www.nature.com/scientificreports

scientific reports

OPEN

A non-sub-sampled shearlet transform-based deep learning sub band enhancement and fusion method for multi-modal images

Sudhakar Sengan[1]✉, Praveen Gugulothu[2], Roobaea Alroobaea[3], Julian L. Webber[4], Abolfazl Mehbodniya[4] & Amr Yousef[5,6]✉

Multi-Modal Medical Image Fusion (MMMIF) has become increasingly important in clinical applications, as it enables the integration of complementary information from different imaging modalities to support more accurate diagnosis and treatment planning. The primary objective of Medical Image Fusion (MIF) is to generate a fused image that retains the most informative features from the Source Images (SI), thereby enhancing the reliability of clinical decision-making systems. However, due to inherent limitations in individual imaging modalities—such as poor spatial resolution in functional images or low contrast in anatomical scans—fused images can suffer from information degradation or distortion. To address these limitations, this study proposes a novel fusion framework that integrates the Non-Subsampled Shearlet Transform (NSST) with a Convolutional Neural Network (CNN) for effective sub-band enhancement and image reconstruction. Initially, each source image is decomposed into Low-Frequency Coefficients (LFC) and multiple High-Frequency Coefficients (HFC) using NSST. The proposed Concurrent Denoising and Enhancement Network (CDEN) is then applied to these sub-bands to suppress noise and enhance critical structural details. The enhanced LFCs are fused using an AlexNet-based activity-level fusion model, while the enhanced HFCs are combined using a Pulse Coupled Neural Network (PCNN) guided by a Novel Sum-Modified Laplacian (NSML) metric. Finally, the fused image is reconstructed via Inverse-NSST (I-NSST). Experimental results prove that the proposed method outperforms existing fusion algorithms, achieving approximately 16.5% higher performance in terms of the QAB/F (edge preservation) metric, along with strong results across both subjective visual assessments and objective quality indices.

Keywords Image fusion, Convolutional neural network, Medical image processing, Shearlet model, Brain image analysis, Margin information retention

Image Fusion (IF) is vital in several image processing tasks such as medical imaging, study, and remote sensing, where images are captured from multiple sources. The main objective of IF is to combine key data from multiple Source Images (SI) into a single image that contains richer data than any individual SI[1]. These SI may be collected at different epochs or from different sensors. As a result, certain regions may be limited in some images, while others may display different physical features. Therefore, IF is predicted to provide a more accurate view and advantage to human and Computer Vision (CV) applications[2,3]. Suppose are typically classified into three levels: pixel, feature, and decision. According to[4], Feature-Level Fusion (FLF) captures differences in factors such as shape, color, texture, and edges and integrates differences based on feature resemblance. In FLF, each SI contributes distinct features. Decision-Level Fusion (DLF) aggregates high-level results from different

[1]Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamil Nadu 627152, India. [2]Department of Computer Science and Engineering, Siddhartha Institute of Technology & Science (SITS), Narapally, Hyderabad, Telangana 500088, India. [3]Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, 21944 Taif, Saudi Arabia. [4]Department of Electronics and Communication Engineering, Kuwait College of Science and Technology (KCST), Doha Area 13060, Kuwait. [5]Electrical Engineering Department, University of Business and Technology, 23435 Jeddah, Saudi Arabia. [6]Engineering Mathematics Department, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt. ✉email: sudhasengan@gmail.com; a.yousef@ubt.edu.sa

algorithms to complete the fusion task. In DLF, each image is processed separately and then combined during the fusion process.

According to Feature Extraction (FE), DLF separates pixels from multiple source images (SIs) and assigns a proper class label to each pixel[5]. Pixel-Level Fusion (PLF) preserves the spatial features of pixels in the SIs. As a result, several PLF-based Image Fusion Methods (IFM) have been recently introduced. Two main types of PLF are the Transform Domain-Based Image Fusion Method (TDBIFM) and the Spatial Domain-Based Image Fusion Method (SDBIFM), classified based on their operational domain. Both TDBIFM and SDBIFM produce fused images (IF)[6,7]. Demonstrative SDBIFM such as Principal Component Analysis (PCA), maximum selection, and simple averaging[8]. While these methods are simple and computationally efficient, they frequently suffer from spectral distortion and reduced contrast in the fused image. Therefore, pure SDBIFM is now rarely used. Unlike SDBIFM, TDBIFM generally follows a three-phase model. First, the source images are transformed into the frequency domain to generate sub-images. These typically include purpose images, signified by Low-Frequency Coefficients (LFC), and detail images, represented by High-Frequency Coefficients (HFC). Second, specific fusion rules are applied to corresponding sub-bands. Finally, the fused image is restored[9].

Spatial Domain Fusion Methods (SDFM) frequently result in reduced contrast and distortion of spectral features in the fused images[10]. Consequently, the transform domain has become the attention of significant IF research. Among the early methods, the Laplacian pyramid and its deviations have been commonly useful for multi-sensor data fusion[11,12]. However, these methods frequently suffer from blocking objects due to a lack of spatial orientation selectivity during the decomposition process[13]. The Wavelet Transform (WT) has also been used in IF methods, where it effectively preserves spectral features and produces superior fusion results. While WT captures 1D point singularities, it fails to address more complex features, such as curves and line individualities[14].

To overcome this limitation, Multi-Scale Geometric Analysis Methodologies (MAGAM), such as the Shearlet Transform (ST) and Contourlet Transform (CT), have been developed[15]. These methods show strong directional selectivity and anisotropic selection. Among them, the Non-Subsampled Contourlet Transform (NSCT) proposes translation invariance and is effective in mitigating the Gibbs phenomenon experimental in previous transforms. However, NSCT incurs a high computation cost, has a large data footprint, and lacks real-time performance efficiency. In comparison, the Shearlet Transform (ST)[16] achieves better fusion results and computational efficiency due to its more flexible design[17]. The Non-Subsampled Shearlet Transform (NSST) further improves directional selectivity and maintains shift invariance. Building upon these improvements, several Medical Image Fusion (MIF) have recently incorporated NSST[18]. Moreover, the integration of Soft Computing-Based Methods (SCBM) in Multi-Modal Medical Image Fusion (MMMIF) has also gained attention alongside SDBIFM + TDBIFM.

Numerous demonstrative models have been successfully applied in MIF, including Dictionary Learning Models (DLM)[19,20], Compressed Sensing (CS)[21], Total Variation (TV)[22], Fuzzy Theory (FT)[23], Gray Wolf Optimization (GWO)[24], Genetic Algorithm (GA)[25], Guided Filter (GF)[26], Sparse Representation (SR)[27], Pulse Coupled Neural Network (PCNN)[28], methods based on local extrema, structure tensor, and Otsu's[29–31]. These models have each contributed to enhancing the accuracy of IF, noise suppression, and structural preservation in multimodal settings.

Assuming the promising results achieved by TDBIFM and SCBM, recent research has focused on hybrid approaches. For example, the integration of interval Type-2 fuzzy sets with NSST has been proposed for multi-sensor IF[32]. An IF based on Particle Swarm Optimization (PSO) and enhanced Fuzzy Logic (FL) within the NSST domain has also been developed[33]. In another study, NSST was combined with GWO to hypothesize a novel IF[34]. The authors in[35] utilized NSST to decompose the SI, followed by sub-IF using sparse representation and dictionary learning. For MIF specifically, NSST-based methods employing PCNN have also been proposed[26,36]. Additionally,[37] introduced a hybrid IF that leverages the advantages of the Spatial Domain (SD) and the Transform Domain (TD). The overall trends and advancements in practical IF for medical imaging have been reviewed in[38,39].

Deep Learning (DL) has made significant improvements in recent years across a wide range of CV and image processing tasks, including super-resolution, image segmentation, saliency detection, and classification[27]. Since 2012, several Deep Convolutional Neural Networks (DCNN) have been introduced, such as ResNet[40], DenseNet[41], AlexNet[42], VGG[43], ZFNet[44], Fully Convolutional Network (FCN)[45], GoogLeNet[46], and U-Net[47]. These networks have proved impressive performance in image classification, segmentation, object detection, and tracking while also presenting novel visions for IF tasks.
The success of DL over conventional Machine Learning (ML) can be attributed to four key factors:

(a) First, advancements in Neural Networks (NNs) have enabled DL to learn high-level features directly from data, disregarding the requirement for complex feature engineering and domain-specific knowledge. This allows DL to propose end-to-end solutions.
(b) Second, the development of Graphics Processing Units (GPU) and supporting GPU-computing libraries has significantly accelerated training speeds—up to 10–30 times faster than on traditional Central Processing Units (CPU). Additionally, open-source mold provides practical implementations for efficient GPU utilization.
(c) Third, publicly available benchmark datasets, such as ImageNet, have enabled researchers to train and validate DL models using large-scale datasets.
(d) Lastly, the introduction of effective optimization methods—such as batch normalization, dropout, the Adam optimizer, and Rectified Linear Unit (ReLU) activation functions—has improved training stability and performance, further contributing to DL's common success.

A unique Multi-Focus Image Fusion Method (MFIFM) based on CNN was proposed in[48], where CNN was used to classify focused regions and generate a decision map. This decision map was then combined with the SI to complete the IF process. Although this method proved superior performance, its applicability was limited to MFIFM tasks. Subsequently, CNNs were extended to all-purpose MIF with promising outcomes. For example,[49] introduced a CNN-based approach that generated Weight Maps (WM) to fuse pixel activity data from input SIs.

There are numerous types of MIF tasks, including the integration of Magnetic Resonance Imaging (MRI), Computed Tomography (CT), T1- and T2-weighted MRI (based on specific relaxation times), Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT). However, a significant limitation of the CNN-based method is its difficulty in handling complex fusion tasks, such as multi-focus and infrared–visible IF.

To improve classification accuracy,[50] invented a DCNN combined with a Discrete Gravitational Search Algorithm (DGSA). This model was evaluated on 4 distinct datasets (I–IV) consisting of modalities such as CT, SPECT, and MRI. Performance was assessed using metrics including sensitivity, precision, accuracy, specificity, spatial frequency, and fusion factor. Despite its effectiveness, deploying this method in real-time applications remains a significant challenge.

The Pulse Coupled Neural Network (PCNN), a biologically inspired neural model developed in recent years[51], has been effectively applied to several image-processing tasks. In IF, PCNN is often combined with transformation methods to form a hybrid IF. Although these methods yield reliable results, they share common limitations, such as contrast reduction and the occasional loss of subtle yet critical feature data[52,53]. In PCNN-based IF, neurons are predominantly activated by single-pixel coefficients derived from either the SD or TD.

To address these problems, the proposed Medical Image Fusion Method (MIFM) incorporates regional energy in the LFC fusion process using the NSST domain. The fusion of HFC within NSST is achieved using PCNN, where the neuron activation is focused by the Novel Sum Modified Laplacian (NSML)[54] as the input stimulus.

Among the available transforms, the Shearlet Transform (ST) and its hybrid variants have emerged in the literature as strong candidates for solving MIF problems. However, TDBIFM methods may still introduce frequency bias in the fused image. In contrast, DL has proved superior performance in a standard range of ML and signal processing tasks, especially when trained on large datasets. The representational power of Deep Neural Networks (DNN) has driven recent progress in denoising and FE[55], and DL is gradually being accepted to address problems involving unknown or highly nonlinear input–output mappings[56].

Recent advancements in MIF have increasingly employed FL to manage uncertainty and preserve structural details across modalities. In[57], the authors proposed a lightweight IFM that combines Laplacian pyramid decomposition with a novel similarity measure based on intuitionistic Fuzzy Set Theory (FST), achieving superior detail retention with low computational complexity. Extending FL to quality evaluation,[58] introduced a multiscale fuzzy assessment model for multi-focus IF, emphasizing sharpness-aware performance metrics. In[59], FST was combined with compensation dictionary learning to improve cross-modal consistency and enhance edge fidelity in complex anatomical regions. Further,[60] developed an intuitionistic fuzzy cross-correlation mechanism to optimize non-membership relations during fusion, while their subsequent FFSWOAF[61] employed Fermatean fuzzy sets integrated with whale optimization to refine fusion rule selection adaptively. Complementing these algorithmic advances,[62] presented a comprehensive survey of fuzzy-driven fusion methods, underscoring the growing impact of fuzzy set extensions—such as intuitionistic and Fermatean models—in enhancing the interpretability and diagnostic relevance of fused outputs.

While several DL and TD methods have been proposed for MIF, existing methods often employ uniform IFM across all decomposed sub-bands or overlook sub-band-specific restoration, resulting in degraded structural preservation and loss of modality-specific detail. In contrast, this work introduces a novel IFM that integrates the NSST with a deep, sub-band-specific enhancement architecture termed the Concurrent Denoising and Enhancement Network (CDEN). The core novelty lies in applying dedicated learning-based enhancement operations separately to Low-Frequency Sub-Bands (LFSB) and High-Frequency Sub-Bands (HFSB) before fusion, rather than performing direct fusion over unprocessed decomposed coefficients as seen in prior NSSTs such as NSST-SNN and FFST-SR-PCNN.

Specifically, the proposed model employs a dual-branch CDEN, where the DNSN branch suppresses sub-band-specific degradations (*e.g.,* Gaussian or sensor noise), and the ENSN branch enhances discriminative details required for modality complementarity. This network is trained with a composite loss that combines mean squared reconstruction error and Softmax-based cross-entropy to optimize perceptual fidelity and semantic separation jointly. For LFSB, the model introduces a novel activity map generation process using AlexNet's early convolutional layers, effectively guiding the fusion based on salient anatomical networks. Unlike entropy- or gradient-based activity measures, this feature-level weighting captures global semantic relevance with minimal computational overhead.

For HFSB, a biologically inspired Pulse Coupled Neural Network (PCNN) fusion method is deployed, where neuron firing is modulated by a Novel Sum-Modified Laplacian (NSML) to enhance edge saliency and suppress irrelevant high-frequency distortions. This combination ensures that sharp anatomical boundaries are retained while suppressing the fusion of objects. Crucially, the entire enhancement-fusion process is applied within the transform domain, preserving directional and scale-specific data extracted by NSST.

Moreover, the proposed method is the first to apply a joint denoising-enhancement network tailored to frequency-separated sub-bands within an NSST, presenting a principled and learnable alternative to hand-crafted or static fusion rules. Quantitative results across multiple medical image modalities demonstrate significant improvements over recent baselines, with a 9.64% gain in QAB/F and a 3.67% increase in SSIM, affirming the architectural advantage of sub-band restoration before fusion.

This integrated model, comprising NSST-based decomposition, frequency-specific enhancement via CDEN, semantic activity weighting from AlexNet, and biologically grounded PCNN fusion, represents a significant advancement in MMIF by addressing noise resilience and feature complementarity within a unified, end-to-end trainable pipeline.

This research aims to develop an IFM by integrating CNN and the NSST, addressing the significance and complexity of the MMIF problem. Initially, the input MMIF images are decomposed into LFSB and HFSB using NSST. The proposed Concurrent De-noising and Enhancement Network (CDEN-NN) is then applied to these sub-bands to perform noise reduction and enhance frequency detail[35].

The fused image is subsequently reconstructed using the inverse NSST. For fusing the enhanced LFSBs from both images, an AlexNet-based IFM is employed, while the HFSBs are fused using the Novel Sum Modified Laplacian (NSML) and a PCNN-based IFM. A comprehensive comparison was conducted between the proposed method and several *State-Of-The-Art* IFMs. The results validate that the proposed model outperforms existing methods in terms of fusion quality and performance metrics[63].

The remainder of this article is organized as follows: Section "Materials and methods" presents the materials and methods used; Section "Proposed CNN-shearlet fusion model" details the proposed methodology; Section "Experimental analysis" provides the experimental results and analysis; Section "Discussion" discusses the contributions and limitations; and Section "Conclusion and future work" concludes the study.

Materials and methods

Non-subsampled shearlet transform

The composite framework of Wavelet Theory (WT), which integrates multiscale analysis with classical geometric selections, includes the Shearlet Transform (ST) as a significant advancement[64]. The Shearlet transform provides an optimally sparse representation of images with distributed discontinuities and achieves near-optimal performance in Nonlinear Approximation (ONA) tasks[50]. Due to its strong directional sensitivity and localized time–frequency features, the ST has been widely applied in image processing tasks such as texture FE, image denoising, and IF.

