities. After fusion, an inverse DWT is performed to reconstruct a single, hybrid image that combines the strengths of both CT and MRI data.

The generated fused image is then processed by a Convolutional Neural Network (CNN), which performs automated feature extraction and classification. The CNN analyzes the fused image to determine whether it appears normal or abnormal. If the image is classified as normal, it is marked accordingly and the process concludes. However, if an abnormality is detected, the system initiates an object detection phase, focusing on identifying the location and characteristics of potential tumors. This includes calculating diagnostic metrics such as the tumor's size, position, and potentially its malignancy score.

This architecture provides a robust pipeline for multimodal medical image fusion and analysis. By merging wavelet transform approaches with modern deep learning strategies, the system enhances diagnostic accuracy while automating key tasks like classification and tumor detection. The integration of decision nodes ensures a logical flow, allowing the system to efficiently handle both normal and abnormal cases.

B.Technologies Used

The implementation of the Wavelet-Based Medical image Fusion System involves a combination of image processing, deep learning, and web deployment technologies.

The backend is constructed using Python-based libraries for handling image processing and fusion, while the frontend interface is built using Streamlit to ensure an interactive and user-friendly experience.

OpenCV

OpenCV is employed to load and preprocess medical images, handling operations like resizing, denoising, and color space conversion to prepare them for the fusion procedure.

NumPy

NumPy supports all numerical operations needed for image processing. It helps in matrix calculations and manipulating pixel data during wavelet transformation.

PyWavelets(pywt)

PyWavelets performs the core wavelet decomposition and reconstruction. It separates images into frequency components and merges coefficients from CT and MRI using fusion rules, followed by reconstruction using IDWT.

Matplotlib

Matplotlib is used to visualize input and output images. It also generates histograms to analyze intensity distributions, helping evaluate the fusion quality.

Streamlit

Streamlit provides the web interface. Users can upload images, trigger the fusion process, and view or download results in real-time. It enables side-by-side image comparison for easy evaluation.

VI. FUSION TECHNIQUES

- **Max Selection**
  - This method creates the output by selecting the maximum value from the corresponding wavelet coefficients of the input images.
  - This approach works on the principle that the most significant features in each image are typically captured by the highest wavelet coefficient values.
  - This method effectively preserves edges and high-frequency details, making it particularly useful for applications requiring enhanced feature retention, such as medical imaging and remote sensing.
  - However, a drawback of max selection fusion is that it may introduce noise, as the highest coefficients may come from images captured at different scales of noise. This can lead to artificial enhancements that do not always represent meaningful information.
- **Averaging**
  - In averaging-based fusion, the wavelet coefficients from both images are averaged to produce the fused coefficients.
  - This method ensures smooth transitions and reduces noise, making it suitable for cases where both images contain similar levels of detail and importance.

- However, one limitation of averaging is that it may lead to a loss of contrast and sharpness in the fused image, as it tends to smooth out details. This can be problematic in applications requiring precise edge detection.

Weighted Fusion

- A weighted combination of the coefficients is used, where different weights are assigned to each image based on their importance.
- The weights can be determined adaptively based on local image features, such as contrast, texture, or entropy, to improve the fusion performance.
- This approach offers greater flexibility in managing the fusion process and can be fine-tuned for particular uses—like improving image contrast while retaining important visual details in medical diagnostics.

VII. IMAGE QUALITY METRICS

The equations used to fuse the images are:

1. **Root Mean Square Error (RMSE)**

$$\text{RMSE}: \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi - yi)^2}$$

Where,

N: Total count of pixels contained within the image.
xi, yi: Pixel values of the original and processed images at position **i**.
Σ (Summation): Adds up the squared differences between corresponding pixel values.
Square Root: Converts squared error into the original unit, making it easier to interpret.

- Measures the difference between the fused image and a reference image.
- A lower RMSE value indicates better fusion quality.
- Does not always align with human perception as it only considers pixel-wise differences.

