Sheffield Hallam University

A Modified Parallel NET (MPNET)-Based Deep Learning Technique for the Segmentation of Visceral and Superficial Adipose Tissues Quantification of CT Scans

ADEKANBI, Josteve, DAS, Debashish, ELMITWALLY, Nouh, ALI, Aliyuda, MADAI, Vince I, BHATIA, Bahadar, MORLESE, John, AFSAL, Muhammed and SHAH, Tanay

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/34830/

This document is the Published Version [VoR]

Citation:

ADEKANBI, Josteve, DAS, Debashish, ELMITWALLY, Nouh, ALI, Aliyuda, MADAI, Vince I, BHATIA, Bahadar, MORLESE, John, AFSAL, Muhammed and SHAH, Tanay (2025). A Modified Parallel NET (MPNET)-Based Deep Learning Technique for the Segmentation of Visceral and Superficial Adipose Tissues Quantification of CT Scans. IEEE Access. [Article]

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

Sheffield Hallam University Research Archive
http://shura.shu.ac.uk

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 December 2024, accepted 24 January 2025, date of publication 4 February 2025, date of current version 14 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3538626

APPLIED RESEARCH

A Modified Parallel NET (MPNET)-Based Deep Learning Technique for the Segmentation and Quantification of Visceral and Superficial Adipose Tissues of CT Scans

JOSTEVE ADEKANBI[1,2], DEBASHISH DAS[1], NOUH ELMITWALLY[1,3], ALIYUDA ALI[4], VINCE I. MADAI[1,5], BAHADAR BHATIA[3,6], JOHN MORLESE[2], MUHAMMED AFSAL[2], AND TANAY SHAH[2]

[1]Faculty of Computing, Engineering and the Built Environment, Birmingham City University, B4 7XG Birmingham, U.K.
[2]Sandwell and West Birmingham Hospitals NHS Trust, B71 4HJ West Bromwich, U.K.
[3]Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt
[4]Department of Computing, College of Business, Technology and Engineering, Sheffield Hallam University, S1 1WB Sheffield, U.K.
[5]QUEST Center for Responsible Research, Berlin Institute of Health (BIH) at Charité, Universitätsmedizin Berlin at Charité, 10117 Berlin, Germany
[6]Space Research Centre, Physics and Astronomy, University of Leicester, LE1 7RH Leicester, U.K.

Corresponding author: Debashish Das (Debashish.Das@bcu.ac.uk)

ABSTRACT This study introduces a modified parallel net (MPNET), a novel deep learning model designed for accurate segmentation and quantification of visceral and superficial adipose tissues. This was used to quantify the visceral and superficial adipose tissues found at the L3 levels of vertebra in CT scans. This will be used to predict the likelihood of the patient developing diabetes or cardiovascular diseases from existing CT scan data. MPNET was compared with state-of-the-art models like UNET, R2UNET, UNET++, and nnUNET. This approach advances the accuracy and efficiency of image segmentation demonstrating a faster learning curve and lower losses at early epochs than traditional models., We developed and validated using a limited dataset of 14 single-slice DICOM files for each patient extracted from the National Health Service UK. The outputs from MPNET not only matched but often exceeded traditional metrics such as the Dice coefficient and IoU in nuanced anatomical delineation, providing greater clinical realism and applicability in segmentation results. As a pilot study, this research paves the way for a forthcoming validation study on a larger and more ethnically diverse dataset.

INDEX TERMS Convolutional neural networks, CT scan analysis, deep learning, image post-processing, segmentation and quantification, visceral and superficial adipose tissues.

I. INTRODUCTION

Cardiovascular (CVD) and cerebrovascular diseases (CBD), such as heart attacks and strokes, are major causes of preventable deaths worldwide, particularly in the UK where CVD remains a leading cause of mortality. A key factor in these conditions is the metabolic syndrome, characterized by a triad of hypertension, type 2 diabetes, and obesity. Early detection and management of metabolic syndrome are crucial, as lifestyle changes can significantly mitigate associated risks.

Traditional metrics like the Body Mass Index (BMI) have been used to assess obesity, but they fall short in accurately predicting cardiovascular risks. More precise indicators such as the total abdominal adipose tissue (TAAT) offer a better assessment. TAAT, particularly the fat surrounding

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

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

intra-abdominal organs, can be quantified using CT scans originally obtained for other clinical reasons, providing a cost-effective method to enhance patient-specific risk assessments without incurring additional NHS costs.

Convolutional Neural Networks (CNNs), particularly, have revolutionized medical imaging, augmenting diagnostic accuracy and introducing automation in complex treatment protocols [1]. The integration of CNNs with advanced architectures like TransU-Net has further expanded their capabilities, facilitating the correlation of pixel dependencies within medical images [2].

Current methods, predominantly manual, are time intensive and subject to inter-operator variability, potentially limiting diagnostic reliability [1], [3].

In this study, we investigated the potential of a modified parallel net (MPNET) a novel CNN architecture designed for accurate segmentation and quantification of visceral and superficial adipose tissues. The NHS has large volumes of abdominal CT scan data of which has been acquired for potentially benign reasons. We decided to use this dataset to accurately segment these adipose tissues in order to develop a model to predict the likelihood of the patient developing diabetes or cardiovascular diseases in the future.

A comparative analysis with existing CNN models like U-Net is presented to explore the strengths and limitations of these approaches in the context of abdominal CT scans. This research seeks to fill gaps in the current understanding of CNN architectures, particularly concerning pixel dependency correlation and the integration of transformer architectures [4].

It is applied to abdominal CT scans to nuanced anatomical variations.

