IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 August 2023, accepted 7 September 2023, date of publication 11 September 2023, date of current version 18 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3313977

TOPICAL REVIEW

Recent Advancements and Future Prospects in Active Deep Learning for Medical Image Segmentation and Classification

TARIQ MAHMOOD [1,2], AMJAD REHMAN [1], (Senior Member, IEEE), TANZILA SABA [1], (Senior Member, IEEE), LUBNA NADEEM [3], AND SAEED ALI OMER BAHAJ [4,5]

[1] Artificial Intelligence and Data Analytics (AIDA) Laboratory, CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia
[2] Faculty of Information Sciences, University of Education, Vehari Campus, Vehari 61100, Pakistan
[3] Department of Telecommunication Engineering, University of Engineering and Technology Taxila, Taxila 47050, Pakistan
[4] MIS Department, College of Business Administration, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[5] Department of Computer Engineering, College of Engineering and Petroleum, Hadhramout University, Mukalla, Hadhramout, Yemen

Corresponding authors: Saeed Ali Omer Bahaj (saeedalibahaj@gmail.com) and Tariq Mahmood (tmsherazi@ue.edu.pk)

This research is supported by the Artificial Intelligence & Data Analytics Lab (AIDA), CCIS, Prince Sultan University, Riyadh, Saudi Arabia. The author would also like to thank Prince Sultan University, Riyadh, Saudi Arabia, for support.

ABSTRACT Medical images are helpful for the diagnosis, treatment, and evaluation of diseases. Precise medical image segmentation improves diagnosis and decision-making, aiding intelligent medical services for better disease management and recovery. Due to the unique nature of medical images, image segmentation algorithms based on deep learning face problems such as sample imbalance, edge blur, false positives, and false negatives. In view of these problems, researchers primarily improve the network structure but rarely improve from the unstructured aspect. The paper tackles these challenges, accentuating the limitations of deep convolutional neural network-based methods and proposing solutions to reduce annotation costs, particularly in complex images, and introduces the improvement strategies to solve the problems of sample imbalance, edge blur, false positives, and false negatives. Additionally, the article introduces the latest deep learning-based applications in medical image analysis, covering segmentation, image acquisition, enhancement, registration, and classification. Moreover, the article provides an overview of four cutting-edge deep learning models, namely convolutional neural network (CNN), deep belief network (DBN), stacked autoencoder (SAE), and recurrent neural network (RNN). The study selection involved searching benchmark academic databases, collecting relevant literature and appropriate indicator for analysis, emphasizing DL-based segmentation and classification approaches, and evaluating performance metrics. The research highlights clinicians' and scholars' obstacles in developing an efficient and accurate malignancy prognostic framework based on state-of-the-art deep-learning algorithms. Furthermore, future perspectives are explored to overcome challenges and advance the field of medical image analysis.

INDEX TERMS Medical imaging, beast cancer, image segmentation, artificial intelligence, deep learning, healthcare, inclusive innovation.

I. INTRODUCTION

Medicine's global significance in enhancing individual health and quality of life is universally acknowledged. Medical research heavily relies on medical image analysis, a pivotal cornerstone for advancing clinical investigations and ensuring accurate diagnoses. The evolution of medical technology has led to diverse imaging modalities, enriching tools for crucial pursuits. The various imaging techniques, including computed tomography (CT) [1], magnetic resonance (MR) imaging [2], positron emission tomography (PET) [3], ultrasound (US), and X-ray imaging [4] are indispensable in clinical diagnosis. Medical images predominantly comprise single-channel grayscale representations

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar.

with diminished contrast and a low signal-to-noise ratio. Conventional encoder-decoder networks fail to capture intricate fine-grained attributes and encompass global contextual semantic characteristics [5]. Consequently, the precision of segmenting medically relevant entities presenting substantial shape disparities becomes a formidable challenge. Image segmentation is a key research direction in medical image processing, focusing on segmenting abnormal tissue and structures in images [6]. This is crucial for evaluating and treating patients and is growing in the medical field. Image segmentation extracts specific tissues or structures from images, providing quantitative information about specific tissues to physicians. The goal of medical image segmentation is to segment the ROI of the image and eliminate unnecessary information [7]. Moreover, the constrained focus on the affected target omits essential boundary features, compounding the intricacy of achieving comprehensive segmentation, particularly for damaged tissue [8]. However, a profound understanding of learning techniques related to medical image processing and computer vision is essential for researchers and clinicians [9], [10], [11]. Medical image processing, notably image segmentation, is gaining prominence for informed patient evaluation and treatment planning. However, massive data is not effectively utilized, as medical test results and images are stored without analysis. This lack of utilization of patient data, especially medical images, leads to the wastage of resources and makes it difficult for clinical workers to uncover hidden trends and patterns in the evolution of diseases. Therefore, applying machine learning-based models to analyze medical images holds significant potential [12]. Intelligent medical imaging diagnosis involves multiple essential steps, including acquiring a substantial volume of high-quality image data, preprocessing the images, detecting and segmenting lesions, and extracting relevant information from segmented lesions for classification [13], [14].

Several methods have been proposed domestically and internationally, from traditional to deep learning-based. Deep learning (DL), a crucial segment of machine learning, has gained prominence in medical image analysis, driving the pursuit of artificial intelligence (AI) in medical imaging [15], [16]. It significantly contributes to computer vision and medical image analysis, using neural networks to aid specialists in diagnosis, reduce radiologists' workload, and enhance efficiency. The advent of big data and AI technology has revolutionized medical image analysis, enabling the analysis of massive amounts of medical image data for intelligent diagnosis, clinical decision-making, and treatment prognosis [17], [18]. DL-based approaches are extensively adopted for pattern recognition, primarily due to their unique feature of being trainable as a complete program [19], [20], [21]. The effectiveness of the deep learning algorithm is predicated on vast and high-quality image datasets to ensure precise learning and training [22], [23]. Utilizing neural networks, deep learning aids in efficient image processing, aiding specialists in diagnosis and enhancing diagnostic accuracy. Unlike traditional methods, deep learning autonomously extracts features from input data, reducing human intervention [24], [25]. Deep learning frameworks, such as CNN, DBN, Stacked Autoencoders [26], and recurrent neural network (RNN), are the three main types of structured networks that form the foundations of deep learning [27], [28].

Figure 1, portrays the step-wise process of medical image analysis, providing a comprehensive overview of the critical stages involved. The diagram emphasizes the sequential tasks that extract meaningful information from medical images, facilitating intelligent diagnosis, decision-making, and prognosis. Various medical imaging techniques have unique features, requiring cautious selection by physicians. Oktay et al. [29] developed an Attention Gate (AG) model for medical images, autonomously identifying target structures and learning significant features through training. This eliminates the need for external modules for localizing tissues or organs using convolutional neural networks. Kamnitsas et al. [30] used dual-channel architecture to combine local and global information, enhancing image segmentation precision by simultaneously processing multiple dimensions of input images. Wang et al. [31] developed an automated wound image analysis system using ConvNet to segment wound regions and assess wound condition. The system uses an SVM classifier to determine infection and the Gaussian process regression algorithm for wound healing evaluation. Despite the initial success of using deep learning in clinical practice, challenges persist in recognizing multiple intersecting diseases and complex lesions, necessitating the development of additional DL-based techniques. These analytical undertakings encompass the identification of obstacles, the development of predictive models, and other vital elements that constitute the foundational components within the context of clinical intelligence-guided decision-making [32], [33], [34].

This study examines the operational utilization of deep learning (DL)-based models in medical image analysis, focusing on critical application domains. It evaluates the limitations and challenges of DL-based models, including architectural simplicity, inadequacies in existing training methodologies, prolonged training periods, unlabeled data annotation, and countering sample biases. The study also identifies intricacies associated with DL-based models, such as reliance on rudimentary deep network architectures, unexplored potential in extant training methods, and the complexity of model parameters. The evolution of updated automatic labeling technology is crucial for addressing adversarial examples and overcoming the effectiveness of prevailing regularization techniques. The study prognosticates the trajectory of DL-based models in medical image analysis, focusing on cross-organizational synergies, expansive image datasets, and continuous methodological

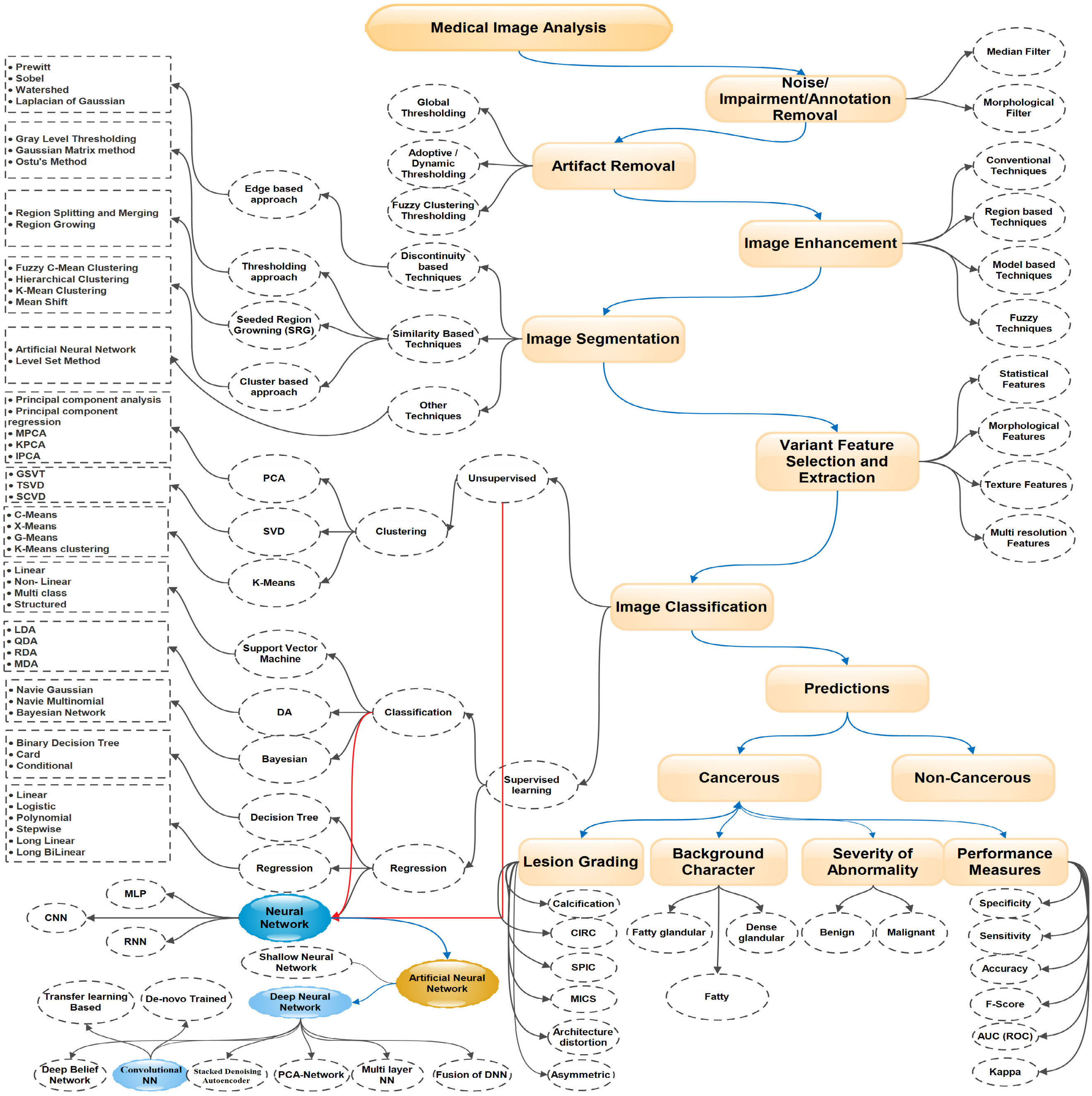


FIGURE 1. The figure depicts the step-wise process of medical image analysis. It highlights the sequential tasks to extract valuable information from medical images, enabling intelligent diagnosis, decision-making, and prognosis.

strategy evolution. It highlights the advancements made in medical imaging and highlights the challenges needed for fully automated cancer tumor identification.

The significant objective of the proposed study are as under:

- The study explores integrating DL-based models in medical image investigation, covering data acquisition, preprocessing, detection, segmentation, and classification methodologies.
- The primary aim of the review study is to guide researchers in developing advanced computer-aided design (CAD) techniques that may aid radiologists in the timely detection, segmentation, and grading cancers abnormalities.
- Deep learning models have advanced medical image analysis, but their implementation in clinical settings faces limitations. Techniques like CNN, LSTM, GAN, attention mechanisms, graph models, and transfer

learning are used, but their implementation remains constrained.
- Deep learning models face challenges in collecting medical imaging data that meets clinical standards due to sensitivity to variations in image accession parameters, resolution, and patient demographics. This results in heterogeneity in data, making it difficult for models to generalize across different data types. Future research should use cognitive thinking and training techniques to improve DL-based models' generalization in medical image applications.
- The study highlights suboptimal surveillance, transfer learning, and multi-task learning strategies for improved classification predictions in limited labeled data.
- This review discusses innovative computer-aided diagnosis (CAD) methods for the early detection of cancerous tumors, focusing on DL-based and hybrid approaches for tumor prognosis.
- The study addresses challenges in medical image analysis and offers future perspectives to overcome them. Deep learning models may not fully understand disease mechanisms, limiting their effectiveness in diagnosis and treatment.

The paper's organization is as under:, Section III a brief overview of pertinent research on advanced techniques and cutting-edge advances in medical image analysis is provided. Section IV is dedicated to the exposition of medical image registration and segmentation. Further insights into medical image grading and recognition are provided in Section V. Deep learning applications and development are outlined in Section Section VI, with a focus on advancements in deep learning and its applications in medical image analysis in Section VII. Finally, Section VIII a summarizes the study's findings

II. LITERATURE REVIEW SELECTION CRITERIA

In this analytical benchmark, the study selection process consisted of four distinct parts. The methodology involved several steps: (1) conducting keyword searches in prominent academic libraries and databases such as IEEE Xplore, ACM Digital Library, Scopus, Google Scholar, Science Direct, PubMed, and Web of Science; (2) collecting pertinent materials and removing duplicate entries; (3) selecting a suitable benchmark for analysis, specifically focusing on DL-based detection, segmentation, and classification approaches applied to medical imaging multi-modalities; and (4) assessing the segmentation and classification architecture performance using benchmark evaluation metrics. The search phrases and logical expressions used in the database are enumerated as follows: The terms "DL-based," "deep convolutional neural network," "convolutional neural network," "CNN," and "DCNN" are commonly used in the field of healthcare and medical imaging. These terms are often associated with analyzing and interpreting various medical images, such as those related to the 'brain,' or 'skin,' or 'breast,' or 'chest,' or 'pulmonary,' or 'lung,' or 'nodules,' or 'tumors,' or 'lesions' or 'cancer,' or 'detection,' or 'diagnosing,' or 'predicting,' or 'identification,' or 'segmentation,' or 'localization,' 'classification,' or ' grading,' 'characterization,' or 'CAD,' or 'Medical Images,' or 'X-ray,' or 'Ultrasound,' or 'histopathological,' 'microscopic,' or 'MRI,' 'MR images,' 'magnetic resonance imaging,' or 'CT,' or 'mammography,' or 'mammogram,.' The study in this area focuses on the detection, segmentation, and grading of distinct medical image modalities. Subsequently, using the aforementioned rigorous scrutiny and selection criteria, 312 research articles deemed most relevant were chosen from those published between the timeframe of 2010 to 2023. The distribution of these articles, categorized by year, is visually shown in Figure 2.

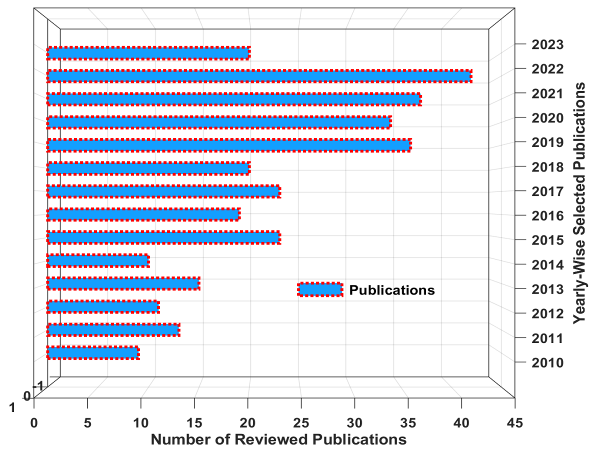


FIGURE 2. Publications focusing on detecting, segmenting, and categorizing cancer lesions were examined yearly.