2. **Peak Signal-to-Noise Ratio (PSNR)**

$$\text{PSNR}: 20log_{10}\left(\frac{MAXI}{\sqrt{MSE}}\right)$$

Where,

MAXI: Maximum possible pixel value in the image (e.g., 255 for 8-bit images).
MSE (Mean Squared Error): Measures the average squared difference between corresponding pixels of two images. Lower MSE means less distortion.

- Assesses the output image's quality by measuring its similarity to a predefined reference image.
- Higher PSNR values indicate better image quality with less distortion.
- Commonly used but may not always reflect perceived image quality accurately.

3. **Structure Similarity Index Measure (SSIM)**

SSIM, or Structural Similarity Index, is a tool that helps us figure out how similar two images are. It measures image quality by comparing a clean, original image to a noisy one, which acts as a distorted version. In this process, the two images are evaluated based on their brightness, contrast, and structure. The Structural Similarity Index (SSIM) between two images, A and B, is calculated using the following formula:

$$\text{SSIM}(x, y) = (2\mu_A\mu_B + C_1)(2\sigma_A\sigma_B + C_2) / (\mu^2_A + \mu^2_B + C_1)(\sigma^2_A + \sigma^2_B + C_2)$$

Where,
The mean values μA and μB depend on the coordinates x and y, while $\sigma^2 A$ and $\sigma^2 B$ represent the variances, and σAB denotes the covariance.

VIII. QUANTITATIVE PERFORMANCE METRICS

Entropy plays a crucial role as a statistical measure to assess the level of randomness or information richness within an image. In the context of medical image fusion, it's particularly valuable for determining how effectively important details from the original images have been preserved. A higher entropy value typically signals a more detailed and informative image, as it suggests a broader range of pixel intensity variations. This is essential when combining data from different imaging techniques—like one that offers detailed anatomical structures and another that highlights soft tissue contrast—ensuring that no critical diagnostic information is lost during the fusion.
To evaluate the performance of the proposed fusion method, several quantitative indicators are considered. Peak Signal-to-Noise Ratio (PSNR) of the output image reaches 31.5 dB, signifying reduced noise and improved clarity when compared to the original images, which recorded 28.3 dB and 27.1 dB respectively. The Structural Similarity Index (SSIM) is 0.92, reflecting a strong preservation of structural details from both input sources—an important factor in maintaining diagnostic reliability.
Moreover, the entropy value of 6.87 highlights that the fused image retains more comprehensive information than either individual input, which had entropy values of 5.34 and 5.12. This confirms the effectiveness of the fusion strategy in conserving critical data. The Edge Preservation Index (EPI) also shows a marked improvement, rising to 0.85 compared to 0.72 and 0.68 from the original images. This indicates that important visual boundaries—such as those around tumors or lesions—remain clear and intact, supporting accurate analysis and clinical decision-making.

Metric	CT Image	MRI Image	Fused Image (Proposed)
Peak Signal-to-Noise Ratio (PSNR) (dB)	28.3	27.1	31.5 (Improved Clarity)
Structural Similarity Index (SSIM)	0.81	0.78	0.92 (Better Structural Retention)
Entropy (Information Content)	5.34	5.12	6.87 (More Image Details Retained)
Edge Preservation Index	0.72	0.68	0.85 (Sharper Edges)

Table -1: Quantitative Performance Metrics

IX. RESULTS

Figures 8 and 9 display the outcomes of the proposed image fusion process applied to medical scans. Earlier research often employed fusion methods based on the Discrete Wavelet Transform (DWT), typically using basic strategies like selecting the maximum, minimum, or average of the wavelet coefficients. In contrast, the objective of this study is to enhance the clarity, completeness, and diagnostic value of the output images by implementing a more selective fusion approach—choosing the sub-band with the most favorable measurement values.