II. RELATED WORKS

The field of medical image segmentation has witnessed transformative progress over the past years, significantly influenced by advancements in deep learning architectures. The U-Net architecture has been a cornerstone in this evolution, providing a robust framework for various segmentation tasks. Its significance is further underscored by subsequent innovations such as the Attention U-Net (AttnUNet), which incorporates an attention mechanism to improve the model's focus on salient features within the image for enhanced segmentation accuracy [5]. Despite their advancements, models like U-Net and AttnUNet often struggle to delineate complex anatomical structures.

Several potential models U-Net, U-Net++, Attention U-Net, nnU-Net, R2U-Net, MPNET which were explored are summarised in Table 1. The advent of hybrid models like TransUNet [2] and Swin-Unet [6] represents a significant leap forward, merging the spatial understanding of Convolutional Neural Networks (CNNs) with the long-range dependency modeling capabilities of Transformers. These models have demonstrated remarkable success in capturing global image contexts, thereby improving encoder strength for medical image segmentation tasks. However, while TransUNet and Swin-Unet excel in global image context interpretation, their segmentation precision is compromised by their standard configurations, which do not optimally balance global and local feature extractions necessary for distinguishing closely adjacent tissue types.

The nnU-Net framework, with its self-configuring capabilities, has set new benchmarks in the automation of network architecture and preprocessing strategy selection for medical image segmentation [7]. This approach significantly reduces the manual effort involved in model configuration, catering to the diverse needs of different segmentation tasks. However, the nnU-Net, while reducing manual configuration through its self-adapting capabilities, does not offer specific optimizations for the L3 vertebral level where a unique balance between precision in local detail extraction and global contextual understanding is required.

Emerging approaches like the token-based MLP-Mixer for abdominal CT multi-organ segmentation [8] and the nested architecture of U-Net++ [7] have introduced novel paradigms in network design and segmentation strategy. Emerging models like the MLP-Mixer and U-Net++ introduce innovative structures which improve general segmentation performance. However, they still fall short when extremely accurate anatomical delineation is necessary to accurately differentiate between visceral and superficial adipose tissues, as their generic training lacks focus on specific anatomical idiosyncrasies.

The dual-encoder networks that synergize Transformer and CNN architectures [9] and the P-UNet with its parallel attention mechanisms [10] signify the ongoing efforts to refine feature detection and segmentation accuracy. These models aim to blend the strengths of local and global feature extraction capabilities, highlighting the community's dedication to overcoming the limitations of existing segmentation models. Yet, the complex challenge of adipose tissue quantification at certain anatomical levels demands a more tailored approach that can navigate the intricate landscape of medical image segmentation with a higher degree of precision and clinical relevance.

A Modified Parallel NET (MPNET)-Based Deep Learning technique, developed within this study, directly addresses these gaps is specifically designed to improve the segmentation and quantification of visceral and superficial fat at the abdominal level, offering targeted solutions to the challenges previously unmet by existing models.

The MPNET model leverages a novel CNN-based architecture that is finely tuned to the unique challenges of this task. Utilizing a comprehensive dataset collected from the National Health Service (NHS), UK, this model not only outperforms existing state-of-the-art models such as UNET, R2UNET, UNET++, and nnUNET in capturing the complex anatomical structures but also enhances the clinical realism of segmentation results.

The incorporation of a unique post-processing algorithm further refines these outcomes, ensuring that the segmented images align more closely with clinical requirements. Consequently, MPNET demonstrates not just competitive performance in traditional metrics like the Dice coefficient and IoU but also excels in producing clinically relevant segmentation results that surpass those of existing models.

Our methodology is described in section III with the experimental work in section IV, discussion in section V and conclusions in section VI.

III. METHODS

A. SUMMARY OF THE APPROACH

This study analyzed CT scans for detecting and quantifying visceral and superficial adipose tissue using advanced deep learning. The process begins with data collection and segmentation, followed by pre-processing to convert images to greyscale. An exploratory data analysis (EDA) was conducted using techniques like pixel intensity distribution and Fourier transform analysis to inform model training.

The data was normalized and encoded before being divided into training, testing, and validation sets in the proportions of 80%, 10%, and 10%, respectively.The training involved two phases with adjustments in epochs, learning rates, and batch sizes, using U-Net variants and MPNET. After model training, a post-processing phase refined the outputs for adipose tissue quantification. This study adhered to rigorous ethical protocols, including approval by the appropriate institutional review boards. All CT scan data was pseudonymised complying with the General Data Protection Regulation (GDPR) to ensure the privacy and security of personal data.

The process of building and experimenting with the models of our work follows a systematic and methodological approach as illustrated in Figure 1.

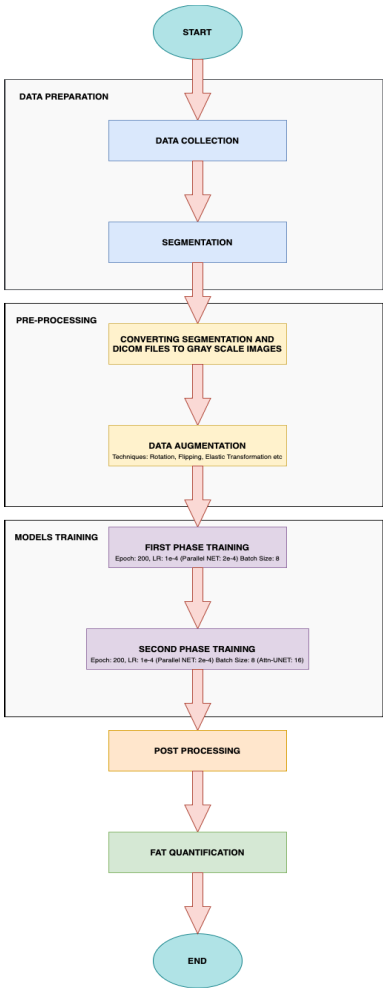


FIGURE 1. Project pipeline illustrating the steps from data collection and segmentation to fat quantification.

