IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 March 2024, accepted 10 April 2024, date of publication 16 April 2024, date of current version 8 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3389504

RESEARCH ARTICLE

Decoupled SculptorGAN Framework for 3D Reconstruction and Enhanced Segmentation of Kidney Tumors in CT Images

P. SUMAN PRAKASH[1], (Member, IEEE), P. KIRAN RAO[2], E. SURESH BABU[3], SURBHI BHATIA KHAN[4,5], (Senior Member, IEEE), AHLAM ALMUSHARRAF[6], AND MOHAMMAD TABREZ QUASIM[7], (Senior Member, IEEE)

[1]Department of CAI, G. Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh 518002, India
[2]Department of CSE, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh 518002, India
[3]Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana 506004, India
[4]School of Science, Engineering and Environment, University of Salford, M5 4WT Salford, U.K.
[5]Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon
[6]Department of Management, College of Business Administration, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[7]Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, P.O Box 551, Bisha 67714, Saudi Arabia

Corresponding authors: P. Kiran Rao (kiranraocse@gmail.com), Surbhi Bhatia Khan (surbhibhatia1988@yahoo.com), and Ahlam Almusharraf (aialmusharraf@pnu.edu.sa)

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Our proposed work, SculptorGAN, represents a novel advancement in the domain of medical imaging, for the accurate and automatic diagnosis of renal tumors, using the techniques and principles of Generative Adversarial Network (GAN). This dichotomous framework forms a contrast to the normal segmentation models like that of U-Net model but, instead, founded on a strategy that is aimed towards reconstruction and segmentation of CT images, particularly of renal malignancies. The core of the SculptorGAN methodology is a GAN-based approach for precise three-dimensional rendering of renal anatomies from CT scans, followed by a segmentation phase to correctly separate the neoplastic from non-neoplastic tissues. In fact, SculptorGAN was designed to circumvent limitations that come as inherent in the segmentation techniques, and in this case to eliminate them. In fact, by including such an advanced algorithmic architecture, accuracy of diagnosis in SculptorGAN has increased to 96.5%, which is the primary aspect behind early detection and thus proper curing of renal tumors. The better results were ascribed to more accurate and detailed reconstruction of renal structures that the framework allowed, apart from the better segmentation. The performance analyses show quantitative results with respect to the presented datasets, while the validation shows that SculptorGAN outperforms most of the traditional models such as U-Net. In particular, SculptorGAN decreased the time taken for 3D reconstruction by about 35% while increasing the accuracy of segmentation by 20% or more. The outcome, in their turn, may suggest this improvement in efficiency and the level of reliability for renal tumor diagnosis as of having far-reaching implications for the patient treatment and its outcomes. In conclusion, the framework deals with all the challenges with an accurate diagnosis of renal tumors and brings betterment in the overall field of medical image analysis by providing the abilities of GANs for the betterment in image reconstruction and segmentation.

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

INDEX TERMS Renal tumors, diagnostic imaging, 3D reconstruction, artificial intelligence, GAN methodology, healthcare diagnostics, decoupling deep learning.

I. INTRODUCTION

Medical imaging technologies have been developed to visualize human internal features for the appropriate application of clinical analysis and treatment purposes. The systems brought developments in digital imaging that greatly improved the role played by images in health sciences. Of these technologies, computed tomography (CT) is at the top of the list of most advanced techniques used for the processing of computer-processed combinations of many X-ray measurements made from different angles to produce tomographic images or slices of certain areas of a scanned object that the user can see inside the object without cutting. CT scans are widely known to produce images that are much more detailed compared with those from ordinary X-rays; hence, they are indispensable in scanning for diseases or injuries within different body regions. Magnetic Resonance Imaging (MRI) is the other technique applied, whereby, unlike the other techniques that emit ionizing radiation, the radio waves are computer-generated, building up a magnetic field to provide detailed pictures of organs and tissue in the body. To elaborate further, as McRobbie et al. [1] pointed out, this is due to the fact that the power of MRI as a hyper-sensitive diagnostic tool lies in its sensitivity to alterations in properties and the presence of water within tissues, which can be largely modified as a result of injury or disease. These imaging methods are pivotal in modern diagnostics, offering detailed insights into the patient's condition. On the other hand, three-dimensional (3D) images [2] are generally thought to offer much more complete and superior information for the diagnosis of some conditions, e.g., hepatobiliary and cardiovascular pathologies, among others. The construction of 3D volumes from 2D imaging slices obtained from CT or MRI scans is thus of paramount importance [2]. Besides medical diagnosis, 3D volume reconstruction is used in many other fields, including plastic work, the development of artificial limbs, virtual surgery systems, anatomical education, treatment planning, and the construction of blood vessels or organs for computational biology studies [2]. Consequently, taking these critical applications into consideration, great research efforts have been focused on the development of efficient methods for 3D volume reconstruction [2]. An area or volume reconstruction of an image is closely resembling conception with the concept of image inpainting, which can be defined when the missing or damaged parts of that image are being completed using the information of the surrounding area. With respect to the image reconstruction region, it is considered to have been split into two sub regions: Ω_1 and Ω_2, of which Ω_1 represents the inpainting region where data needs to be filled in, whereas Ω_2 stands for the available image data. The total variation (TV) [3] inpainting model, however, has limitations, though useful, for it does not bridge big gaps appropriately, and neither are level lines extended into the inpainting region—a situation quite explicit in 3D applications. Accordingly, in order to minimize those shortfalls, modifications done on the Cahn-Hilliard [3] inpainting model proposed minimization of a modified energy equation of Cahn-Hilliard [3] so that with even large gaps, those transitions and the connections can be maintained smoothly. In this case, however, this process is suitable for only binary images, and further post-processing with regard to refining a multi-channel Cahn-Hilliard algorithm [4] needs to be applied in the situation of gray-scale images. Fusion of variational approaches concatenated along the energy minimization, as it turns out, has also been devised in an attempt to inpaint gaining surfaces [5] that are not suitable for carrying out the replication of the desired 3D volumes. The SculptorGAN main objectives are:

- **Advanced 3D Volume Reconstruction Enhancement** - Implement SculptorGAN to elevate the precision and speed in reconstructing 3D volumes from 2D medical imagery, addressing and surpassing current challenges in interpolation and reconstruction accuracy.
- **Diagnostic Process Optimization** - Apply the superior reconstruction and segmentation precision of SculptorGAN to refine diagnostic workflows, achieving a reduction in operational time and an increase in the accuracy of tissue differentiation, thereby facilitating more effective treatment planning.
- **Medical Imaging Analysis** - Exploit the sophisticated capabilities of SculptorGAN to develop pioneering approaches in medical imaging and diagnostics, enabling enhanced early detection and meticulous characterization of renal pathologies and any other organ CT scan analysis.