A. MEDICAL IMAGING MODALITIES

Each imaging modality utilized in the healthcare field possesses unique data and features. The comprehensive array of electromagnetic (EM) scanning methodologies used for monitoring and diagnosing various medical conditions within the human body is detailed in Table 1 and illustrated in Figure 3. Each scanning modality has a unique frequency and wavelength, manifesting dissimilar characteristics [8]. In response to interacting with an object, electromagnetic wavelengths endure dispersion, reflection, or reception. The Magnetic Resonance Imaging (MRI) process produces a magnetic field that induces the alignment of protons within the body. These protons are stimulated upon exposure to a radio-frequency signal, resulting in a rotational motion at an angle to the magnetic field. Notably, the frequency spectrum between 3×1016 and 3×1019 Hz, relevant to computed tomography (CT) and X-ray imaging, is associated with significant radiation levels, posing potential public health risks. Gamma radiation, utilized in nuclear imaging techniques such as Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT),

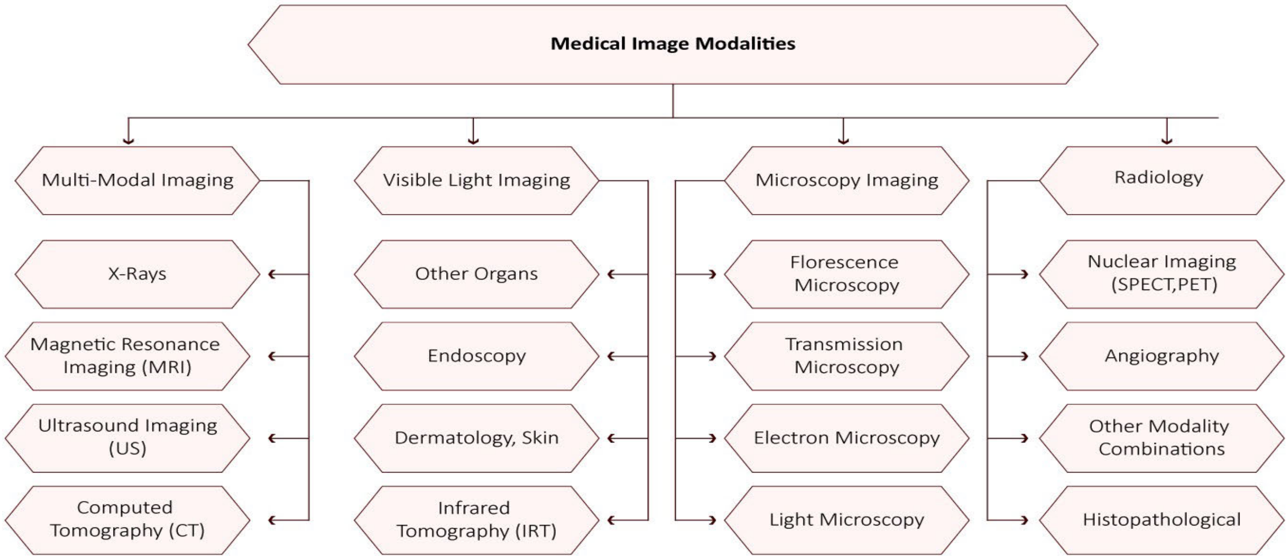


FIGURE 3. The comprehensive array of electromagnetic (EM) scanning methodologies used for monitoring and diagnosing various medical conditions within the human body.

aids in the diagnosis of physiological processes within biological tissues. These characteristics are exhibited by gamma radiation, which has a frequency greater than 1019 hertz and a wavelength less than ten picometers. MRI and Ultrasound imaging methods are founded on non-ionizing radiation principles; X-rays, SPECT, CT, and PET rely on ionizing radiation, as evidenced by prior research [9].

B. OPEN SOURCE DATASETS

Deep learning researchers face challenges in obtaining sufficient data for training deep learning systems. However, numerous imaging data sources are available for experiments. Deep learning segmentation networks require large amounts of labeled training data, but medical image annotation is challenging due to the need for medical specialists. Public datasets for medical image segmentation tasks are smaller than general computer vision tasks, containing hundreds of thousands to millions of annotated images. Understanding image segmentation requires understanding high-quality datasets and using public datasets as benchmarks to compare the performance and accuracy of medical image segmentation models. In Table 2, the study compiles open-source datasets for medical image segmentation tasks, categorizing commonly used datasets.

III. CUTTING-EDGE ADVANCES AND TECHNIQUES IN MEDICAL IMAGE ANALYSIS

Medical images are crucial for diagnosing diseases and guiding treatment protocols, providing valuable insights for healthcare professionals. Advancements in image processing techniques have improved image quality, enabling essential operations like contrast amplification and interpretation. Integrating automated and semi-automated image analysis techniques has increased efficiency, precision, and dependability in image interpretation tasks. Deep learning algorithms are increasingly used in medical image analysis, including diagnosing, partitioning, and categorizing. However, there is a significant divide between theory and practice due to limited multimodal medical image samples, the relationship between deep learning model efficacy and data set magnitude, and the generalization and robustness of the model. Existing small-sample learning methods are hampered by the paucity of images, and researchers should continue investigating effective fusion methods for diverse modalities.

A. RESTORATION AND RECONSTRUCTION OF MEDICAL IMAGES

Deep learning technology relies on massive and high-quality medical image data to achieve accurate image diagnosis. However, medical imaging is inherently prone to noise and artifacts due to equipment limitations and acquisition time constraints. Additionally, certain imaging techniques require a trade-off between imaging resolution and acquisition, such as in CT imaging, where radiation exposure necessitates a reduction in the number of projection acquisitions, or in resonance imaging, where organ auto-artifacts due to body movement require a reduction in the sampling rate of K space to decrease acquisition time. However, a low sampling rate can significantly impact image reconstruction quality, necessitating additional work, such as noise reduction and super-resolution, to obtain high-quality images. Consequently, restoration and reconstruction efforts, including removing artifacts and image reconstruction, are often necessary. Deep learning has been used to decrease medical image impairments, particularly in low-resolution images [60].

TABLE 1. Comparative analysis of famous medical imaging modalities.

Medical Images.	Risk Catagory	Descriptions
X-Rays [35, 36]	non-invasive k	Radiography, the primary medical imaging modality, were first discovered by Sir Wilhelm Roentgen in 1895. Radiology, also known as diagnostic radiology, has played a crucial role in establishing the field and introducing radiologists, healthcare professionals interpreting and analyzing medical images. X-ray imaging aids in diagnosing biological anatomy, identifying healthy bones, irregularities, or fractures.
CT [37, 38]	non-invasive	In the mid-1970s, CT became a popular medical screening tool in hospitals. CT images are created by spinning X-ray tubes and directing X-rays at different positions. A straight identifier group is placed directly in front of the X-beam emitter, integrating data points to create a tomographic depiction of the subject. CT scans accurately represent cellular locations within an individual's anatomical structure, including skull, bone structure, abnormalities, infarction, bleeding, and brain tumors.
MRI [39]	non-invasive	MRI scans provide detailed insights into blood circulation and cerebral processes. Still, they cannot detect brainwave activity or facilitate communication between tumor cells. MRI techniques use a magnetic flux with a magnitude ranging from 10,000 to 60,000 times Earth's magnetic field. Protons, abundant in biological substances, are used due to their nuclear magnetic resonance characteristics. Protons undergo electron acceptance from radio signals with resonance frequencies of 63 MHz when situated within a magnetic flux of 1.5 Tesla (T). The subject is inside a magnetic field, while a transmitter generates radio frequency waves.
Ultrasound [40]	non-invasive	The United States uses a frequency range of sound energy from 20 kilohertz to several gigahertz, making it more cost-effective and quality than X-rays and CT scans. This alternative method detects pressure waves from a paperweight collision with the ground. Ultrasonic waves of specified intensity visually represent an individual's anatomical structure. The ultrasonic transmitter generates short sound waves that interact with the photographed cell, generating echoes when sound frequencies propagate through the individual's cellular composition and undergo reflection due to their physical structure. Pulse echo imaging is the operational mechanism of ultrasound technology.
PET [41]	Invasive	PET scans can visualize cerebral activity and glucose metabolism, indicating brain function. They also provide specific information on brain tumor cells using various shades.
SPECT [42]	Invasive	SPECT imaging analyzes blood circulation to identify brain regions with the highest activity, aiding in identifying medical disorders like dementia, epilepsy, and arterial occlusion. Using gamma rays, SPECT is comparable to conventional nuclear medicine planar imaging and provides comprehensive three-dimensional data.

The CNN-DAE (Convolution Noise Reduction Automatic Encoder) is an early-stage model that removes noise from medical images. It relies on stacked convolutional layers for encoding and decoding, allowing it to learn noise-free images from noisy images. However, this model is less robust and more sensitive to changes in noise types. Chen et al. [61] introduced a solution to this problem through the RED-CNN noise reduction model. This approach combines residual networks with convolutional autoencoders, creating a deep network that effectively utilizes skip connections to decrease noise in low-resolution CT images.

Kang et al. [62] applied the directional wavelet transformation on low-resolution CT images and noise reduction using a deep CNN model on the wavelet coefficient image. Despite the significant improvement in noise reduction achieved by these network structures compared to conventional methods, detail loss, and texture absence persist. To assess the noise reduction technique's success, compare the original image with the denoised image and use metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), or mean square error (MSE). The program terminates if an image with noise reduction is suitable for output. Alternatively, the noise reduction algorithm settings can be modified and repeated until the picture is sufficiently denoised.

1) MEDICAL IMAGE NOISE REDUCTION

The optimization objective of minimizing the average squared difference between de-noised medical images and standard medical images presents challenges. Researchers propose improving loss function and model architecture methodologies to optimize noise reduction efficacy in low-dose CT imaging. The WGAN-VGG model, which uses perceptual loss and the Wasserstein generative adversarial network framework, addresses noise by improving the resemblance between denoised and genuine images [63].

Algorithm 1, illustrating image noise reduction with Different Approaches. Moreover, the Algorithm 2 offers a

TABLE 2. Most famous open source datasets.

Ref.	Datasets	Descriptions
[43]	UK Biobank	Utilizes publicly accessible medical image data.
[44]	Cancer	Provides multiple links to cancer-related datasets.
[45]	ABIDE	Autism Spectrum Disorder and control imaging data.
[46]	Give A-Scan	Comprises lung patient images and clinical data.
[47]	FastMRI	Includes 1,500 completely sampled knee MRIs and DICOM images from 10,000 clinical knee MRIs applying 3 and 1.5 Tesla magnets.
[48]	OpenNeuro	Neuroimaging data gathered using various techniques.
[49]	CTMedical	Cancer imaging archive CT images are in the dataset.
[50]	ADNI	Contains image data for Alzheimer's disease neuroimaging initiative.
[51]	Medical-Imaging	CT, MRI, PET images aid in diagnosis.
[52]	TCIA	Offers extensive cancer imaging data archive.
[53]	ML datasets	Includes several links to machine learning datasets.
[54]	Medical Imaging Repository	Open-access medical image archives include NIH, OASIS, UCI, ML Repository, and others.
[55]	DDSM, MICCAI	Breast Mammography 2620 patients, Histopathology 165 images.
[56]	CAMELYON17	Breast WSI 899 images.
[57]	DRIVE, STARE	Eye SLO 400 patients.
[58]	CHASEDB1	Eye SLO 28 images.
[59]	HVSMR 2018	Cardiac CMR 4 patients.

comprehensive elucidation of the various methodologies used in the denoising and enhancement of medical images. The primary aim of this algorithm is to achieve a reliable level of accuracy in the classification process. This algorithm likely provides an extensive overview of diverse techniques employed in medical image processing, specifically focusing on strategies for reducing noise and enhancing images. The ultimate goal of this algorithmic approach is to enhance the precision and reliability of the classification accuracy achieved using these medical images. This advancement ultimately contributes to more precise medical diagnoses and analyses, all while maintaining the originality and authenticity of the content.

The SMGAN model, an extension of the WGAN-GP model, incorporates multi-scale degree structure loss and L1 norm loss to minimize noise. However, the representation capabilities of the generative adversarial network are compromised by the presence of a gradient penalty. Ma et al. [64] proposed a resolution by adopting a residual generator structure based on the least square GAN (LS-GAN) framework. This generator acquires knowledge about noise patterns and applies denoising techniques to enhance picture quality by eliminating network input effects from both the synthesizer and output. Some scholars have also used 3D residual networks for denoising in both projection and image domains. Yin et al. [65] introduced an alternative approach using a consensus neural network model for unsupervised strategy in reducing picture noise. However, acquiring precise noisy pictures for clinical applications may pose challenges. Future research in medical picture noise reduction using deep learning should focus on unsupervised or self-supervised supervision models relying solely on noisy images.

2) SUPER-RESOLUTION RECONSTRUCTION OF MEDICAL IMAGES

Obtaining clinically useful high-resolution medical images can be challenging due to limitations in the collection equipment. The pursuit of generating high-resolution images from one or multiple low-resolution images has become a significant research area. Researchers are exploring deep learning technology to achieve this goal. The prosperous application of DL models in natural image super-resolution reconstruction has led to the gradual advancement of medical image super-resolution reconstruction research. However, medical images differ significantly from natural images. Their super-resolution reconstruction requires processing individual image slices and integrating information between slices, as depicted in Figure 4.

Oktay et al. [66] utilized Deep Residual Convolutional Networks to produce 3D high-resolution MR images by reconstructing multiple 2D cardiac magnetic resonance images. This method improved interlayer resolution rates. Pham et al. [67] extended the SRCNN model to 3D for the super-resolution reconstruction of brain MR images. McDonagh et al. [68] suggested a context-sensitive residual

Algorithm 1 Deep Learning Algorithm for Medical Image Noise Reduction

Input: Medical image with noise
Output: Denoised medical image

1 Preprocess image by applying filters such as median filter, Gaussian filter, or wavelet transform; Apply a deep learning model for noise reduction. Choose from available models such as:
- CNN-DAE (Convolution Noise Reduction Automatic Encoder): uses stacked convolutional layers for encoding and decoding, allowing it to learn noise-free images from noisy images. However, it is less robust and more sensitive to changes in noise types.
- RED-CNN noise reduction model: combines residual networks with convolutional autoencoders, creating a deep network that effectively utilizes skip connections to decrease noise in low-resolution CT images.
- WGAN-VGG model: utilizes perceptual loss and implements the Wasserstein generative adversarial network model to mitigate noise in the image data by enhancing the similarity between denoised and authentic images.
- SMGAN model: integrates the multi-scale degree structure loss and L1 norm loss into the objective function and utilizes information from adjacent slices to reduce noise.
- LS-GAN residual generator structure: uses a basis generative confrontation network to subtract the consequence of the network input from the synthesizer and the network output to learn noise and denoise the image.
- 3D residual networks in the projection and image domains for denoising.
- Consensus neural network model for an unsupervised image noise reduction strategy.

Algorithm 2 Denoising Medical Image Algorithm

Input: Medical image with noise
Output: Denoised medical image

1 Preprocess image by applying filters such as median filter, Gaussian filter, or wavelet transform;
2 Apply a deep learning model for noise reduction. Choose from available models such as:
- Gaussian Filter: Utilizes a Gaussian convolution kernel to perform Gaussian blurring, reducing high-frequency noise while preserving edges.
- Median Filter: Replaces each pixel value with the median value of its neighbors, effectively reducing salt-and-pepper noise.
- Wavelet Transform: Decomposes the image into different frequency components using wavelets, facilitating noise reduction in specific frequency bands.
- CNN-DAE (Convolution Noise Reduction Automatic Encoder): …
- RED-CNN noise reduction model: …
- WGAN-VGG model: …
- SMGAN model: …
- LS-GAN residual generator structure: …
- 3D residual networks in the projection and image domains for denoising.
- Consensus neural network model for an unsupervised image noise reduction strategy.

network architecture to develop high-resolution MR images that exhibit clear boundaries and textures. Zheng et al. [69] introduced an MR high-resolution reconstruction model with multiple Dense modules and multi-way branch combinations, leading to better weight results and generalization ability. Zhao et al. [70] proposed a channel-separable high-resolution reconstruction model for brain MR images, which utilizes one channel adopting the residual structure and another adopting the dense connection structure. This method effectively uses the features to enhance the reconstruction grade of high-pixel images.

Tanno et al. [71] utilized 3DSubpixelCNN and variational inference to achieve ultra-high resolution reconstruction of MR diffusion tensor images. Peng et al. [72] presented the Spatially Aware Interpolation Network (SAIN), which used various features to enhance the reconstruction quality of super-resolution images at 2x, 4x, and 6x resolutions, leading to better results. Shi et al. [73] introduced the MultiScale Global Local Residual Learning (MGLRL) model, combining a multi-scale global and local phase residual network to improve the super-resolution reconstruction of MR images and improve the reconstruction details. Lyu et al. [74] employed a GAN-based approach for super-resolution reconstruction of multi-contrast MR images.

Medical imaging noise reduction and low and high-resolution image sample pairs are mandated to train DL models for super-resolution image reconstruction. Usually, downsampling is used to create high/low-resolution image pairs for training purposes. However, the high and low-resolution image correspondence can vary for different medical imaging modes with distinct imaging principles. Therefore, researchers use artificial downsampling to obtain training data. Despite this approach, creating realistic high/low-resolution image sample pairs for super-resolution reconstruction using DL can be challenging. This is because the relationship between low and high-resolution images in the actual image acquisition may differ, making it difficult to achieve accurate and realistic super-resolution reconstruction results.

3) MEDICAL IMAGE RECONSTRUCTION

Reconstructing medical images involves gathering raw data and transforming it into clinically useful images, such as with CT scans. MR images begin as projection images, and an

Algorithm 3 Medical Image Super-Resolution Reconstruction Using DL

Input: Medical image Reconstruction
Output: Reconstructed medical image

1. Preprocess image by applying filters such as median filter, Wiener Filter, Anisotropic Diffusion Filter, Gaussian filter, or Total Variation (TV) Filter; Apply a deep learning model for image Reconstruction.
   - Collect low-resolution medical images.
   - Artificially downsample the images to create high/low-resolution pairs.
   - Apply noise reduction techniques to low-resolution images.
   - Train the DL model using the high/low-resolution pairs.
   - Use the DL model to generate high-resolution images from new low-resolution images.
   - Evaluate the quality of the generated images and refine the DL model as needed.

algorithm is needed to translate K-space data into images suitable for medical diagnosis. Practical considerations often necessitate reducing acquisition time or the number of projections or K-space fills, as in the case of CT scans to minimize radiation exposure or MR imaging to avoid patient discomfort and motion artifacts. However, decreasing the amount of data collected inevitably impacts the overall quality of image reconstruction. Algorithm 3, Illustrating medical image reconstruction using different approaches. Algorithm 3 demonstrates a dual-pronged approach aimed at elevating medical image quality. The process commences with the application of noise reduction filters, including median, Wiener, anisotropic diffusion, Gaussian, and total variation, to preprocess the image. Subsequently, a deep learning model is trained using matched high and low-resolution medical image pairs. The primary objective of this model is to anticipate intricate details of high-resolution images based on the corresponding low-resolution inputs. The ensuing step involves the assessment of the quality of high-resolution images produced by the model. If deemed necessary, the model can be fine-tuned for improvements. This iterative procedure culminates in enhanced precision for medical image analysis and diagnosis.