B. DATA PREPARATION

1) DATASET

This study utilised fourteen (14) single-slice DICOM files from the Sandwell and West Birmingham Hospitals NHS Trust, taken at the L3 vertebra level, with corresponding RGB segmentation images for adipose tissue differentiation. Each DICOM file metadata had values for slice thickness and pixel scaling later used for quantification. The DICOM files were converted to greyscale to delineate superficial (SAT) and visceral (VAT) adipose tissues for deep learning model analysis.

2) SEGMENTATION

In the segmentation refinement, reference maps were crucial for accurately identifying visceral adipose tissue (VAT) and superficial adipose tissue (SAT) in DICOM files. Using ITK-SNAP, we delineated VAT and SAT from DICOM files, producing corresponding NRRD output segmentation files. Figure 2 illustrates this process, with SAT marked in red and VAT in green, showcasing the detail captured in a DICOM slice.

C. PRE-PROCESSING

1) TRANSFORMATION OF NRRD AND DICOM FILES INTO GREYSCALE IMAGES

To integrate into the deep learning workflow, both NRRD segmentation and DICOM CT scan files were converted into 512×512 pixel greyscale images. Custom scripts, using the `nibabel` library, processed NRRD files, coding superficial adipose tissue (SAT) as white (255), visceral adipose tissue (VAT) as grey (127), and the background as black (0). This color scheme simplified segmentation maps for one-hot encoding in model training.

Similarly, DICOM images were normalized to the same resolution, standardizing pixel intensities, using a script with the `pydicom` library. This ensured uniform image dimensions and tissue representation, critical for accurate comparison and overlay in subsequent analysis.

TABLE 1. Gap analysis summary.

Author & Year	Dataset	Algorithm/Model	Pre-processing Techniques	Evaluation Measure	Future Works
Tajbakhsh et al., 2022	Various	CNNs	Diverse	Accuracy, IoU	Explore CNNs in new imaging modalities
Chen et al., 2021	-	TransUNet	-	-	Integration with transformers
Ronneberger et al., 2015	Biomedical	U-Net	Standard scaling	Dice Coefficient	Improvements in U-Net architecture
Cao et al., 2021	-	Swin-Unet	-	-	Pure transformer-based models
Weisman et al., 2023	CT Scans	CNNs	Impact of settings	Accuracy	Training variability
Kim et al., 2020	Abdominal CT	3D CNN	Patch-based	Accuracy	Deep learning for 3D segmentation
Pan et al., 2022	Abdominal CT	MLP-Mixer	Token-based	Accuracy	MLP in medical imaging
Hong et al., 2023	-	Transformer-CNN	Dual encoding	Accuracy	Transformer-CNN synergy
Zhang et al., 2022	-	P-UNet	Attention-based	Accuracy	Attention mechanisms in CNNs
Oktay et al., 2018	Pancreatic	Attention U-Net	Focus areas	Dice Coefficient	Attention in segmentation
Isensee et al., 2018	Medical Images	nnU-Net	Self-adapting	Dice Coefficient	Automated framework adaptations
Zhou et al., 2018	Medical Images	UNet++	Nested architecture	Dice Coefficient	Nested U-Net advancements

2) DATA AUGMENTATION

Data augmentation was key to increasing the dataset from 14 to 1032 examples, simulating a range of anatomical variations and imaging conditions found in clinical settings (validated by radiologists). Techniques included scaling and translation (adjusting size and position of structures), shearing and rotation (altering shape and orientation), and flipping (mimicking different patient positions). Coarse dropout introduced missing data areas, training the model on incomplete patterns.

Perspective changes, poisson noise and blur effects simulated depth and focus variations, while elastic transformations added local geometric distortions, akin to patient movement. Contrast and brightness adjustments accounted for variations in CT scan quality.

These augmentation methods significantly expanded our dataset, enhancing the deep learning models' ability to recognize various adipose tissue deposits and adapt to diverse imaging conditions, thereby improving their generalisability and robustness.

D. MODELS

1) UNET

The U-Net architecture is designed specifically for medical image segmentation. It has a symmetric structure with an encoder path to capture context and a corresponding decoder path that enables accurate localization. The model uses a series of convolutional and max pooling layers to downsample the image, followed by upsampling layers to construct the segmentation output. Skip connections between downsampling and upsampling layers help in preserving the spatial context which is often lost during pooling operations [4].

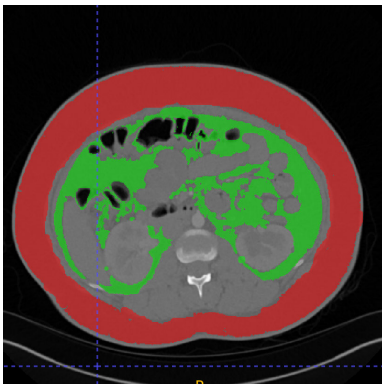


FIGURE 2. Example Segmentation overlay on DICOM image of a single CT slice through the abdomen (ITK-SNAP). Green: Visceral adipose tissue, Red: Superficial adipose tissue. Other: Background.

2) UNET++

U-Net++ improves upon the original U-Net by introducing nested and dense skip pathways. These pathways are designed to bridge the semantic gap between the feature maps of the encoder and decoder sub-networks. This is achieved by connecting the corresponding layers in the encoder and decoder with a series of convolutional operations, which helps in improving the flow of information and gradients throughout the network, leading to enhanced feature representation and segmentation results [7].