The Discrete Shearlet Transform (DST) is constructed within the model of composite wavelet theory, incorporating multiscale decomposition and directional sensitivity. The DST system, as $\mathcal{SH}(\psi)$ is generated from a mother shearlet function $\psi \in L^2\left(\mathbb{R}^2\right)$ as Eq. (1)

$$SH(\psi) = \left\{\psi_{j,l,k} = 2^{\frac{3j}{2}} \psi\left(G^l S^j x - k\right) : j \geq 0, -2^j \leq l \leq 2^j, k \in \mathbb{Z}^2\right\} \tag{1}$$

where: j → The scale index, governing resolution refinement, → the shear index, controlling directional selectivity, $k \in \mathbb{Z}^2$ → the translation index, determining spatial location, ψ → the mother shearlet function, localized in space and frequency.

The matrices S and G are defined as Eq. (2):

$$S = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, G = \begin{pmatrix} 1 & 1 \\ 0 & 1' \end{pmatrix} \tag{2}$$

where, S→ The anisotropic scaling matrix responsible for frequency refinement and redundancy. G→ The shear matrix that introduces directional selectivity by controlling the orientation angle in the transform domain.

Furthermore, distributed discontinuities in the SD are captured using the shift parameter 'k'. When applying the Fourier transform to the SD Shearlet atom $\psi_{j,l,k}(x)$, the following expression is attained Eq. (3):

$$\hat{\psi}_{j,l,k}\left(w\right) = 2^{-\frac{3j}{2}} \psi\left(wS^{-j}G^{-l}\right) e^{2\pi i w S^{-i} G^{-l} k} \tag{3}$$

where: $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ →The frequency vector, $\hat{\psi}$ →The Fourier transform of the mother Shearlet 'ψ', S^{-j}, G^{-l} →The inverse anisotropic scaling and shear matrices, $\langle\cdot,\cdot\rangle$→ The standard inner product.

The frequency support of $\hat{\psi}_{j,l,k}$, which symbolizes its directional localization, is bounded as follows Eq. (4):

$$\begin{aligned} &\text{Supp}\left(\hat{\psi}_{j,l,k}\left(\omega_1, \omega_2\right)\right) \subset \left\{(\omega_1, \omega_2) \in \mathbb{R}^2 : \omega_1 \in \left[-2^{2j-1}, -2^{2j-4}\right]\right. \\ &\left.\cup \left[2^{2j-4}, 2^{2j-1}\right], \left|\frac{\omega_2}{\omega_1} - l2^{-j}\right| \leq 2^{-j}\right\} \; for \; \omega_1 > 0, \omega_2 > 0 \end{aligned} \tag{4}$$

This support region forms a trapezoidal segment in the frequency domain, centered along the slope $l2^{-j}$ with the angle controlled by the shear parameter 'l'. These anisotropic and directional properties enable the shearlet system to represent edges and curves more effectively than traditional transforms, providing a sparse computation of images with distributed singularities (Fig. 1a,b).

Furthermore, distributed discontinuities in the SD are captured using the shift parameter 'k'. When applying the Fourier transform to the SD Shearlet atom as $\psi_{j,l,k}(x)$, Eq. (5):

$$\widehat{\psi}_{j,l,k}(w) = 2^{-\frac{3j}{2}} \psi\left(wS^{-j}G^{-l}\right) e^{2\pi i w S^{-i} G^{-l} k} \tag{5}$$

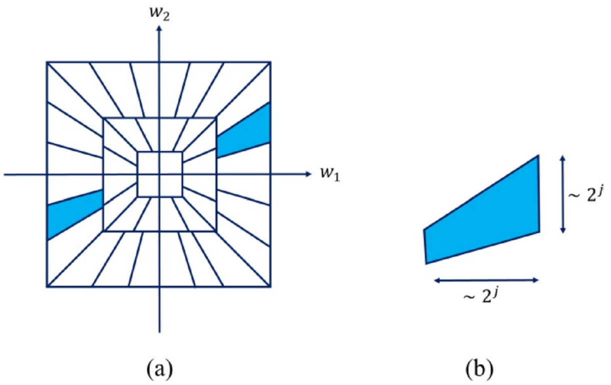

Fig. 1. (**a**) Frequency domain partitioning; (**b**) Shearlet frequency distribution.

where: $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ →The frequency vector, $\widehat{\psi}$ →The Fourier transform of the mother Shearlet ψ, S^{-j}, G^{-l} →The inverse anisotropic scaling and shear matrices, $\langle \cdot, \cdot \rangle$→ The standard inner product.

The frequency support of $\psi_{j,l,k}$, which symbolizes its directional localization, is bounded as follows Eq. (6):

$$\mathrm{Supp}\left(\hat{\psi}_{j,l,k}\left(\omega_1,\omega_2\right)\right) \subset \left\{(\omega_1,\omega_2) \in \mathbb{R}^2 : \omega_1 \in \left[-2^{2j-1}, -2^{2j-4}\right] \cup \left[2^{2j-4}, 2^{2j-1}\right], \left|\frac{\omega_2}{\omega_1} - l2^{-j}\right| \leq 2^{-j}\right\},\ for\ \omega_1 > 0, \omega_2 > 0 \tag{6}$$

This support region forms a trapezoidal segment in the frequency domain, centered along the slope $l2^{-j}$, with the angle controlled by the shear parameter ‘*l*’. These anisotropic and directional features enable the Shearlet to represent edges and curves more effectively than traditional transforms, providing a sparse computation of images with distributed singularities.

As the number of coefficients $N \to \infty$, the asymptotic error of the Shearlet Transform (ST) approaches $N^{-2}(\mathrm{Log}N)^{3}$ [67], enabling highly accurate selection of interferometric borders. In the frequency domain, the Shearlet also forms a Parseval frame[65]. The vital frequency support is defined by trapezoidal regions of size $2^j \times 2^{2j}$, oriented along zero-crossing lines with slopes of -2^{-j}. In the SD, the orientation corresponds to the slopes of 2^{-j}. Shearlet elements can be uniquely distinguished based on their scale, location, and directional orientation. However, the transform exhibits rapid degradation in the spatial domain[66].

The Shearlet Transform is particularly effective in interferogram filtering due to its high directional selectivity. Nevertheless, its implementation involves subsampling operations, which introduce spectral aliasing in the frequency domain and make the transform shift-variant in practice[67]. Directional filtering is achieved via shifted window functions, but the resulting subsampling often leads to reform objects such as Gibbs distortion[68]. To address these limitations, the NSST was developed, inspired by the Non-Subsampled Contourlet Transform. NSST replaces subsampling with convolution-based directional filtering, thereby eliminating spectral aliasing and ensuring shift-invariance. This improvement significantly reduces pseudo-Gibbs phenomena, resulting in more visually intuitive and diagnostically useful fused images. The NSST decomposition process consists of two main stages, as shown in Fig. 2 [69].

Step 1: Multiscale Decomposition	The image is decomposed into HFC and LFC using a Non-Subsampled Pyramid (NSP). Do this step iteratively until the image is decomposed into *j* scales
Step 2: Direction Localization	Non-subsampled shearing Filter Banks (NSSFB), which apply the shearing filter’s 2-D convolution and the HFC on the cartesian domain, are the foundation of direction localization. The NSST is shift-invariant because the convolution operation prevents subsampling

Due to several specific advantages, NSST was selected over alternative multi-scale transforms such as DWT and NSCT. Unlike DWT, NSST provides shift-invariance, which prevents pseudo-Gibbs phenomena at tissue boundaries in medical images. Compared to NSCT, NSST proposals have comparable directional selectivity with lower computational complexity (approximately 40% faster processing time in typical implementations)[70]. Most importantly, NSST achieves superior sparse representation of curvilinear structures (with asymptotic approximation error of O(N^{-2}) (Log N)^3) compared to O(N^{-1}) for wavelets, making it particularly effective for preserving anatomical boundaries in medical images[71]. These advantages make NSST an optimal transform basis for the proposed neural development method.

AlexNet

The 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) marked a significant breakthrough in visual object recognition, as the winning model introduced a deeper and wider CNN compared to the earlier LeNet[72].

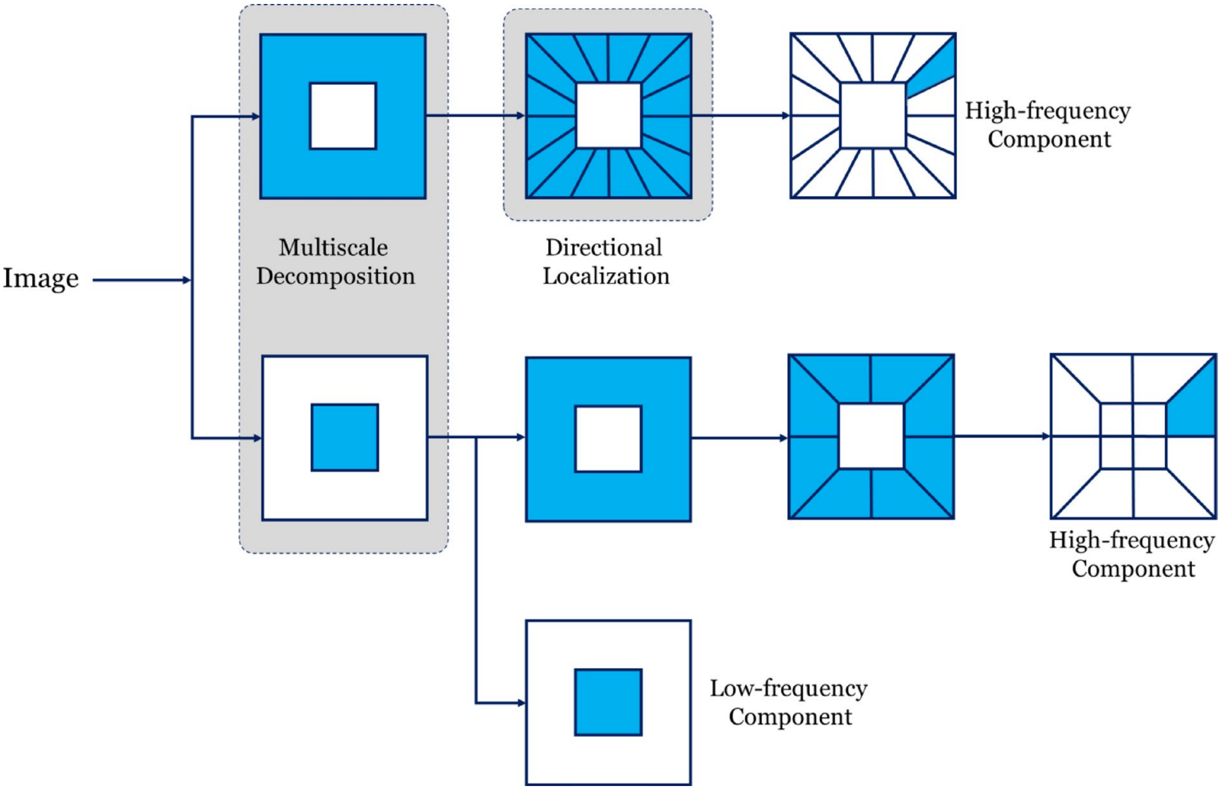

Fig. 2. NSST's Decomposition process.

AlexNet significantly outperformed all traditional ML and CV in the ILSVRC, achieving unprecedented recognition accuracy (Fig. 3). The architecture consists of 8 learned layers, including 5-convolutional layers, 3-max-pooling layers, 2-local Response Normalization (LRN) layers, 2-Fully Connected (FC) layers, and a final SoftMax output layer. In the 1st convolutional layer, 64 large kernels are used to extract low-level features from the input image. This is followed by overlapping max-pooling layers applied after the 2nd and 3rd convolutional layers to reduce spatial dimensions while preserving feature integrity.

The 3rd convolutional layer uses a standard kernel size with 192 filters, followed by the 4th and 5th convolutional layers with 384 filters each, enabling deeper FE. These layers are connected directly without intermediate pooling. Another overlapping max-pooling layer follows the 5th convolutional layer. The output of this pooling layer is passed to the 1st-FC layer, while the 2nd-FC layer connects to a 1000-way SoftMax classifier corresponding to the 1000 class labels in the ILSVRC dataset.

AlexNet was selected for the LFSB fusion task over new models based on several domain-specific considerations[73]. While more recent networks propose superior depth, experimental evaluations presented that AlexNet provides an optimal balance between performance and efficiency for MIF. The key advantage in this context is AlexNet's larger kernels in early layers (11×11, 5×5), which effectively capture the global network data predominant in low-frequency components[74].

The five distinct convolutional layers also provide an ideal multi-level FE for generating the WM labelled in the IF process. The relatively simple model also facilitates faster training and testing, making it more suitable for clinical deployment where computational resources may be limited[75].

Proposed concurrent de-noising and enhancement network (CDEN)

The input to the proposed CDEN is a noisy decomposed frequency as $\tilde{x} = x + n$, where 'x' → the clean signal; 'n' the noise component or signal perturbations introduced during decomposition[76]. It is essential to note that this configuration is employed exclusively during the training phase, enabling the model to learn the effective separation of clean and noisy components.

Targeted enhancement of each frequency sub-band is vital, as it specifies the distinct challenges associated with different spectral components in medical images. The LFSB preserves the primary structural information but frequently suffers from contrast degradation during multiscale decomposition. In contrast, the HFSB contains fine edge and texture features that are vital for clinical interpretation but are more vulnerable to noise contamination and structural distortion[77–79].

De-noising sub-network (DNSN)

The first component of the proposed CDEN-NN is the De-noising Sub-Network (Fig. 4)), which aims to learn a mapping function as $g_{\Theta_1}(\tilde{x}) = \overline{x}$, where $\tilde{x}$ → the noisy input; $\overline{x}$ → the predicted clean signal; Θ_1 → the set of

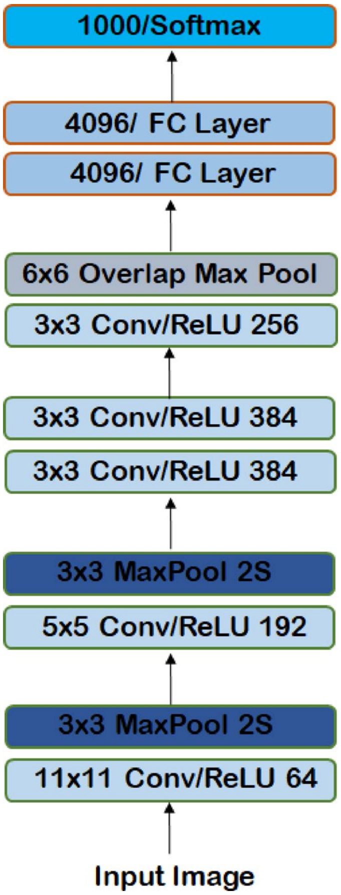

Fig. 3. AlexNet used System.

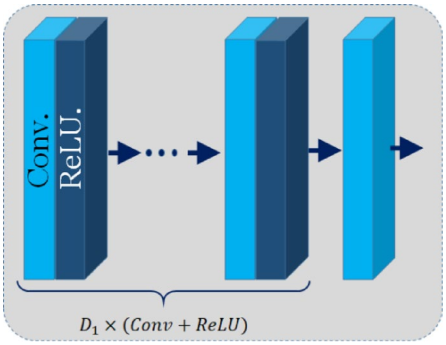

Fig. 4. De-noising subnetwork.

trainable parameters within the DNSN[80]. The training objective is to minimize the discrepancy between the predicted clean signal $'\overline{x}'$ and the ground-truth clean signal 'x', thereby optimizing 'θ_1' to accurately reconstruct the latent clean representation from noisy decomposed sub-band inputs. To achieve this, the network minimizes the Average Mean Squared Error (AMSE) loss defined as Eq. (7)

$$\text{Arg\_Min}_{\Theta_1}\mathcal{L}_1\left(\Theta_1\right) = \frac{1}{2N}\sum\nolimits_{i=1}^{N}\left\|\overline{\mathbf{x}}_i - \mathbf{x}_i\right\|_2^2 = \frac{1}{2N}\left\|\overline{\mathbf{X}} - \mathbf{X}\right\|_F^2 \tag{7}$$

where: Θ_1 →The parameter set of the DNSN, x_i →The ground truth (clean signal) for the i-th input, $\overline{x}_i$ →The network-predicted denoised signal, N →The sum of training samples, $\|\cdot\|_2$ →The Euclidean norm, $\|\cdot\|_F$→ The Frobenius norm.