Current research in medical image reconstruction focuses on finding suitable reconstruction algorithms that can produce high-quality images even with low data sampling rates. Medical graphics often employ deep learning models, with reconstruction techniques broadly classified as original or digital. The former involves directly applying data to produce images, which are then improved through post-processing methods to enhance their quality. Examples of this class of methods include traditional iterative alternating direction algorithms that use ADMM multiplier hyperparameters in

Algorithm 4 Medical Image Reconstruction Algorithm

Input: Medical image Reconstruction
Output: Reconstructed medical image

1. Preprocess image by applying filters such as median filter, Wiener Filter, Anisotropic Diffusion Filter, Gaussian filter, or Total Variation (TV) Filter;
2. Apply a deep learning model for image Reconstruction.
   - Median Filter: Replaces each pixel value with the median value of its neighbors, effectively reducing salt-and-pepper noise.
   - Wiener Filter: Uses a statistical approach to restore images by minimizing noise while preserving image details.
   - Anisotropic Diffusion Filter: Enhances image quality by diffusing noise while preserving edges and fine structures.
   - Gaussian Filter: Utilizes a Gaussian convolution kernel to perform Gaussian blurring, reducing high-frequency noise while preserving edges.
   - Total Variation (TV) Filter: Reduces noise while preserving image structures by minimizing the total variation of pixel intensities.
   - Collect low-resolution medical images.
   - Artificially downsample the images to create high/low-resolution pairs.
   - Apply noise reduction techniques to low-resolution images.
   - Train the DL model using the high/low-resolution pairs.
   - Use the DL model to generate high-resolution images from new low-resolution images.
   - Evaluate the quality of the generated images and refine the DL model as needed.

optimization algorithms to reproduce undersampled K-space data and develop MR images directly [75]. Adler et al. [76] proposed the duality learning model for CT reconstruction, which replaces filtered back-projection. Cheng et al. [61] presented a method for achieving fidelity of projection data to CT images using the primitive-dual network (PD-Net), which enables rapid MR image reconstruction. Zhang et al. [77] proposed the joint spatial-Radon domain reconstruction net (JSR-Net), which uses a DCNN model to reconstruct CT images and their corresponding Radon projection transformations, producing superior results compared to the PD-Net. Currently, the primary reconstruction technique is post-processing models for image artifact removal, a digital reconstruction method.

A U-Net model structure with a residual module is suggested to comprehend the pseudo-correlation between the reconstructed image and the original undersampled image shadow, followed by a two-way U-Net model proposal.

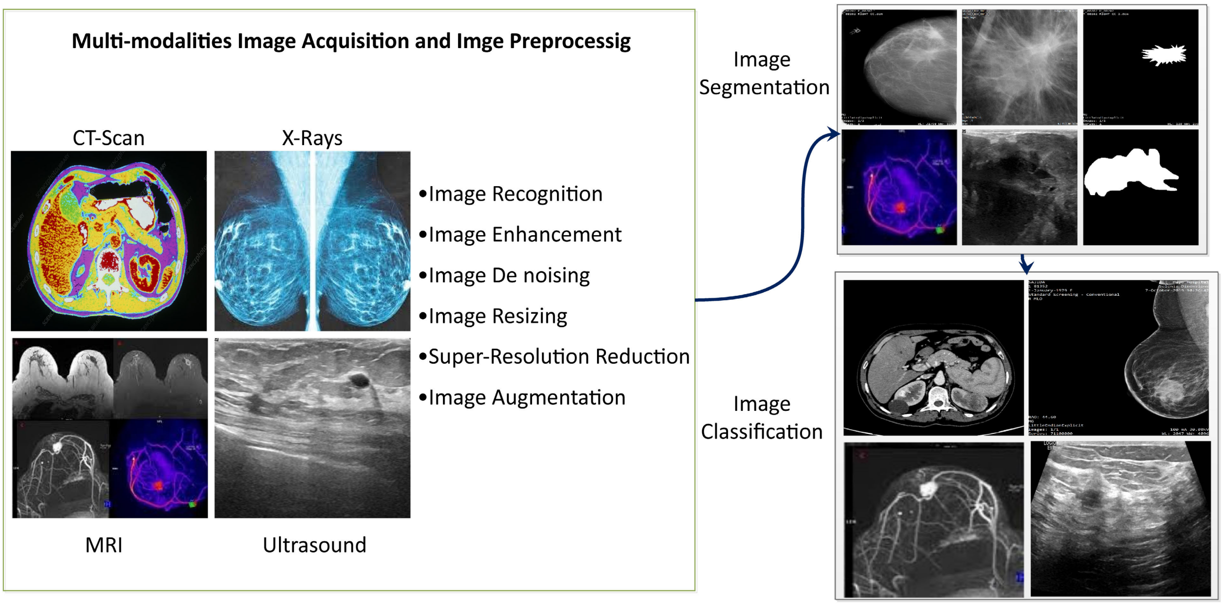


FIGURE 4. The diagram illustrates advanced deep-learning techniques for Cancer classification using various imaging modalities, segmentation, and feature extraction. Additionally, the study comprehensively examines existing diagnostic methods, highlighting new research areas for radiologists and academics to develop an efficient deep-learning-based prediction system. The ultimate aim is to enable early intervention and improve the accuracy of cancer prognosis.

Improving the fidelity of MR image reconstruction can be achieved by reconstructing both the phase image and magnitude image, as cited in [78]. A deep cascaded CNN model is employed to retain the temporal relationship of dynamic MR image acquisition. Additionally, Han et al. [79] utilized a domain adaptation fine-tuning method and a network to MR images for CT image reconstruction to reduce the weight of dynamic MR images through rapid acquisition construction durability. Precise reconstruction at large sampling rates can be achieved using the KIKI-Net proposed by Eo et al. [80] improved recurrent, residual network combined with the residual concatenation of poor and dense blocks trained with complex images as suggested by Bao et al. [81]. Dai et al. [82] designed a depth-based multi-scale atrous convolution residual convolutional network to enhance the accuracy of MR image reconstruction. Yang et al. [83] proposed utilizing a Deep Anti-Aliasing Generative Adversarial Network (DAGAN) in MRI reconstruction, inspired by the effectiveness of GANs in the realm of vision, to remove aliasing artifacts. In contrast, Quan et al. [84] proposed a very low sampling rate periodic loss Ref inGAN model to enhance the precision of MR image reconstruction. Mardani et al. [85] showed that inverse problem-solving could reduce image noise, perform image super-resolution and image reconstruction, and use these models interchangeably. This article will not delve into the specifics of each model. The following is the key point to achieving Medical image reconstruction.

- Gather raw data for medical images.
- Refining the raw data into clinically valuable images, such as CT scans or MR images
- Translate K-space data into images suitable for medical diagnosis.
- Reduce the acquisition time, the number of projections, or K-space fills to minimize radiation exposure or avoid patient discomfort and motion artifacts.
- The decrease in data collection impacts the overall image reconstruction quality.
- Research focuses on finding suitable reconstruction algorithms that produce high-quality images with low data sampling rates.
- Reconstruction techniques can be broadly classified as original or digital.
- Original techniques involve directly applying data to produce images, which are then improved through post-processing methods to enhance their quality.
- The current primary reconstruction technique is digital, using post-processing models for image artifact removal.
- Various deep learning models are used for image reconstruction, including U-Net with a residual module, Deep Cascaded CNN, and DAGAN.
- A depth-based multi-scale atrous convolutional residual convolutional network is used to enhance the accuracy of MR image reconstruction.
- Inverse problem-solving techniques are used to reduce image noise and perform image super-resolution and reconstruction.

B. MEDICAL IMAGE SYNTHESIS

1) MEDICAL IMAGE DATA EXPANSION

Currently, there are two primary approaches to generating clinical medical images. The first purpose is to expand the dataset by obtaining many medical images to train deep learning models, which can enhance clinical diagnostic and predictive accuracy. Despite using different data augmentation methods, such as translation, rotation, shear, and adding noise, these methods do not effectively produce a broad range of diverse data. There is room for improvement in enhancing deep learning models' prediction accuracy and generalization ability. The second approach involves synthesizing analog images to integrate information from other medical imaging modalities, which can enhance the precision of clinical diagnosis. However, obtaining multimodal image information about a patient is challenging, so time-based image synthesis provides a valuable solution. Furthermore, some emerging imaging technologies have strict requirements, and only a few hospitals and research institutions can meet them. Hence, image synthesis becomes essential to make scarce image data available. Due to the success of GAN models in generating natural images, researchers have recently focused on applying GAN's derivative models to synthesize medical images. One approach to expanding medical image datasets is synthesizing stripless GAN models, primarily generating medical images from noise data. Typically, a deep convolutional generative adversarial network (DCGAN) is utilized to enhance the baseline model in this technique.

Kitchen et al. [86] employed a DCGAN to generate synthetic images of prostate lesions. They trained their algorithm on a dataset of real medical images of prostate lesions and used it to develop synthetic images that appeared identical to the real images. Deecke et al. [87] suggested an AnoGAN standard that utilizes DCGAN to generate diverse retinal images that can aid in detecting diseases. The model was specifically developed to develop synthetic images that nearly resemble retinal images, to utilize them for data augmentation and enhance the accuracy of disease detection methods. Chuquicusma et al. [88] employed the DCGAN model to render synthetic lung nodule data, achieving realism that even experienced clinical radiologists could not distinguish from real lung nodules. This demonstrates the potential of generative models like DCGAN for generating realistic medical imaging data, which can be used to train and enhance computer-aided diagnosis systems' performance. FridAdar et al. [89] employed DCGAN to create synthetic samples of Type 3 liver injuries, such as cysts, hemangiomas, and transferases. The goal was to enhance the accuracy of liver disease classification by utilizing these synthetic samples for data augmentation. The results indicated the significance of this method, showing improved performance in liver lesion classification tasks. This highlights the potential of using generative models like DCGAN in medical imaging research. Bermudez et al. [90] employed the training strategy of DCGAN to produce high-quality T1-weighted MR images of the human brain. The study's primary objective was to generate synthetic images that closely approximate actual images, which could be used to augment medical imaging datasets and improve the accuracy of disease detection methods. This approach successfully generated realistic and high-quality brain MR images using DCGAN training.

Despite achieving significant success in medical image synthesis, DCGAN has limitations in generating and resolving high-resolution images. Several improved GAN models have been suggested to address this problem and improve the quality of medical image synthesis. Based on the Laplacian pyramid, Baur et al. [91] used LAPGAN to generate high-resolution skin lesions images. The approach gradually changes the scale to generate images that improve the accuracy of skin disease classification. Additionally, Korkinof et al. [92] employed a PGGAN to synthesize high-resolution mammogram X-ray images at 1280×1024 pixels, highlighting the advantage of using PGGAN in high-resolution image synthesis.

2) MEDICAL IMAGE MODALITY CONVERSION

Medical image modality conversion synthesis can be classified into two types. The first class is unimodal conversion, which involves transforming images from one modality to another. Kang et al. [93] and Wolterink et al. [94] employed DL models to transform low-resolution to available CT images and 3T MRI images to 7T MRI images, respectively, to improve image quality. The other type is cross-modal one-to-one conversion, which addresses issues such as low contrast of CT in soft tissue and radiation exposure. Nie et al. [95] proposed a perceptual generation model to tackle this issue, which employs a cascaded 3D fully convolutional network and paired image training to yield CT images from MR images. This approach enhances the authenticity of the synthesized CT images, improving their usefulness in medical imaging tasks.

The prevalent deep learning network architecture used for medical image synthesis is the encoding-decoding structure, with Pix2Pix [96] and CycleGAN [97] being typical examples. For instance, Maspero et al. [98] utilized the Pix2Pix network structure to transform MR images into CT images for radiation dose calculation during radiotherapy and chemotherapy. Additionally, Choi and Lee [99] employed the Pix2Pix model to generate brain MRI images with clearer structural information from PET images. Despite its effectiveness in converting multimodal images, the Pix2Pix model presents a challenge in clinical settings as it necessitates the spatial alignment of source and target images, which is often difficult to obtain. The CycleGAN model is commonly used for medical image generation to address mismatched source and target images.

Wolterink et al. [100] utilized CycleGAN to synthesize unpaired head MRI images. They compared the resulting composite image to its corresponding CT image, finding that the composite image was more realistic. CycleGAN has emerged as a prevalent method for transferring multi-modal medical images, including synthesizing cardiac MR images to CT images [101], abdominal MR images to CT images, and brain CT images to MR images. Nevertheless, it is important to note that CycleGAN may not always preserve the structural boundaries in the generated images. To mitigate this issue, Hiasa et al. [102] suggested the addition of a gradient consistency loss to the CycleGAN model. This loss assesses the consistency of per-pixel gradients at the structural boundary between the original and synthetic images, leading to enhanced quality of the generated images. The following are key important steps to achieve the Medical image modality conversion.

1) Medical imaging modality conversion synthesis can be classified into unimodal and cross-modal one-to-one.
2) Unimodal conversion involves transforming images from one modality to another.
3) DL models are used for unimodal conversions, such as transforming low-resolution CT images to higher quality images or 3T MRI images to 7T MRI images.
4) Cross-modal one-to-one conversion addresses issues such as low contrast of CT in soft tissue and radiation exposure.
5) Perceptual generation models, such as the cascaded 3D fully convolutional network, synthesize CT images from MR images to enhance image authenticity and usefulness.
6) Encoding-decoding structure is the overall deep learning network architecture used for medical image synthesis.
7) Typical examples of encoding-decoding structures include Pix2Pix and CycleGAN.
8) Pix2Pix model transforms MR images to CT images for radiation dose calculation during radiotherapy and chemotherapy or generates brain MRI images with clearer structural information from PET images.
9) Pix2Pix model presents a challenge in clinical settings as it necessitates the spatial alignment of source and target images, which is often difficult to obtain.
10) CycleGAN model is commonly used for medical image generation to address mismatched source and target images.
11) CycleGAN model transfers multi-modal medical images, including synthesizing cardiac MR images to CT images, abdominal MR images to CT images, and brain CT images to MR images.
12) CycleGAN may not always preserve the structural boundaries in the generated images.
13) Gradient consistency loss is added to the CycleGAN model to mitigate this issue and enhance the quality of the generated images.

IV. MEDICAL IMAGE REGISTRATION AND SEGMENTATION

Medical image registration and segmentation are two important tasks in medical image analysis; the combination can improve the accuracy of the two tasks. The alternating training process of registration and segmentation are shown in Figure 5. However, the existing joint registration and segmentation framework of single-modal images is difficult to apply to multi-modal images. Aiming at the above problems, a CT-MR image-based joint registration and segmentation framework based on modality consistent supervision and multi-scale modality independent neighborhood descriptor is proposed, which consists of a multimodal image registration network and two segmentation networks. The deformation field generated by the multi-modal registration is used to establish the corresponding deformation relationship between the segmentation network results of the two modalities, and the modality consistency supervision loss is constructed, which improves the accuracy of multi-modal segmentation by supervising each other between two segmentation networks; in the multimodal image registration network, a multi-scale modality independent neighborhood descriptor is constructed to enhance the representation ability of crossmodal information and the descriptor is added to the registration network as a structural loss term to constrain the local structure correspondence of multimodal images more accurately. In various medical image examination tasks, obtaining high-quality data is often the first step. The data must then be processed for registration, segmentation, and analysis tasks. Certainly, this section will focus on using DL in medical image rosters and segmentation.

A. MEDICAL IMAGE REGISTRATION

Medical image registration is a significant preprocessing step in image processing, implicating the alignment of images from different sources in spatial positions, such as in multimodal image fusion analysis, atlas establishment, and tumor detection. Deep learning has been classified into three categories in medical image registration research. The first type uses a DL model to learn a similarity measure. Then it uses traditional optimization methods to learn deformations for registration, but this method is slow and less efficient. The second type is supervised deep learning, which involves network training using real variables corresponding to the registration pair's shape field [103]. The registration framework for supervised learning is exhibited in Figure 6. The third type is unsupervised deep learning-based registration. This article investigates supervised and unsupervised learning techniques for medical image normalization. Image registration involves aligning images obtained at specific times and by various imaging machines.