3) ATTENTION U-Net (AttnUNet)

Attention U-Net adds an attention mechanism to the standard U-Net architecture. This mechanism helps the model to focus on target structures of varying shapes and sizes. During the upsampling phase, the attention gates selectively highlight salient features that are passed through the skip connections.

This allows the model to suppress irrelevant regions in the input while highlighting features that are useful for a specific task [5].

4) nnU-Net

nnU-Net is a self-adapting framework that automatically configures itself to the dataset at hand. It adapts its preprocessing, network architecture, training, and inference strategies to the specific requirements of the dataset. This adaptability is achieved through an extensive analysis of dataset properties, enabling nnU-Net to achieve robust performance across different medical segmentation challenges without any manual intervention in designing the network architecture [21].

5) R2U-Net

R2U-Net integrates recurrent convolutional layers within the U-Net architecture. These recurrent layers can capture fine-grained details over successive iterations, making it more adept at handling complex structures within medical images. The recurrent layers are combined with residual connections, which improve gradient flow and enable the training of deeper networks by mitigating the vanishing gradient problem [22].

6) MPNET

The architecture of the MPNET for segmentation was ingeniously designed to exploit the benefits of parallel processing in convolutional neural networks. It stands out for its unique approach to feature extraction and merging, setting it apart from traditional architectures like the U-Net, which follows a more linear progression of convolutions and pooling.

At its core, the MPNET employs multiple convolutional paths at each depth level, each with different kernel sizes. This design choice enables the network to capture a diverse range of features from the input data. For instance, at the first depth level, two parallel paths are employed: one with a 3×3 kernel and another with a 5×5 kernel. These paths are then merged, harnessing the strengths of both small and large kernels in capturing fine and coarse features, respectively.

The model intensifies this parallel processing approach as it deepens, consistently using dual-path convolutions at each depth level. The use of the PReLU (Parametric Rectified Linear Unit) activation function after each convolution enhances the model's capacity to learn non-linear transformations more effectively than standard ReLUs.

A distinctive feature of this architecture is the increased dropout rate (0.65) at the deepest layer. This decision is motivated by the need to prevent overfitting, especially given the model's increased complexity due to the parallel paths.

During the upsampling phase, the MPNET introduces skip connections, a technique reminiscent of the U-Net. These connections help recover spatial information lost during downsampling. However, unlike the U-Net, the MPNET merges the upsampled output with feature maps from both paths of the corresponding depth level, further enriching the feature representation.

In the final stage, a convolutional layer with a softmax activation function generates the segmentation map. This layer is critical for classifying each pixel into one of the predefined categories, which is essential in applications like medical image segmentation.

Mathematical Formulation: The operations within MPNET can be represented mathematically as follows:

Let I be the input image, P_{A_k} and P_{B_k} be the k-th parallel paths with kernel sizes of 3×3 and 5×5 respectively, M_k be the merged output, and U_k be the upsampled feature map at each depth level k.

For each parallel block at depth k:

$$P_{A_k} = PReLU(Conv2D(I, kernel\_size = (3, 3))) \quad (1)$$

$$P_{B_k} = PReLU(Conv2D(I, kernel\_size = (5, 5))) \quad (2)$$

$$M_k = Concatenate([P_{A_k}, P_{B_k}]) \quad (3)$$

For upsampling and merging with previous layers:

$$U_k = UpSampling2D(M_{k+1}) \quad (4)$$

$$S_k = Concatenate([U_k, M_k]) \quad (5)$$

The final segmentation map Y is obtained through a 1×1 convolution on the last merged feature map M:

$$Y = Softmax(Conv2D(M_{final}, kernel\_size = (1, 1))) \quad (6)$$

E. METRICS

1) DICE COEFFICIENT

The Dice Coefficient, also known as the Sørensen–Dice index, measures the similarity between two samples and is widely used in image segmentation to quantify the performance of models. Although it's favoured for its straightforward calculation and interpretation, the coefficient has limitations, particularly in handling imbalanced data and small object segmentation [23]. These issues can lead to misleading high scores despite poor model performance on minor yet important details. Recognizing these limitations, we employed the Dice Coefficient alongside other metrics, allowing us to mitigate its biases and provide a more balanced evaluation of our models' segmentation capabilities. Mathematically, for a smoothing factor smooth to prevent division by zero, it is given by equation 7.

$$Dice = \frac{2 \cdot |Y_{true} \cap Y_{pred}| + \text{smooth}}{|Y_{true}| + |Y_{pred}| + \text{smooth}} \quad (7)$$

where Y_{true} represents the ground truth mask and Y_{pred} the predicted mask.

2) DICE LOSS

Dice Loss is a metric derived from the Dice Coefficient, used as a loss function for training segmentation models.