The Frobenius norm formulation provides a compact method for expressing the overall discrepancy between the predicted and true signal matrices. These matrices are defined as Eq. (8):

$$X = [x_1, x_2, \ldots, x_N], \overline{X} = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N] \tag{8}$$

where, Each Column $\mathrm{x_i} \in \mathbb{R}^{\mathrm{d}}$ → a single input signal sample. d →The feature measurement.

The proposed DNSN comprises a sequence of convolutional layers as D_1 (Conv + ReLU blocks), followed by a final convolutional layer without activation. Each Conv + ReLU block contains 64 filters of size $3 \times 3 \times c$, followed by a Rectified Linear Unit (ReLU) to introduce nonlinearity. In the 1st layer, the number of input channels '*c*' may vary (1, 2, or 3) depending on the number of input polarizations used[81], whereas all subsequent layers are standardized with $c = 64$. Since the target clean signal may include negative values, the final convolutional layer omits the ReLU activation to preserve the signal's full range. Zero padding is applied throughout the subnetwork to ensure that the predicted output as '$\bar{X}$' has the exact spatial dimensions as the ground-truth signal as '*X*'.

EnhanceNet sub-network (ENSN)

The ENSN generates a probability vector $\mathbf{y} = h_{\Theta_2}(\overline{\mathbf{x}})$, from each input, '$\overline{\mathbf{X}}'$. The predicted probability, which indicates that the original signal belongs to the matching class individually, is exposed by[82] each element in '*y*'. '*y*', the output vector is compelled to be near '$\hat{\mathbf{y}}'$, the one-hot vector, which stands for the input signal's real label, to maximize performance[83].

The objective function of the second network, known as the ENSN, is designed to optimize classification accuracy by minimizing the cross-entropy between the predicted and true label distributions Eq. (9):

$$L_2(\Theta_2) = \frac{1}{N}\sum_{i=1}^{N} H(\hat{y}_i, y_i) \tag{9}$$

where Θ_2→ The set of trainable parameters in the ENSN, $\hat{y}_i \in \{0,1\}^C$ →The one-hot ground truth label vector for the i-th sample, $y_i \in [0,1]^C$ →The predicted probability distribution output from the network for the i-th sample, C is the number of classes, $H(\hat{y}_i, y_i)$ →The Cross-Entropy Function (CEF), defined as Eq. (10):

$$H(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_{c=1}^{C} \hat{y}_i \mathrm{Log}(y_i) \tag{10}$$

This loss penalizes deviations between the predicted class probabilities and the ground truth, encouraging the network to produce highly confident and correct classifications.

The training procedure for the feature improvement problem could be split into two parts. The DNSN is individually trained in the first part. For ENSN, the 1st sub-network generates training samples after training. $\overline{\mathbf{X}}$, the hidden new signals are used to train the 2nd sub-network. First, *x* + *n*, a noisy feed, is fed into the DNSN during the test process. The extraction subnetwork is then fed with the predicted new signal $\overline{\mathbf{X}}$ to produce the predicted label. Both sub-networks are trained simultaneously by integrating their Loss Functions (LF)[84].

The total loss for the CDEN is defined as Eq. (11):

$$\mathcal{L}_{\mathrm{CDEN}}(\Theta_1, \Theta_2) = \frac{1}{2N} \parallel \overline{\mathbf{X}} - \mathbf{X} \|_F^2 + \gamma \frac{1}{N} \sum_{i=1}^{N} H(\widehat{\mathbf{y}}, \mathbf{y}) \tag{11}$$

where:$\bar{X}$, X→ The denoised prediction matrix and ground truth matrix respectively, Eq. (8), $\mathrm{H}(\hat{y}_\mathrm{i}, \mathrm{y_i})$ →The cross-entropy loss between the predicted class probability vector and the one-hot ground truth label (Eq. 10), Θ_1, Θ_2→ The trainable parameters of DNSN and ENSN. $\gamma \in \mathbb{R}^+$→'R'egularization coefficient that balances the contribution of the enhancement loss relative '*t*' the denoising loss.

To achieve simultaneous signal denoising and enhancement, this study introduces an enhancement subnetwork, EnhanceNet[56], whose network (Fig. 5). EnhanceNet is primarily composed of a pointwise 1×1 convolutional layer followed by a modified Inception Module (IM) variant. At the base of this network lies a squeeze convolutional block, which includes a 1×1 convolution layer, a Batch Normalization (BN) layer, and a ReLU6 activation function. This module is measured to reduce channel dimensionality and efficiently extract essential features. The modified IM implementation supports only asymmetric kernels of sizes $1 \times 1, 1 \times 2,$ and 1×3, which significantly reduces computational overhead between successive convolutional layers. EnhanceNet further integrates two convolution methods[86]: (1) a standard convolution block composed of a 1×3 Conv2D layer, followed by BN and ReLU6 activation, and (2) a depthwise separable convolution block that includes a 1×3 depthwise Conv2D layer, BN, and ReLU6. These components collectively improve FE while maintaining computational efficiency.

Following the depthwise convolution block, a channel attention mechanism is introduced to the network to enhance the extraction of salient feature data across channels. This module is fundamental as the output features at this stage consist of multiple channels with enriched contextual information. Drawing inspiration from the inverted residual network[87], the attention block is strategically placed after the 1×3 depthwise convolutional layer to recalibrate channel-wise responses selectively. The input to the channel attention module is the feature map generated by the second 1×3 depthwise convolution layer, enabling the model to emphasize informative channels while suppressing less relevant ones[88].

The Global Average Pooling Operation (GAPO) is employed in the channel attention module in Fig. 5 for compressing the FM with the size H × W in each channel into 11.

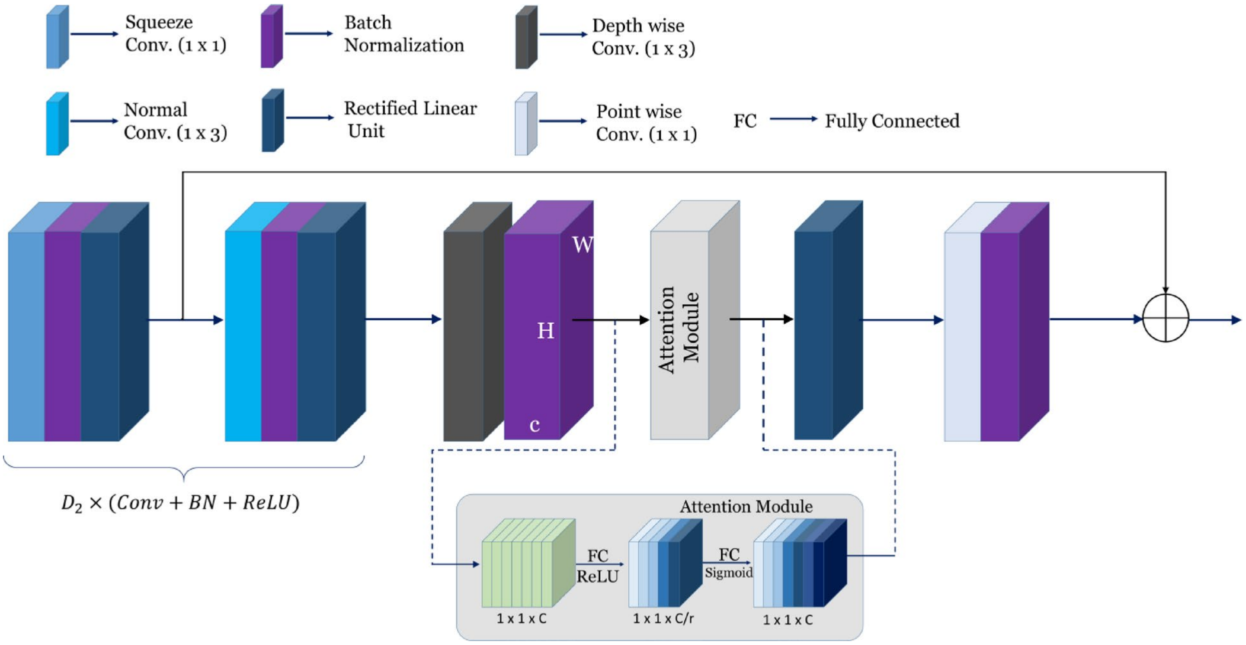

Fig. 5. EnhanceNet Model.

The GAPO's output represents the corresponding channel's global data. The output of the GAPO for the c^{th} channel is computed as follows Eq. (12):

$$x_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} y_c(i,j) \tag{12}$$

where: $y_c(i,j)$ →The activation value at spatial position, (i,j) →The feature map (FM) corresponding to the c^{th} channel, H, W→ The height and width of the feature map.

This operation condenses spatial data from each feature map into a single scalar value, ‘x_c’, representing the global descriptor for the c^{th} channel. The resulting vector $\mathbf{x} = [x_1, x_2, \ldots, x_c]$ is then passed through two FC layers. The 1st FC layer reduces the dimensionality from ‘*c*’ to ‘*c/r*’, where ‘*r*’ is a channel reduction ratio (tunable hyperparameter) to limit overfitting and computational complexity. This compressed extraction is subsequently activated by a ReLU function and passed to a 2nd FC layer, followed by a sigmoid activation to generate the channel attention weights.

The ReLU-AF covers the first FC layer [89], while the sigmoid AF covers the final one. The channel attention module processes the global descriptor vector $\mathbf{x} \in \mathbb{R}^c$ (attained via GAPO) using two FC layers, a ReLU activation, and a sigmoid function to generate a channel-wise attention weight vector $\mathbf{z} \in [0{,}1]^c$, as follows Eq. (13):

$$\mathbf{z} = \sigma\left(\mathbf{W}_2 \cdot \delta\left(\mathbf{W}_1 \cdot \mathbf{x}\right)\right) \tag{13}$$

where: $\mathbf{W}_1 \in \mathbb{R}^{\frac{c}{r} \times c}$→ The weight matrix of the 1st FC layer, $\delta(\cdot)$ →The ReLU activation function, $\mathbf{W}_2 \in \mathbb{R}^{c \times \frac{c}{r}}$ →The weight matrix of the 2nd FC layer, $\sigma(\cdot)$ →the sigmoid activation function that outputs the attention weights $z_c \in (0{,}1)$ for each channel ‘*c*’.

Each attention weight as ‘z_c’, then used to scale the corresponding feature map ‘y_c’ from the previous convolutional layer, resulting in the reweighted output Eq. (14):

$$o_c = z_c \cdot y_c \tag{14}$$

where: y_c→The feature map of the c^{th} channel, z_c →The scalar attention weight for that channel, o_c →The reweighted feature map emphasizes more informative channels.

This mechanism allows the model to selectively enhance discriminative feature maps while suppressing less relevant ones, thereby improving representation quality for downstream classification tasks.

The channel attention mechanism further improves FE by amplifying the channels that carry more discriminative information. It is followed by a pointwise convolutional block comprising a 1× 3 Conv2D layer and a Batch Normalization (BN) layer, with the output passed by a ReLU6 activation function ‘o_c’[90]. To preserve gradient flow and support deeper network training, EnhanceNet includes an optional residual connection, which helps mitigate the vanishing gradient problem in deep architectures[91]. EnhanceNet is designed to be parameter-efficient by employing squeezed convolutions and depthwise separable convolutions, which significantly reduce the model's computational complexity. Additionally, a 1 × 1 standard convolutional layer is employed to refine feature map clusters and enhance local FS. By delaying spatial down-sampling, the network maintains large activation maps that contribute to improved representational accuracy. The use of heterogeneous kernel sizes further enables multi-scale FE, enhancing the model's capacity for robust and fine-grained feature learning.

LFC fusion method using AlexNet

LFSB and HFSB components primarily differ in content and fusion challenges, necessitating specialized mechanisms for addressing these differences. LFCs primarily contain structural and intensity data with higher Signal-to-Noise Ratios (SNR), making them suitable for deep FE by convolutional models[92], such as AlexNet, which captures multi-scale structural relationships. Conversely, HFCs contain edge and texture details with lower SNR, making them particularly susceptible to noise amplification during fusion[93]. The PCN with NSML input is designed to address this challenge through its bioinspired temporal linking mechanism, which effectively decides between meaningful edges and noise in HFCs. This specialized treatment of each component type enables optimal balance between structural integrity and detail preservation, which is particularly crucial for multimodal medical images where complementary data appear in different frequency bands[94].

In general, the LFC of a source image retains the principal structural components, while the HFC preserves finer details such as edges and textures. The traditional IMF frequently applies simple weighted averaging or maximum-value selection methods to the LFC, which neglects the contextual relationships among pixels and may result in suboptimal integration. To address these limitations, this study employs AlexNet for multi-layer FE from the source images. Subsequently, an Adaptive Selection Algorithm (ASA) is employed to generate optimized Weight Maps (WM), resulting in more effective and context-aware fusion compared to classical IFM[95].

The LFC fusion process involves FE maps from AlexNet at multiple layers and computes activity-level maps that guide the generation of WM. This method is ruled by Eqs. (15) to (17) as follows Eq. (15):

$$f_k^{(n,m)} = F_n\left(I_k\right) \tag{15}$$

where: I_k→ k^{th} source image, $F_n(\cdot)$→ The transformation (Convolution + Activation) applied by the n^{th} layer of AlexNet, $f_k^{(n,m)} \in \mathbb{R}^{H \times W \times m}$→ The resulting feature map of spatial dimensions $H \times W$ with $m = 64 \cdot 2^{n-1}$ channels.

The activity level map $A_k^n(x, y)$ for each spatial location (x, y) is computed by applying the L_1-norm across the depth dimension of the feature map Eq. (16):

$$A_k^n(x,y) = \|f_k^{(n,m)}(x,y)\|_1 = \sum_{c=1}^{m} \left| f_k^{(n,m)}(x,y,c) \right| \tag{16}$$

To improve robustness to local variation and ensure spatial smoothness, the activity level map is smoothed using a block-wise average filter of radius '*r*' as Eq. (17):

$$\hat{A}_k^n(i,j) = \frac{1}{(2r+1)^2} \sum_{\beta=-r}^{r} \sum_{\theta=-r}^{r} A_k^n(i+\beta, j+\theta) \tag{17}$$

where: $r \in \mathbb{Z}^+$→ Controls the neighborhood size (set to $r = 1$ in proposed test results). $\hat{A}_k^n(i,j)$ →The smoothed activity level map, which reflects the intensity of FE in the neighborhood around (i, j).

These refined activity level maps are later used to construct adaptive WM that guide the fusion of LFSB.

The IFM employs multi-layer FE from AlexNet to compute activity-level maps that guide low-frequency fusion. The *k*-SI's FM at the *n* layer is $f_k^{n,m}$, and the FM's dimension is *m*, $m = 64 \times 2^{n-1}$, $k = 2$, where F_n specifies the layer in the AlexNet, and $n \in \{1, 2, 3,4, 5\}$, and [ReLU1, ReLU2,…,ReLU5] activation layer are represented by $n \in \{1, 2, 3,4, 5\}$. To make the IFM resilient to misregistration, l_1-norm generates the activity level map known as $A_k^n(x, y)$ at position (i, j). The final activity level map $A_k^n(x, y)$ is computed using the block-based average operator, where '*r*' specifies the block size and is set to 1 to preserve more data[96].