Training the network model aims to reduce the discrepancy between the true deformation field and the deformation field produced by the network. Once obtained, the deformation field is applied to the image requiring registration to achieve the desired outcome. In supervised learning for medical image registration, the variable label for the shape field can be obtained through two methods by exploiting the deformation domain obtained from classical registration algorithms as the label or by simulating changes in the target image and using the deformation parameters as the real labels and

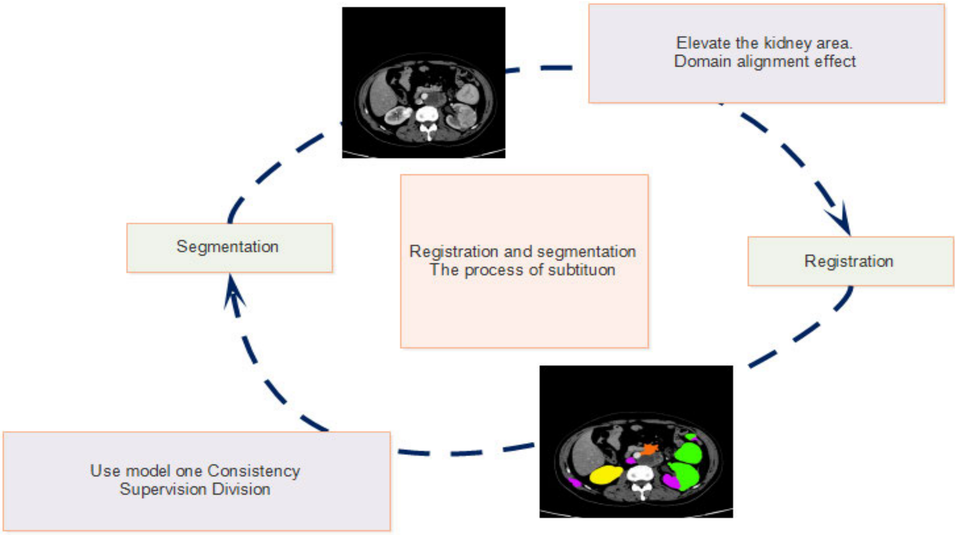


FIGURE 5. **The alternating training process of registration and segmentation.**

the deformation map as the image for registration. In rigid registration based on supervised learning, various studies have employed convolutional neural networks (CNNs) combined with regression techniques. Miao et al. [104] utilized a deep residual regression network and a correction network to create a fetal brain map by employing the "coarse registration first, then fine registration" approach for the rigid-body registration of T1- and T2-weighted MR images of the fetal brain. They used geodesic distance loss during the process. Similarly, Salehi et al. [105] employed a deep residual regression network and a correction network with the "coarse registration first, then fine registration" approach to performing rigid-body registration of T1- and T2-weighted MR images of the fetal brain, utilizing geodesic distance loss. Their goal was to establish a fetal brain map. Later, Zheng et al. [106] used domain adaptation, pre-trained networks, and pairwise domain adaptation modules to design 2D and 3D radiographic image registration that could adjust the differences between simulated training data and real test data, improving registration robustness.

Nonlinear registration presents a more significant challenge in simulating the nonlinear variation of the shape field than in simulating the rigid deformation field. As a result, most nonlinear registration based on supervised learning employs the classical method to obtain the deformation field, which is then used as a label to train the model. Recently, Yang et al. [107] utilized the U-Net network algorithm as the baseline structure and the differential isochronous embryo algorithm to obtain the deformation field, which was then used as a label for end-to-end registration of 2D and 3D brain MR images. However, own to the difficulty in simulating the nonlinear deformation field, weakly supervised registration and double supervised registration have been introduced as training concepts. Weakly supervised registration employs anatomical structure labels for precise matching and learning of the deformation field. For instance, Hu et al. [108] labeled the structural features of prostate ultrasound and MR images to train a CNN model that learned the deformation field, which was then applied to the grayscale image to achieve registration between MR and ultrasound images.

Hering et al. [109] utilized similarity measurement and organizational structure segmentation labels as part of the registration network's training process to enhance the accuracy of cardiac MR image registration. They implemented a double-supervised registration approach that trained the model with two types of supervised loss functions. Cao et al. [110] employed generative networks to transform MR images into their corresponding CT images and vice versa. During registration, the model assesses the similarity loss between the original and the generated MR image and the similarity loss between the original CT image and the synthesized CT image. Besides, Fan et al. [111] integrated both supervised and unsupervised model loss to achieve precise registration of brain MR images. The precision of supervised learning in medical image registration highly depends on the labels' accuracy. Consequently, creating reliable labeling and designing an appropriate loss function are crucial challenges to overcome in supervised learning for medical image registration. The spatial transformation network (STN) has made unsupervised deep high-degree learning models a burning research issue in medical image registration, as depicted in Figure 7. The Displacement Deformation Field (DDF) represents the spatial transformations that align images through neural network predictions, as shown in Figure 7. This field encapsulates the spatial transformations required to align images through the predictions of neural networks. It plays a pivotal role in achieving image registration within the framework. Yoo et al. [112] combined the convolutional autoencoder (CAE) and STN models to achieve the registration of neural tissue microscopic images.

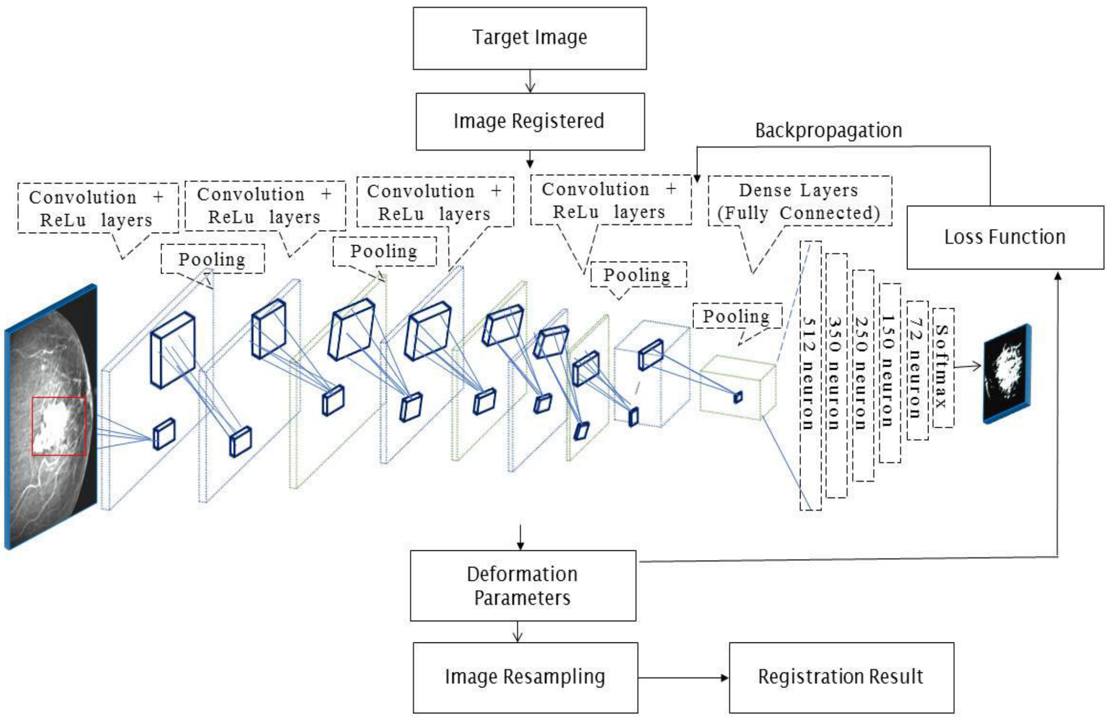


FIGURE 6. Supervised DL medical image registration framework.

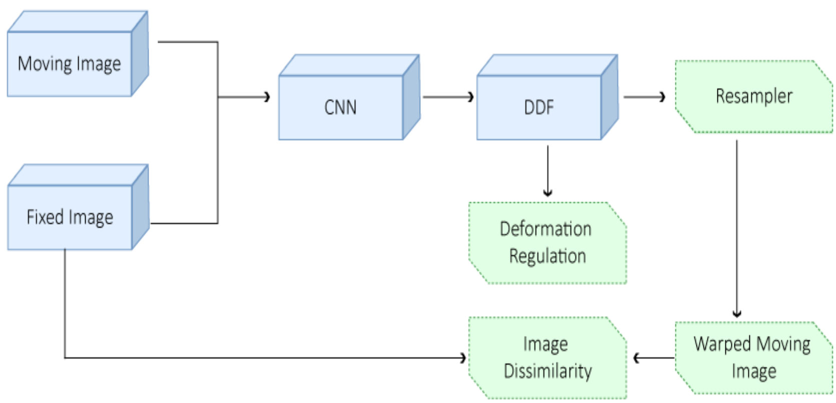


FIGURE 7. Unsupervised deep learning image registration network framework.

The CAE extracts the features of the image to be registered and the target image, based on which the similarity loss is calculated, demonstrating improved matching quasi-results. In 2018, Balakrishnan et al. [113] proposed the VoxelMorph network structure, based on the U-Net model and STN module, to realize the non-linear registration of MR images. The model was then refined by introducing the segmentation labeling auxiliary loss to enhance the matching standard Dice score [114]. Kuang et al. [115] proposed the empty inter-transformation module, which replaces the U-Net network structure, to register brain MR images while reducing the model parameters accurately.

Zhang et al. [116] introduced transformation smoothing loss, reverse consistency loss, and anti-folding loss, in addition to similarity loss, to enhance the accuracy of unsupervised registration. The transformation smoothing loss and anti-folding loss retain the balance of the deformation field. In contrast, the reverse consistency loss ensures that the deformation field is reversible when the image and target image are swapped in the registration map. Tang et al. [117] utilized unsupervised networks to implement end-to-end registration of brain MR images. The network model simultaneously learns affine and nonlinear transformation parameters, i.e., change parameters. In addition to unsupervised registration based on the CNN model, registration with GAN models has become a research trend in medical image registration. The conditional generative adversarial networks are used for registration, with the generator generating the transformation parameters or registered images and the discriminator comparing the registered images to identify similarities. Usually, the STN module is integrated between the generator and the discriminator for end-to-end training. Currently, the GAN model has numerous applications in medical image registration, such as prostate MR image and ultrasound image registration [118], multimodal retinal maps using Cycle GAN as the baseline model [119], and CT image and MR image registration [120], among others. The GAN model can serve as regularization in medical image matching, adjusting the deformation field and registering images, or performing image transformation and cross-domain registration to enhance registration accuracy. Table 3 provides an overview of typical unsupervised and supervised registration models.

B. MEDICAL IMAGE SEGMENTATION

The ability to perform medical image segmentation is essential in computer-aided diagnosis as it quantifies specific areas of interest. Recently, there has been a growing tendency to adopt natural image segmentation models for medical images with deep learning in semantic segmentation. This trend aims to improve the accuracy and efficiency of medical image segmentation for better diagnosis and treatment planning. The mainstream network frameworks adopted in medical image segmentation are CNN, FCN, U-Net, SegNet, DeepLabV2, RNN, and GAN models. The 2.5D CNN is a frequently employed model for medical image segmentation, which utilizes 2D convolutions on transverse, shapeless, and coronal planes to carry out the segmentation process. It improves segmentation accuracy using neighborhood information in 3D space while saving on computing costs [131].

1) FULLY CONVOLUTIONAL NETWORK (FCN)

The initial deep-learning approach to image segmentation employs the sliding window technique for target delineation. However, this method generates excessive candidate regions, causing redundant computations and inefficiencies. The problem intensifies with larger image block sizes, impacting segmentation accuracy and imposing inherent constraints. In 2015, Long et al. [132] introduced the FCN as a significant departure from traditional image segmentation methods. FCN uses convolutional operations exclusively, accommodating various input image sizes and producing corresponding output. Despite its adaptability advantage over conventional networks, FCN's pixel-wise classification ignores inter-pixel relationships and global context. The upsampling process, relying on feature map expansion, could lead to losing image details and blurred outcomes. Various deep learning standards for semantic segmentation, including the FCN and its variants such as the Parallel FCN [133], Focus FCN [134], Multi-branch FCN [123], and Cycle Ring FCN [124], have been extensively employed in a variety of medical image segmentation tasks, exhibiting impressive performance.

2) U-NET

In the realm of medical image segmentation, the U-Net network, pioneered by Ronnebeger et al., [135], serves as a crucial complement to the FCN approach. U-Net represents an evolved version of FCN with a comparable architecture, featuring convolutional and pooling layers but omitting traditional fully connected layers. Structurally divided into encoder and decoder phases, the U-Net design incorporates downsampling, upsampling, and bypass connections, symmetrically arranged to form a U-shaped configuration. The downsampling component extracts basic image features, while the upsampling part handles the encoding and decoding intricate network structural information. Notably, the traditional fully connected layer is replaced by a convolutional layer, aligning U-Net with the FCN framework. A critical architectural innovation, the bypass connection, bridges global network insights with local nuances, enhancing information integration. The encoder's core task involves capturing high-dimensional image attributes, achieved through sequential convolutional and pooling stages that reduce spatial dimensions. The decoder then rescales the output feature map, aligning it with the input image's scale, and assigns high-dimensional features to individual pixels, thereby achieving precise pixel-level image segmentation. Gradually increasing receptive field dimensions and feature abstraction are achieved through additional convolutional layers. The skip connection optimally combines fundamental insights from downsampling with enriched upsampling insights, enhancing segmentation precision. U-Net and its variants, such as Nested U-Net, V-Net [127], and Cyclic Residual U-Net [126], have become the mainstream models for medical image segmentation. These models have performed better than other segmentation approaches, making them the preferred choice for many medical imaging tasks.

3) U-NET++

Badrinarayanan et al. [136] introduced an enhanced version of the U-Net architecture called U-Net++, which improves the skip connection element of the original U-Net. This enhancement incorporates a convolution operation denoted as X and employs a dense connection approach rather than the simple sequential connection of the initial U-Net. This dynamic approach allows the network to learn the significance of depth-related features during training, facilitating the adaptation of downsampling layers for optimized network parameters and performance. U-Net++ also features a shared downsampling component across the network, streamlining training by requiring a single training iteration for the downsampling network. Symmetrical restoration of features from different depths is achieved through corresponding

TABLE 3. Typical deep learning medical image segmentation method.

Dataset	Model	Organ	Loss	Precision	Ref.
Internal	DCNN	Brain, Breast, Angiography Images.	Cross-Entropy	—	[121]
PROMISE12, ISLES, MSD Pancreas dataset	3D-FCN	Intervertebral Disc MR	Weighted Cross-Entropy	Dice 0.912	[122]
SpineWeb	Multistream 3D-FCN	Brain MR	Likelihood	Dice 0.95 ASD 0.127 MHD 9.62	[123]
MICCAI 2017 Challenge Dataset	Recurrent FCN	Cardiac MR	Cross-Entropy	Dice 0.90 APD 2.05	[124]
Pancreatic CT Images	U-Net	Multiple organs	Combo	Dice 0.92	[125]
DRIVE, STARE , CHAOS	R2U-Net	Multiple Organs	Binary Cross-Entropy	Dice 0.86	[126]
PROMISE12 Challenge Dataset	V-Net	Prostate MR	Dice	Dice 0.87 HD 5.71mm	[127]
DRIVE, ISIC 2018	Bi-LSTM	Dice	Binary Cross-Entropy	F1-Score 0.99	[128]
BRATS 2013, BRATS 2015	SegNet, GAN	Multi-organ, Brain MR	Multi-scale L1 Norm	Dice 0.84 (2013) Dice 0.85 (2015)	[129]
INbreast , DDSM	CGAN	Mammography Images	Dice	Dice 0.94 (INbreast)	[130]

downsampling layers. Additionally, U-Net++ introduces deep supervision, connecting the output of each network layer to the final output, providing better control and insight into the network's intermediate stages.

4) SEGNET

The encoding network of SegNet [136] closely resembles the convolutional layer structure of VGG-16, lacking a fully connected layer, and prioritizes feature extraction. The decoder uses max-pooling indices to perform non-linear upsampling on input feature maps, reducing learning, boundary division, and training parameters. The sparse feature maps are then subjected to trainable convolution operations to generate dense maps. The final softmax layer determines the utmost probability of each pixel in all categories. SegNet retains the maximum aggregated index and applies it to the decoding network, achieving enhanced performance and superior efficiency compared to other segmentation networks.

5) DEEPLAB

The DeepLab-v1 [137] network merges the FCN architecture with a conditional random field (CRF) model to address precision issues in FCN segmentation. It incorporates a fully connected CRF model, refining initial FCN segmentation outputs and incorporating a hole algorithm. The network also augments to capture more contextual information, reducing resolution degradation during progressive convolution and pooling stages. The integration of dilated convolutions accelerates processing speed. The DeepLab-v2 framework adopts the CRF model for enhanced segmentation precision and integrates the hole algorithm to expand the receptive field. It also utilizes the Atrous Spatial Pyramid Pooling (ASPP) module, which conducts parallel sampling of feature maps at varying rates and consolidates outputs to yield enriched spatial information. The network replaces the conventional VGG-16 module with a ResNet module, enhancing the segmentation process. DeepLab-v3 [138] advances beyond its predecessors by highlighting the utility of atrous convolutions and doubling the sampling rate within the cascade module. The ASPP module, inherited from DeepLab-v2, is expanded to enhance performance further. DeepLab consistently outperforms previous iterations on the PASCAL VOC 2012 dataset in segmentation accuracy.

6) RECURRENT NEURAL NETWORK (RNN)

The RNN is a deep neural network model for processing sequential input. It uses a sequence of neurons activated sequentially, followed by an iterative process. Using various techniques, RNN-based image reconstruction methods can be categorized into direct and indirect decoding. The RNN segmentation model uses slices as sequence information to accurately segment medical images by considering contextual relationships between slices. Standard algorithms include the CW-RNN (clockwork RNN) [139] and the up-and-down LSTM model [128], which captures the correlation between adjacent slices and sharpens segmentation edges. Choy et al. [140] introduced the 3D-R2N2 image reconstruction network using a 3D cyclic neural network architecture. This network accepts multiple perspective images and encodes them into feature vectors. The input is then fed into the LSTM network to capture multi-view information, and a 3D-DCNN model is used to produce the three-dimensional shape. The 3D-R2N2 method effectively leverages multi-view information, using neural networks to dynamically acquire knowledge about the 3D model's attributes and establish a correspondence between the image and the 3D model. Zou et al. [141] introduced the 3D primitive recurrent neural network (3D-PRNN) for medical image reconstruction, which predicts the shape center of the primitive and generates a feature vector from a depth image. The loop generator repeatedly predicts primitives and associated features, forming a three-dimensional (3D) model.