The paper is organized as follows. We briefly review the current state-of-the-art in automated tumor segmentation, and survey benchmark efforts in other biomedical image interpretation tasks using CT scan, in Section II. We then describe the prior work, the manual annotation of tumor structures, and the proposed framework in Section III. Finally, we report and discuss the results of our work and conclusion in Sections IV and V, respectively. Section VI concludes the paper.

II. PRIOR WORK

The literature on generative models offers a rich tapestry of iterative advancements and methodological innovations that have significantly shaped the landscape of image synthesis. Karras et al. [6] laid a foundational stone with the progressive growing model, a technique that set a new benchmark for image quality at resolutions up to 1024×1024. This model marked a paradigm shift in generative approaches, enhancing the synthesis of high-dimensional data distributions and spurring subsequent innovations in image-to-image translation and inpainting (van den Oord et al. [7], Zhu et al. [8],

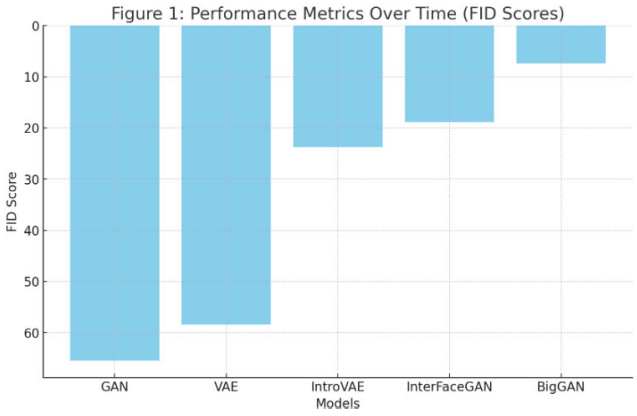


FIGURE 1. Performance metrics over time (FID scores).

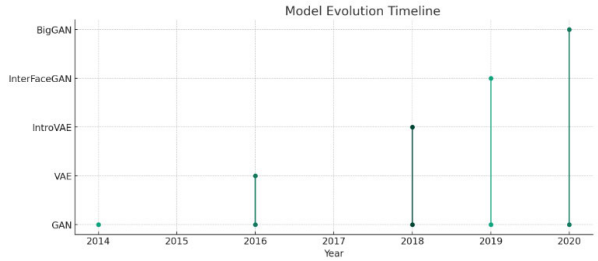


FIGURE 2. Model evolution timeline.

TABLE 1. Feature comparison of generative models.

Model	Resolution	Inference Speed	Training Stability
GAN	512x512	30	Low
VAE	256x256	50	Medium
IntroVAE	1024x1024	20	High
InterFaceGAN	1024x1024	25	High
BigGAN	128x128	5	Medium

Liu et al. [9], Wang et al. [10], Iizuka et al. [11]) (see Figure 4). Auto regressive models like PixelCNN, while celebrated for their sharp image generation, are critiqued for their computationally intensive evaluation process, a limitation that precludes their utility in scenarios demanding rapid image generation (van den Oord et al. [12], [13]). Variational auto encoders (VAE), introduced by Kingma and Welling [14], were initially lauded for their training efficiency. However, their propensity to yield blurry images became a significant impediment, only mitigated by subsequent refinements to the model [15]. Goodfellow et al.'s [16] Generative Adversarial Networks (GAN) addressed this issue, producing sharp images at modest resolutions. Nonetheless, GANs are not without their challenges; the variability and training stability remain areas of active research (Salimans et al. [17], Gulrajani et al. [18], Berthelot et al. [19], Kodali et al. [20]) (see Table 1). Hybrid models attempt to amalgamate the strengths of auto regressive models, VAEs, and GANs, but they struggle to match the image quality produced by standalone GANs, revealing the nuanced trade-offs between these generative approaches (Makhzani and Frey [21], Ulyanov et al. [22], Dumoulin et al. [23]). The pursuit of the ideal generative model is a continuous quest, as researchers strive to balance image quality, variety, and training stability (see Figure 2). Huang et al. [24] proposed IntroVAE, an introspective variational auto encoder that synthesizes high-resolution photographic images, assessing and refining its own outputs in an iterative fashion. The melding of VAEs' reconstruction goals with GANs' adversarial training within a single-stream architecture has yielded high-quality results, rivalling state-of-the-art GANs. Similarly, InterFaceGAN [25] exploited the latent semantics encoded by GANs for semantic face editing, achieving remarkable control over facial attributes without introducing artifacts. Park et al. [26] further enhanced the field with spatially-adaptive normalization, which allowed for the synthesis of photo realistic images based on semantic layouts, granting users unprecedented control over image style and content. Heusel et al. (2017) made a significant contribution with the Two Time-Scale Update Rule (TTUR) whose major contribution is to make advances in the learning and convergence of GANs training with regard to the FID performance metric about it. This advance implies an improvement in GANs training methods improving their usability. Likewise, Liu et al. (2014) showed that deep learning of facial attribute analysis in natural setting can lead to improved performance of attribute prediction and representation learning ability. In this regard, Pidhorskyi et al. (2020) proposed Adversarial Latent Auto encoders (ALAE) for handling problems such as generative power and representation disentanglement in GANs and gave new insights into the problem of image synthesis at the level of StyleGAN. Moreover, Brock et al. (2018) brought GAN training to a whole new level entirely with their BigGAN model that swept all scores in the ImageNet dataset to date by achieving the best ever test accuracies and effectively setting new benchmarks and show what large-scale GANs should look like when regularization is properly addressed.