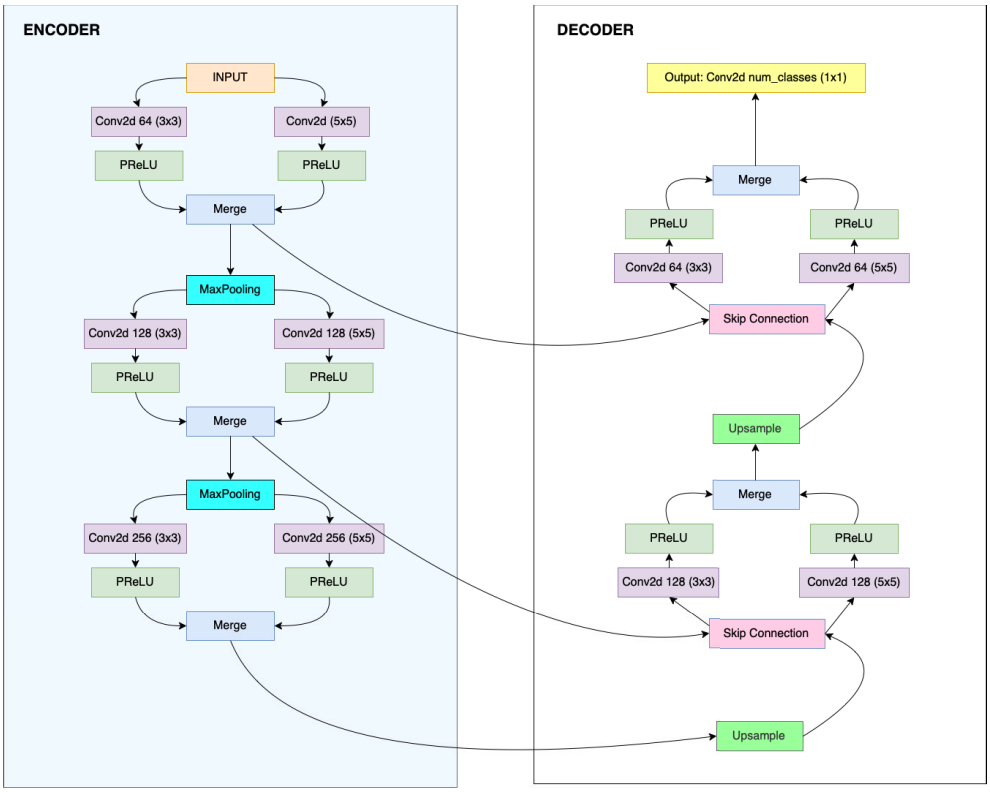


FIGURE 3. MPNET architecture.

It is particularly useful when dealing with imbalanced datasets [24]. The Dice Loss formula is stated in equation 8.

$$DiceLoss = 1 - DiceCoefficient \tag{8}$$

The closer the Dice Coefficient is to 1, the lower the Dice Loss and the better the model's segmentation performance.

3) INTERSECTION OVER UNION (IoU)

Intersection Over Union (IoU) is another common metric for evaluating the performance of image segmentation models. It measures the overlap between the predicted segmentation and the ground truth, normalized by the union of both [25]. The IoU formulation is state in equation 9.

$$IoU = \frac{|Y_{true} \cap Y_{pred}| + \text{smooth}}{|Y_{true} \cup Y_{pred}| + \text{smooth}} \tag{9}$$

IoU is an essential metric for assessing how well the model predictions align with the actual data, with a higher IoU indicating better performance.

4) COMBINED DICE AND CROSS-ENTROPY LOSS

To harness the strengths of both Dice Loss and Categorical Cross-Entropy, a combined loss function is employed. The Categorical Cross-Entropy is sensitive to the imbalanced class distribution but does not explicitly consider the overlap between the segmentation masks, which the Dice Loss compensates for [26]. The combined loss is formulated as equation 10.

$$Loss = \alpha \cdot DiceLoss + (1 - \alpha) \cdot CategoricalCrossEntropy \tag{10}$$

where α is a weighting coefficient dictating the balance between the two loss components.

Following the recommendations by Taha and Hanbury, who advocate for a tailored approach in selecting segmentation evaluation metrics, we employed the Dice Coefficient and IoU for their complementary perspectives on model accuracy in overlap and union of segmented areas. Additionally, the combined use of Dice and Cross-Entropy in our loss function aligns with their guidance on addressing class imbalance and segmentation precision, which are critical in the medical imaging context.

F. POST-PROCESSING

In the post-processing stage, this study introduced an algorithm designed to refine the segmentation results, specifically targeting the accurate delineation between visceral and superficial adipose tissues. The algorithm operates by discerning and modifying only the white pixels that represent superficial adipose tissue within the interior regions, ensuring the preservation of the outer white boundary and pre-existing grey areas indicative of visceral adipose tissue.

The process begins by establishing a binary mask over the white pixels (superficial adipose tissue) in the greyscale

image. Contours are identified within this mask using a contour detection method. The contours that are not part of the outer boundary are then filled, transforming the interior white pixels to grey to represent visceral adipose tissue, while keeping the border intact.

Mathematically, this can be represented as follows:

Let I be the greyscale image, M be the binary mask, and C be the set of contours detected. For each contour $c \in C$ with a parent contour (indicating it is not the boundary), the algorithm performs the following operation:

$$I(x,y) = \begin{cases} 127 & \text{if } M(x,y) = 1 \text{ and } I(x,y) \neq 0 \\ I(x,y) & \text{otherwise} \end{cases} \tag{11}$$

where $I(x,y)$ is the pixel intensity at position (x,y), 127 represents the greyscale value for visceral adipose tissue, and 0 denotes the background.

This refined image I is then utilised in the subsequent quantification process. The algorithm ensures that the segmentation output reflects a more accurate representation of the different adipose tissue regions, which is critical for accurate adipose tissues quantification. The use of this method significantly enhances the quality of the segmentation, contributing to the overall reliability of the study's findings.

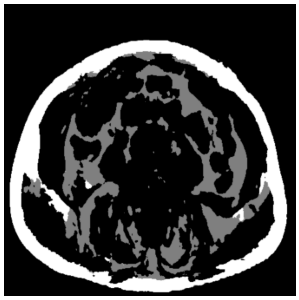


FIGURE 4. Example segmentation image before post-processing.

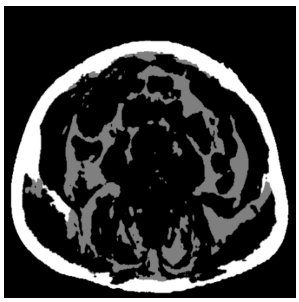


FIGURE 5. Example segmentation image after post-processing.