To construct the adaptive WM used for fusing the LFC, an Adaptive Selection Algorithm (ASA) is employed. The method begins by computing the ratio of the smoothed activity level maps between the two source images EQU (16):

$$t_n(i,j) = \frac{\hat{A}_1^n(i,j)}{\hat{A}_2^n(i,j)} \tag{18}$$

This ratio $t_n(i, j)$ is used to compute the WM as $W_1^n(i, j)$ and $W_2^n(i, j)$ for each image as follows: Eqs. (19) and (20):

$$W_1^n(i,j) = \frac{t_n^3(i,j)}{1 + t_n^3(i,j)} \tag{19}$$

$$W_2^n(i,j) = \frac{1}{1 + t_n^3(i,j)} \tag{20}$$

These formulations ensure that if $t_n(i,j) \to 0$, more weight is assumed to the second source image, and vice versa.

Since the feature maps from AlexNet are downsampled due to pooling operations, the weight maps must be upsampled to match the original spatial resolution of the source images[97].

This is done using nearest-neighbor upsampling Eqs. (21) and (22):

$$\hat{W}_k^n(i+p, j+q) = W_k^n(i,j) \tag{21}$$

$$p, q \in \left\{0,1,\ldots,2^{n-1}-1\right\} \tag{22}$$

where: $\hat{W}_k^n$→ The upsampled weight map, $k \in \{1,2\}$→ The image index, The upsampling factor 2^{n-1} aligns with the receptive field size of the AlexNet layer 'n'

Using these upsampled WM[98], the fused LFC at each layer is computed by weighted averaging Eq. (23):

$$L_{\text{Fused}}^n\,(i,j) = L_1(i,j)\cdot\hat{W}_1^n(i,j) + L_2(i,j)\cdot\hat{W}_2^n(i,j) \tag{23}$$

Finally, across the multiple layers $n \in \{1,2,3,4,5\}$, the fused coefficient at each spatial location is selected using a maximum activity rule Eq. (24):

$$L_{\text{Fused}}\,(i,j) = \max_n\left[L_{\text{Fused}}^n\,(i,j)\right] \tag{24}$$

This step ensures that all layers' most salient fused features contribute to the final low-frequency fusion result.

HFC-fusion method using pulse-coupled neural network (PCNN)

The PCNN is a biologically inspired Feedback Neural Network (FNN) on the visual cortex of mammals[99]. It consists of a single-layer, 2-D array of neurons, where each neuron corresponds 1-to-1 with a pixel in the input image. In this network, spatially adjacent neurons interact within a defined local neighborhood, enabling localized feature improvement. From Fig. 6, each PCNN neuron comprises three core components: the receptive field, which receives external stimuli; the linking or modulation field, which facilitates inter-neuronal communication; and the pulse generator, which directs the neuron's firing behavior based on internal dynamics and external input. Also, according to[100], there is a split of the feed signal into the input and $F_{i,j}$ and linking $L_{i,j}$ Inputs.

The PCNN for IF is formulated as follows, where each neuron corresponds to a pixel at a spatial location (i,j), and the PCNN evolves over discrete time steps $n \in \{1,2,\ldots,n_{\max}\}$ as implicit in Eqs. (25) to (29).

$$F_{i,j}[n] = S_{i,j} + e^{-\alpha_F}F_{i,j}[n-1] + V_F\sum_{k,l}M_{i,j,k,l}Y_{i,j}[n-1] \tag{25}$$

$$L_{i,j}[n] = e^{-\alpha_L}L_{i,j}[n-1] + V_L\sum_{k,l}W_{i,j,k,l}Y_{i,j}[n-1] \tag{26}$$

$$U_{i,j}[n] = F_{i,j}[n]\left(1+\beta L_{i,j}[n]\right) \tag{27}$$

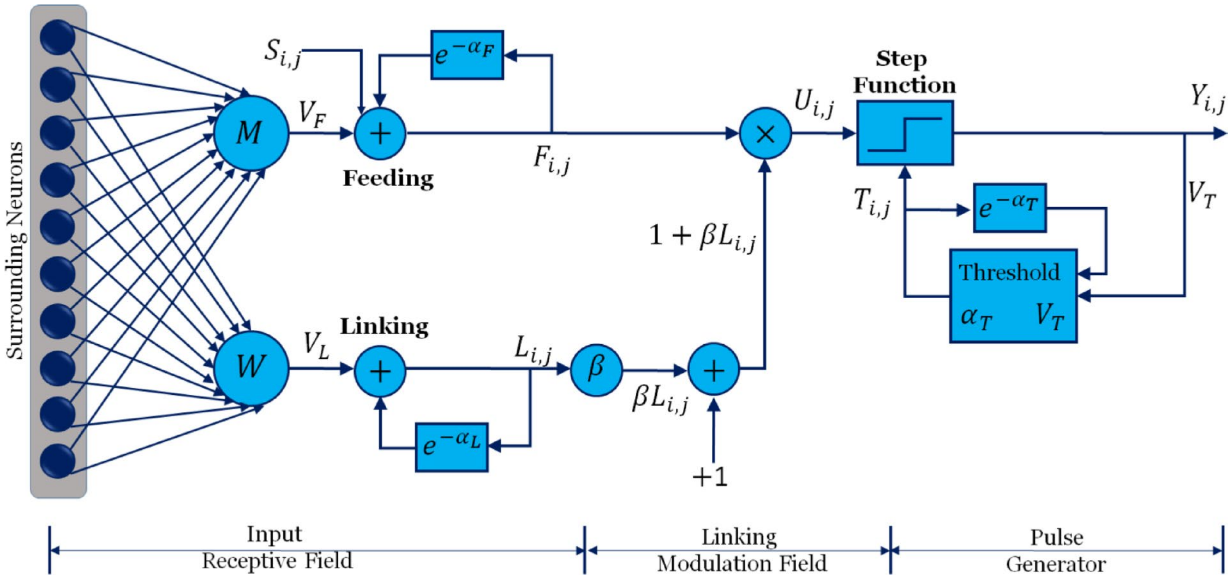

Fig. 6. The PCNN model.

$$T_{i,j}[n] = e^{-\alpha_T} T_{i,j}[n-1] + V_T Y_{i,j}[n] \tag{28}$$

$$Y_{i,j}[n] = \begin{cases} 1, U_{i,j} > T_{i,j} \\ 0, \text{ Otherwise} \end{cases} \tag{29}$$

where, $F_{i,j}[n]$ →Feeding input at position (i,j) at iteration 'n', composed of the static input $S_{i,j}$, its decayed past value, and the effect from neighboring neuron firings via synaptic weights $M_{i,j,k,l}$. $L_{i,j}[n]$ →Linking input, decayed over time and influenced by neighboring outputs via weights $W_{i,j,k,l}$. $U_{i,j}[n]$ →Internal activity (modulated signal) combining feeding and linking with a linking strength co-efficient β. $T_{i,j}[n]$ →Dynamic threshold, decays over time and increases upon firing. $Y_{i,j}[n]$ →Output pulse; a binary indicator of neuron firing at time 'n'. $\alpha_F, \alpha_L, \alpha_T$ →Decay coefficients for feeding, linking, and threshold signals. V_F, V_L, V_T →Normalization constants scaling the influence of previous outputs. (k, l)→Indices of neighboring pixels in the spatial neighborhood. n_{Max} →Maximum number of iterations.

Six parameters, including three degeneration factors $(\alpha_F, \alpha_L, \alpha_T)$ and three normalizing constants (V_F, V_L, V_T) for the feeding $(F_{i,j})$, linking $(L_{i,j})$, and threshold $(T_{i,j})$ inputs are acquired using the PCNN, which consists of a feeding and linking field. $U_{i,j}$ stands for the neuron's internal activity (linking modulation) in Eqs. (26) to (28) (T_{ij}, Y_{ij}) for the dynamic threshold and the neurons' pulse output. An important factor that alters the linking field's weight is the linking parameter 'β'. The HFC's Novel Sum-Modified LAPLACIAN (NSML) measures and assesses contrast levels to satisfy the HVS requirement.

Compute the NSML for the HFSB as follows: Eqs. (30) and (31):

$$NSML(i,j) = \sum_a \sum_b w(i,j) \cdot F(i+a, j+b) \tag{30}$$

$$\begin{gathered} F(i,j) = \mid 2HF_Z^{NSST}(i,j) - HF_Z^{NSST}(i-1,j) - HF_Z^{NSST}(i+1,j) \\ \mid + \mid 2HF_Z^{NSST}(i,j) - HF_Z^{NSST}(i,j-1) - HF_Z^{NSST}(i,j+1) \mid \end{gathered} \tag{31}$$

where:$HF_Z^{NSST}(i,j)$ →HFC from modality Z (*e.g.*, CT or MRI) at pixel location (i,j), attained by NSST. $F(i,j)$ →Local contrast magnitude at pixel (i,j) computed using modified Laplacian operators. $w(a,b)$ →Normalized window function used in local contrast aggregation, has the following Eq. (32):

$$w(i,j) = \begin{bmatrix} 1/15 & 2/15 & 1/15 \\ 2/15 & 3/15 & 2/15 \\ 1/15 & 2/15 & 1/15 \end{bmatrix} \tag{32}$$

To activate the PCNN, use the following Eq. (33) and set up the neuron's pulse with the help of each HFSB's NSML:

$$\left. \begin{array}{c} F_{i,j}^Z[n] \quad = NSML_{i,j}^Z \\ L_{i,j}^Z[n] \quad = e^{-\alpha L} L_{i,j}^Z[n-1] + V_L \sum_{k,l} W_{i,j,k,l}^Z Y_{i,j}^Z[n-1] \\ U_{i,j}^Z[n] \quad = F_{i,j}^Z[n] \left(1 + \beta L_{i,j}^Z[n]\right) \\ T_{i,j}^Z[n] \quad = e^{-\alpha_T} T_{i,j}^Z[n-1] + V_T Y_{i,j}^Z[n] \\ Y_{i,j}^Z[n] \quad = \begin{cases} 1, & U_{i,j}^Z > T_{i,j}^Z \\ 0, & \text{otherwise} \end{cases} \end{array} \right\} \tag{33}$$

where: Z →Source modality (*e.g.*, CT or MRI). (i,j) →Pixel coordinates. $F_{i,j}^Z[n]$ →Feeding input derived from NSML for modality Z. $L_{i,j}^Z[n]$ →Linking input aggregating neighborhood activity. $U_{i,j}^Z[n]$ →Internal activity (modulated product of feeding and linking inputs). $T_{i,j}^Z[n]$ →Dynamic threshold for neuron firing. $Y_{i,j}^Z[n]$ →Binary neuron output (firing decision). α_L, α_T →Decay rates for linking input and threshold. V_L, V_T →Normalization constants. β →Linking modulation coefficient. $W_{i,j,k,l}^Z$ →Synaptic weights connecting neighboring neurons.

In the iterative process of the PCNN, a neuron's firing is found by whether its internal activity $U_{i,j}^Z[n]$ exceeds the dynamic threshold $T_{i,j}^Z[n]$.

This condition generates the binary firing output $Y_{i,j}^Z[n] \in \{0,1\}$ as defined in Eq. (29). Over n_{Max} iterations, the total firing time map is computed by collecting the number of times each neuron fires, Eq. (34):

$$t_{i,j}^Z[n] = t_{i,j}^Z[n-1] + Y_{i,j}^Z[n], \text{ for } n = 1,2,\ldots,n_{\text{Max}} \tag{34}$$

where: $t_{i,j}^Z[n]$ →The cumulative firing count at pixel (i,j) up to iteration 'n', $Y_{i,j}^Z[n] = 1$ if $U_{i,j}^Z[n] > T_{i,j}^Z[n]$, and 0 otherwise, This summation reflects the number of times a neuron at location (i,j) was activated across the PCNN evolution.

At the decision of the iterations (*i.e.*, when $n = n_{\text{Max}}$), the final firing times $t_{i,j}^X[n_{\text{Max}}]$ and $t_{i,j}^Y[n_{\text{Max}}]$ are used as a decision measure to select the most salient co-efficient from the HFSB of the input images[101].

The fusion rule is defined in Eq. (35):

$$HF_F^{NSST}(i,j) = \begin{cases} HF_X^{NSST}(i,j), & \text{If } t_{i,j}^X[n_{\text{Max}}] \geq t_{i,j}^Y[n_{\text{Max}}] \\ HF_Y^{NSST}(i,j), & \text{Otherwise} \end{cases} \tag{35}$$

This decision rule ensures that for each pixel (i, j), the fused HFC is taken from the source image whose PCNN neuron exhibits robust and more reliable firing behavior, indicating higher local saliency, such as edges or texture elements[102].

Proposed CNN-shearlet fusion model

The proposed IMM (Fig. 7) is written as follows for MMIFM and for implementing the abovementioned concept: Let's use 'X' as the first image and 'Y' as the second.

Step 1:	Decompose the CT and MRI images ($\boldsymbol{X} = \boldsymbol{X}_{\boldsymbol{ij}}$ and $\boldsymbol{Y} = \boldsymbol{Y}_{\boldsymbol{ij}}$) to produce LFC $\{\boldsymbol{L}_{\boldsymbol{x}}, \boldsymbol{L}_{\boldsymbol{y}}\}$ and a series of HFC $\{\boldsymbol{c}, \boldsymbol{H}_{\boldsymbol{y}}\}$ for each K-scale and l-direction, where 1 ≤ k ≤ K
Step 2:	Input the LFC and HFC using the loss function in EQU (5) to the denoising network for obtaining noise-free $\{L_{\overline{x}}, L_{\overline{y}}\}$ and noise-free HFC $\{H_{\overline{x}}, H_{\overline{y}}\}$
Step 3:	Improve features with EnhanceNet, and the EnhanceNet generates a probability vector $\{L_x, L_y\}$ and $\{H_x, H_y$ for each noise-free input, $\{L_{\overline{x}}, L_{\overline{y}}\}$ and $\{H_{\overline{x}}^{l,k}, H_{\overline{y}}^{l,k}\}$. The predicted probability that each part of x, y is the original signal belongs to the parallel region of interest. Each element in x, y is the original signal's predictable probability that belongs to the corresponding region of interest. Using objective function 6, the output vectors must be near the one-hot vectors that indicate the input signal's accurate label to maximize performance[103]
Step 4:	Assuming the FE from both images, the DL fuses the LFC. The AlexNet that could select the LFC adaptively generates the WM $L_{\text{Fused}}\ (i,j) = Max\,[L_{\text{fised}}^{n}\ (i,j), n \in \{1,2,3,4,5\}]$ (34)
Step 5:	Compute the NSML for the HFSB. To activate the PCNN, increase the neuron pulse and compute the NSML of each HFSB
	If $n = n_{Max}$, the iteration ends, and the firing timings based on the next fusion rule are used to fuse the HFSB image coefficients. Using EQU (35), the firing times $t_{i,j}^{Z}[n]$ in 'n' iteration is evaluated $H_{Fused} = \begin{cases} H_x; & \text{if } t_{i,j}^{X}\ [n_{max}] \geq t_{i,j}^{Y}\ [n_{max}] \\ H_y; & \text{if } t_{i,j}^{X}\ [n_{max}] < t_{i,j}^{Y}\ [n_{max}] \end{cases}$ (35)
Step 6:	To generate the IF, perform INSST to combine the LFC and HFC, EQU (36) $Image_{Fused} = NSST^{-1}\ (L_{Fused}, H_{Fused})$ (36)

Experimental analysis

The experimental study used a simulation tool of MATLAB R2022a and Python 3.11 software on a 64-bit Windows 10 operating system. The system was specified with a GTX 1080 GPU with 2560 CUDA cores and 8 GB of GDDR5X GPU memory. Moreover, the system has a 3.70 GHz Core i7 CPU and 32 GB of RAM. To assess the effectiveness of the proposed NSST + CNN for MMIF, its performance was compared with five different ultramodern IFS, including:

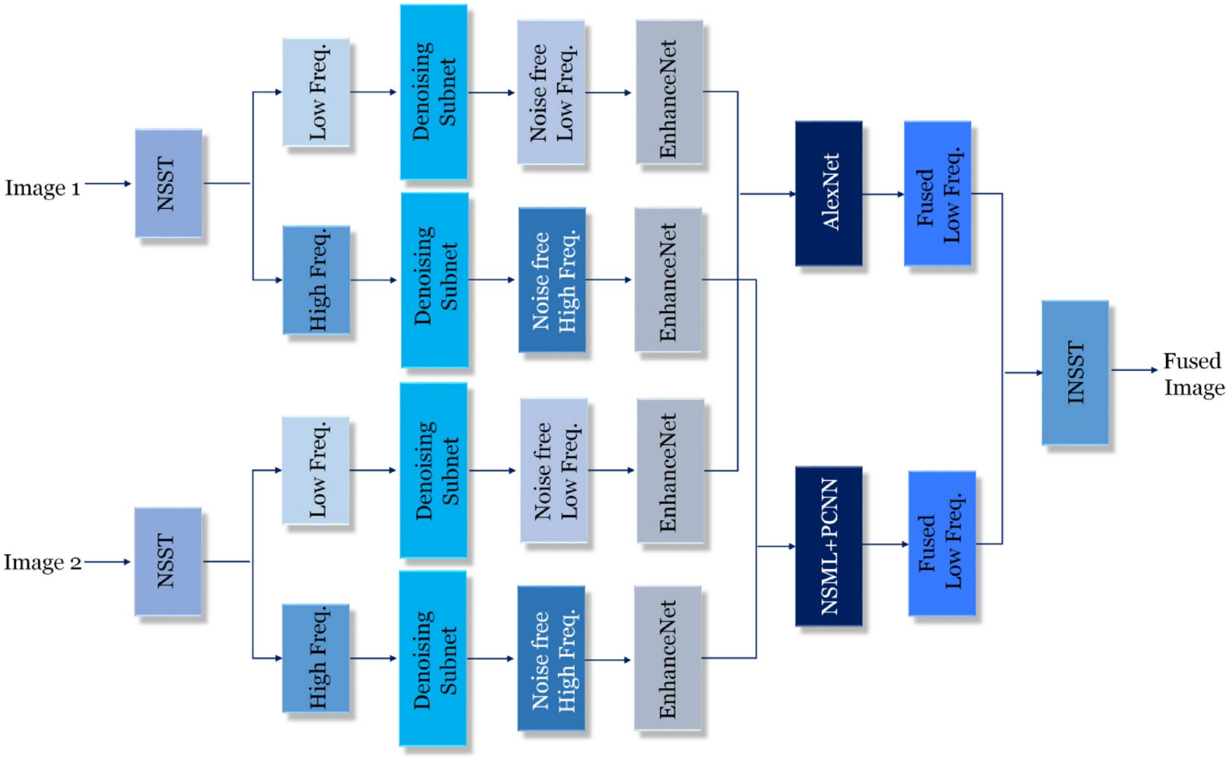

Fig. 7. The Proposed CNN-Shearlet IF Model.

A. **Method 1:** The Singular Value Decomposition (SVD) on the ST domain was used in SVD-Shearlet, a novel MIF[104], to enhance the information content of an image using IF. The proposed method uses ST first to transform the SI into a Shearlet image. Subsequently, using the SVD in the LFSB, the changed sub-bands are selected by local features. SVD is used to process the composition of different HFSB coefficients. The LFSB and HFSB are then fused. Lastly, the Inverse Shearlet Transform reconstructs the IF (IST).
B. **Method 2:** A new IMF for MR and CT medical images is proposed by NSST-SNN[105], which uses NSST and spiking neural network features. The NSST is used to effectively create image coefficients, such as a new image with distinct features. First, the NSST decomposes the SI of MR and CT images into some sub-images. The LFC is fused using regional energy. Moreover, HFC that used a pulse-coupled NN, a type of NN with bioinspiration, is fused. Moreover, it maintains the SI's edges and detailed information. The INSST is finally used to generate IF.
C. **Method 3:** An IMF relying upon sparse representation and Fast Finite Shearlet Transform (FFST) is presented in FFST-SR-PCNN to address the problem of poor edge clarity, which makes it difficult for current algorithms to maintain the details of the SI. Initially, FFST decomposes the SI into LFC and HFC. Secondly, the LFC is trained using the K-SVD to generate the overcomplete dictionary D, and then the LFC is sparsely encoded using the OMP algorithm to complete the LFC fusion. The fusion coefficient of the HFC is then selected based on the number of ignitions, and an HFC is then used to excite a PCNN. The FFST inverse transform is then used to reconstruct the fused LFC and HFC into the MIF.
D. **Method 4**: A novel IMF to CT and MRI medical imaging called CNN + DCSCM is based on CNNs and a Dual-Channel Spiking Cortical Model (DCSCM). Initially, the SI gets decomposed into an LFC and a series of HFCs using NSST. Second, the HFC gets fused through DCSCM, in which the modified average gradient of an HFC is used as the input stimulus. Third, the LFC gets fused by CNN, wherein WM is generated by a series of FM and an adaptive selection rule. In the end, inverse NSST is used to recreate the IF.
E. **Method 5:** RBM-EMBO is a novel IMF built upon concepts from DL and best thresholding. The best threshold for fusion rules in ST is decided using an Enhanced Monarch Butterfly Optimization (EMBO). Then, using FM and the data specified by the DL method's extraction part, LFSB and HFSB are fused. Therefore, the IFM was shown using the Restricted Boltzmann Machine (RBM).

Applying 180 pairs of SI, the effectiveness of the proposed model is tested. The Harvard Medical School's Whole Brain Atlas was used to compile the complete SI sample. A sample image set comprising CT, MRI, SPECT, and PET images was used in this study. Each SI's gray and color scale spatial resolution is 512 × 512 pixels. The predictable 6 pairs of multimodal brain images, including CT, MRI, SPECT, and PET images, serve as a sample set for the experimental study results. The images in Pair I and Pair II are MR-T1 and MR-T2, MRI and PET are the images in Pair II, MR-T2, and SPECT are the images in Pair III, CT and MR-T2 are the images in Pair IV, MR-T2 is the image in Pair V, and MR-T1 and MR-T2 are the images in Pair VI.

Assessment metrics

Throughout this paper, four metrics, Margin Information Retention (QAB/F) and Spatial Frequency (SF), Mutual Information (MI), Image Quality Index (IQI), and Image Entropy (IE), and each is drawn from one of the five types selected for performing analysis on the combined results in an objective method.

A. Image Entropy (IE): IE is a tool that measures the volume of data in a SI. If a fused image's entropy increases better than the SI, the fused images have more data.

The entropy is said by Shannon information as Eq. (37),

$$EN = -\sum_{i=0}^{L-1} p(i)\mathrm{Log}_2 p(i) \tag{37}$$

where, p (i)→ The gray level's probability ranging between [0,…, L − 1].

B. Spatial Frequency: One more index that assesses an image's overall activity and level of clarity is called Spatial Frequency (SF)[106].

It is defined as Eq. (38)

$$SF = \sqrt{RF^2 + CF^2} \tag{38}$$

where, RF→ The row frequency. CF→ The column frequency

These are assessed as Eqs. (39) and (40)

$$RF = \sqrt{\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j-1) - I(i,j))^2} \tag{39}$$

$$CF = \sqrt{\frac{1}{(M-1)N}\sum_{i=2}^{M}\sum_{j=1}^{N}(I(i,j) - I(i-1,j))^2} \tag{40}$$

where, I(i, j)→The fused image's gray level value. {M, N}→The image's size. The better the fusion quality, the higher the spatial frequency.

C. Mutual Information (MI): A fused image's quality is measured as Eqs. (41) to (42).

$$MI = I\left(x_X; x_F\right) + I\left(x_Y; x_F\right) \tag{41}$$

Where

$$I\left(x_R; x_F\right) = \sum_{u=1}^{L}\sum_{v=1}^{L} h_{R,F}(u,v)\mathrm{Log}_2\frac{h_{R,F}(u,v)}{h_R(u)h_F(v)} \tag{42}$$

where $h_R\left(u\right), h_F(v)$ stand as the normalized gray-level histograms for (x_R,x_F). x_R, x_F →Joint gray-level histogram is known as $h_{R,F}$. The reference image→ x_R, The fused image→ x_F.

If the fused image's MI value is higher, it has more texture and detailed data.

D. Image Quality Index: The reference image (R) and fused image (F)'s Quality Index (IQI) is defined as Eq. (42)

$$IQI = \left(\frac{\sigma_{FR}}{\sigma_F\sigma_R}\right)\cdot\left(\frac{2\mu_F\mu_R}{\mu_F^2+\mu_R^2}\right)\cdot\left(\frac{2\sigma_F\sigma_R}{\sigma_F^2+\sigma_R^2}\right) \tag{43}$$

where, μ_F, μ_R →The mean (σ_F^2, σ_R^2)→ The fused and reference images' variance. (X, Y)→The two reference images. I→one-fused image, is employed in this study.

In order to evaluate the quality metric (Q_0), $IQI(X,F)$ and $IQI(Y,F)$ are averaged, Eq. (44).

$$IQI = \frac{IQI(X,F) + IQI(Y,F)}{2} \tag{44}$$

E. Margin Information Retention: How effectively the IFM performs in aspects of edge strength is provided as Eq. (45)

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} w^A(i,j) + w^B(i,j)} \tag{45}$$

where, $Q^{AF}(i,j) = Q_g^{AF}(i,j) + Q_\alpha^{AF}(i,j)$ →The values for edge strength and positioning preservation are exposed by $Q_g^{AF}\ (i,j)$ and $Q_\alpha^{AF}\ (i,j)$. $Q^{BF}\ (i,j)$ →likewise assessed.

To ensure statistical validity and robust performance evaluation, fivefold cross-validation was implemented on the dataset of 180 image pairs. The dataset was randomly partitioned into five equal subsets, with four subsets used for training and one for testing in each fold. The reported performance metrics represent the average across all five folds, ensuring the evaluation is not biased by any particular training–testing split. This cross-validation method confirms the generalizability of the proposed model across different image combinations within the neuroimaging domain[107].

Table 1 and Fig. 8 present the comparative performance of the proposed CNN + Shearlet fusion model against existing approaches. The proposed model proves superior performance across multiple evaluation metrics, particularly in preserving detailed features while maintaining overall image quality.

- **Image Entropy (IE)**, the proposed method, achieves consistently higher values, indicating enhanced data content in the fused images. This is particularly evident in Pair IV (CT and MR-T2 images), where the proposed method achieves an IE of 5.67, compared to 5.64 for the best alternative method.
- **Spatial Frequency (SF)**, which measures an image's overall activity level and clarity, significantly improves with the proposed approach. The proposed model achieves an SF value of 44.26 for Pair IV, signifying superior preservation of high-frequency components, which are critical for diagnostic detail. The higher SF values across all image pairs confirm the method's effectiveness in maintaining spatial detail distribution—a crucial factor for clinical reading.
- For edge data preservation, as measured by QAB/F, the proposed model achieves an average development of 16.5% compared to other algorithms. This substantial enhancement can be attributed to the directional FE capabilities of the EnhanceNet component, which effectively preserves boundary data between different tissue types, as highlighted in the red regions of Fig. 8.
- **Image Quality Index (IQI)** values, which measure the similarity between source and IF, are consistently higher for the proposed model across all image pairs. With IQI values approaching unity (ranging from 0.47 to 0.58), the proposed model maintains better fidelity to the source images while effectively combining complementary data.
- **Mutual Information (MI)** metrics further confirm the superior data transfer achieved by the proposed model, with consistently higher values indicating better preservation of source image content in the fusion result.

A graphical comparison in Fig. 8 reveals the precise advantages of the proposed model. Method 1 (SVD-Shearlet) displays blocking objects and fails to preserve detailed feature information, particularly in the highlighted regions. Method 2 (NSST + SNN) proves inadequate edge definition, particularly in Pairs I, III, IV, V, and VI, where acceptable anatomical boundaries appear blurred. Methods 3 (FFST + SR + PCNN) and Methods 4 (CNN + DCSCM) validate improved performance but still display insufficient edge clarity, as quantitatively confirmed by their lower QAB/F scores in Table 1. Method 5 (RBM + EMBO), despite presenting competitive

	Performance metrics	M-I	M-II	M-III	M-IV	M-V	Proposed
Pair I	SF	27.21845	26.71028	25.2282	24.63622	28.04914	29.50575
	IE	5.530153	5.390857	5.522727	5.423427	5.593807	5.621603
	MI	2.681849	3.317116	3.326134	2.751444	4.004049	3.371646
	QAB/F	0.492576	0.499366	0.53098	0.493955	0.581161	0.607259
	IQI	0.299386	0.38638	0.446002	0.468387	0.481861	0.499578
Pair II	SF	34.61929	29.69883	29.73172	35.83296	36.34113	36.61484
	IE	5.192045	5.157247	5.160005	5.235329	5.236708	5.295376
	MI	3.516035	3.416098	3.444848	3.664879	3.955672	4.295266
	QAB/F	0.624446	0.4686	0.484407	0.693935	0.671656	0.758013
	IQI	0.323787	0.351264	0.434651	0.43794	0.461067	0.475495
Pair III	SF	23.47241	18.70367	18.51058	22.91968	23.65064	23.0385
	IE	4.725461	4.803013	4.902207	4.973075	4.979334	5.012434
	MI	2.865491	2.771283	2.904214	2.963306	3.094433	4.059216
	QAB/F	0.70518	0.420859	0.45258	0.687569	0.72661	0.758968
	IQI	0.349036	0.355083	0.424042	0.442714	0.461916	0.466372
Pair IV	SF	39.86013	43.60193	38.75468	41.16822	42.45722	44.26287
	IE	5.570468	5.601764	5.547022	5.609084	5.642291	5.67433
	MI	2.215796	2.519001	2.198609	2.332919	2.564832	2.622439
	QAB/F	0.589754	0.592194	0.556124	0.652241	0.668367	0.691282
	IQI	0.344156	0.418525	0.477617	0.453853	0.463083	0.483028
Pair V	SF	20.18893	16.04187	16.21586	19.86111	20.16134	19.48555
	IE	5.161279	5.091471	5.128072	5.190347	5.209868	5.254426
	MI	2.726725	2.526215	2.67368	2.837059	2.941876	3.640797
	QAB/F	0.709954	0.392533	0.485998	0.699027	0.707408	0.718548
	IQI	0.314451	0.390729	0.460006	0.44823	0.456505	0.474222
Pair VI	SF	35.04895	38.07888	28.28465	32.16331	35.79689	36.34007
	IE	4.365073	4.499595	4.539697	4.545002	4.633481	4.708168
	MI	2.870689	4.281474	2.62212	2.768525	3.143341	3.467976
	QAB/F	0.645664	0.693829	0.437727	0.565778	0.560686	0.67399
	IQI	0.305857	0.36113	0.51687	0.544348	0.562701	0.581585

Table 1. Objective IF evaluation.

performance in some metrics, generates images with low contrast in key regions, particularly noticeable in the upper proper portions of the IF.

In contrast, the proposed CNN + Shearlet model maintains optimal contrast levels, preserves fine structural details, and avoids introducing objects such as dark lines or blocking effects. The superior performance is particularly evident in the highlighted regions of Fig. 8, where the proposed method maintains clear tissue boundaries while preserving textural details in adjacent areas.

The comprehensive quantitative and qualitative analysis confirms that the proposed CNN + Shearlet fusion model outperforms existing methods across all evaluation metrics, significantly improving edge preservation (QAB/F) and feature detail retention (SF and IE).

Figure 9 presents the graphical analysis of the Average Running Time (ART) for the proposed method compared to several existing IFMs. Experimental results indicate that the proposed model achieved an ART of approximately 0.53 *Sec.* While this is slightly higher than Method 5, which reported an ART of 0.34 *Sec.*, it remains significantly faster than the other evaluated methods.

Beyond execution speed, the proposed method proved superior fusion performance across a range of input images. The consistency and quality of the fused outputs underscore the model's robustness and adaptability in handling diverse medical imaging modalities. In contrast, other approaches exhibited variability and instability, particularly when confronted with heterogeneous inputs. The proposed model achieved optimal thresholding performance, highlighting its potential for deployment in advanced medical imaging applications and intelligent diagnostic systems.