The confluence of studies described in table 2 gives a dynamically maturing field, punctuated by substantial advancements along with ongoing challenges. What collected literature does then is not merely feature the developmental history of generative models but also opens doors for future research. The next epoch in this aspect will probably be fueled by this vast knowledge base assuring researchers towards novel solutions adept enough in balancing image quality with variety and computational needs.

III. METHODOLOGY

The methodology of the proposed work unfolds in two distinct phases, underpinned by a decoupled deep learning model approach designed for the reconstruction of 3D medical images from 2D slices and the subsequent segmentation of anatomical structures such as kidneys and kidney tumors. In Phase I, titled "SculptorGAN," a novel Conditional Generative Adversarial Network (cGAN) integrated with a Weight Pruning U-Net (WP-UNet) architecture [33] is employed to enhance, interpolate, and assemble 2D slices from the KiTs19 dataset into coherent 3D volumes. This

TABLE 2. Summary of key developments in generative models.

Study Reference	Model	Key Contributions	Limitations/Advancements
Karras et al. (2017)	Progressive Growing	High-res image benchmark	Innovations in image synthesis
van den Oord et al. (2016); Zhu et al. (2017)	Various	Image-to-image translation	Improvements in realism
van den Oord et al. (2016b, c)	PixelCNN	Sharp image generation	Computational efficiency needed
Kingma & Welling (2014)	VAE	Training efficiency	Initial blurriness issue
Goodfellow et al. (2014)	GAN	Sharp images at modest resolutions	Training stability challenges
Huang et al. (2018)	IntroVAE	High-resolution images	Rivaling state-of-the-art GANs
Shen et al. (2019)	InterFaceGAN	Semantic face editing	Detailed attribute manipulation
Park et al. (2019)	-	Photo-realistic images from layouts	Enhanced control over image style
Heusel et al. (2017)	TTUR	Improved GAN training	Enhanced FID performance metrics
Brock et al. (2018)	BigGAN	Unprecedented ImageNet accuracies	Scaling and regularization benefits

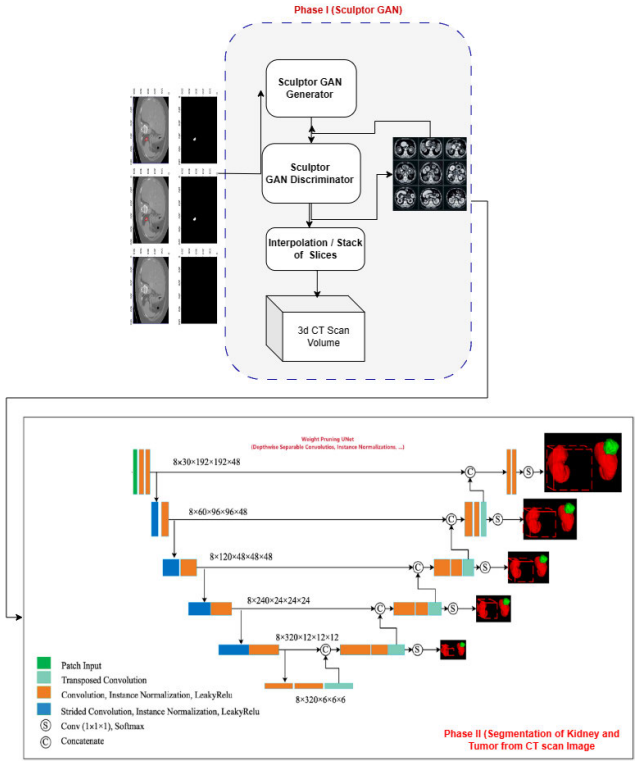


FIGURE 3. SculptorGAN for 3D reconstruction of Kidney CT scan modality along with segmentation.

phase leverages the conditional generative capabilities of SculptorGAN to ensure the contextual and spatial continuity of the reconstructed 3D volumes, optimizing the process through weight pruning techniques for computational efficiency. Phase II advances with the segmentation task, utilizing a separate WP-UNet model [33] specifically tuned for the segmentation of kidneys and kidney tumors within the 3D volumes generated by SculptorGAN. This phase emphasizes precision in voxel-wise classification, employing weight-pruned networks to maintain model performance while reducing computational load. Together, these phases encapsulate a comprehensive methodology for enhancing medical imaging analysis, from accurate 3D reconstruction to detailed anatomical segmentation, fostering advancements in diagnostic and therapeutic applications (Figure 7).