G. QUANTIFICATION

Quantifying adipose tissue from CT images involves a systematic process of translating image data into measurable volumes of visceral and superficial adipose tissues. We begin with the segmentation of CT images, where adipose tissues are identified and marked with specific greyscale values: 127 for visceral adipose tissue and 255 for superficial adipose tissue. These values are crucial for distinguishing between the different types of adipose tissues.

The algorithm calculates the count of pixels corresponding to each type of adipose tissue. Because the CT images are resized to 512 × 512 for inference, it is essential to adjust these pixel counts to account for any changes in image scale. This adjustment maintains the accuracy of the quantification process.

The next step involves converting the adjusted pixel counts into physical volumes. This is where the pixel spacing of the original CT images (DICOM files) becomes vital. Pixel spacing refers to the physical distance between the centres of adjacent pixels, measured in millimetres. By knowing the pixel spacing and the slice thickness, the study can accurately determine the actual physical volume each pixel represents.

To calculate the physical volume occupied by each type of adipose tissue, the adjusted pixel counts are multiplied by the volume of a single pixel (derived from pixel spacing and slice thickness). This calculation yields the total volume occupied by visceral and superficial adipose tissue in cubic millimetres, which are then converted to cubic centimetres.

1) MATHEMATICAL REPRESENTATION

The mathematical representation of the process is as follows:

Let P_x and P_y be the pixel spacings in the x and y dimensions, R_x and R_y the resized image dimensions, O_x and O_y the original image dimensions, and T the slice thickness.

The scaling factors are:

$$S_x = \frac{O_x}{R_x}, \quad S_y = \frac{O_y}{R_y} \tag{12}$$

The area per pixel:

$$A = P_x \times P_y \tag{13}$$

The corrected volume for each adipose tissue type in cubic millimeters:

$$V_{\text{visceral, mm}^3} = \text{Count}_{\text{visceral}} \times S_x \times S_y \times A \times T \tag{14}$$

$$V_{\text{superficial, mm}^3} = \text{Count}_{\text{superficial}} \times S_x \times S_y \times A \times T \tag{15}$$

And the conversion to cubic centimeters:

$$V_{\text{visceral, cm}^3} = \frac{V_{\text{visceral, mm}^3}}{1000} \tag{16}$$

$$V_{\text{superficial, cm}^3} = \frac{V_{\text{superficial, mm}^3}}{1000} \tag{17}$$

2) SCIENTIFIC FOUNDATION AND RELATED WORK

The methodology for converting pixel count to physical volume was substantiated by recent studies in the field. The research by Parikh et al. [27] delineated a similar approach, validating a method that calculated adipose tissue volume using voxel counts within CT images, reinforcing the accuracy of such methods for clinical assessments [27]. Moreover, contemporary studies highlighted the application

TABLE 2. Hyperparameters for U-Net, U-Net++, Attention U-Net, nnU-Net, R2U-Net, and MPNET during two training phases.

Model	Learning Rate	Batch Size		Epoch
		Phase 1	Phase 2	
U-Net	1e-4	8	4	200
U-Net++	1e-4	8	4	200
Attention U-Net	1e-4	16	8	200
nnU-Net	1e-4	8	4	200
R2U-Net	1e-4	8	4	200
MPNET	**2e-4**	**8**	**4**	**200**

of deep learning models for segmentation tasks, with subsequent pixel counting for volume calculations, showcasing the relevance of this approach [28], [29]

IV. EXPERIMENT

A. SETUP

The experimental setup for this study was conducted on Cluster 3, a high-performance computing facility provided by Birmingham City University. This section briefly recaps the computational environment used and its rationale in the context of the deep learning models and datasets employed in this research.

1) COMPUTATIONAL ENVIRONMENT

Cluster 3's multi-node architecture comprised of CPU and GPU nodes, all configured with PowerEdge R740 servers. The CPU nodes provide substantial memory capacity, each powered by approximately 180GB RAM for handling large datasets efficiently.

Complementing the CPU resources, the GPU nodes are equipped with two NVIDIA Tesla T4 GPUs per node, leveraging their 320 Turing Tensor cores and 2560 CUDA cores for rapid deep learning computation, critical for model training and inference tasks.

The system runs on CentOS-8.2.2004-x86_64, ensuring a robust platform for conducting complex machine learning experiments.

B. TRAIN, VALIDATION, TEST SPLIT

The dataset, comprising 1032 examples, was systematically divided into distinct sets for training, validation, and testing. This stratification was crucial for the evaluation of the deep learning models, ensuring that each had exposure to a comprehensive range of data during training while also setting aside unbiased subsets for performance assessment.

A total of 825 examples (approximately 80% of the dataset) were allocated for the training set, which is used to fit the models and adjust the weights. The validation set, consisting of 103 examples, corresponds to 10% of the dataset, serving as a benchmark for tuning model hyperparameters and preventing overfitting. The remaining 104 examples (also around 10%) were designated as the test set, providing a final, independent evaluation of the model's performance and its ability to generalise to new data.