Comparison with recent fuzzy-based fusion methods

To further validate the effectiveness and generalizability of the proposed CNN + Shearlet fusion model, this section presents a comparative analysis against recent *state-of-the-art* fuzzy-based methods published after 2020. The following advanced fuzzy-based IMF were implemented and evaluated using identical test conditions and assessment metrics:

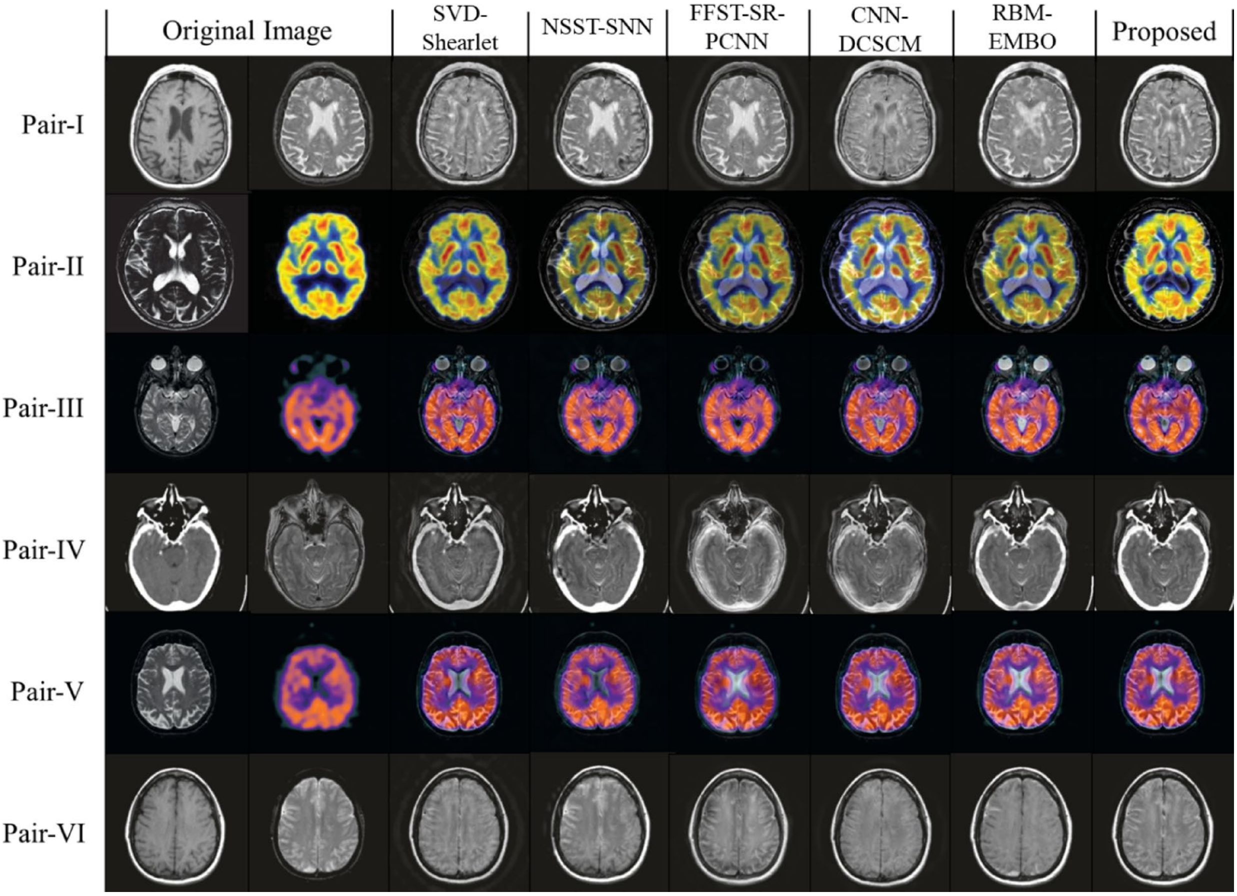

Fig. 8. Comparison of image pair IF methods.

(a) **Intuitionistic Fuzzy Sets and Joint Laplacian Pyramid (IFSJLP):** Jiang et al. (2023) proposed a lightweight multimodal medical IF that integrates a newly defined similarity measure from intuitionistic FST with Laplacian pyramid decomposition. Their method effectively captures and merges fuzzy features within HFSB and LFSB, achieving high fidelity and edge preservation.

(b) **Multiscale Fuzzy Quality Assessment (MFQA):** The Authors developed a multi-focus IF based on a sophisticated multiscale quality evaluation mechanism. Their method utilizes FL to determine optimal fusion weights across various decomposition levels, with a focus on preserving edges and enhancing contrast.

(c) **Fuzzy Compensation Dictionary Learning (FCDL):** Researchers presented an IF combining FST with compensation dictionary learning. This method addresses the limitations of traditional dictionary learning by incorporating fuzzy membership functions that adaptively manage local feature differences, particularly at boundaries between different tissues in medical images.

The quantitative comparison in Table 2 and Fig. 10 reveals several key insights regarding the performance of the proposed CNN + Shearlet method compared to recent fuzzy-based methods. Regarding structural preservation metrics, the proposed method consistently outperforms the fuzzy models. The SSIM values, which measure the structural similarity between the fused and reference images, average 0.925 for the proposed method, compared to 0.912, 0.903, and 0.903 for the FCDL, MFQA, and IFSJLP. This 1.3%-2.2% improvement in SSIM is particularly significant for MIF applications, where preserving anatomical structures is critical for accurate diagnosis. The superior structural preservation can be attributed to the directional selectivity of the Shearlet transform combined with the FE sizes of the proposed CDEN + NN. The feature preservation capability, measured by FSIM, follows a similar trend, with the proposed method achieving an average of 0.888 compared to 0.877 for the FCDL. This improvement is most pronounced in Pair II (CT and MRI) and Pair V (MR-T2), which contain complex tissue boundaries and fine anatomical details. The fuzzy-based methods, particularly IFSJLP and MFQA, have problems with FS in these challenging cases due to their limited directional sensitivity. All methods perform comparably for information transfer metrics (MI and IE), but the proposed method maintains a slight edge with an average MI of 3.58 compared to 3.40 for FCDL, the next best performer. The MFQA proves competitive performance in MI for specific image pairs (notably Pair II and Pair V), likely due to its multiscale quality assessment method that optimizes information retention across different scales. However, it fails to maintain this advantage consistently across all test cases. The edge selection capability, measured by QAB/F and SF, shows the proposed method's significant advantage in handling directional features. The average QAB/F of 0.70 for the proposed method represents a 1.4% improvement over FCDL (0.69) and 4.5% over IFSJLP and MFQA (0.67).

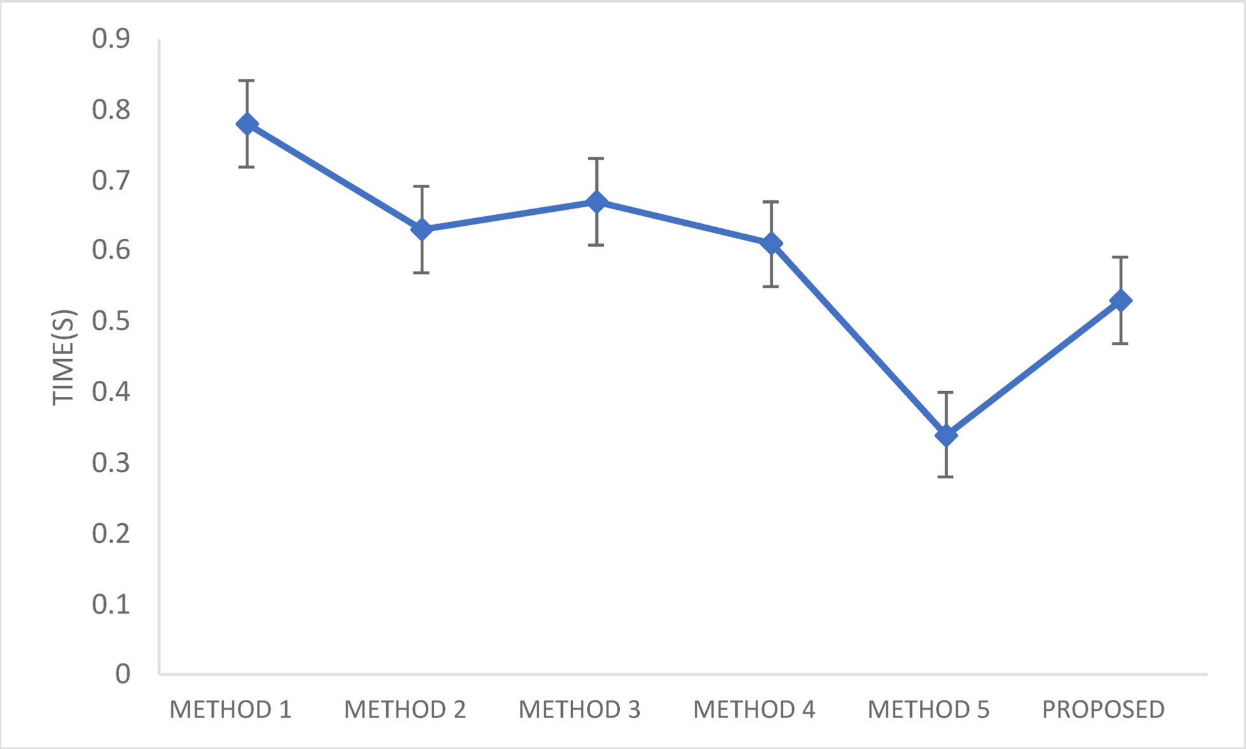

Fig. 9. Analysis of average running time (ART).

This advantage is most evident in Pair I and IV, which contain numerous fine-edge structures in brain ventricles and tissue boundaries. Signal fidelity, measured by PSNR, determines that the proposed method achieves an average of 36.78 dB, compared to 36.18 dB for FCDL, 35.84 dB for MFQA, and 35.84 dB for IFSJLP. The 0.6–0.94 dB improvement indicates better noise suppression while maintaining signal integrity. This is particularly important for clinical applications where noise objects could mask subtle pathological features. Interestingly, the performance patterns across different methods vary by image pair. For Pair III (MR-T2 and SPECT), the FCDL proves competitive performance in IE (5.02 vs. 5.01) and maintains close performance in SSIM (0.904 vs. 0.916). This proposes that the fuzzy compensation dictionary learning approach has particular strengths in handling the significant intensity differences characteristic of functional–anatomical image pairs. However, it still falls behind in edge preservation (SF and QAB/F) and overall signal fidelity (PSNR). The IFSJLP proves relatively robust performance in IE across several image pairs, reflecting its ability to preserve the overall information content. However, its performance in structural and feature preservation metrics (SSIM and FSIM) reliably lags behind that of the other methods, indicating limitations in maintaining the spatial integrity of important anatomical structures. The MFQA method validates competitive performance in Spatial Frequency (SF) for Pair II and Pair IV, indicating good detail preservation in specific cases. However, its inconsistent performance across other metrics and image pairs recommends that its quality assessment mechanism may not be adequately adaptive to the varied challenges of different modality combinations. Overall, the proposed CNN + Shearlet method proves consistent superiority across all evaluation metrics, with an average improvement of 3.67% over the baseline IFSJLP, compared to 1.96% and 0.57% improvements for FCDL and MFQA, respectively. This comprehensive performance advantage stems from the synergistic combination of precise directional decomposition provided by NSST and adaptive FS presented by the DL components, which addresses global and local features of the fusion challenge.

Ablation study

A comprehensive ablation study was conducted to systematically evaluate the contribution of each component in the proposed CNN + Shearlet fusion model. Based on the model, four model variants were implemented and compared:

A. **Baseline (NSST only):** This variant uses only the NSST for the decomposition of input images into LFSB and HFSB components, followed by conventional fusion rules (maximum selection for HFSB coefficients and weighted averaging for LFSB coefficients) without any CNN-based processing. The fused components are then reconstructed using inverse.
B. **NSST. 2. NSST + DNSN:** This variant incorporates the Denoising Subnet, as shown in Fig. 4 with NSST, but removes the EnhanceNet components from the model. LFSB and HFSB components from each source image undergo denoising in this configuration, but no FS is performed. The denoised components are then fused using AlexNet for low frequencies and NSML + PCNN for high frequencies.

Metrics	IFSJLP (Jiang et al. 2023)	MFQA (Li et al. 2024)	FCDL[76]	Proposed CNN-Shearlet method
Pair I				
SF	28.67	28.93	29.32	29.51
IE	5.59	5.54	5.58	5.62
MI	3.25	3.14	3.29	3.37
QAB/F	0.57	0.56	0.60	0.61
IQI	0.48	0.47	0.49	0.50
PSNR	35.26	34.89	35.87	36.42
SSIM	0.895	0.887	0.912	0.924
FSIM	0.864	0.858	0.872	0.883
Pair II				
SF	35.84	36.42	36.16	36.61
IE	5.25	5.22	5.28	5.30
MI	3.94	4.08	3.98	4.30
QAB/F	0.71	0.74	0.73	0.76
IQI	0.45	0.47	0.46	0.48
PSNR	37.12	37.45	37.28	38.16
SSIM	0.914	0.921	0.918	0.936
FSIM	0.876	0.884	0.879	0.895
Pair III				
SF	22.87	22.75	22.96	23.04
IE	4.95	4.98	5.02	5.01
MI	3.72	3.67	3.98	4.06
QAB/F	0.73	0.71	0.75	0.76
IQI	0.46	0.44	0.46	0.47
PSNR	34.65	34.38	35.27	35.54
SSIM	0.892	0.885	0.904	0.916
FSIM	0.871	0.863	0.882	0.889
Pair IV				
SF	43.24	43.87	43.65	44.26
IE	5.63	5.61	5.64	5.67
MI	2.51	2.55	2.49	2.62
QAB/F	0.64	0.67	0.67	0.69
IQI	0.46	0.48	0.47	0.48
PSNR	39.12	39.43	39.26	40.08
SSIM	0.927	0.935	0.932	0.943
FSIM	0.886	0.892	0.889	0.901
Pair V				
SF	19.26	19.38	19.12	19.49
IE	5.20	5.19	5.24	5.25
MI	3.35	3.48	3.42	3.64
QAB/F	0.70	0.69	0.71	0.72
IQI	0.45	0.47	0.46	0.47
PSNR	33.18	33.45	33.26	33.92
SSIM	0.878	0.881	0.884	0.896
FSIM	0.842	0.847	0.851	0.865
Pair VI				
SF	35.86	35.74	36.21	36.34
IE	4.64	4.62	4.70	4.71
MI	3.28	3.31	3.25	3.47
QAB/F	0.65	0.63	0.66	0.67
IQI	0.56	0.57	0.57	0.58
PSNR	35.71	35.45	36.12	36.58
SSIM	0.913	0.907	0.921	0.932
FSIM	0.875	0.871	0.886	0.896
Average				
Continued				

Metrics	IFSJLP (Jiang et al. 2023)	MFQA (Li et al. 2024)	FCDL[76]	Proposed CNN-Shearlet method
SF	30.96	31.18	31.24	31.54
IE	5.21	5.19	5.24	5.26
MI	3.34	3.37	3.40	3.58
QAB/F	0.67	0.67	0.69	0.70
IQI	0.48	0.48	0.49	0.50
PSNR	35.84	35.84	36.18	36.78
SSIM	0.903	0.903	0.912	0.925
FSIM	0.869	0.869	0.877	0.888
Avg. improvement (%)	Baseline	+ 0.57%	+ 1.96%	+ 3.67%

Table 2. Comparison of proposed CNN + Shearlet method with recent fuzzy-based methods (including additional metrics).