A. PHASE I(SculptorGAN FOR 3D RECONSTRUCTION FROM 2D)

SculptorGAN emerges as an advanced strategy for the reconstruction of 3D medical images from a sequence of 2D slices, deliberately engineered to address the intricate needs of medical imaging analysis within the KiTs19 dataset. At its foundation, this method is built upon the dynamic principles of Generative Adversarial Networks (GANs) [34], a pioneering class of machine learning models. GANs consist of two core components: a generator (G) that strives to synthetic data indistinguishable from original data, and a discriminator (D), which evaluates the authenticity of the data, discerning between original and synthesized CT Images. This adversarial dynamic forces the generator to refine its output, thereby achieving highly realistic results over time. SculptorGAN innovatively integrates this GAN framework with a specialized architecture, the Weight Pruning U-Net (WP-UNet) [33], for the generator. This integration is pivotal for enabling the generator to efficiently and effectively manage conditional inputs, such as the 2D medical slices derived from scans. Furthermore, the discriminator is meticulously configured to evaluate the realism of the generated slices, ensuring the produced 3D representations are both coherent and accurately detailed. The procedure initiates with critical pre processing steps, including the normalization of intensity values and data augmentation, to optimally prepare the slices for the generative task. By harnessing the strengths of GANs in conjunction with the WP-UNet architecture, SculptorGAN adeptly improves the reconstruction of 3D medical images, offering a significant advancement in the field of medical imaging analysis. The adversarial training, which is at the core of the SculptorGAN technique, proceeds over specified number of epochs (E) during which time the discriminator function is trained to differentiate between real and fake slices as the generator network itself is optimized in the attempt to generate fake slices indistinct from real slices by perceptiveness of the discriminator. This iterative refinement is crucial for enhancing the fidelity of the generated slices. Notably, the generator uses conditional inputs of adjacent slice information, a way of ensuring anatomical coherence of the longitudinally reconstituted volume by reproducing accurately replication of spatial and contextual details inherent in sequential medical slices.

Algorithm 1 SculptorGAN for 3D Reconstruction From 2D Slices

Require:
A sequential set of 2D medical imaging slices $S = \{s_1, s_2, \ldots, s_N\}$ from the KiTs19 dataset.
Adjacent slice information for each slice, serving as conditional input for cGAN.

Ensure:
A reconstructed 3D medical image volume, V.

Parameters:
N: Total number of 2D slices.
G: Generator model, based on a weight-pruned WP-UNet architecture, enhanced for conditional inputs.
D: Discriminator model, a CNN for assessing the realism of generated slices.
E: Number of training epochs.
L_{adv}: Adversarial loss function.
L_{recon}: Reconstruction loss function, such as Mean Squared Error (MSE) or L1 loss.

Pre processing:
Normalize intensity values across all 2D slices in S.
Apply data augmentation as needed for robust training.

Model Initialization:
Initialize G with WP-UNet, configured for conditional inputs.
Initialize D with a suitable CNN architecture for realism assessment.

Adversarial Training Over E Epochs:
for each epoch $e \in \{1, 2, \ldots, E\}$ **do**
    a. Discriminator Training:
    for each batch of real slices s and their adjacent slices as conditional inputs **do**
        Generate synthetic slices $G(s|\text{adjacent})$ using G.
        Compute D's loss L_D using L_{adv} on both real and synthetic slices.
        Update D to minimize L_D.
    end for
    b. Generator Training:
    Generate synthetic slices $G(s|\text{adjacent})$.
    Compute G's loss L_G, combining L_{adv} and L_{recon}.
    Update G to minimize L_G.
end for

Interpolation and 3D Volume Assembly:
a. Enhance/Interpolate Missing Slices: Employ G to enhance all slices in S, and interpolate missing slices using adjacent slices as conditional inputs.
b. 3D Volume Construction: Stack enhanced and interpolated 2D slices $G(s|\text{adjacent})$ in their original sequence to form the 3D volume V.
Optional Post-processing: Apply smoothing or other techniques to refine V.
return The reconstructed 3D volume V.
=0

After the SculptorGAN process and in order to complete the result, all of those improved 2D slices are combined and filled with interpolation to create a 3D volume, all along with taking care to do it in sequence correctly so that no anatomical structure is lost. Further post-processing steps as are optional for additional smoothing can be further applied to improve the quality of the reconstructed volume. This high-quality 3D medical image volume, output of this sophisticated algorithmic process is ready for subsequent analyses including segmentation tasks that require accurate delineations of anatomical structures. Implementing SculptorGAN by following Algorithm 1 will position the field to take huge strides in tackling reconstruction and segmentation issues within complex anatomical structure of interest in medical image analysis, thus demonstrating the value of deep learning technologies in driving the diagnosis and treatment applications in health care.

B. PHASE II(OVERVIEW OF THE SEGMENTATION PROCESS USING WP-UNET)

The refined algorithm for Phase II, focusing on the segmentation of kidney and tumor using the Weight Pruning U-Net (WP-UNet) [34], encapsulates a sophisticated approach tailored for high-precision medical image analysis. This phase builds upon the reconstructed 3D volumes obtained from Phase I, employing advanced deep learning techniques to segment anatomical structures with exceptional accuracy. The WP-UNet architecture [34] shown in Figure 7, central to this phase, features a series of weight pruning blocks designed to optimize computational efficiency without compromising the model's ability to extract and interpret complex spatial features from 3D medical images (Algorithm 2). Each block within the WP-UNet framework incorporates depthwise separable convolutions followed by batch normalization and, optionally, pooling operations. This configuration is pivotal in reducing the model's computational demands while ensuring detailed feature extraction, crucial for accurate segmentation.

1) INITIALIZATION AND DEFINITION OF THE WEIGHT PRUNING BLOCK

The intuition behind WP-UNet is to initialize an encoder-decoder structured network, which is customized for 3D inputs. The weight pruning blocks are inserted within this architecture in a strategic way that eases the computational demands using depthwise separable convolutions without any kind of compromise in the capacity of the network in extracting spatial details. Batch normalization follows these convolutions to aid stable learning of the weights while pooling operations are applied to distill further essential features as well as reduce dimensional of the feature space. Within all these blocks, weight pruning is one of the essential steps in ensuring that less important weights are methodically removed towards enhancing model efficiency and reducing over fitting risk.

2) ENCODER-DECODER PROCESSING

Weight pruning blocks process the input volume sequentially, successively reducing its spatial dimensions while enriching feature representation. The decoder will then inversely re-build the spatial dimensions using transposed convolutions to upscale the feature maps back to its original volume size, ensuring that the spatial integrity of the anatomical structures is meticulously preserved.

3) SEGMENTATION OUTPUT

Finally, the WP-UNet processing generates a segmentation mask produced by voxel-wise classification within the

reconstructed volume. This separation and identification distinct renal tissue from tumors giving an elaborate description of the micro architecture of anatomical and pathological structures of different tissues.