7) OTHERS APPROACHES

The U-Net network has gained significant interest recently due to its excellent segmentation effect. Researchers have proposed new methods based on the traditional U-Net network, such as 3D U-Net, V-Net, SegNet, and DeepLab series networks [142]. Medical imaging devices generate three-dimensional images in medical imaging, and researchers have proposed 3D U-net, V-Net, and Drozdzal networks [143]. SegNet and DeepLab series networks have improved the FCN network structure, while the encoder structure of FCN is typically a classic classification network. In 2016, Paszke et al. [144] proposed the Efficient Neural Network (ENet), which can perform real-time pixel-by-pixel semantic segmentation. ENet is 18 times faster, requires 1/75 fewer floating-point operations, has 1/79 fewer parameters, and provides similar or higher accuracy than existing models. Pohlen et al. [145] proposed a network architecture FRRN, which uses two branches to combine multiple scale context information to address the issue of semantic segmentation relying too much on pre-trained networks for rich features. This approach improves the network's ability to mine context information globally. Zhao et al. [146] proposed a Pyramid Scene Parsing Network (PSPNet) to achieve high-precision segmentation in complex scenes. PSPNet aggregates context information based on different regions, improving the network's ability to mine context information globally. To solve the problem of the full-resolution residual network being computationally intensive and slow on full-scale images, PSPNet uses four different maximum pooling operations corresponding to different window sizes and step sizes.

Based on these models, Chen et al. [147] suggested a bidirectional context LSTM model, the BDCLSTM, which bi-directionally learns the context relation in cross-section, sagittal, and coronal directions, with better results than the pyramid with multi-scale segmentation LSTM standards. The basic idea of GAN-based segmentation is creating early segmentation results with the distinction and then refining the segmentation results with the discriminator. The synthesizer frequently incorporates the FCN or U-Net network system, whereas the classifier utilizes an ordinary classification network framework, such as ResNet or VGG. GAN-based medical image segmentation has been used to chop tissue and organs [148].

8) COMPARISON OF IMAGE SEGMENTATION ALGORITHMS

This study analyzes image segmentation techniques and methodologies to assess their effectiveness in segmenting medical images based on semantic information. The thorough examination provides valuable insights into the effectiveness of each approach in capturing and analyzing semantic information in medical images. Table 4 Table 1 displays the technical characteristics and advantages of deep learning-based image segmentation methods. The performance of the algorithms is evaluated on various datasets, including PASCAL VOC 2012, FCN, DeepLab-v1, DeepLab-v2, and DeepLab-v3. The mIoU values for PSPNet and PSPNet are 62.2%, 72.6%, and 72.6%, respectively. On the CityScapes dataset, the mIoU values are 65.3%, 63.1%, 70.4%, 58.3%, 71.8%, and 81.2%. The mIoU values reach 35.1% and 45.7% on the PASCAL-CON-TEXT dataset. On the CamVid dataset, the mIoU values are 55.6% and 51.3%. Currently, most algorithms use PASCAL VOC 2012 as the test data set for static semantic segmentation, while many analyze dynamic scenes or real-time image semantic segmentation using CityScapes. DeepLab V3 and PSPNet algorithms have a reasonable recognition rate for objects of different scales in the image data, and their target segmentation results are closer to real segmentation edges.

C. TECHNICAL CHALLENGES IN SEGMENTING MEDICAL IMAGES

Medical image segmentation is more complex than natural due to unique characteristics and distinct properties. This presents challenges due to the distinctive attributes inherent in medical images, making them more challenging to understand and interpret.

- The dataset size is limited, but medical image data is challenging due to labeling demands and privacy concerns. Model interpretability becomes less important in large datasets, allowing for easier training. In scarce datasets, prior knowledge is crucial for learning crucial features, while parameter count must be constrained to mitigate overfitting risks.
- Medical images' small targets require high precision segmentation, requiring high-resolution input to ensure accurate segmentation. This is due to irregular shapes, blurred boundaries, and complex gradients.
- Context information in medical images is crucial for diagnosing human diseases. However, organ structures' semantics are limited, necessitating the model to utilize low-resolution information in training for accurate target recognition.
- Medical images are predominantly three-dimensional, necessitating three-dimensional convolution to extract three-dimensional information from the data, increasing the number of parameters and making it simple to overfit.
- Medical images contain multiple data modalities, including MRI and PET images in the OASIS-3 dataset. Models trained on specific data types may not apply to other types, requiring them to derive features from various modalities to enhance generalization capability.

Due to these characteristics inherent to medical images, the segmentation of medical images requires using an

TABLE 4. Comparing deep learning-based image segmentation methods' technical features and benefits.

Algorithms	Feature	Benefits	Drawbacks
FCN [132]	Fully convolutional network framework with skip connections	no input image size limit, high-level features fused with low-level features	insufficient image context information, low segmentation accuracy.
U-Net [135]	Implementing a network for decoding based on the convolutional network	Effective blending of low- and high-level characteristics	The segmentation effect at object boundaries is undesirable
U-Net++ [136]	Enhance its bypass connection based on U-Net.	The gap between downsampling and skip connection features is narrowed, and segmentation precision is high.	The network's architecture is intricate, and the number of parameters is extensive.
SegNet [136]	Maximum code pooling layer index refers to the top layer storage sample in a feature code map coder.	It minimizes sampling precision by avoiding borders	The segmentation accuracy is insufficient to meet real-time requirements.
DeepLab-V1 [137]	Optimize feature extraction using dilated convolution and connected CRF	expanding receptive field for more information extraction	The DCNN network reduces spatial resolution and requires a lot of storage.
DeepLab-V2	Introduce multiscale spatial pyramid pooling by leveraging DeepLab-v1.	expanding receptive field for more information extraction	Incapability to detect nuanced object boundaries and recover processing detail
DeepLab-V3 [138]	DeepLab-v2 improve atrous convolution and spatial pyramid pooling.	expanding receptive field for more information extraction	outputs deficient magnification.
ENet [144]	Encoder structures should be larger than decoder structures	reduce model parameters while maintaining accurate segmentation	boundary is irregular and not continuous.
ERRN [145]	Implement a high-resolution data stream for full-resolution information transmission	Increase border segmentation precision,	as simultaneous positioning and segmentation are challenging due to factors.
PSPNet [146]	A spatial pyramid module aggregates features at multiple sizes	improves picture context and local and global characteristics	loses target portion border information.

encoder-decoder structured network model. The increased complexity and difficulty associated with medical image segmentation technology is the primary factor generating significant interest in medical image segmentation within the field of image segmentation.

D. PROPOSED REGISTRATION-SEGMENTATION OPTIMIZATION DEEP LEARNING ARCHITECTURE

The current joint registration segmentation algorithm encounters difficulties when handling multimodal data. Researchers have proposed a new framework, the Multi-scale Neighborhood Descriptor and Consistency Supervision-based Joint Registration and Segmentation Network (MC-RSNet), that employs multi-scale neighborhood descriptors and modality consistency supervision to overcome this issue. The MC-RSNet integrates two segmentation networks and a multimodal registration network, allowing the segmentation networks to operate sequentially while the segmentation and registration processes work in parallel. By doing so, the MC-RSNet achieves simultaneous improvements in both registration and segmentation, addressing the limitations of existing joint registration segmentation algorithms. In summary, MC-RSNet is a novel framework that can simultaneously address the challenge of applying joint registration segmentation algorithms to multimodal data by achieving improved performance in registration and segmentation processes. The network comprises two branches: the segmentation sub-network and the registration sub-network.

The network takes M pairs of floating images I_m and fixed images If and N pairs of real segmentation labels L_m^{gt} and L_f^{gt}, $N < M$ as input to generate an optimized registration and segmentation model. I_m and I_f are input during image registration; the output is the deformation field σ. The original images I_m and I_f from the two modalities are input for image segmentation. The segmentation results in L_m^{seg} and L_f^{seg} of the two modal images are output. Consequently, the co-registration segmentation network is a well-designed model that can handle the task of registration and segmentation simultaneously. The modality consistency supervision loss helps to ensure the accuracy and reliability of the model's output. The study proposes a joint training framework for registration and segmentation using a pre-trained registration sub-network and segmentation sub-network. The pre-training process involves unsupervised training for the registration sub-network and supervised training for the segmentation sub-network on M pairs of data with N pairs of segmentation labels. The joint training process involves alternating between registration and segmentation while keeping the other fixed to achieve mutual improvement through complementary constraints. The authors utilize the complementary characteristics of the two networks to achieve consistency approximation between modalities. The pre-trained model is only used in the first round of training, and subsequent alternate training uses optimized networks from the previous step. Figure 5 shows the alternation process. The authors provide specific training steps for their proposed framework.

- Step1: The initial step in the joint framework is to pre-train all M pairs of medical images independently using the registration sub-network and N pairs of medical images with segmentation labels using the segmentation sub-network. The pre-trained network is then utilized as the initial network in the joint framework.
- Step2: The joint registration and segmentation network takes all M pairs of medical images and uses E_{msMIND} to ensure structure alignment. The framework also uses E_{dice} and $Econsistency$ to align the segmentation labels, whether real or not.
- Step3: Input all M pairs of medical images into the segmentation part of the joint framework, and use E_{dice} and $Econsistency$ constraints to segment the labels in both cases with and without ground truth labels, respectively.
- Step4: To obtain the final registration and segmentation model, the second and third steps are run alternately until the network becomes stable. However, since the registration and segmentation sub-network is not trained, the network parameters are randomly initialized, and the registration and segmentation results obtained may not be ideal. Directly using the joint optimization of registration and segmentation sub-networks may not be optimized correctly, leading to the gradual deterioration of results due to mutual constraints. To overcome this issue, the pre-training registration and segmentation sub-network provide better initial parameters, ensuring no deterioration during joint training. The joint training process is fine-tuned and optimized based on pre-training. It is important to note that only N pairs of images in all M pairs of CT-MR images contain real labels, and the upper-performance limit of the MCRSNet model proposed in this paper is reached when $N = M$.

The proposed method uses joint training instead of multi-task learning, as multi-task learning requires changes in the structure of the registration and segmentation network for new tasks. Joint training allows for registration and segmentation to be achieved simultaneously without altering the sub-network structure, allowing for any new deep learning registration and segmentation network to be directly combined. The joint training framework requires the input of two modalities' images to output the deformation and segmentation labels for the two modal images. The joint alternate optimization of registration and segmentation is mainly achieved through deformation fields and segmentation labels. The proposed framework is a plug-and-play network that allows for both deformation fields and segmentation label cases to be combined and applied in the joint framework. This approach allows for optimizing specific tasks based on existing models, avoiding changing the network structure and ensuring that the interaction between registration and segmentation improves the effect while reducing workload.

V. MEDICAL IMAGE TARGET RECOGNITION AND CLASSIFICATION

A. MEDICAL IMAGE TARGET RECOGNITION

Clinical diagnosis involves identifying and locating potential lesions in medical images. This process can be achieved through medical image target recognition, which utilizes algorithms to analyze and classify regions of interest within an image. The standard approach of handwritten marker recognition is a laborious and time-consuming procedure. However, with deep learning standards, images can be split between smaller blocks and processed through convolutional neural networks (CNN) and other models to create a binary classification model. This model then determines whether the block belongs to the target region. Deep learning algorithms have shown significant potential in accurate and efficient medical image recognition, as demonstrated in Figure 5, and have been utilized in object detection in recent studies [171].

The measurement discipline has progressed quickly, introducing Fast R-CNN and Mask R-CNN models. These models have enabled medical image input models to detect potential target areas accurately. However, these models consist of one area proposal module and one classification module, both requiring iterative updates, making the model speed insufficient to meet real-time clinical requirements. The YOLO (you only look once) and SSD (single shot multibox detector) models have been introduced to address this timeliness issue, solving the target detection speed problem. These models have been further improved upon by Lin T Y et al. [172] with the development of the RetinaNet

TABLE 5. Summary of medical image classification based on deep learning.

Organ	Used Dataset	Approaches	Target	Precision	Ref.
Lung	LIDC-IDRI	CNN	Nodule Classification	Sensitivity 0.85	[149]
	LIDC-IDRI	Knowledge Transfer Learning	Nodule classification	AUC 0.95	[150]
	Chest X-ray	CNN	Lung Cancer Grading	AUC 0.87	[151]
	HUG-HRCT Scans	Transfer Learning	Lung Cancer Grading	F1-Score 0.88	[152]
	LIDC CT Images	Transfer Learning	Lung Cancer Detection	AUC 0.81	[153]
	LIDC-IDRI	Knowledge Transfer Learning	Pulmonary Nodules classification	AUC 0.95	[154]
	LIDC-IDRI	Manifold Learning	Pulmonary Nodules classification	ACC 0.90	
Breast	MIAS, BCDR	Transfer Learning	Breast Cancer Grading	AUC 0.99 (MIAS) AUC 0.95 (BCDR)	[155]
	Histopathology	Deep Belief Networks	Breast Grading	ACC 0.86	[156]
Skin	SIC, ISBI (2016-17),	Knowledge Transfer Learning	Dermatology Classification	AUC 0.91 AUC 0.91	[157] [158]
	ISIC	ML-based Algorithms	Dermatology Classification	OCA 0.97	[159] [158]
	ISIC	Hybrid Features learning model	Dermatology Classification	AUC 0.78	[160]
Fundus	UCSD NEI Internal dataset	Transfer Learning (AlexNet, CNN)	Retinopathy Grading, Glaucoma Classification	ACC 0.96, AUC 0.97 ACC 0.91, AUC 0.96	[161] [162]
Lever	MICCAI-Sliver07 3Dircadb01	Deep Belief Networks	Liver segmentation	Dice 0.94 Dice 0.91	[163]
	MICCAI-Sliver07	Deep Stacked Auto-Encoder	Liver segmentation	Dice 0.90	[164]
	MICCAI-Sliver07	CNN-LivSeg	Liver segmentation	ACC 0.95	[165]
	MICCAI-Sliver07	Gaussian-weight initialization of CNN	Liver segmentation	Dice 0.95	[166]
Vertebrae	MICCAI CSI	Overlapping patch-based convNet	Vertebrae segmentation	ACC 0.98	[167]
	MICCAI CSI	Vertebrae segmentation (SVseg)	Vertebrae segmentation	Dice 0.77	[168]
	VerSe, CSI-Seg,	Stacked sparse autoencoder	Vertebrae segmentation	ACC 0.90	[169]
	MICCAI CSI	Patch-based deep belief networks	Vertebrae segmentation	Dice 0.86	[170]

model. This model has been extended to recognize diseases such as breast tumors in mammography images [102] and detect lung nodules in CT images [173]. It should be noted that the aforementioned models only detect targets in 2D images and do not assume the spatial facts between slices in a 3D image. To improve the recognition accuracy, models based on RNN and LSTM differential models have been applied to medical images [174].

Insufficient data is a significant challenge in medical image target recognition. Transfer learning has been gradually developed for medical image recognition to overwhelm this issue. This includes models that are transferred from ImageNet data to detect lung nodules [131], breast cancer [175], and colorectal polyps. Transfer learning based on clinical experience and knowledge has also been applied to object detection in medical images. The AGCL model, for instance, utilizes attention-based curriculum learning to detect tumors in chest radiographs..In contrast, the CASED model utilizes adaptive curriculum sampling to address extreme data imbalance and identify pulmonary nodules in CT images. The feature pyramid network (FPN) model is another example that exploits multi-scale attention to detect tumors using images of different contrasts.. The current research is on amalgamating artificial intelligence technology and clinical diagnosis to achieve image classification and object recognition in medical imaging. The author provides a few instances of classification recognition to gain insight into its development direction. In Table 5 presents a comparative analysis of frequently utilized medical image datasets and deep learning models for tumor classification, including an evaluation of their respective classification performance.

B. MEDICAL IMAGE CLASSIFICATION

Medical image grading and recognition are crucial in CAD systems. Before the advent of deep learning, the classification of medical images relied on artificially defined image features such as texture, shape, and grayscale histogram. Following feature selection, machine learning models such as support vector machines, logistic regression, and random forest were commonly used for image classification. Radiomics methods are a typical example of this approach and have been used for tumor classification, staging, and treatment prognosis prediction. However, manually defining features and the feature selection process greatly affect the reliability and robustness of classification. With the fast growth of deep learning models, especially CNNs, neural network models that automatically extract, select, and classify information have become the norm. Different variants of CNN-based models have been widely utilized for clinical disease diagnosis based on medical imaging, such as the Kaggle public datasets for fundus mapping. Shanthi et al. [161] used a modified advanced AlexNet for diabetic retinopathy classification, achieving an accuracy of approximately 96.6%. VGG has been used to categorize benign and malignant pulmonary nodules using chest radiography, attaining a precision of up to 99

ResNet and VGG are the most typically utilized CNN variants for medical image classification, achieving the best performance in tumor detection, neurological disease classification, cardiovascular disease detection, and other research fields. However, medical image data is limited compared to natural image data, making it difficult to meet the model's training requirements. Therefore, knowledge transfer has become the mainstream method for training medical image classification models to improve accuracy. Knowledge transfer can be achieved by transferring parameters of natural image-trained network models to extract medical image features, combined with traditional machine learning methods, or by using these parameters as initialization for the medical image-trained model and fine-tuning for medical image classification [157]. In addition to raw image transfer, other medical image datasets can also be utilized. Multi-task learning can be applied to share data information and overcome the classification defects caused by insufficient data. [175]

Incorporating clinical knowledge into medical image classification can enhance the accuracy of diagnosis by considering factors such as a physician's experiential learning style, diagnostic methods, and areas and features of diagnostic concern. One approach uses a "curriculum learning" model, where the image classification tasks are divided from easy to difficult. The model first learns simple tasks before progressing to more complex ones based on the clinician's diagnosis experience experiment [129]. Another approach is interrupt mode, where the entire medical image is quickly browsed. Then certain slices are selected for diagnosis, developing whole focal and partial classification models. Attention-based CNN models, such as attention-based CNN for glaucoma detection (AG-CNN) [176] and LACNN (lesion aware CNN) [177], have also been developed to improve classification accuracy by focusing on certain regions of the image. In addition, a trend in medical image classification is the fusion of deep learning features with manually defined features based on empirical characteristics used by doctors, such as tumor size, shape, and boundary information. Researchers have compared the results of using artificial versus deep learning features and achieved improved accuracy in skin cancer and glaucoma classification. However, effectively fusing traditional artificial features with deep learning features remains a challenge in model design.