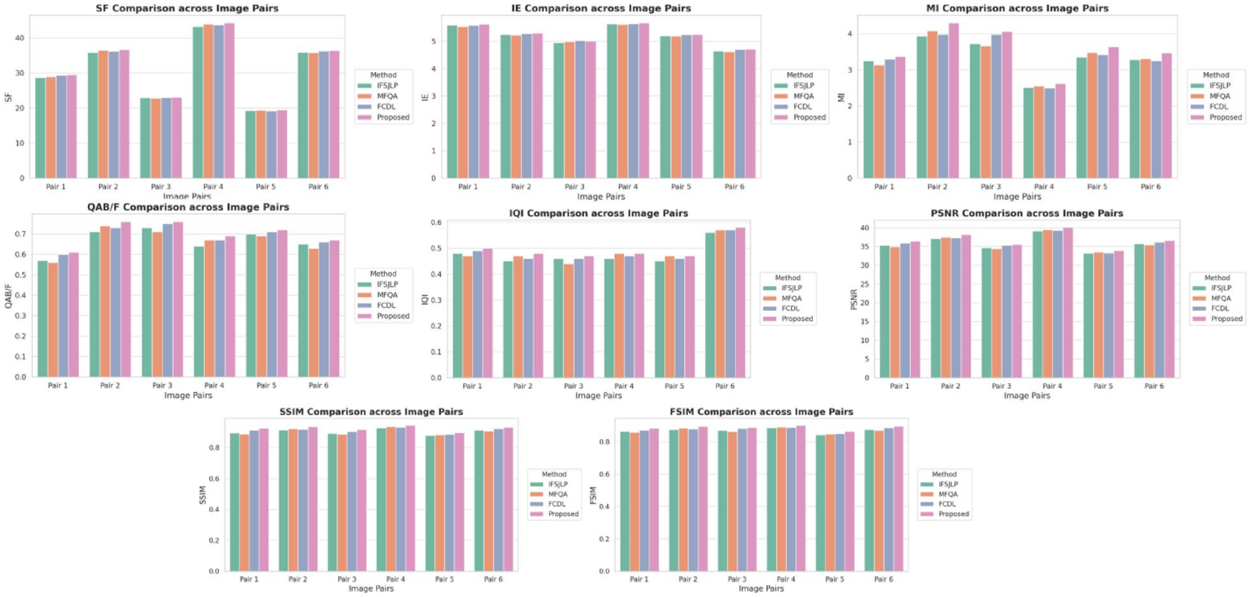

Fig. 10. SOTA model comparison.

C. **NSST + ENSN:** This variant incorporates the EnhanceNet exposed in Fig. 5 with NSST but removes the Denoising Subnet components. The decomposed frequency components are directly fed to the EnhanceNet without prior denoising, followed by the same fusion method as the complete model. This configuration tests the importance of FS independent of denoising.
D. **Full Model (NSST + CDEN):** The complete proposed method, as depicted in Fig. 6, incorporates NSST decomposition, followed by sequential Denoising and Enhancement of both frequency bands from both source images, and finally fusion using AlexNet for LFSB and NSML + PCNN for HFSB.

Table 3 presents the quantitative performance metrics for all four variants across the 6 test image pairs.

The ablation study results reveal several dynamic insights into the proposed model. The baseline NSST achieves acceptable performance due to the directional selectivity of the Shearlet transform; however, it has limitations in handling noise and preserving fine details. This is evident in its lower performance across all metrics, particularly in PSNR (33.42 dB) and structural similarity (SSIM: 0.864).

The NSST + DNSN difference proves significant improvement over the baseline in signal fidelity metrics (PSNR: + 1.85 dB on average), confirming the effectiveness of the denoising subnet in reducing noise objects while preserving the underlying signal model. The denoising process, governed by the loss function defined in EQU (5), effectively learns the mapping from noisy decomposed frequencies to their new counterparts. However, this variant still shows limited feature emphasis and contrast improvement capability without the enhancement stage.

The NSST + ENSN optional proves notable improvements in structural preservation (SSIM: 0.904) and feature similarity (FSIM: 0.871) compared to the baseline. This confirms that the channel attention mechanism of EnhanceNet effectively highlights diagnostically relevant features and enhances feature difference through

Metrics	Baseline (NSST only)	NSST + DNSN	NSST + ENSN	Full model (NSST + CDEN)
Pair I				
SF	26.83	28.47	28.74	29.51
IE	5.32	5.48	5.53	5.62
MI	2.85	3.16	3.21	3.37
QAB/F	0.52	0.56	0.58	0.61
IQI	0.43	0.46	0.47	0.50
PSNR	32.67	34.92	34.14	36.42
SSIM	0.861	0.892	0.901	0.924
FSIM	0.826	0.854	0.868	0.883
Pair II				
SF	33.42	34.86	35.57	36.61
IE	5.04	5.17	5.22	5.30
MI	3.58	3.92	4.02	4.30
QAB/F	0.64	0.69	0.71	0.76
IQI	0.42	0.45	0.46	0.48
PSNR	34.27	36.23	35.47	38.16
SSIM	0.878	0.904	0.912	0.936
FSIM	0.841	0.868	0.879	0.895
Pair III				
SF	20.95	21.84	22.17	23.04
IE	4.76	4.89	4.94	5.01
MI	3.45	3.72	3.87	4.06
QAB/F	0.65	0.71	0.73	0.76
IQI	0.41	0.44	0.45	0.47
PSNR	32.16	34.02	33.65	35.54
SSIM	0.857	0.887	0.893	0.916
FSIM	0.828	0.854	0.867	0.889
Pair IV				
SF	40.81	42.43	43.05	44.26
IE	5.39	5.52	5.59	5.67
MI	2.23	2.42	2.47	2.62
QAB/F	0.59	0.64	0.66	0.69
IQI	0.43	0.45	0.46	0.48
PSNR	36.54	38.42	37.93	40.08
SSIM	0.891	0.914	0.921	0.943
FSIM	0.853	0.877	0.885	0.901
Pair V				
SF	17.83	18.65	18.92	19.49
IE	5.01	5.12	5.17	5.25
MI	3.06	3.37	3.42	3.64
QAB/F	0.62	0.67	0.69	0.72
IQI	0.41	0.44	0.45	0.47
PSNR	31.24	32.86	32.47	33.92
SSIM	0.843	0.872	0.881	0.896
FSIM	0.823	0.841	0.852	0.865
Pair VI				
SF	33.15	34.69	35.24	36.34
IE	4.45	4.56	4.63	4.71
MI	2.94	3.21	3.27	3.47
QAB/F	0.57	0.62	0.64	0.67
IQI	0.51	0.54	0.56	0.58
PSNR	33.62	35.17	34.95	36.58
SSIM	0.882	0.907	0.917	0.932
FSIM	0.847	0.867	0.878	0.896
Average				
SF	28.83	30.16	30.62	31.54
Continued				

Metrics	Baseline (NSST only)	NSST + DNSN	NSST + ENSN	Full model (NSST + CDEN)
IE	5.00	5.12	5.18	5.26
MI	3.02	3.30	3.38	3.58
QAB/F	0.60	0.65	0.67	0.70
IQI	0.43	0.46	0.48	0.50
PSNR	33.42	35.27	34.77	36.78
SSIM	0.864	0.896	0.904	0.925
FSIM	0.832	0.857	0.871	0.888
The improvement over baseline (%)	Baseline	+ 5.53%	+ 6.12%	+ 9.64%

Table 3. Results of the ablation study.

its adaptive weighting, EQU (10) to EQU (12). However, without prior denoising, its performance in noise-sensitive metrics, such as PSNR (34.77 dB), remains suboptimal, as it must operate on potentially noisy inputs.

The complete model (NSST + CDEN) consistently outperforms all other variants across all evaluation metrics, with particularly significant improvements in PSNR (+ 3.36 dB over the baseline), SSIM (+ 0.061 over the baseline), and FSIM (+ 0.056 over the baseline).

Discussion

Comparative analysis with existing methods

The proposed IFM distinguishes itself from existing CNN + PCNN using the synergistic integration of NSST, a CDEN, and a bio-inspired PCNN guided by the Novel Sum Modified Laplacian (NSML). While previous studies, such as[48], have verified notable success using CNNs for MFIFM, their utility was constrained to focus classification and binary decision maps, limiting generalizability across broader imaging modalities. Similarly, the CNN-based pixel activity computation for WM generation in[49] presented improved saliency encoding yet required robustness in infrared–visible fusion tasks.

In contrast, the proposed model displays versatility across varied imaging domains, including medical modalities such as MRI, CT, SPECT, and PET. The integration of NSST ensures superior frequency localization and directional sensitivity, overcoming the limitations of the TDBIFM, which challenges directional aliasing and frequency bias. Moreover, adopting NSML for HFSB fusion preserves intricate structural data, surpassing conventional statistical measures in PCNN-driven fusion.

Furthermore, unlike[50], which integrated Deep CNNs with a Discrete Gravitational Search Algorithm (DGSA) but faced barriers in real-time applicability due to its computational burden, the proposed model retains computational tractability via channel attention mechanisms and lightweight dual-branch training. Including bio-inspired dynamics using PCNN improves edge preservation while mitigating the common problem of contrast degradation obtained in existing research models[51–53].

Recent innovations in fuzzy-logic-based fusion, including intuitionistic[57,59] and Fermatean fuzzy systems[61], proposal interpretability, and structural fidelity. However, they frequently face challenges due to overdependence on rule-based heuristic parameters. This model addresses this by embedding uncertainty handling within a data-driven, End-To-End that dynamically learns spatial and frequency-domain fusion cues.

Implications for medical and multimodal image fusion

This model substantially improves the integration of MIF modalities, where preserving texture granularity and diagnostic features is essential. The dual-branch denoising and improvement model ensures signal reconstruction integrity, especially in noisy or compressed attainment environments such as low-dose CT and thermal-visible fusion. The channel-wise attention modulation adapts fusion weights in real-time, supporting clinical applicability across anatomy-specific variations.

Using NSST and PCNN augments visual interpretability and modality consistency, a long-standing challenge in anatomical alignment during fusion. The consistent spatial feature activation across networks enhances the robustness of decision-making in applications such as tumor detection, organ delineation, and pathology mapping.

Methodological limitations

Despite the performance gains, the proposed model has several limitations. First, the reliance on a dual-network model increases the training time and parameter load, which may constrain deployment in low-resource or edge-computing environments. Second, although NSST proposes high directional selectivity, it introduces redundancy that may not be optimal for all fusion scenarios, particularly when real-time constraints dominate.

Additionally, while PCNN-based fusion enhances perceptual quality, its neuron firing threshold and iterative nature introduces sensitivity to initial parameters, which can potentially affect convergence stability. The NSML's performance depends on accurate high-frequency FE, which may degrade under extreme noise or motion blur conditions not seen in training datasets.

Future work

To overcome these limitations, future research will optimize the model using pruning and quantization methods to reduce model complexity without compromising fusion quality. Adaptive NSST parameter tuning via metalearning will be investigated to improve cross-domain generalizability. Also, this work plans to explore hybrid models that combine fuzzy-set-based interpretability with DL adaptability, potentially enabling rule-guided yet data-driven fusion control.

Incorporating real-time performance benchmarking and edge-device deployment approaches will also be vital to translating the model into clinical and field use cases. Finally, user-interpretable quality evaluation metrics, particularly those tailored for medical diagnosis, will be integrated to validate fusion efficacy beyond conventional metrics.

Conclusion and future work

Image Fusion remains a challenging domain, particularly for preserving critical details from Source Images in multi-modal medical applications. To address this, the present study proposed a CNN and Shearlet-based fusion model tailored for MMMI. The model incorporates a frequency-level denoising and enhancement network to mitigate quality degradation introduced during NSST decomposition. Enhanced fusion methods were applied to LFSB and HFSB, resulting in improved overall fusion quality.

Extensive experiments were tested using standard benchmark datasets to evaluate the proposed model across multiple performance metrics, including edge information preservation (QAB/F), Image Quality Index (IQI), Spatial Frequency (SF), Mutual Information (MI), and Image Entropy (IE). Comparative analysis demonstrated that the proposed model consistently outperformed existing *state-of-the-art* methods, producing fused images with superior structural detail and diagnostic relevance. These findings recommend that the proposed method has a robust probability of integration into clinical imaging workflows that require accurate MMIF.

Limitations of the work

One limitation of the current study is the relatively small dataset comprising 180 multimodal image pairs, which, while diverse in modality and structure, may not comprehensively reflect the variability encountered in larger clinical imaging repositories.

Although the proposed CNN + NSST-based fusion model proved superior performance across all evaluation metrics within this dataset, future work will validate the method on larger and more diverse datasets, potentially involving cross-institutional medical imaging archives, to measure scalability and clinical generalizability. The recommended method must be used with real-time applications and other types of MMIF in the future.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 21 January 2025; Accepted: 28 July 2025
Published online: 12 August 2025

References

1. James, A. P. & Dasarathy, B. V. Medical image fusion: A survey of the state of the art. *Inf. Fusion* **19**, 4–19 (2014).
2. Wang, Z., Ziou, D., Armenakis, C., Li, D. & Li, Q. A comparative analysis of image fusion methods. *IEEE Trans. Geosci. Remote Sens* **43**(6), 1391–1402 (2005).
3. Biswas, B. et al. Medical image fusion by combining SVD and shearlet transform. In: *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India* 148–153 (2015).
4. Ullah, H. et al. Multimodality medical images fusion based on local-features fuzzy sets and novel sum-modified-Laplacian in non-subsampled shearlet transform domain. *Biomed. Signal Process. Control* **57**, 101724 (2020).
5. Fu, J., Li, W., Du, J. & Xiao, B. Multimodal medical image fusion via laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy. *Comput. Biol. Med.* **126**, 104048 (2020).
6. Li, X. & Zhao, J. A novel Multi-Modal Medical Image Fusion algorithm. *J. Ambient Intell. Humaniz. Comput.* **12**, 1995–2002 (2021).
7. Almasri, M. M. & Alajlan, A. M. Artificial intelligence-based multimodal medical image fusion using hybrid S^2 optimal CNN. *Electronics* **11**(14), 2124 (2022).
8. He, H. A deep research in classic classification network. *MS E* **740**(11), 012152 (2020).
9. Kong, W., Li, C. & Lei, Y. Multimodal medical image fusion using convolutional neural network and extreme learning machine. *Front. Neurorobot.* **16**(16), 1050981 (2022).
10. Piella, G. & Heijmans, H. A new quality metric for image fusion. In: *Proceedings of the IEEE International Conference on Image Processing* III-173–III-176 (2003).
11. Vincent, A. C. S. R. & Sengan, S. Edge computing-based ensemble learning model for health care decision systems. *Sci. Rep.* **14**, 26997 (2024).
12. Toet, A. Image fusion by a ratio of low-pass pyramid. *Pattern Recogn. Lett.* **9**(4), 245–253 (1989).
13. Li, H., Manjunath, B. S. & Mitra, S. K. Multisensor image fusion using the wavelet transform. *J. Graph. Model Image Process.* **57**, 235–245 (1995).
14. Singh, S., Gupta, D., Anand, R. S. & Kumar, V. Nonsubsampled shearlet-based CT and MR medical image fusion using biologically inspired spiking neural network. *Biomed. Signal Process. Control* **18**, 91–101 (2015).
15. Liu, Y. et al. Multi-focus image fusion with a deep convolutional neural network. *Inform Fusion* **36**, 191–207 (2017).
16. Labate, D., Lim, W. Q., Kutyniok, G. & Weiss, G. Sparse multidimensional representation using shearlets. *Proc. SPIE* **5914**, 59140U (2005).
17. Tan, L. & Yu, X. Medical image fusion based on fast finite shearlet transform and sparse representation. *Comput. Math. Methods Med.* **2019**, 3503267 (2019).
18. Wang, Z. & Ma, Y. Medical image fusion using m-PCNN. *Inf. Fusion* **9**, 176–185 (2008).
19. Zhu, Z. Q., Chai, Y., Yin, H. P., Li, Y. X. & Liu, Z. D. A novel dictionary learning approach for multi-modality medical image fusion. *Neurocomputing* **214**, 471–482 (2016).
20. Robert Vincent, A. C. S. & Sengan, S. Effective clinical decision support implementation using a multi filter and wrapper optimisation model for Internet of Things based healthcare data. *Sci. Rep.* **14**, 21820 (2024).