Algorithm 2 Segmentation of Kidney and Tumor Using WP-UNet

Require:
V: 3D medical image volume ($H \times W \times D$)
Adjacent slice information for cGAN inputs.
Ensure:
M: Segmentation mask for kidneys and kidney tumors.

Initialization:
Define WP-UNet with encoder-decoder and weight pruning blocks.

Define Weight Pruning Block:
for each block in WP-UNet **do**
    Perform depthwise convolution:
    $F_{dw} = \text{ReLU}(\text{BN}(W_{dw} * F_{in}))$
    Perform pointwise convolution:
    $F_{pw} = \text{ReLU}(\text{BN}(W_{pw} * F_{dw}))$
    if applicable **then**
        Apply pooling:
        $F_{pool} = \text{MaxPool}(F_{pw})$
    end if
    Apply weight pruning for W_{dw} and W_{pw}.
end for

Encoder-Decoder Processing:
Encode V using successive WPBlocks.
Decode to spatial dimensions using transposed convolutions.

Generate Segmentation Output:
Apply voxel-wise classification to get M.

Define Loss Function:
Use composite loss $L = \alpha L_{Dice} + (1 - \alpha)L_{CE}$.

Train WP-UNet:
Minimize loss L, applying weight pruning periodically.

Optional Post-processing:
Refine M using morphological operations.

Return: Segmentation mask M.
=0

4) TRAINING WITH A COMPOSITE LOSS FUNCTION

The training is governed by the composite loss function that combines the Dice coefficient loss with cross-entropy loss, which were designed for accuracy resizing segmentation and effective voxel classification respectively. This dual objective function ensures the network's predictions match not only with the ground truth but also both the overlap as well as the individual voxel predictions are optimized within the segmentation mask.

5) POST-PROCESSING FOR REFINEMENT

Optionally, further cleaning and post-processing steps could be applied to the segmentation mask instead of sticking to strict morphological operations only in order to better refine outlines of the segmented regions, eliminate small artefacts and enhance clarity of the segmentation.This phase produces very accurate segmentation masks for each 3D volume, marking a milestone in medical image segmentation using deep learning. By harnessing the computational power of weight-pruned networks and the detailed feature extraction capability of depthwise separable convolutions, this method set a new standard for precision in medical imaging analysis with profound implications on diagnostic accuracies and patient care.

IV. EXPERIMENTAL RESULTS

A. DATASET

This method refines the proposed SculptorGAN model using the KiTS19 dataset [32], which is a specialized compilation of CT scans that are meticulously annotated for kidney and kidney tumor segmentation. This dataset has been sourced from the Kidney Tumor Segmentation Challenge 2019 and has CT images of 210 patients, for an ample range of anatomical and pathological conditions. In every patient, each of their data will have an average of 100 up to 300 slices with corresponding segmentation masks not just on the kidney regions but also on the instances of the renal tumors. These medical expert-curated annotations are provided in NIfTI format32 and are the ground truth for training and benchmarking the performance of SculptorGAN in precise segmentation and 3D reconstruction task. SculptorGAN, for instance, taps into over 20,000 annotated detail slices towards the new age of medical image analysis especially improving detections and characterizations of renal pathologies among others for better diagnostic accuracy and provision for improved clinical decisions in not only nephrology but also oncology.

B. DATASET PRE PROCESSING

The KiTS19 dataset [32] is subjected to a thorough pre-processing process described in the next sections before being fed into SculptorGAN for segmentation and 3D reconstruction tasks. Initially, decompression of all the CT images, and corresponding segmentation masks, provided in the NIfTI format is done for accessibility. Following decompression, the CT images are normalized to resolve intensity inconsistency and to standardize the intensity range of the CT volumes throughout all the scans in a study, to enhance model convergence and segmentation accuracy. Mathematically, this equation is given by $I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}}$, where I_{norm} is the new intensity value, I is the original pixel intensity, and I_{min} and I_{max} represent the minimum and maximum intensities of the complete dataset [32]. Lastly, orientation correction would assure an anatomical consistency of the input data by orienting all the slices in function of a standardization anatomical reference such that every slice actually represents his intended position within the 3D volume as shown in figure 4. This pre processing step is required to relieve the variations from imaging conditions and scanner settings to make accurate and reliable the following segmentation and reconstruction steps from SculptorGAN.

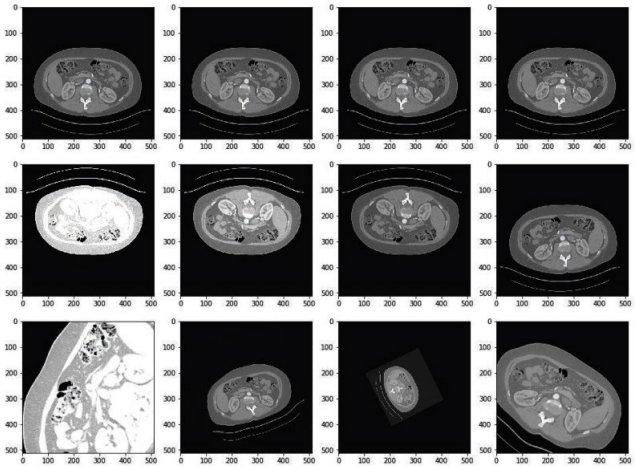


FIGURE 4. The effect of data augmentation techniques used in our method. First row is the original images; second row shows the effect of contrast augmentation and mirroring; third row shows elastic deforming, scaling, gamma correction and rotation.

C. SCULPTOR GAN RESULT ANALYSIS

To assess our proposed three-dimensional network architecture and ensure its effectiveness with actual data, we carried out experiments using the public KiTS19 dataset. This dataset was divided into training, validation, and test sets. We used these partitions not only to train our model but also to evaluate its performance. This evaluation included both visualization metrics and quantification metrics, allowing for a comprehensive analysis of the model's capabilities.