After applying DWT to each input, PSNR and SNR values are evaluated for all sub-bands. The sub-band that exhibits the highest quality metric is selected for inclusion in the final output. Once the optimal sub-bands are chosen, the Inverse Discrete Wavelet Transform (IDWT) is performed to reconstruct the image back into the spatial domain, resulting in a single, enhanced image. As shown in Figure 8, this method ensures that the most informative components from the input images are retained, offering a more accurate and visually rich output for clinical analysis.

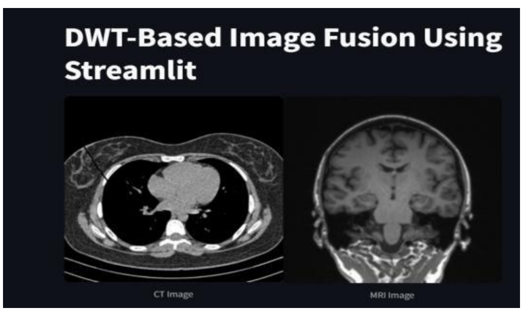

Figure 8: DWT based image fusion

The first image displays two separate medical images:

- **Left Image (CT Scan)**:
  - CT (Computed Tomography) imaging is widely used for capturing detailed anatomical structures, particularly bones and dense tissues.
  - It provides high-resolution images of skeletal structures and helps in detecting fractures, tumors, and internal bleeding.
  - However, it lacks the ability to show soft tissue contrast effectively.
- **Right Image (MRI Scan)**:
  - MRI (Magnetic Resonance Imaging) captures detailed soft tissue structures using strong magnetic fields and radio waves.
  - It provides high contrast between different types of soft tissues, making it useful for diagnosing conditions related to the brain, muscles, ligaments, and nerves.
  - However, MRI scans do not provide clear details of bone structures.

Since CT and MRI scans provide complementary information, they are commonly used together for better medical analysis. The goal of image fusion is to combine their advantages into a single image.

Figure 9: fused image

Figure 9 demonstrates CT-MRI image fusion by selecting sub-bands with higher PSNR to retain key features, followed by inverse wavelet transform to convert the image for clinical use. Figure 10 highlights segmentation of fused images to detect tumors and fractures using methods like thresholding and deep learning. The histogram shows pixel intensity distribution, indicating image quality and detail retention.

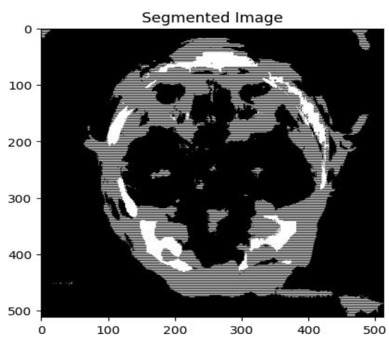

Figure 10: segmented image

X. CONCLUSION AND FUTURE WORK

Medical scan fusion strategies encompass various methodologies, including domain-based approaches— spatial and frequency methods—and learning-based techniques that leverage machine learning and deep learning algorithms. Spatial techniques focus on pixel manipulation, while frequency methods analyze transformed representations for structural integrity. Strategy-based techniques optimize fusion decisions through weighted or decision-based methods. These techniques significantly enhance image quality, providing a clearer and more comprehensive view of a patient's condition, which is crucial for clinical decision-making. The integration of multiple imaging modalities ensures that valuable information is effectively utilized, ultimately supporting accurate diagnoses and personalized treatment plans. This systematic approach to medical imaging fosters improved patient outcomes and facilitates advancements in diverse applications, including tumor detection, treatment planning, and pediatric imaging.

REFERENCES

[1]. Y. Ji, X. Li, M. Wang, F. Zhou, and H. Tan, "An extended difference-of-Gaussians method with edge-preserving strategy for image fusion," *Expert Systems With Applications*, vol. 227, 2023.

[2]. R. Zhu, X. Li, X. Zhang, and M. Ma, "A synchronized-anisotropic diffusion framework for combining CT and MR scans," *IEEE Access*, pp. 91336–91350, 2020.