VI. DEEP LEARNING IN MEDICAL IMAGE ANALYSIS: APPLICATIONS AND DEVELOPMENT TRENDS

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is a feedforward neural network that processes input image pixel matrices by passing them through a layer of filters to select features. The input feature matrix is then compressed through a pooling layer, resulting in a smaller feature matrix and reduced computational complexity. A CNN comprises one or more convolutional layers and a fully connected layer at the top, analogous to a traditional neural network. These layers are made up of associated weights and pooling layers. CNNs can provide superior image and speech recognition results compared to other deep learning architectures.

1) APPLICATION STATUS OF CNN MODEL

The field of medical images has produced a substantial body of research using CNN, which is widely considered the most popular model in this area. Expert teams have leveraged this model to study different parts of the human body domestically and internationally [178]. This paper summarizes CNN's application in the medical brain, eyes, and chest images. CNN's application in medical image recognition dates back to 1995 when it was used in the fusion of the dual matching method and artificial visual neural network technology for pulmonary nodule detection. Generally, neural network technology is used for grayscale imaging and has the potential for clinical applications, as shown by tests conducted at the DEC Alpha studio. Since then, the use of CNN in medical images has gradually developed. Regarding CNN's application for image segmentation, Hao Chen, Qi Dou, and colleagues from the Chinese University of Hong Kong conducted several studies. They introduced the VoxResNet residual network, which discusses deep residual learning in volumetric brain segmentation tasks, and extended it to handle 3D variants for volumetric data. Furthermore, the authors put forth an automated contextual iteration of VoxResNet, which enhances the efficacy of rudimentary visual cues, implied structural cues, and comprehensive contextual amalgamation. This study uncovers the potential of 3D deep learning in enhancing recognition performance for volumetric image segmentation. The authors subsequently presented a framework based on 3D CNNs for automatic liver segmentation in CT volumes with abnormalities. This approach resulted in improved accuracy in liver segmentation. The authors have devised a sturdy and effective two-step methodology for the automated identification of cerebral microvessels from SWI images. Additionally, they have introduced a fresh technique for the automated detection of pulmonary nodules utilizing 3D CNNs, which effectively minimizes false positives. The Diagnostic Image Analysis Group at Radboud University Medical Center in Nijmegen, Netherlands, has researched CNN-based image classification. They have developed a deep learning system that utilizes multi-stream multi-scale convolutional networks. This system can classify all relevant nodule types for treatment, demonstrating the model's usability and accuracy. Regarding clinical image dataset training, Mark Cicero and other researchers trained a CNN model using chest X-rays of 35,038 patients from 2005 to 2015 to detect and rule out common pathologies in chest X-rays, demonstrating the CNN's high sensitivity and clinical usability. While domestic research on CNN's application in medical image analysis started in 2015, the theoretical exploration of CNN models for a single disease is yet in its earlier stages. More experiments are needed to verify their accuracy and clinical usability. Li Wen's study shows that CNN-based dynamic learning features can enhance tumor segmentation accuracy. Deep learning models can significantly improve the auxiliary diagnosis system's accuracy, sensitivity, and specificity compared to traditional algorithms.

B. STACKED AUTOENCODERS

The Autoencoder (AE) is a type of neural network utilized for unsupervised learning, featuring a solitary, hidden layer [169], [179]. The Stacked Autoencoder is a deep network composed of simple structures stacked on each other. The training process for the Autoencoder utilizes a layer-by-layer pre-training algorithm, where training is carried out through error reconstruction. The Bengio scholars introduced the Stacked Autoencoder, composed of three layers: an input layer, a hidden layer, and an output layer connected sequentially. The goal of the Autoencoder model during training is to fit the input data by setting the network output to be equal to the input, using the backpropagation algorithm for training. Despite being based on a supervised learning algorithm, the Autoencoder model does not require the original data to have classification labels, making the entire training process unsupervised [180], [181]. To construct a Stacked Autoencoder, the first autoencoder is trained, its hidden layer is used as input to train the second autoencoder, and so on. Multiple autoencoders are trained and stacked sequentially to form the Stacked Autoencoder model. The last layer of the Stacked Autoencoder represents the abstract feature obtained after multiple transformations of the input data. Finally, the problem's unique characteristics connect various output layers, and the output layer weights are educated using a supervised learning method to get the definitive classification conclusion.

1) APPLICATION STATUS OF STACKED AUTOENCODER MODEL

Compared to previous models, the stacked autoencoder model has a relatively limited scope of application [182]. However, some scholars have conducted theoretical research and experiments on its use in medical image analysis, particularly in digital pathology, microscopes, and various human body organs, such as the brain, heart, kidney, liver, and abdomen. For example, Hoo-Chang and Matthew used deep learning methods to recognize organs in magnetic resonance medical images. The utilization of stacked autoencoders was suggested by Dinggang and colleagues as a means of amalgamating various tiers of characteristics for the classification of AD/MCI, resulting in notable diagnostic precision. Other research teams have proposed combining deep learning with state-space modeling to diagnose mild cognitive impairment based on rs-fMRI and evaluated their method on multiple datasets. The use of stacked autoencoders has gone through various development stages and has greatly improved the accuracy of medical image recognition. However, training time remains challenging, and obtaining large medical image datasets is important for further improvements [183]. Deep learning has quickly become a popular research topic in medical image analysis, with numerous models applied to various tasks. The scarcity of extensive training datasets has been identified as a hurdle; however, many images are presently accessible in Western healthcare facilities, and public datasets are progressively becoming more accessible. Systematically obtaining medical records and annotations remains a challenge.

C. DEEP BELIEF NETWORK (DBN)

DBN is a type of deep network consisting of multiple Restricted Boltzmann Machines (RBM) layers [163], [184], [185]. Using unsupervised pre-training and fine-tuning is employed to train the entire network. During pre-training, each layer is trained separately with unlabeled data. This approach is limited to only one Boltzmann machine. The lower layer is restricted using a bottom-up approach, where the output of the lower Boltzmann machine is used as input to the upper layer's restricted Boltzmann machine. Although pre-training provides a better initial value for the network, it is not optimal [12], [156], [170], [186]. After completing the pre-training phase, the neural network is subjected to fine-tuning, which involves the application of the backpropagation algorithm, often accompanied by utilizing the dropout technique. The emergence of deep belief networks represents a significant milestone in deep learning. The successful implementation of these techniques has been observed in various fields, such as speech and image manipulation, especially in the context of extensive datasets. Ahmad et al., [163] proposed an automatic feature learning algorithm using DBN in CT scan images for liver segmentation. The DBN-DNN method saves memory and computational time using blocks for feature learning and an automatic active contour method for refinement. The algorithm achieves satisfactory results on healthy and pathological liver images, with a 94.80% Dice similarity coefficient on mixed images and 91.83% on pathological images. Hirra et al., [156] proposed a new patch-based deep learning method (Pa-DBN-BC) for accurate breast cancer detection and classification in histopathology imaging. It uses DBN and LR algorithms to extract features from histopathology patches. After training and testing on a slide histopathology dataset, the method achieved 86% accuracy, outperforming traditional methods.

1) APPLICATION STATUS OF DEEP BELIEF NETWORK MODEL

Deep Bayesian Networks (DBNs) have gained significant interest in recent years, particularly in the human body's neural, thoracic, and cardiovascular systems. Tom Brosch and his team used a deep generative model to reduce the dimensionality of input photos, enabling 3D medical image training with a maximum resolution of $128 \times 128 \times 128$. DBNs have shown potential in analyzing brain morphometric data, such as schizophrenia, and have been shown to perform lung tissue classification and airway recognition in CT images. Researchers have also investigated deep-learning algorithms for diagnosing lung cancer, using a fusion of deep-learning techniques and level ensembles to partition cardiac electromagnetic resonance data for medical picture segmentation. DBNs have evolved from their initial applications in voice recognition and natural language processing to include image processing, indicating a shift in focus within the field.

D. TRANSFORMER-BASED APPROACHES

This study explores ViT and Swin Transformer-based models for medical image segmentation, classification, and registration. ViT divides an input image into fixed-size slices, which are then linearly mapped using matrix multiplication. Position encoding information is added to each slice before sending it to the Transformer encoder, consisting of L standard Transformer modules. Swin Transformer reduces the computational complexity of self-attention by introducing a moving window mechanism based on the visual Transformer. Medical image segmentation is crucial for classification work and determining the size, shape, and other characteristics of segmented lesions for benign and malignant tumors and preoperative analysis. In 2021, many studies have combined visual Transformers with medical image segmentation, achieving performance close to or even better than CNN on different data sets. The Transformer-based model, originally designed for language translation tasks, has been repurposed to tackle medical image segmentation and classification intricacies. It dissects medical images into consistent patches, infusing them with two-dimensional positional indicators to maintain spatial context.

VII. DISCUSSION AND ANALYSIS

Medical image segmentation and localization detection are essential areas of exploration in the realm of medical image processing and analysis. These domains hold key significance as they facilitate the extraction of relevant regions of interest from medical imaging data, automating tasks like segmenting from lung, breast, kidney, heart, brain, and vertebral body positioning and labeling. This automation substantially alleviates the arduous data screening process for medical professionals, aids in lesion analysis and localization, and expedites the formulation of treatment strategies. The intricate nature of the human body's anatomy and the diverse medical imaging devices introduce challenges, including potential interferences during the imaging process. Unlike natural images, medical images possess distinct traits such as considerable sample variations, elevated noise levels, and diminished contrast. Consequently, this frequently results in imprecise boundaries and a decline in overall model performance, posing a challenge in meeting clinical requisites. This research underscores the pivotal role of automated medical image segmentation, enabling physicians to focus intently on specific areas of concern, thereby enhancing pathological examination and diagnosis. The variability introduced by distinct imaging devices and potential interferences during imaging accentuates the necessity for tailored approaches, as the adaptability of many proposed networks to diverse medical images is constrained. Accurate segmentation is paramount for meticulous medical diagnosis, given that inaccuracies can impede precise medical assessments and delay optimal patient treatment. As primary networks, multiple algorithms incorporate global context attention mechanisms to mitigate these challenges by capturing mesoscale information and leveraging synergies between sub-networks and backbone networks. The decoder integrates a multi-branch attention gate model in the information fusion process, effectively harmonizing profound semantic comprehension with nuanced feature details. The outcome is a holistic, efficient, and accurate automated medical image segmentation methodology. The study also delves into network interpretability, evaluating the capacity to convey information and the feasibility of compressing channels. Rigorous compression tests are conducted on the model's Encoder, Decoder, and Bottleneck components to gauge the extent of data redundancy. Data sensitivity remains a concern given the lack of medical imaging data and the pronounced variations among different centers. Prioritizing data preprocessing techniques and expanding the dataset through example augmentation are vital strategies. The notion of set division is explored, shedding light on its repercussions for model performance concerning data quality. Furthermore, this study expounds upon three archetypal deep learning-based models' foundational principles and evolutionary trajectories. It encapsulates the growing integration of deep learning into medical image analysis, driven by advancements in artificial intelligence research. Worth noting is the attainment of recognition accuracy rates exceeding 90% for individual pathological slices through CNNs and stacked autoencoders trained on medical image datasets. However, the challenge persists in accurately discerning intersecting diseases due to the intricate and variable sizes and shapes of tumors. Furthermore, deep learning models come with their own set of challenges that warrant targeted attention.''

In the future phase, the study will further investigate preserving the existing segmentation effect by using three-dimensional spatial information about anatomical components as a constraint for the network. While most existing networks face challenges balancing accuracy and computational overhead, one can enhance the network's complexity to improve segmentation accuracy. Alternatively, sacrificing precision within an acceptable range can lead to more significant savings in computational overhead.

A. CHALLENGES IN DEEP LEARNING MODELS

- Despite the emergence of numerous deep learning models since Hinton proposed the concept, most of these models still rely on simple models stacked on each other, making a novel and more effective model a focus of attention.

- The interpretability of deep learning models is still challenging, hindering their adoption in clinical practice and making it challenging for clinicians to understand the reasoning behind their decisions.
- Deep learning models may suffer from biases in the training data, leading to discriminatory outcomes and disparities in healthcare.
- Current models still rely heavily on unsupervised learning methods, and there is a need to move towards true unsupervised learning.
- As the problems that deep learning models are used to solve become increasingly complex, longer training times are required due to the increased number of benchmark parameters. To address this, algorithm improvements are necessary to increase training speed and reduce time spent on training.
- With the growing volume of unlabeled data, an updated automated labeling technology is necessary as manual labeling is no longer efficient in the modern information society.
- The scarcity of labeled data can significantly affect the precision and consistency of predictions, and ongoing research interests include investigating the potential of weak.
- The Small changes in actual samples can cause classifiers to misclassify them, and current common regularization methods are insufficient to address this problem.

B. PROSPECTS OF DEEP LEARNING

Deep learning-based medical image segmentation methods eliminate human involvement and play a crucial role in medical image processing. However, comparing the literature on each method reveals challenges in the development and evolution of these networks.

- Medical image resolution is increasing, but current hardware struggles with processing high-resolution images. Cropping and training images limit the network's ability to extract more spatial data.
- Acquiring medical imaging datasets is challenging due to varying data labeling requirements. Most datasets are conventional, and the quantity of training data directly impacts the model's training effect. Insufficient training data can lead to overfitting, affecting performance on other data sets.
- Medical image datasets often have sample imbalance, leading to model bias in deep neural networks. This is particularly problematic for diagnosing and treating diseases. To overcome these challenges, researchers are increasingly focusing on medical image processing and discovering innovations to improve the accuracy of medical image segmentation.
- Supervised-training models struggle with large training data, making image segmentation under semi-supervised or unsupervised conditions a primary research focus in the future without labeled data.
- Generative adversarial networks generate datasets for medical image analysis. Combining GAN framework image data with original data enhances model performance. However, equitably partitioning work between original and derived data is a pressing issue that must be resolved in the near future.
- Collaboration between hospitals, vendors, and machine learning experts is crucial for deep learning in medical image analysis. This technology relies on patient data, which is essential for accurate models. Collecting patient data can be challenging, but collaboration between hospitals, vendors, and experts can provide access to data that machine learning researchers may not otherwise have.
- Deep learning relies on large image datasets, but medical data is challenging to annotate due to its high cost, time consumption, and expert learning requirements. Shared data resources are needed for medical service providers to develop more accurate deep-learning models by accessing annotated medical images.
- Deep learning techniques for analyzing medical images face challenges due to the difficulty in acquiring labeled data. Unsupervised or semi-supervised learning methods are being studied to overcome this obstacle. However, developing accurate models using these methods is complex, and a complete solution has not yet been found. Despite these challenges, deep learning in medical image analysis holds practical and social value in high-precision intelligent identification, analysis, and diagnosis.

Deep learning in medical imaging has led to computer vision specialists participating in image semantic segmentation tasks. The medical imaging community constantly develops new theories and technologies to expand application opportunities. Computer-aided diagnosis, an emerging field combining clinical medicine, graphic imaging, and computer science, uses deep learning models and PACS to automatically describe and label medical images, creating an intelligent database. This technology can aid physicians in screening, reduce reading time, improve judgment accuracy, work efficacy, and decrease missed clinical detection rates. There is still much to discover in medical image investigation and intelligent diagnosis, with enormous potential in the field.

VIII. CONCLUSION

This study analyzes deep learning models used in medical image analysis, including recognition, registration, segmentation, and classification. Medical image segmentation is crucial for quantitative analysis, 3D reconstruction, and registration. Segmentation of medical images based on deep learning is significant to medical research and clinical disease diagnosis. It plays a critical role in clinical processes like radiation therapy, image-guided surgery, and pathological diagnosis, emphasizing the importance of ongoing research and development. Deep learning-based CNNs are widely used in image segmentation, but medical images are mostly single-channel grayscale with low contrast and signal-noise ratio. Simple encoder-decoder networks can extract fine-grained and global context semantic features simultaneously, making it difficult to segment medical targets with significant shape differences accurately. Additionally, damage tragedy is adhered to surrounding tissue, making convolution difficult to extract boundary features and segment damaged issues. While DL-based models like FCN, U-Net, SegNet, Deep Lab, ENet ERRN, and PSPNet in medical image segmentation and CNN, SAE, DBN, RNN attention mechanism, graph model, and transfer learning have made significant progress in medical image segmentation and classification, their application in clinical settings remains limited due to limitations. Collecting medical imaging data that meets clinical standards can be challenging, and supervised learning is the most commonly used technique. However, the scarcity of labeled data can significantly impact the precision and consistency of predictions. An ongoing research area is exploring the potential of weak supervision, transfer learning, and multi-task learning approaches to improve classification predictions using limited labeled data.