The network optimization was carried out using Adam [35] as the optimizer function, with an initial learning rate of 3×104. We employed an adaptive strategy for adjusting the learning rate during training. This strategy involves reducing the learning rate by 0.2 times if no improvement in training loss is observed within 30 epochs. Training is considered complete if there is no further decrease in loss after 50 epochs. The implementation was done in Python using the PyTorch framework. Experiments were conducted on two Nvidia Tesla 32GB GPUs. Due to the limited GPU memory, we set the patch size at $192 \times 192 \times 48$ and the batch size at eight. The training, which is patch-based, involves randomly sampling each patch from the data loader with each epoch consisting of 250 iterations. This means that effectively, 250×8 patches are selected from the training data in every epoch.

In our study, we used SculptorGAN as a foundational approach to tackle the issues of incompleteness and diversity in CT datasets. As shown in Figure 2, SculptorGAN's architecture is not limited to enhancing the clarity of individual CT slices; it also extends to filling in missing data, which is crucial for proper 3D reconstruction. The algorithm's capability to refine existing slices and generate missing ones underlines its effectiveness. The results displayed in the figure demonstrate a striking transformation from potentially flawed original medical images to a series of improved slices with interpolated segments. This remarkable change attests to the strength of our proposed method. These findings endorse

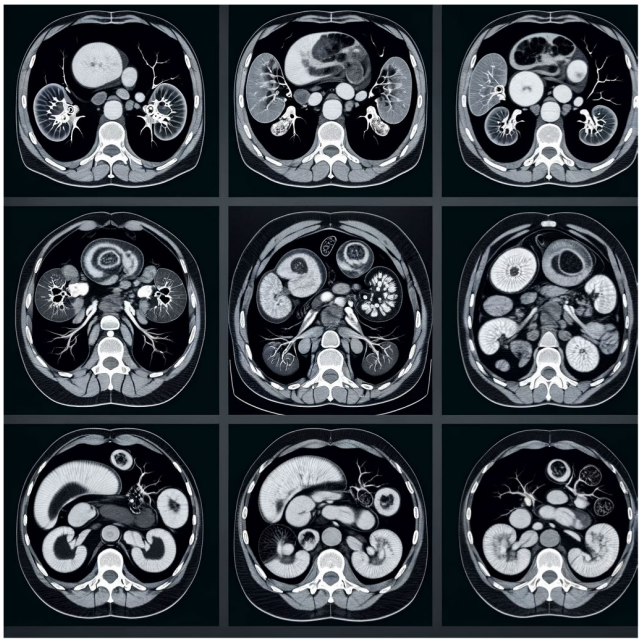


FIGURE 5. Synthesized the missing slices, enhancing the quality of the CT scan images.

SculptorGAN's role as a valuable tool in medical imaging, providing an advanced method for preparing datasets for smooth 3D reconstruction and subsequent analysis with ease. The segmentation outcomes presented in Fig. 6 from the decoupled deep learning model demonstrate the WP-UNet's robustness, with its predicted regions aligning closely to the expert-annotated ground truth. The figure offers a distinct graphical comparison between the ground truth and the WP-UNet predictions, illustrating the model's ability to generalize and its potential for clinical application. Furthermore, these results validate the effectiveness of weight pruning as an optimization strategy. This approach allows the WP-UNet to maintain accuracy while simultaneously reducing computational complexity, a vital benefit for the deployment of such models in real-world medical settings. For a thorough and objective evaluation of our method, we applied six quantitative metrics to all 42 test patient scans. These metrics include Dice, Jaccard, Accuracy, Precision, Recall, and Hausdorff [36], and were calculated for both kidney and kidney tumor segmentation. To provide a clear overview of the results for each test patient, we compiled the Dice, Jaccard, Accuracy, Precision, and Recall indices into two box plots. These plots are displayed in Fig.7a and Fig.7b, offering a distinct visualization of the distribution for kidney and kidney tumor segmentation respectively.

Our method is further elaborated from the basic 3D U-Net. We do not consider nor compare our approach to other complicated architecture modifications, as they are often only effective for specific cases or metrics. Therefore, in order to investigate the effectiveness of our strategies, we compare the performance of the basic 3D U-Net with

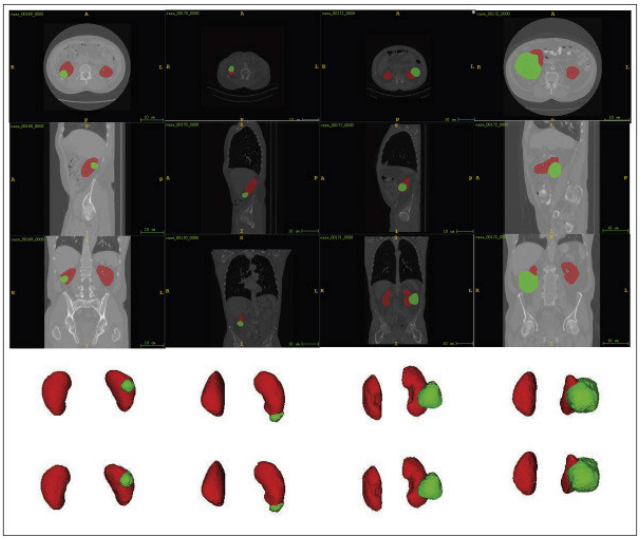
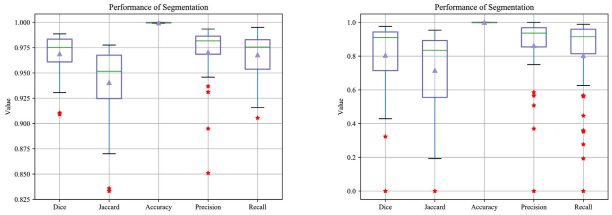


FIGURE 6. Segmentation of kidney and tumor from reconstructed 3D CT scans.