[3]. X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "MCFNet: A hierarchical concatenation fusion architecture for medical image synthesis," *IEEE Sensors Journal*, pp. 7107–7119, 2019.

[4]. S. Singh and R. S. Anand, "A hybrid decomposition approach using CNN mapping and structural clustering for multimodal medical fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3855–3865, 2020.

[5]. L. Wang, J. Zhang, Y. Liu, J. Mi, and J. Zhang, "A fusion technique using Gabor features and deep fuzzy logic for integrating multimodal medical data," *IEEE Access*, vol. 9, pp. 67634–67647, 2021.

[6]. Y. Ling, Y. Wang, W. Dai, J. Yu, P. Liang, and D. Kong, "MTANet: A deep learning-based attention network for joint image classification and segmentation," unpublished, 2023.

[7]. X. Li, X. Guo, P. Han, X. Wang, H. Li, and T. Luo, "Image fusion through Laplacian-based multilevel decomposition," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6880–6890, 2020.

[8]. F. Zhao, G. Xu, and W. Zhao, "An adaptive decomposition approach for combining medical scans from different imaging methods," *IEEE Access*, vol. 7, pp. 44002–44009, 2019.

[9]. G. Wang, W. Li, X. Gao, B. Xiao, and J. Du, "Combining functional and structural imaging using gradient-enhanced decomposition," *IEEE Trans. Instrum. Meas.*, vol. 71, Art. no. 2508714, 2022.

[10]. Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Fusion strategies for multimodal medical imaging using CNN-based segmentation," in *Proc. IEEE ISBI*, 2018.

[11]. Q. Zuo, J. Zhang, and Y. Yang, "DMC-Fusion: A deep cascade fusion network with classifier-based feature enhancement for multi-modal medical data," *IEEE J. Biomed. Health Inform.*, vol. 25, Sep. 2021, doi: 10.1109/JBHI.2021.3083752.

[12]. Y. Zhao, Q. Zheng, P. Zhu, X. Zhang, and W. Ma, "TUFusion: A transformer-driven universal framework for multimodal medical image integration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, Mar. 2024, doi: 10.1109/TCSVT.2023.3296745.

[13]. K. Shi, A. Liu, J. Zhang, Y. Liu, and X. Chen, "Medical imaging fusion using bidirectional feature interaction across multiple layers," *IEEE Sensors Journal*, vol. 24, Jun. 2024, doi: 10.1109/JSEN.2024.3393619.

[14]. S. Iqbal, A. N. Qureshi, M. Alhussein, I. A. Choudhry, K. Aurangzeb, and T. M. Khan, "Deep learning-based integration of visual and textural cues for medical image retrieval," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3310245.

[15]. W. Li, L. Jia, and J. Du, "Multi-modal sensor data fusion using saliency detection and guided filtering," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2953786.

[16]. R. Shen, I. Cheng, and A. Basu, "Volumetric medical data fusion via cross-scale coefficient selection," *IEEE Trans. Biomed. Eng.*, vol. 60, Apr. 2013, doi: 10.1109/TBME.2012.2211017.

[17]. R. R. Nair and T. Singh, "Color medical image fusion using multiresolution techniques across multiple sensors," in *Proc. 10th Int. Conf. Adv. Comput. (ICoAC)*, 2018, doi: 10.1109/ICoAC44903.2018.8939112.

[18]. W. Kong, Q. Miao, and Y. Lei, "Fusion of sensor-derived medical images based on local variation in a non-subsampled domain," *IEEE Trans. Instrum. Meas.*, 2018, doi: 10.1109/TIM.2018.2865046.

[19]. M. T. Irshad and H. U. Rehman, "Compass-based adaptive medical image fusion for multimodal sensor inputs," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3054843.

[20]. Y. Tong and J. Chen, "A sensor network-based algorithm for multi-focus image integration," *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.286602.