1) TRAINING PHASES

Two primary training variations were employed:

1) Phase 1 (Initial Training): The first training phase proceeded over 200 epochs with a learning rate set at 1e-4 (as seen in Table 2 for the majority of models. An exception was made for the Parallel Net, which operated with a slightly higher learning rate of 2e-4 to accommodate its unique architecture. All models were trained with a batch size of 8 except Attention U-Net which used a batch size of 16, setting the stage for baseline performance metrics.
   Each model showcased unique strengths and weaknesses. As seen in Table 4 The MPNET model demonstrated the highest training accuracy (0.9799) and Dice coefficient (0.9726), indicating its effectiveness in segmenting adipose tissue in CT images. However, in the validation set, as seen in Table 3 it exhibited a higher loss (0.0839) compared to other models, suggesting some challenges in generalising to unseen data as.
   The U-Net model, while not reaching the peak performance of MPNET in training, showed a more consistent performance between training and validation, with a minimal loss increase from 0.0594 to 0.0681. This consistency is indicative of its generalisability.
   R2U-Net, U-Net++, and nnU-Net had a more balanced performance between training and validation, though with lower accuracy and Dice coefficients compared to MPNET and U-Net. The Attention U-Net model showed the lowest accuracy in both training (0.8554) and validation (0.8460), pointing to potential difficulties in learning from the dataset in this phase.

2) Phase 2 (Hyperparameter tuning): In Phase 2, the study refined the training process by selectively adjusting the batch sizes for models with an initial accuracy above 80%. Through an iterative approach of trial and error, the batch size was reduced to 4 for most models. This adjustment allowed for more frequent updates to the model weights, thereby facilitating nuanced learning. However, the Attention U-Net, which incorporates an attention mechanism, demonstrated improved performance with a larger batch size of 8. This specific example suggests that the model benefited from a broader variety of examples per update, optimizing its attention-driven feature selection.
   This heuristic approach to hyperparameter tuning was pragmatically employed to balance computational efficiency with the empirical observation of performance improvements. These adjustments were driven by direct experimentation, where different settings were tested and those yielding the best results were implemented. This method reflects a practical perspective, emphasizing outcomes over theoretical predictions, and is indicative of a real-world scenario where optimal

TABLE 3. Performance metrics for U-Net, U-Net++, Attention U-Net, nnU-Net, R2U-Net, and MPNET during two validation phases, detailing loss, accuracy, dice Coefficient, and IoU.

Model	Loss (Validation)		Accuracy (Validation)		Dice Coefficient (Validation)		IoU (Validation)	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
U-Net	0.0681	0.0600	0.9666	0.9724	0.9577	0.9625	0.9453	0.9496
U-Net++	0.0781	0.0663	0.9616	0.9677	0.9509	0.9588	0.9363	0.9461
Attention U-Net	0.2713	0.0957	0.8460	0.9633	0.8042	0.9323	0.7588	0.9052
nnU-Net	0.1158	0.1096	0.9409	0.9441	0.9241	0.9286	0.9014	0.9064
R2U-Net	0.0875	0.0746	0.9565	0.9629	0.9436	0.9522	0.9263	0.9372
MPNET	**0.0839**	**0.0807**	**0.9625**	**0.9631**	**0.9557**	**0.9559**	**0.9447**	**0.9443**

TABLE 4. Performance metrics for U-Net, U-Net++, Attention U-Net, nnU-Net, R2U-Net, and MPNET during two training phases, detailing loss, accuracy, dice coefficient, and IoU.

Model	Loss (Train)		Accuracy (Train)		Dice Coefficient (Train)		IoU (Train)	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
U-Net	0.0594	0.0473	0.9705	0.9779	0.9612	0.9678	0.9487	0.9557
U-Net++	0.0698	0.0568	0.9654	0.9720	0.9544	0.9630	0.9397	0.9512
Attention U-Net	0.2646	0.0854	0.8554	0.9691	0.8140	0.9375	0.7663	0.9103
nnU-Net	0.1075	0.1002	0.9451	0.9489	0.9284	0.9330	0.9057	0.9118
R2U-Net	0.0842	0.0694	0.9582	0.9654	0.9451	0.9544	0.9274	0.9399
MPNET	**0.0408**	**0.0437**	**0.9799**	**0.9784**	**0.9726**	**0.9708**	**0.9633**	**0.9611**

TABLE 5. Confidence intervals (CI) for Dice and IoU, along with *p*-values (vs MPNET), based on 100 test examples split into 10 groups.

Model	CI (Dice)	CI (IoU)	p-value (Dice)	p-value (IoU)
U-Net	(0.95898, 0.96389)	(0.94531, 0.95121)	0.0446	0.0118
U-Net++	(0.95236, 0.95834)	(0.93890, 0.94593)	0.0001	0.0001
Attention U-Net	(0.92737, 0.93350)	(0.89920, 0.90704)	3.11×10^{-12}	1.70×10^{-13}
nnU-Net	(0.92393, 0.93296)	(0.90101, 0.91313)	9.67×10^{-11}	1.29×10^{-10}
R2U-Net	(0.94892, 0.95377)	(0.93381, 0.93963)	6.35×10^{-7}	2.76×10^{-7}
MPNET	**(0.96299, 0.96813)**	**(0.95159, 0.95763)**	–	–

parameters are often determined through hands-on testing and observation.

2) EMPIRICAL HYPERPARAMETER TUNING: OBSERVATIONS AND MODEL PERFORMANCE ACROSS VALIDATION PHASES

Initial hyperparameters were set experimentally, and adjustments were made based on the first phase's results. The modifications demonstrate how architectural characteristics interact with hyperparameter settings. Detailed findings from this analysis are outlined below.

1) U-Net
   - Changes in metrics: Loss decreased from 0.0681 to 0.0600, accuracy increased from 0.9666 to 0.9724, Dice coefficient improved from 0.9577 to 0.9625, and IoU rose from 0.9453 to 0.9496.
   - Architectural influence: U-Net's architecture, which efficiently learns features with fewer parameters, benefits from smaller batch sizes enhancing its generalisation capabilities.