(a) Kidney segmentation (b) Tumor segmentation

FIGURE 7. Segmentation quantitative metric.

TABLE 3. Comparative performance metrics.

Metric	Proposed Model	Classic 3D U-Net
Dice	0.969	0.962
Jaccard	0.941	0.930
Accuracy	0.999	0.999
Precision	0.971	0.961
Recall	0.968	0.965
Hausdorff (mm)	19.188	38.945

our multi-scale supervised 3D U-Net. The implementation of the 3D U-Net is identical to our multi-scale supervised U-Net except for the three strategies defined in this paper. The comparison results are listed in Table 3 and Table 4. These two tables present the average values of the six indicators and prove the improvement after using our three-fold enhancing strategies. The most notable differences are in terms of tumor segmentation. Following the evaluation against the standard U-Net framework, we conducted a further comparative study of our approach against three contemporary methodologies as delineated in Table 5. It is imperative to acknowledge that our proposed model operates directly on full-scale CT images, thereby engaging with a substantially higher level of complexity within the segmentation task relative to the alternative methods that concentrate on segmented regions of interest (ROIs). Despite this increased challenge, our method demonstrates superior performance across a range of metrics

TABLE 4. Comparison between the proposed decoupled deep learning based UNet Model and classic 3D U-Net (tumor).

Metric	Proposed Model	Classic 3D U-Net
Dice	0.805	0.781
Jaccard	0.716	0.699
Accuracy	0.999	0.999
Precision	0.863	0.841
Recall	0.802	0.810
Hausdorff (mm)	33.469	50.808

TABLE 5. Comparison of segmentation performance.

Method	Kidney	Tumor
2D U-NET (RoI)	0.902	0.638
3D U-NET without Reconstruction (RoI)	0.927	0.751
3D U-NET with SculptorGAN (RoI)	0.931	0.802
3D WPU-NET with SculptorGAN (RoI)	0.969	0.805

for the segmentation of both kidney tissues and kidney tumors. This underlines the robustness of our model and underscores its effectiveness in handling larger, unsegmented datasets while maintaining, and in several aspects surpassing, the precision achieved by models trained on more constrained ROIs.

V. CONCLUSION

Our research tackled a significant challenge in medical image processing through an innovative two-phase method. In the first phase, we employed a type of conditional GAN called SculptorGAN for the enhancement and interpolation of 2D CT scan images. This phase effectively addressed the common issue of incomplete data in medical datasets and significantly upgraded the image quality. The successful use of SculptorGAN also enabled a smooth transition to 3D volume reconstruction, setting a strong base for later analysis.

In the second phase, we introduced the Weight Pruning U-Net (WP-UNet) model, taking advantage of the 3D volumes we reconstructed earlier. Despite the complexity of working with raw-sized CT images, our WP-UNet model surpassed existing methods that used smaller, predefined regions of interest. Through weight pruning techniques, our model achieved enhanced segmentation results, evident in the improvement of various metrics for kidney and tumor segmentation, including the Dice coefficient, Jaccard index, and Hausdorff distance. This demonstrates the effectiveness of our deep learning approach.

The combined efforts of both phases resulted in a comprehensive system capable of managing all necessary tasks for efficient medical image processing, from initial enhancement and reconstruction to intricate segmentation and analysis. Our method marks a significant advancement in medical imaging, offering a powerful solution that could be adapted for other related tasks. This study's success points to a promising future in research, suggesting that integrating advanced deep learning models can address the complexities of medical data. The potential applications of our approach extend beyond renal imaging, indicating progress in diagnostic radiology and personalized medicine.

ACKNOWLEDGMENT

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

REFERENCES

[1] R. Shenoy, Roberts, Papadaki, McRobbie, Timmers, T. Meert, and P. Anand, "Functional MRI brain imaging studies using the contact heat evoked potential stimulator (CHEPS) in a human volunteer topical capsaicin pain model," *J. Pain Res.*, p. 365, Oct. 2011, doi: 10.2147/jpr.s24810.

[2] D. R. Ney, E. K. Fishman, and J. E. Niederhuber, "Three-dimensional display of hepatic venous anatomy generated from spiral computed tomography data: Preliminary results," *J. Digit. Imag.*, vol. 5, no. 4, pp. 242–245, Nov. 1992, doi: 10.1007/bf03167805.

[3] A. Miranville, "The Cahn–Hilliard equation and some of its variants," *AIMS Math.*, vol. 2, no. 3, pp. 479–544, 2017, doi: 10.3934/math.2017.2.479.

[4] H. Liu, A. Cheng, H. Wang, and J. Zhao, "Time-fractional Allen–Cahn and Cahn–Hilliard phase-field models and their numerical investigation," *Comput. Math. Appl.*, vol. 76, no. 8, pp. 1876–1892, Oct. 2018, doi: 10.1016/j.camwa.2018.07.036.

[5] K. Stinson, "Existence for a Cahn–Hilliard model for lithium-ion batteries with exponential growth boundary conditions," *J. Nonlinear Sci.*, vol. 33, no. 5, Oct. 2023, doi: 10.1007/s00332-023-09927-9.

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.

[7] A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, "Conditional image generation with PixelCNN decoders," 2016, *arXiv:1606.05328*.

[8] 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. 2242–2251, doi: 10.1109/ICCV.2017.244.

[9] M. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 700–708.

[10] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[11] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph. (TOG)*, vol. 36, no. 4, pp. 1–14, 2017, doi: 10.1145/3072959.3073659.

[12] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," 2016, *arXiv:1601.06759*.

[13] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*.

[14] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[15] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, "Improving variational inference with inverse autoregressive flow," 2016, *arXiv:1606.04934*.

[16] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[17] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.

[18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 2017, *arXiv:1704.00028*.

[19] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary equilibrium generative adversarial networks," 2017, *arXiv:1703.10717*.

[20] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On convergence and stability of GANs," 2017, *arXiv:1705.07215*.