In clinical settings, interpretability is crucial for medical imaging. Current deep learning models struggle to answer questions about features, and researchers suggest using visualization and parametric analysis to interpret patterns. However, there is still a gap in creating interpretable imaging markers that meet clinical requirements. Investigating methods to enhance the interpretability and robustness of models is a crucial focus of medical imaging research. Moreover, enhancing model predictions' robustness remains a challenge, as most models perform well on a single dataset but cannot make accurate predictions on other datasets without training. Factors like acquisition parameters, equipment, and time can affect image performance, leading to poor generalization and robustness. To address this issue, research on enhancing the generalization capability of deep learning models in medical image applications through cognitive thinking structures and training techniques is essential for future investigation.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT

This research is supported by the Artificial Intelligence & Data Analytics Lab (AIDA), CCIS, Prince Sultan University, Riyadh, Saudi Arabia. The author would also like to thank Prince Sultan University, Riyadh, Saudi Arabia, for support.

REFERENCES

[1] S. Iqbal et al., "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Arch. Comput. Methods Eng.*, vol. 30, no. 5, pp. 3173–3233, 2023.

[2] R. A. Ramadan, A. Y. Khedr, K. Yadav, E. J. Alreshidi, M. H. Sharif, A. T. Azar, and H. Kamberaj, "Convolution neural network based automatic localization of landmarks on lateral X-ray images," *Multimedia Tools Appl.*, vol. 81, no. 26, pp. 37403–37415, Nov. 2022.

[3] A. Imran, J. Li, Y. Pei, F. Akhtar, T. Mahmood, and L. Zhang, "Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network," *Vis. Comput.*, vol. 37, no. 8, pp. 2407–2417, Aug. 2021.

[4] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Arch. Comput. Methods Eng.*, vol. 30, no. 5, pp. 3173–3233, 2023.

[5] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1746–1751.

[6] S. M. Fati, E. M. Senan, and A. T. Azar, "Hybrid and deep learning approach for early diagnosis of lower gastrointestinal diseases," *Sensors*, vol. 22, no. 11, p. 4079, May 2022.

[7] F. Shi, L. Li, Q. Li, H. Zhu, X. Liang, H. Li, and J. Tian, "Deep learning in medical image analysis: A review," *Med. Image Anal.*, vol. 78, no. 1, 2022, Art. no. 102122.

[8] J. Alyami, A. Rehman, F. Almutairi, A. M. Fayyaz, S. Roy, T. Saba, and A. Alkhurim, "Tumor localization and classification from MRI of brain using deep convolution neural network and salp swarm algorithm," *Cognit. Comput.*, vol. 1, pp. 1–11, Jan. 2023.

[9] M. M. Adnan, M. S. M. Rahim, A. R. Khan, A. Alkhayyat, F. S. Alamri, T. Saba, and S. A. Bahaj, "Automated image annotation with novel features based on deep ResNet50-SLT," *IEEE Access*, vol. 11, pp. 40258–40277, 2023.

[10] S. Ali, J. Li, Y. Pei, R. U. Khurram, and T. Mahmood, "A comprehensive survey on brain tumor diagnosis using deep learning and emerging hybrid techniques with multi-modal MR image," *Arch. Comput. Methods Eng.*, vol. 29, no. 7, pp. 4871–4896, 2022.

[11] S. Daud, M. Ullah, A. Rehman, T. Saba, R. Damaševičius, and A. Sattar, "Topic classification of online news articles using optimized machine learning models," *Computers*, vol. 12, no. 1, p. 16, Jan. 2023.

[12] G. Hinton, S. Osindero, and Y.-W. Teh, "Fast and deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[13] X. Feng, M. Li, X. Zhou, X. Huang, X. Yu, M. Wang, Z. Hu, C. Fang, and Z. Chen, "Deep learning in medical image analysis," *Comput. Biol. Med.*, vol. 139, no. 1, 2022, Art. no. 105058.

[14] K. U. Rehman, J. Li, Y. Pei, A. Yasin, S. Ali, and T. Mahmood, "Computer vision-based microcalcification detection in digital mammograms using fully connected depthwise separable convolutional neural network," *Sensors*, vol. 21, no. 14, pp. 1–25, 2021.

[15] S. Z. Kurdi, M. H. Ali, M. M. Jaber, T. Saba, A. Rehman, and R. Damaševičius, "Brain tumor classification using meta-heuristic optimized convolutional neural networks," *J. Personalized Med.*, vol. 13, no. 2, p. 181, Jan. 2023.

[16] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, "SIGNFORMER: DeepVision transformer for sign language recognition," *IEEE Access*, vol. 11, pp. 4730–4739, 2023.

[17] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.

[18] M. Yaqub, F. Jinchao, K. Arshid, S. Ahmed, W. Zhang, M. Z. Nawaz, and T. Mahmood, "Deep learning-based image reconstruction for different medical imaging modalities," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–18, Jun. 2022.

[19] G. Yang, W. Shi, S. Liu, X. Liu, J. Yang, S. Zhang, and S. Ji, "Deep learning in medical image analysis: An overview," *Frontiers Med.*, vol. 9, no. 1, p. 150, 2022.

[20] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. Alsafari, and I. A. Hameed, "A machine learning and blockchain based efficient fraud detection mechanism," *Sensors*, vol. 22, no. 19, p. 7162, Sep. 2022.

[21] T. Mahmood, J. Li, Y. Pei, F. Akhtar, M. U. Rehman, and S. H. Wast, "Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach," *PLoS ONE*, vol. 17, no. 1, pp. 1–30, 2022.

[22] M. Ghassemi, T. Naumann, P. Schulam, and A. L. Beam, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 23, no. 1, pp. 50–66, 2022.

[23] S. Iqbal, A. N. Qureshi, A. Ullah, J. Li, and T. Mahmood, "Improving the robustness and quality of biomedical CNN models through adaptive hyperparameter tuning," *Appl. Sci.*, vol. 12, no. 22, p. 11870, Nov. 2022.

[24] J. Waleed, A. T. Azar, S. Albawi, W. K. Al-Azzawi, I. K. Ibraheem, A. Alkhayyat, I. A. Hameed, and N. A. Kamal, "An effective deep learning model to discriminate coronavirus disease from typical pneumonia," *Int. J. Service Sci., Manage., Eng., Technol.*, vol. 13, no. 1, pp. 1–16, Mar. 2023.

[25] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An automated in-depth feature learning algorithm for breast abnormality prognosis and robust characterization from mammography images using deep transfer learning," *Biology*, vol. 10, no. 9, p. 859, 2021.
[26] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
[27] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, no. 1, pp. 1–29, 2021.
[28] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran, and M. Yaqub, "An automatic detection and localization of mammographic microcalcifications ROI with multi-scale features using the radiomics analysis approach," *Cancers*, vol. 13, no. 23, p. 5916, 2021.
[29] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[30] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
[31] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel, and H. Lee, "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2415–2418.
[32] S. Ali, S. Sahiba, M. Azeem, Z. Shaukat, T. Mahmood, Z. Sakhawat, and M. S. Aslam, "A recognition model for handwritten persian/arabic numbers based on optimized deep convolutional neural network," *Multimedia Tools Appl.*, no. 82, no. 10, pp. 14557–14580, 2023.
[33] K. Javaid, A. Siddiqa, S. A. Z. Naqvi, A. Ditta, M. Ahsan, M. A. Khan, T. Mahmood, and M. Adnan Khan, "Explainable artificial intelligence solution for online retail," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 4425–4442, 2022.
[34] Z. Ahmad, J. Li, and T. Mahmood, "Adaptive hyperparameter fine-tuning for boosting the robustness and quality of the particle swarm optimization algorithm for nonlinear RBF neural network modelling and its applications," *Mathematics*, vol. 11, no. 1, p. 242, 2023.
[35] M. A. Haidekker, "X-ray projection imaging," in *Medical Imaging Technology*. New York, NY, USA: Springer, 2013, doi: 10.1007/978-1-4614-7073-1_2.
[36] A. Jabbar, S. Naseem, T. Mahmood, T. Saba, F. S. Alamri, and A. Rehman, "Brain tumor detection and multi-grade segmentation through hybrid caps-VGGNet model," *IEEE Access*, vol. 11, pp. 72518–72536, 2023.
[37] V. I. Mikla and V. V. Mikla, *Medical Imaging Technology*. Amsterdam, The Netherlands: Elsevier, 2013.
[38] M. S. M. Rahim, A. Norouzi, A. Rehman, and T. Saba, "3D bones segmentation based on ct images visualization," *Biomed. Res.*, vol. 28, no. 8, pp. 3641–3644, 2017.
[39] National Electrical Manufacturers Association Medical Imaging & Technology Alliance. (2022). *Medical Image Modalities*. [Online]. Available: https://www.medicalimaging.org/about-mita/modalities
[40] C. F. Dietrich, L. Bolondi, F. Duck, D. H. Evans, C. Ewertsen, A. G. Fraser, O. H. Gilja, C. Jenssen, E. Merz, C. Nolsoe, D. Nürnberg, H. Lutz, F. Piscaglia, A. Saftoiu, P. Vilmann, Y. Dong, C. R. Hill, "History of ultrasound in medicine from its birth to date, on occasion of the 50 years anniversary of EFSUMB," *Med. Ultrasonogr.*, vol. 24, no. 4, pp. 434–450, 2022.
[41] S. Iqbal, A. N. Qureshi, J. Li, I. A. Choudhry, and T. Mahmood, "Dynamic learning for imbalance data in learning chest X-ray and CT images," *Heliyon*, vol. 9, no. 6, pp. 1–20, 2023.
[42] F. E.-Z.-A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egyptian Informat. J.*, vol. 17, no. 1, pp. 99–124, Mar. 2016.
[43] *U.K. Biobank*. Accessed: Feb. 15, 2023. [Online]. Available: https://www.ukbiobank.ac.uk
[44] *Dataset Directory*. Accessed: May 23, 2023. [Online]. Available: https://www.acrdsi.org/DSI-Services/Dataset-Directory
[45] *Autism Brain Imaging Data Exchange*. Accessed: Feb. 15, 2023. [Online]. Available: http://fcon_1000.projects.nitrc.org/indi/abide
[46] *Give A Scan*. Accessed: Feb. 15, 2023. [Online]. Available: http://giveascan.org
[47] *Welcome to the Fastmri Dataset*. Accessed: Feb. 15, 2023. [Online]. Available: https://fastmri.med.nyu.edu
[48] *Openneuro*. Accessed: Feb. 14, 2023. [Online]. Available: https://openneuro.org
[49] K. S. Mader. *CT Images From Cancer Imaging Archive With Contrast and Patient Age*. Accessed: May 23, 2023. [Online]. Available: https://www.kaggle.com/kmader/siim-medical-images
[50] *The Alzheimer's Disease Neuroimaging Initiative*. Accessed: Feb. 14, 2023. [Online]. Available: http://adni.loni.usc.edu
[51] *List of Open Access Medical Imaging Datasets*. Accessed: May 23, 2023. [Online]. Available: https://www.radrounds.com/profiles/blogs/list-of-open-access-medical-imaging-datasets
[52] *TCIA Collections*. Accessed: Feb. 15, 2023. [Online]. Available: https://www.cancerimagingarchive.net
[53] *Machine Learning Datasets*. Accessed: May 23, 2020. [Online]. Available: https://www.datasetlist.com
[54] S. Aylward. *Open-Access Medical Image Repositories*. Accessed: May 23, 2023. [Online]. Available: http://www.aylward.org/notes/open-access-medical-image-repositories
[55] K. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, M. Sallam, K. Chang, and K. Woods, "The digital database for screening mammography," in *Proc. 3rd Int. Workshop Digit. Mammography*, vol. 58, 1996, p. 27.
[56] P. Bándi et al., "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 550–560, Feb. 2019.
[57] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
[58] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.
[59] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, "Interactive whole-heart segmentation in congenital heart disease," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-24574-4_10.
[60] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 241–246.
[61] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder–decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
[62] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.
[63] C. You, W. Cong, G. Wang, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, and Y. Zhang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018.
[64] Y.-J. Ma, Y. Ren, P. Feng, P. He, X.-D. Guo, and B. Wei, "Sinogram denoising via attention residual dense convolutional neural network for low-dose computed tomography," *Nucl. Sci. Techn.*, vol. 32, no. 4, pp. 1–14, Apr. 2021.
[65] X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y. Chen, H. Shu, L. Luo, and J.-L. Coatrieux, "Domain progressive 3D residual convolution network to improve low-dose CT imaging," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2903–2913, Dec. 2019.
[66] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O'Regan, and D. Rueckert, "Multi-input cardiac image super-resolution using convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2016, pp. 246–254.
[67] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 197–200.
[68] S. McDonagh, B. Hou, A. Alansary, O. Oktay, K. Kamnitsas, M. Rutherford, J. V. Hajnal, and B. Kainz, "Context-sensitive super-resolution for fast fetal magnetic resonance imaging," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. Cham, Switzerland: Springer, 2017, pp. 116–126.
[69] C. Li, W. Li, C. Liu, H. Zheng, J. Cai, and S. Wang, "Artificial intelligence in multiparametric magnetic resonance imaging: A review," *Med. Phys.*, vol. 49, no. 10, pp. e1024–e1054, 2022.

[70] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, "Channel splitting network for single MR image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5649–5662, Nov. 2019.

[71] R. Tanno, D. E. Worrall, A. Ghosh, E. Kaden, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2017, pp. 611–619.

[72] C. Peng, W.-A. Lin, H. Liao, R. Chellappa, and S. K. Zhou, "SAINT: Spatially aware interpolation NeTwork for medical slice synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7747–7756.

[73] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu, "Super-resolution reconstruction of MR image with a novel residual learning network algorithm," *Phys. Med. Biol.*, vol. 63, no. 8, Apr. 2018, Art. no. 085011.

[74] Q. Lyu, H. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, and G. Wang, "Multi-contrast super-resolution MRI through a progressive network," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2738–2749, Sep. 2020.

[75] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 521–538, Mar. 2020.

[76] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.

[77] H. Zhang, B. Dong, and B. Liu, "JSR-Net: A deep network for joint spatial-radon domain CT reconstruction from incomplete data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3657–3661.

[78] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1985–1995, Sep. 2018.

[79] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magn. Reson. Med.*, vol. 80, no. 3, pp. 1189–1205, Sep. 2018.

[80] T. Eo, Y. Jun, T. Kim, J. Jang, H. Lee, and D. Hwang, "KIKI-Net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2188–2201, Nov. 2018.

[81] L. Bao, F. Ye, C. Cai, J. Wu, K. Zeng, P. C. M. van Zijl, and Z. Chen, "Undersampled MR image reconstruction using an enhanced recursive residual network," *J. Magn. Reson.*, vol. 305, pp. 232–246, Aug. 2019.

[82] Y. Dai and P. Zhuang, "Compressed sensing MRI via a multi-scale dilated residual convolution network," *Magn. Reson. Imag.*, vol. 63, pp. 93–104, Nov. 2019.

[83] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.

[84] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[85] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[86] A. Kitchen and J. Seah, "Deep generative adversarial neural networks for realistic prostate lesion MRI synthesis," 2017, *arXiv:1708.00129*.

[87] L. Deecke, R. Vandermeulen, L. Ruff, S. Mandt, and M. Kloft, "Image anomaly detection with generative adversarial networks," in *Machine Learning and Knowledge Discovery in Databases*. Cham, Switzerland: Springer, 2019, pp. 3–17.

[88] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 240–244.

[89] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[90] C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, and B. A. Landman, "Learning implicit brain MRI manifolds with deep learning," *Proc. SPIE*, vol. 10574, pp. 408–414, Mar. 2018.

[91] C. Baur, S. Albarqouni, and N. Navab, "MelanoGANs: High resolution skin lesion synthesis with GANs," 2018, *arXiv:1804.04338*.

[92] D. Korkinof, T. Rijken, M. O'Neill, J. Yearsley, H. Harvey, and B. Glocker, "High-resolution mammogram synthesis using progressive generative adversarial networks," 2018, *arXiv:1807.03401*.

[93] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1358–1369, Jun. 2018.

[94] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[95] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 170–178.

[96] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[97] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[98] M. Maspero, M. H. F. Savenije, A. M. Dinkla, P. R. Seevinck, M. P. W. Intven, I. M. Jurgenliemk-Schulz, L. G. W. Kerkmeijer, and C. A. T. van den Berg, "Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy," *Phys. Med. Biol.*, vol. 63, no. 18, Sep. 2018, Art. no. 185001.

[99] H. Choi and D. S. Lee, "Generation of structural MR images from amyloid PET: Application to MR-less quantification," *J. Nucl. Med.*, vol. 59, no. 7, pp. 1111–1117, Jul. 2018.

[100] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, and Seevinck, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simulation Synth. Med. Imag.*, 2017, pp. 14–23.

[101] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," in *Simulation and Synthesis in Medical Imaging*. Cham, Switzerland: Springer, 2017, pp. 3–13.

[102] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using CycleGAN: Effects of gradient consistency loss and training data size," in *Proc. Int. Workshop Simulation Synth. Med. Imag.*, 2018, pp. 31–41.

[103] X. Cheng, L. Zhang, and Y. Zheng, "Deep similarity learning for multimodal medical images," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 6, no. 3, pp. 248–252, May 2018.

[104] S. Miao, Z. J. Wang, Y. Zheng, and R. Liao, "Real-time 2D/3D registration via CNN regression," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1430–1434.

[105] S. S. M. Salehi, S. Khan, D. Erdogmus, and A. Gholipour, "Real-time deep pose estimation with geodesic loss for image-to-template rigid registration," 2018, *arXiv:1803.05982*.