21. Ding, S. F., Du, P., Zhao, X. Y., Zhu, Q. B. & Xue, Y. BEMD image fusion based on PCNN and compressed sensing. *Soft Comput.* **23**, 10045–10054 (2019).
22. Zhao, W. D. & Lu, H. C. Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation. *IEEE Trans. Instrument. Meas.* **66**, 2283–2294 (2017).
23. Yang, S., Wang, M., Lu, Y., Qi, W. & Jiao, L. Fusion of multiparametric SAR images based on SW-nonsubsampled contourlet and PCNN. *Signal Process.* **89**, 2596–2608 (2009).
24. Daniel, E. Optimum wavelet-based homomorphic medical image fusion using hybrid genetic-grey wolf optimization algorithm. *IEEE Sens. J.* **18**, 6804–6811 (2018).
25. Arif, M. & Wang, G. J. Fast curvelet transform through genetic algorithm for multimodal medical image fusion. *Soft Comput.* **24**, 1815–1836 (2020).
26. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **7**, 3142–3155 (2017).
27. Liu, Y. et al. Deep learning for pixel-level image fusion: recent advances and prospects. *Inform Fusion* **42**, 158–173 (2018).
28. Xu, X. Z., Shan, D., Wang, G. Y. & Jiang, X. Y. Multimodal medical image fusion using PCNN optimized by the QPSO algorithm. *Appl. Soft Comput.* **46**, 588–595 (2016).
29. Du, J., Fang, M. E., Yu, Y. F. & Lu, G. An adaptive two-scale biomedical image fusion method with statistical comparisons. *Comput. Meth. Prog. Biol.* **196**, 105603. https://doi.org/10.1016/j.cmpb.2020.105603 (2020).
30. Du, J., Li, W. S. & Tan, H. L. Three-layer image representation by an enhanced illumination-based image fusion method. *IEEE J. Biomed Health.* **24**, 1169–1179 (2020).
31. Du, J., Li, W. S. & Tan, H. L. Three-layer medical image fusion with tensor-based (2020c).
32. Jiang, Q., Jin, X., Hou, J. Y., Lee, S. & Yao, S. W. Multi-sensor image fusion based on interval type-2 fuzzy sets and regional features in non-subsampled shearlet transform domain. *IEEE Sens. J.* **18**, 2494–2505 (2018).
33. Gao, Y., Ma, S. W., Liu, J. J., Liu, Y. Y. & Zhang, X. X. Fusion of medical images based on salient features extraction by PSO optimized fuzzy logic in NSST domain. *Biomed. Signal Process.* **69**, 102852 (2021).
34. Asha, C. S., Lal, S., Gurupur, V. P. & Saxena, P. U. P. Multi-modal medical image fusion with adaptive weighted combination of NSST bands using chaotic grey wolf optimization. *IEEE Access* **7**, 40782–40796 (2019).
35. Singh, S. & Anand, R. S. Multimodal medical image sensor fusion model using sparse K-SVD dictionary learning in nonsubsampled shearlet domain. *IEEE Trans. Instrument. Meas.* **69**, 593–607 (2020).
36. Yin, M., Liu, X. N., Liu, Y. & Chen, X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Trans. Instrument. Meas.* **68**, 49–64 (2019).
37. Zhu, R., Li, X. F., Zhang, X. L. & Wang, J. HID: the hybrid image decomposition model for MRI and CT fusion. *IEEE J. Biomed Health.* **26**, 727–739 (2022).
38. Hermessi, H., Mourali, O. & Zagrouba, E. Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Process.* **183**, 108036 (2021).
39. Azam, M. A., Khan, K. B., Salahuddin, S. & Rehman, E. A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques, and quality metrics. *Comput. Biol. Med.* **144**, 105253 (2022).
40. Szegedy, C. et al. A. Rabinovich Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (2015).
41. He, K., Zhang, X., Ren, S. & Sun J. Deep Residual Learning for Image Recognition. arXiv (2015) arXiv:1512.03385
42. Zhang, N., Ruan, S., Lebonvallet, S., Liao, Q. & Zhu, Y. Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation. *Comput. Vis Image Understand* **115**(2), 256–269 (2011).
43. Zeiler, M. D. R. Fergus Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision* 818–833 (Springer, 2014).
44. Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. *Neural Inf. Process. Syst.* **25**.
45. Huang, G., Liu, Z. & Van Der Maaten, L. K.Q. Weinberger Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4700–4708 (2017).
46. Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
47. Long, J. & Shelhamer, E. T. Darrell Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3431–3440 (2015).
48. Liu, Y., Chen, X., Ward, R. & Wang, Z. J. Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* **23**, 1882–1886 (2016).
49. Wang, K., Zheng, M., Wei, H., Qi, G. & Li, Y. Multi-modality medical image fusion using convolutional neural network and contrast pyramid. *Sensors* **20**, 2169 (2020).
50. Parvathy, V. S., Sivakumar, P. & Sampson, J. A novel approach in multimodality medical image fusion using optimal shearlet and deep learning. *Int. J. Imaging Syst. Technol.* **30**, 847–859 (2020).
51. He, C. T., Liu, Q. X., Li, H. L. & Wang, H. X. Multimodal medical image fusion based on IHS and PCA. *Proc. Eng.* **7**, 280–285 (2010).
52. Geng, P., Zheng, X., Zhang, Z., Shi, Y. & Yan, S. Multifocus image fusion with PCNN in shearlet domain. *Res. J. Appl. Sci. Eng. Technol.* **4**, 2283–2290 (2012).
53. Geng, P., Wang, Z., Zhang, Z. & Xiao, Z. Image fusion by pulse couple neural network with shearlet. *Opt. Eng.* **51**, 067005 (2012).
54. Chai, Y., Li, H. & Zhang, X. Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain. *Optik* **123**, 569–581 (2012).
55. Jogin, M. et al. Feature extraction using Convolution neural networks (CNN) and deep learning. 2319–2323. https://doi.org/10.1109/RTEICT42901.2018.9012567 (2018).
56. Wei, D. An online extraction algorithm for image feature information based on convolutional neural network. *Mobile Inf. Syst.* **2022**, 11 (2022).
57. Häuser, S. & Steidl, G. Fast finite shearlet transform. arXiv 2012 arXiv:1202.1773.
58. Lim, W. Q. The discrete shearlet transform: A new directional transform and compactly supported shearlet frames. *IEEE Trans. Image Process.* **19**, 1166–1180 (2010).
59. Easley, G., Labate, D. & Lim, W. Q. Sparse directional image representations using the discrete shearlet transform. *Appl. Comput. Harmon. Anal.* **25**, 25–46 (2008).
60. Ranganath, H. S. Kuntimad, G., Johnson, J. L. Pulse coupled neural networks for image processing. In: *Proceedings of the IEEE Visualize the Future* 37–43 (1995).
61. Yang, Y. et al. Multimodal medical image fusion based on fuzzy discrimination with structural patch decomposition. *IEEE J. Biomed Health.* **23**, 1647–1660 (2019).
62. Biswas, B., Ghoshal, S., Chatterjee, P., Chakrabarti, A. & Dey, K. N. Medical image fusion by combining SVD and shearlet transform. In: *2015 2nd International Conference on Signal Processing and Integrated Networks*.
63. Tan, L. & Yu, X. Medical image fusion based on fast finite shearlet transform and sparse representation. *Comput. Math. Methods Med.* **3**(2019), 3503267 (2019).
64. Hou, R. et al. Brain CT and MRI medical image fusion using convolutional neural networks and a dual-channel spiking cortical model. *Med Biol Eng Comput* **57**, 887–900 (2019).

65. Vladimir P, Costas X (2004) Evaluation of image fusion performance with visible differences. In: *8th European Conference on Computer Vision, ECCV 2004*, Lecture Notes in Computer Science, vol 3023, pp 380–391
66. Zheng, Y., Essock, E. A., Hansen, B. C. & Haun, A. M. A new metric based on extended spatial frequency and its application to DWT-based fusion algorithms. *Inform. Fusion* **8**, 177–192 (2007).
67. Hossny, M., Nahavandi, S. & Creighton, D. Comments on 'information measure for performance of image fusion. *Electron. Lett.* **44**, 1066–1067. https://doi.org/10.1049/el:20081754 (2008).
68. Wang, Z. & Bovik, A. C. A universal image quality index. *IEEE Signal Process. Lett.* **9**, 81–84 (2002).
69. Jin, X. et al. Multifocus color image fusion based on NSST and PCNN. *J Sens* **2016**, 8359602 (2016).
70. Jiang, Q. et al. A lightweight multimode medical image fusion method using similarity measure between intuitionistic fuzzy sets joint laplacian pyramid. *IEEE Trans. Emerg. Top. Comput. Intell.* **7**(3), 631–647 (2023).
71. Li, J. et al. Multi-focus image fusion based on multiscale fuzzy quality assessment. *Digital Signal Process.* **153**, 104592 (2024).
72. Jie, Y., Li, X., Tan, T., Yang, L. & Wang, M. Multi-modality image fusion using fuzzy set theory and compensation dictionary learning. *Opt. Laser Technol.* **181**, 112001 (2025).
73. Haribabu, M. & Guruviah, V. An improved multimodal medical image fusion approach using intuitionistic fuzzy set and intuitionistic fuzzy cross-correlation. *Diagnostics* **13**(14), 2330 (2023).
74. Haribabu, M. & Guruviah, V. FFSWOAFuse: Multi-modal medical image fusion via fermatean fuzzy set and whale optimization algorithm. *Comput. Biol. Med.* **189**, 109889 (2025).
75. Haribabu, M., Guruviah, V. & Yogarajah, P. An overview of recent advancements in multimodal medical image fusion techniques for better diagnosis. *Curr. Med. Imaging Rev.* **19**(7), 673–694 (2023).
76. Zhou, L., Cai, J. & Ding, S. The identification of ice floes and calculation of sea ice concentration based on a deep learning method. *Remote Sens.* **15**(10), 2663 (2023).
77. Cai, J. et al. Broken ice circumferential crack estimation via image techniques. *Ocean Eng.* **259**, 111735 (2022).
78. Wang, W. et al. Low-light image enhancement based on virtual exposure. *Signal Process. Image Commun.* **118**, 117016 (2023).
79. Zhou, G., Li, J., Tian, Z., Xu, J. & Bai, Y. The extended stumpf model for water depth retrieval from satellite multispectral images. *IEEE J. Select Top. Appl. Earth Observ. Remote Sens.* **17**, 6779–6790. https://doi.org/10.1109/JSTARS.2024.3368761 (2024).
80. Zhou, G., Qian, L. & Gamba, P. A novel iterative self-organizing pixel matrix entanglement classifier for remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **62**, 1–21. https://doi.org/10.1109/TGRS.2024.3424227 (2024).
81. Wang, P. et al. Server-initiated federated unlearning to eliminate impacts of low-quality data. *IEEE Trans. Serv. Comput.* **17**(3), 1196–1211 (2024).
82. Zhou, Z. et al. Resource-saving and high-robustness image sensing based on binary optical computing. *Laser Photon. Rev.* **19**(7), 2400936 (2025).
83. Wang, W., Yin, B., Li, L., Li, L. & Liu, H. A low light image enhancement method based on Dehazing physical model. *Comput. Model. Eng. Sci.* **143**(2), 1595–1616 (2025).
84. Lu, L. et al. Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging. *Adv. Photon.* **6**(4), (2024).
85. Zhang, Z., Liu, C., Hao, F. & Liu, Z. Style-transfer-based unsupervised change detection from heterogeneous images. *IEEE Trans. Aerosp. Electron. Syst.* **61**(3), 6537–6550 (2025).
86. Zhuang, J., Chen, W., Guo, B. & Yan, Y. Infrared weak target detection in dual images and dual areas. *Remote Sens.* **16**(19), 3608 (2024).
87. Cao, Y. et al. IBPL: Information bottleneck-based prompt learning for graph out-of-distribution detection. *Neural Netw.* **188**, 107381 (2025).
88. Zhang, J., Li, Y., Fukuda, T. & Wang, B. Urban safety perception assessments via integrating multimodal large language models with street view images. *Cities* **165**, 106122 (2025).
89. Wang, W., Yuan, X., Wu, X. & Liu, Y. Fast image dehazing method based on linear transformation. *IEEE Trans. Multimedia* **19**(6), 1142–1155 (2017).
90. Yin, L. et al. Convolution-transformer for image feature extraction. *Comput. Model. Eng. Sci.* **141**(1), 87–106 (2024).
91. Zheng, Y. et al. Fast-zoom and high-resolution sparse compound-eye camera based on dual-end collaborative optimization. *Opto-Electron. Adv.* **8**, 240285 (2025).
92. Zhou, G. et al. True2 orthoimage map generation. *Remote Sens.* **14**(17), 4396 (2022).
93. Chen, J. et al. Specular removal of industrial metal objects without changing lighting configuration. *IEEE Trans. Ind. Inform.* **20**(3), 3144–3153 (2024).
94. Li, S., Chen, J., Peng, W., Shi, X. & Bu, W. A vehicle detection method based on disparity segmentation. *Multimedia Tools Appl.* **82**(13), 19643–19655 (2023).
95. Tu, B. et al. Anomaly detection in hyperspectral images using adaptive graph frequency location. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15 (2024).
96. Wang, Z., Wang, C., Li, X., Xia, C. & Xu, J. MLP-Net: Multilayer perceptron fusion network for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **63**, 1–13 (2025).
97. Wang, Z., Zhang, Z., Qi, W., Yang, F. & Xu, J. FreqGAN: Infrared and visible image fusion via unified frequency adversarial learning. *IEEE Trans. Circuits Syst. Video Technol.* **35**(1), 728–740 (2025).
98. Wang, T. et al. TASTA: Text-assisted spatial and temporal attention network for video question answering. *Adv. Intell. Syst.* **5**(4), 2200131 (2023).
99. Deng, S. et al. Learning to compose diversified prompts for image emotion classification. *Comput. Vis. Media* **10**(6), 1169–1183 (2024).
100. Shi, G. et al. One for all: A unified generative framework for image emotion classification. *IEEE Trans. Circuits Syst. Video Technol.* **34**(8), 7057–7068 (2024).
101. Li, R. et al. UE-extractor: A grid-to-point ground extraction framework for unstructured environments using adaptive grid projection. *IEEE Robot. Autom. Lett.* **10**(6), 5991–5998 (2025).
102. Chen, X. & Jing, R. Video super-resolution based on deformable 3D convolutional group fusion. *Sci. Rep.* **15**(1), 9050 (2025).
103. Al-Selwi, M. et al. Enhancing object pose estimation for RGB images in cluttered scenes. *Sci. Rep.* **15**(1), 8745 (2025).
104. Yao, F., Zhang, H., Gong, Y., Zhang, Q. & Xiao, P. A study of enhanced visual perception of marine biology images based on diffusion-GAN. *Complex. Intell. Syst.* **11**(5), 227 (2025).
105. Wang, B. et al. Single-shot super-resolved fringe projection profilometry (SSSR-FPP): 100,000 frames-per-second 3D imaging with deep learning. *Light Sci. Appl.* **14**(1), 70 (2025).
106. Gu, K. et al. Perceptual information fidelity for quality estimation of industrial images. *IEEE Trans. Circuits Syst. Video Technol.* **35**(1), 477–491 (2025).
107. Chen, J. et al. A flow feedback traffic prediction based on visual quantified features. *IEEE Trans. Intell. Transp. Syst.* **24**(9), 10067–10075 (2023).

Acknowledgements

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-17).

Author contributions

Conceptualization : Sudhakar SenganMethodology : Sudhakar Sengan, Amr YousefSoftware : Praveen Gugulothu, Julian L. WebberValidation : Julian L. Webber, Amr YousefFormal analysis : Sudhakar SenganInvestigation : Sudhakar Sengan, Amr YousefResources : Praveen Gugulothu, Amr YousefData Curation : Roobaea Alroobaea, Abolfazl MehbodniyaWriting - Original Draft : Sudhakar SenganWriting - Review & Editing : Sudhakar Sengan, Abolfazl MehbodniyaVisualization : Roobaea Alroobaea, Amr Yousef.

Funding

This research was funded by Taif University, Taif, Saudi Arabia, Project Number (TU-DSPP-2024-17).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.S. or A.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.