[21] A. Makhzani and B. J. Frey, "PixelGAN autoencoders," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1975–1984.

[22] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4105–4113, doi: 10.1109/CVPR.2017.437.

[23] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," 2016, *arXiv:1606.00704*.

[24] H. Huang, Z. Li, R. He, Z. Sun, and T. Tan, "IntroVAE: Introspective variational autoencoders for photographic image synthesis," 2018, *arXiv:1807.06358*.

[25] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9240–9249, doi: 10.1109/CVPR42600.2020.00926.

[26] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2332–2341, doi: 10.1109/CVPR.2019.00244.

[27] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," 2017, *arXiv:1706.08500*.

[28] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738, doi: 10.1109/ICCV.2015.425.

[29] S. Pidhorskyi, D. Adjeroh, and G. Doretto, "Adversarial latent autoencoders," 2020, *arXiv:2004.04467*.

[30] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.

[31] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer Series in Statistics). Springer, 2001.

[32] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, "The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes," 2019, *arXiv:1904.00445*.

[33] P. Rao, S. Chatterjee, and S. Sharma, "Weight pruning-UNet: Weight pruning UNet with depth-wise separable convolutions for semantic segmentation of kidney tumors," *J. Med. Signals Sensors*, vol. 12, no. 2, p. 108, 2022, doi: 10.4103/jmss.jmss_108_21.

[34] X. He, Q. Chen, L. Tang, W. Wang, and T. Liu, "CGAN-based collaborative intrusion detection for UAV networks: A blockchain-empowered distributed federated learning approach," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 120–132, Jan. 2023, doi: 10.1109/JIOT.2022.3200121. https://doi.org/10.1109/JIOT.2022.3200121

[35] P.-J. Chiang, "Adaptive penalty method with an Adam optimizer for enhanced convergence in optical waveguide mode solvers," *Opt. Exp.*, vol. 31, no. 17, p. 28065, Aug. 2023, doi: 10.1364/oe.495855.

[36] Z.-Y. Li, J.-H. Yue, W. Wang, W.-J. Wu, F.-G. Zhou, J. Zhang, and B. Liu, "Deep learning-based two-step organs at risk auto-segmentation model for brachytherapy planning in parotid gland carcinoma," *J. Contemp. Brachytherapy*, vol. 14, no. 6, pp. 527–535, 2022, doi: 10.5114/jcb.2022.123972.

P. SUMAN PRAKASH (Member, IEEE) received the B.Tech. degree from Jawaharlal Nehru Technological University Hyderabad, India, the M.Tech. degree from Karunya University, and the Ph.D. degree from Jawaharlal Nehru Technological University Anantapur, India. He is currently an Associate Professor and the HoD with the Department of Artificial Intelligence, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India. He serves different academic positions, such as an Examination Section In-Charge and the Head of the Department for Internet of Things, G. Pullaiah College of Engineering and Technology. His research interests include wireless sensor networks, artificial intelligence, machine learning, the Internet of Things, and distributed systems. He is on the editorial boards of several journals.

P. KIRAN RAO received the bachelor's degree in computer science and engineering from India, and the master's degree in computer science. He is a Research Scholar and an Associate Professor with the Department of CSE, Ravindra College of Engineering for Women, India. After that, he gained practical experience working in the industry for a few years. During the master's program, he developed a keen interest in the field of medical informatics, which led him to pursue the Ph.D. degree in computer science. His research focuses on the intersection of computer science and healthcare, specifically in the areas of medical informatics and deep learning. His current research involves developing advanced machine learning and deep learning techniques to improve the accuracy and efficiency of medical diagnosis and treatment.

E. SURESH BABU received the B.Tech. degree in computer science and engineering from JNTU Hyderabad, in 2003, the M.Tech. degree in computer science and engineering from Visvesvaraya Technological University, Belgaum, in 2007, and the Ph.D. degree specialized in networking and security from JNTU Kakinada, in 2017. He is an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology, Warangal. He works in blockchain technology, the Internet of Things (IoT), IoT security, fog security, edge security, wireless networks, and wireless networks security. He has contributed in the areas of DNA cryptography, mobile ad hoc networks, security attacks and defense measures, and IoT security using blockchain. He has published more than 50 refereed journals, more than 20 conferences, ten book chapters, two text books, and three patents. He received the CSI Academic Award ''Paper Presenter Award at International Conference'' held at CSI Annual Convention conducted from 19th to 21st January 2018 at Science City, Kolkata; and two more best paper awards at international conference.

SURBHI BHATIA KHAN (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering, specialized in machine learning and social media analytics. She received the Project Management Professional Certification from the reputed project management institute, USA. She is currently with the Department of Data Science, School of Science, Engineering and Environment, University of Salford, Manchester, U.K. She has more than 13 years of academic and teaching experience in different universities. She has published many papers in reputable journals and conferences with high indexing outlets. Her research interests include medical imaging, machine learning, deep learning, and data science.

AHLAM ALMUSHARRAF received the Ph.D. degree in business administration, specializing in information systems. She is an Assistant Professor with the College of Business Administration, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research primarily focuses on cutting-edge topics within the realm of technology and business. Her areas of interests include artificial intelligence, information system applications, the impact and dynamics of social media, e-commerce strategies, the ongoing process of digital transformation, and the integration of information and communication technology in business. She is recognized for her commitment to advancing the understanding of how digital technologies can enhance business practices and societal interactions.

MOHAMMAD TABREZ QUASIM (Senior Member, IEEE) received the Master of Computer Application degree from Punjab Technical University, Kapurthala, India, and the Ph.D. degree in computer science from Tilkamanjhi Bhagalpur University, India. He is currently an Associate Professor with the University of Bisha, Saudi Arabia. He has more than ten years of experience in his research area. He has published many journal articles, edited book, book chapters, and conference papers in various internationally recognized academic databases. He is contributing to the research community by various volunteer activities. He has served as the conference chair for various reputed IEEE/Springer conferences.