[106] J. Zheng, S. Miao, Z. J. Wang, and R. Liao, "Pairwise domain adaptation module for CNN-based 2-D/3-D registration," *J. Med. Imag.*, vol. 5, no. 2, 2018, Art. no. 021204.

[107] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—A deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, Sep. 2017.

[108] Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Label-driven weakly-supervised learning for multimodal deformarle image registration," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1070–1074.

[109] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, "Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking," *Bildverarbeitung für die Medizin 2019*, H. Handels, T. Deserno, A. Maier, K. Maier-Hein, C. Palm, and T. Tolxdorff, Eds. Wiesbaden, Germany: Springer-Vieweg, 2019, doi: 10.1007/978-3-658-25326-4_69.

[110] X. Cao, J. Yang, L. Wang, Z. Xue, Q. Wang, and D. Shen, "Deep learning based inter-modality image registration supervised by intra-modality similarity," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2018, pp. 55–63.

[111] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Med. Image Anal.*, vol. 54, pp. 193–206, May 2019.

[112] I. Yoo, D. G. C. Hildebrand, W. F. Tobin, W.-C. A. Lee, and W.-K. Jeong, ''ssEMnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features,'' in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2017, pp. 249–257.

[113] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, ''VoxelMorph: A learning framework for deformable medical image registration,'' *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[114] L. R. Dice, ''Measures of the amount of ecologic association between species,'' *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[115] D. Kuang and T. Schmah, ''FAIM—A ConvNet method for unsupervised 3D medical image registration,'' in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2019, pp. 646–654.

[116] J. Zhang, ''Inverse-consistent deep networks for unsupervised deformable image registration,'' 2018, *arXiv:1809.03443*.

[117] J. Tang, J. Huang, Y. Wang, Y. Chen, D. Zeng, X. Yang, and X. Chen, ''Utilizing unsupervised networks for end-to-end registration of brain MR images,'' *Neurocomputing*, vol. 408, no. 1, pp. 39–50, 2020.

[118] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, ''Adversarial image registration with application for MR and TRUS image fusion,'' in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2018, pp. 197–204.

[119] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, ''Deformable medical image registration using generative adversarial networks,'' in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1449–1453.

[120] C. Tanner, F. Ozdemir, R. Profanter, V. Vishnevsky, E. Konukoglu, and O. Goksel, ''Generative adversarial networks for MR-CT deformable image registration,'' 2018, *arXiv:1807.07349*.

[121] P. Moeskops, J. M. Wolterink, B. H. M. Van Der Velden, K. G. A. Gilhuijs, and Leiner, ''Deep learning for multi-task medical image segmentation in multiple modalities,'' in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2016, pp. 478–486.

[122] X. Li, Q. Dou, H. Chen, C.-W. Fu, X. Qi, D. L. Belavý, G. Armbrecht, D. Felsenberg, G. Zheng, and P.-A. Heng, ''3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images,'' *Med. Image Anal.*, vol. 45, pp. 41–54, Apr. 2018.

[123] G. Zeng and G. Zheng, ''Multi-stream 3D FCN with multi-scale deep supervision for multi-modality isointense infant brain MR image segmentation,'' in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 136–140.

[124] R. P. K. Poudel, P. Lamata, and G. Montana, ''Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation,'' in *Reconstruction, Segmentation, and Analysis of Medical Images* (Lecture Notes in Computer Science), vol. 10129, M. Zuluaga, K. Bhatia, B. Kainz, M. Moghari, and D. Pace, Eds. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-52280-7_8.

[125] S. A. Taghanaki, Y. Zheng, S. Kevin Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, ''Combo loss: Handling input and output imbalance in multi-organ segmentation,'' *Comput. Med. Imag. Graph.*, vol. 75, pp. 24–33, Jul. 2019.

[126] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, ''Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation,'' 2018, *arXiv:1802.06955*.

[127] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[128] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, ''Bi-directional ConvLSTM U-Net with Densley connected convolutions,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, p. 20.

[129] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, ''DeTrac: Transfer learning of class decomposed medical images in convolutional neural networks,'' *IEEE Access*, vol. 8, no. 99, pp. 74901–74913, 2020.

[130] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, and J. Torrents-Barrena, ''Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network,'' *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112855.

[131] P. Moeskops, J. M. Wolterink, B. H. M. Van Der Velden, K. G. A. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, ''Deep learning for multi-task medical image segmentation in multiple modalities,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 478–486.

[132] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[133] L. Bi, J. Kim, A. Kumar, M. Fulham, and D. Feng, ''Stacked fully convolutional networks with multi-channel learning: Application to medical image segmentation,'' *Vis. Comput.*, vol. 33, nos. 6–8, pp. 1061–1071, Jun. 2017.

[134] X.-Y. Zhou Celia Riga, S.-L. Lee, and G.-Z. Yang, ''Towards automatic 3D shape instantiation for deployed stent grafts: 2D multiple-class and class-imbalance marker segmentation with equally-weighted focal U-Net,'' 2017, *arXiv:1711.01506*.

[135] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Germany. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[136] V. Badrinarayanan, A. Kendall, and R. Cipolla, ''SegNet: A deep convolutional encoder–decoder architecture for image segmentation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[137] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[138] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking atrous convolution for semantic image segmentation,'' 2017, *arXiv:1706.05587*.

[139] Y. Xie, Z. Zhang, M. Sapkota, and L. Yang, ''Spatial clockwork recurrent neural network for muscle perimysium segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 185–193.

[140] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, ''3D-R2N2: A unified approach for single and multi-view 3D object reconstruction,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 628–644.

[141] C. Zou, E. Yumer, J. Yang, D. Ceylan, and D. Hoiem, ''3D-PRNN: Generating shape primitives with recurrent neural networks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 900–909.

[142] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, ''3D U-Net: Learning dense volumetric segmentation from sparse annotation,'' in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Athens, Greece. Cham, Switzerland: Springer, Oct. 2016, pp. 424–432.

[143] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[144] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, ''ENet: A deep neural network architecture for real-time semantic segmentation,'' 2016, *arXiv:1606.02147*.

[145] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, ''Full-resolution residual networks for semantic segmentation in street scenes,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3309–3318.

[146] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ''Pyramid scene parsing network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[147] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, ''Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, p. 20.

[148] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, ''SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation,'' *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018.

[149] C. Shen, J. Zhao, X. Zeng, H. Qi, J. Lu, Z. Xue, and H. Lu, ''Interpretable multimodal fusion neural networks for chronic obstructive pulmonary disease diagnosis,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2041–2053, 2019.

[150] F. Li, T. Li, H. Chen, J. Zhang, and L. Lu, ''Automatic pulmonary nodule detection and classification using transfer learning on LIDC-IDRI dataset,'' *J. X-Ray Sci. Technol.*, vol. 29, no. 3, pp. 455–467, 2021.

[151] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, ''Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification,'' 2018, *arXiv:1801.09927*.

[152] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, ''Multisource transfer learning with convolutional neural networks for lung pattern analysis,'' *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 76–84, Jan. 2017.

[153] G. Maicas, A. L. Martel, N. Diaz-Rodriguez, J. Carballido, and M. Gomez-Rio, ''LIDC and internal transfer learning for lung cancer detection,'' in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2021, pp. 221–230.

[154] C. Shen, J. Zhao, J. Lu, H. Qi, and Z. Xue, "Pulmonary nodule classification using knowledge transfer learning on LIDC-IDRI dataset," *Comput. Biol. Med.*, vol. 133, no. 1, 2021, Art. no. 104395.

[155] M. Alkhaleefah, S.-C. Ma, Y.-L. Chang, B. Huang, P. K. Chittem, and V. P. Achhannagari, "Double-shot transfer learning for breast cancer classification from X-ray images," *Appl. Sci.*, vol. 10, no. 11, p. 3999, Jun. 2020.

[156] I. Hirra, M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, A. M. Alghamdi, and A. S. Alfakeeh, "Breast cancer classification from histopathological images using patch-based deep learning modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021.

[157] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Comput. Med. Imag. Graph.*, vol. 71, no. 1, pp. 19–29, Jan. 2019.

[158] D. Sun, N. Zhang, Q.-J. Xu, R.-B. Huang, and L.-S. Zheng, "A novel 3D silver (I) supramolecular framework assembled from hybrid ligands incorporating Ag···C interactions," *J. Organometallic Chem.*, vol. 695, nos. 10–11, pp. 1598–1602, May 2010.

[159] S. S. Zareen, S. Guangmin, Y. Li, M. Kundi, S. Qadri, S. F. Qadri, M. Ahmad, and A. H. Khan, "A machine vision approach for classification of skin cancer using hybrid texture features," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Jul. 2022.

[160] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.

[161] T. Shanthi and R. S. Sabeenian, "Modified alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, Jun. 2019.

[162] Y. Chai, H. Liu, and J. Xu, "Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models," *Knowl.-Based Syst.*, vol. 161, pp. 147–156, Dec. 2018.

[163] M. Ahmad, D. Ai, G. Xie, S. F. Qadri, H. Song, Y. Huang, Y. Wang, and J. Yang, "Deep belief network modeling for automatic liver segmentation," *IEEE Access*, vol. 7, pp. 20585–20595, 2019.

[164] M. Ahmad, J. Yang, D. Ai, S. F. Qadri, and Y. Wang, "Deep-stacked auto encoder for liver segmentation," in *Proc. 12th Chin. Conf. Adv. Image Graph. Technol. (IGTA)*, Beijing, China. Cham, Switzerland: Springer, 2018, pp. 243–251.

[165] M. Ahmad, Y. Ding, S. F. Qadri, and J. Yang, "Convolutional-neural-network-based feature extraction for liver segmentation from CT images," in *Proc. 11th Int. Conf. Digit. Image Process. (ICDIP)*, Aug. 2019, pp. 829–835.

[166] M. Ahmad, S. F. Qadri, S. Qadri, I. A. Saeed, S. S. Zareen, Z. Iqbal, A. Alabrah, H. M. Alaghbari, and S. M. Mizanur Rahman, "A lightweight convolutional neural network model for liver segmentation in medical diagnosis," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, Mar. 2022.

[167] S. F. Qadri, L. Shen, M. Ahmad, S. Qadri, S. S. Zareen, and S. Khan, "OP-convNet: A patch classification-based framework for CT vertebrae segmentation," *IEEE Access*, vol. 9, pp. 158227–158240, 2021.

[168] S. F. Qadri, L. Shen, M. Ahmad, S. Qadri, S. S. Zareen, and M. A. Akbar, "SVseg: Stacked sparse autoencoder-based patch classification modeling for vertebrae segmentation," *Mathematics*, vol. 10, no. 5, p. 796, Mar. 2022.

[169] S. F. Qadri, H. Lin, L. Shen, M. Ahmad, S. Qadri, S. Khan, M. Khan, S. S. Zareen, M. A. Akbar, M. B. Bin Heyat, and S. Qamar, "CT-based automatic spine segmentation using patch-based deep learning," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–14, Mar. 2023.

[170] S. F. Qadri, D. Ai, G. Hu, M. Ahmad, Y. Huang, Y. Wang, and J. Yang, "Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images," *Appl. Sci.*, vol. 9, no. 1, p. 69, Dec. 2018.

[171] F.-F. Yang, Z.-Q. Tu, Y.-M. Fang, Y. Li, Y. Peng, T. Dong, C. Wang, S.-X. Lin, N.-Y. Zhan, Z.-M. Ma, Y.-Z. Feng, S.-Y. Tan, and X.-M. Lai, "Monitoring of peptide-specific and gamma interferon-productive T cells in patients with active and convalescent tuberculosis using an enzyme-linked immunosorbent spot assay," *Clin. Vaccine Immunol.*, vol. 19, no. 3, pp. 401–410, Mar. 2012.

[172] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[173] K. Bahrami, F. Shi, I. Rekik, and D. Shen, "Convolutional neural network for reconstruction of 7T-like images from 3T MRI using appearance and anatomical features," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 39–47.

[174] R. Santeramo, S. Withey, and G. Montana, "Longitudinal detection of radiological abnormalities with time-modulated LSTM," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2018, pp. 326–333.

[175] M. Alkhaleefah, P. K. Chittem, V. P. Achhannagari, S.-C. Ma, and Y.-L. Chang, "The influence of image augmentation on breast lesion classification using transfer learning," in *Proc. Int. Conf. Artif. Intell. Signal Process. (AISP)*, Jan. 2020, pp. 1–5.

[176] L. Li, M. Xu, X. Wang, L. Jiang, and H. Liu, "Attention based glaucoma detection: A large-scale database and CNN model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10563–10572.

[177] M. P. Vigouroux, V. Delaye, N. Bernier, R. Cipro, D. Lafond, G. Audoit, T. Baron, J. L. Rouvière, M. Martin, B. Chenevier, and F. Bertin, "Strain mapping at the nanoscale using precession electron diffraction in transmission electron microscope with off axis camera," *Appl. Phys. Lett.*, vol. 105, no. 19, pp. 271–286, Nov. 2014.

[178] I. Oksuz, B. Ruijsink, E. Puyol-Antón, J. R. Clough, G. Cruz, A. Bustin, C. Prieto, R. Botnar, D. Rueckert, J. A. Schnabel, and A. P. King, "Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning," *Med. Image Anal.*, vol. 55, pp. 136–147, Jul. 2019.

[179] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[180] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 214, p. 1062, Apr. 2016.

[181] S. F. Qadri, Z. Zhao, D. Ai, M. Ahmad, and Y. Wang, "Vertebrae segmentation via stacked sparse autoencoder from computed tomography images," in *Proc. 11th Int. Conf. Digit. Image Process. (ICDIP)*, vol. 11179, Aug. 2019, pp. 1206–1211.

[182] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016.

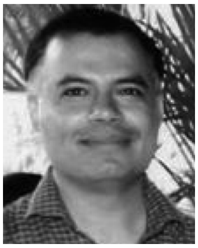
[183] M. Ahmad, S. F. Qadri, M. U. Ashraf, K. Subhi, S. Khan, S. S. Zareen, and S. Qadri, "Efficient liver segmentation from computed tomography images using deep learning," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, May 2022.

[184] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[185] S. F. Qadri, M. Ahmad, D. Ai, J. Yang, and Y. Wang, "Deep belief network based vertebra segmentation for CT images," in *Proc. 13th Conf. Image Graph. Technol. Appl. Image Graph. Technol. Appl. (IGTA)*, Beijing, China. Cham, Switzerland: Springer, Apr. 2018, pp. 536–545.

[186] J. Chen, X. Zhang, Y. Cheng, and N. Xi, "Surface EMG based continuous estimation of human lower limb joint angles by using deep belief networks," *Biomed. Signal Process. Control*, vol. 40, pp. 335–342, Feb. 2018.

TARIQ MAHMOOD received the master's degree in computer science from the University of Lahore, Pakistan, and the Ph.D. degree in software engineering from the Beijing University of Technology, China. He is currently an Assistant Professor/HOD with the Faculty of Information Sciences, University of Education, Vehari Campus, Vehari, Pakistan. He is a renowned expert in image processing, healthcare informatics and social media analysis, adhoc networks, and WSN. He has contributed more than 30 research papers in well-reputed international journals and conferences. His research interests include image processing, social media analysis, medical image diagnosis, machine learning, and data mining. He aims to contribute to interdisciplinary research of computer science and human-related disciplines. He is an Editorial Member and a Reviewer of various journals, including *PloS One*, *The Journal of Supercomputing*, *Journal of Digital Imaging*, and *International Journal of Sensors, Wireless Communications and Control.*



AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degrees (Hons.) from the Faculty of Computing, Universiti Teknologi Malaysia, with a specialization in forensic documents analysis and security, in 2010 and 2011, respectively. He is currently a Senior Researcher with the Artificial Intelligence and Data Analytics Laboratory, College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh, Saudi Arabia. He is the author of more than 200 ISI journal articles and conferences. He is a PI in several funded projects and also completed projects funded from MOHE Malaysia and Saudi Arabia. His research interests include data mining, health informatics, and pattern recognition. He received the Rector Award for the 2010 Best Student from Universiti Teknologi Malaysia.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently an Associate Chair with the Information Systems Department, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia, where she is also the Leader of the Artificial Intelligence and Data Analytics Research Laboratory. Her research interests include medical imaging, pattern recognition, data mining, MRI analysis, and soft computing. She is an active Professional Member of ACM, AIS, and IAENG organizations. She is the PSU Women in Data Science (WiDS) Ambassador at Stanford University and the Global Women Tech Conference. She was a recipient of the Best Student Award from the Faculty of Computing, UTM, in 2012.

LUBNA NADEEM received the Ph.D. degree in telecommunication engineering from the University of Engineering and Technology (UET) Taxila, Pakistan, in May 2022. From 2019 to 2022, she was a Research Associate with the Communication Systems Research Group, School of Electronic Engineering and Computer Science (EECS), Queen Marry University of London (QMUL), U.K. She has been a Lecturer with the Telecommunication Engineering Department, UET Taxila, since March 2014. She is currently associated with the Agile Creative Technologies for Smart Electromagnetic Novel Applications (ACTSENA) Research Group, UET Taxila. Her main research interests include application-based machine learning and deep learning, 5G wireless communications, radio resource allocation and optimization in cellular networks, the Internet of Things with machine learning, and deep learning.

SAEED ALI OMER BAHAJ received the Ph.D. degree from Pune University, India, in 2006. He is currently an Associate Professor with the Department of Management Information Systems, College of Business Administration, Al-Kharj. He is also an Associate Professor with the Computer Engineering Department, Hadhramout University, Yemen, and the MIS Department, COBA, Prince Sattam bin Abdulaziz University. His main research interests include artificial intelligence, information management, forecasting, information engineering, big data, and information security.