2) **U-Net**++
   - Changes in metrics: Loss reduced from 0.0781 to 0.0663, accuracy improved from 0.9616 to 0.9677, Dice coefficient increased from 0.9509 to 0.9588, and IoU advanced from 0.9363 to 0.9461.
   - Architectural influence: The complex nested connections of U-Net++ more effectively process the refined gradient information from smaller batch sizes, enhancing model learning and performance.

3) Attention U-Net
   - Changes in metrics: Significant reductions in loss from 0.2713 to 0.0957, with accuracy soaring from 0.8460 to 0.9633, Dice coefficient climbing from 0.8042 to 0.9323, and IoU jumping from 0.7588 to 0.9052.
   - Architectural influence: The attention mechanisms of this model benefit from a larger batch size that still provides a diverse set of examples, optimizing its attention-driven feature selection capabilities.

4) nnU-Net
   - Changes in metrics: Minor improvements with loss marginally decreasing from 0.1158 to 0.1096, accuracy rising from 0.9409 to 0.9441, Dice coefficient slightly improving from 0.9241 to 0.9286, and IoU increasing from 0.9014 to 0.9064.

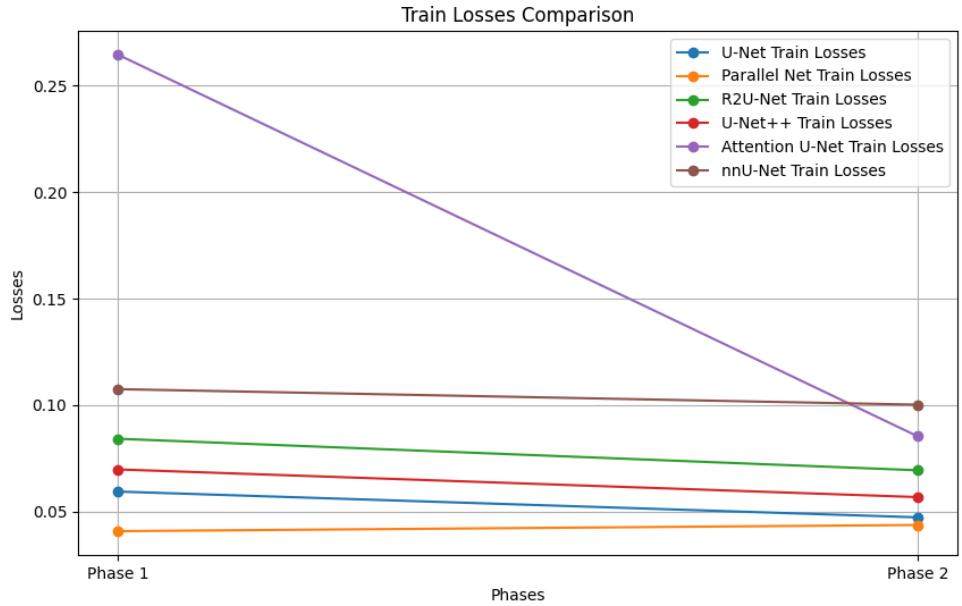


FIGURE 6. Best train losses comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

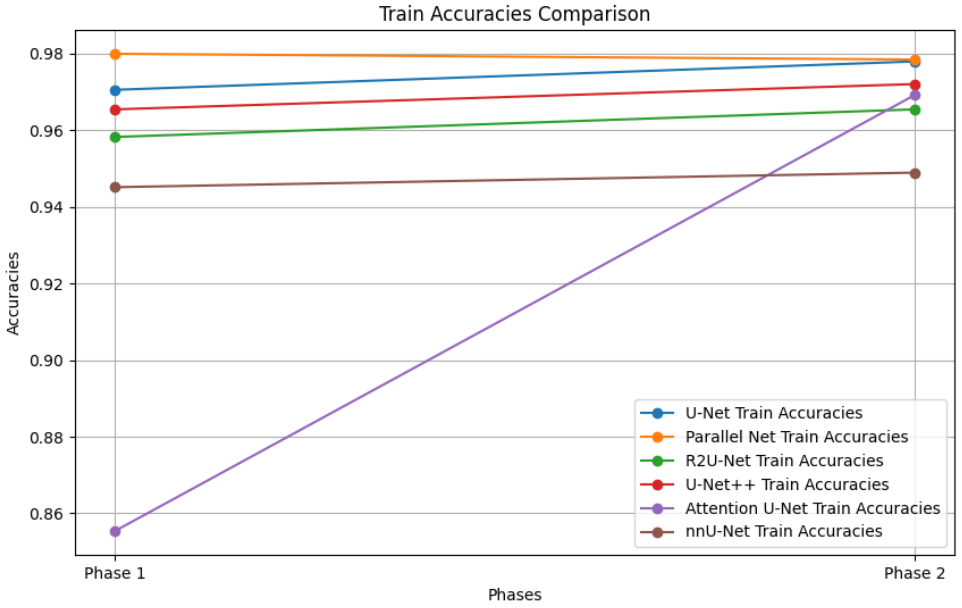


FIGURE 7. Best train accuracies comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

- Architectural influence: As a self-configuring system, nnU-Net subtly adjusts to changes in batch size, optimizing performance through its internal parameter adjustments.

5) R2U-Net

- Changes in metrics: Loss dropped from 0.0875 to 0.0746, accuracy increased from 0.9565 to 0.9629, Dice coefficient grew from 0.9436 to 0.9522, and IoU improved from 0.9263 to 0.9372.
- Architectural influence: The recurrent layers in R2U-Net benefit from smaller batch sizes, allowing for frequent updates that help integrate and refine contextual information more effectively.

6) MPNET

- Changes in metrics: Minimal improvement with loss slightly reducing from 0.0839 to 0.0807, accuracy remaining nearly the same from 0.9625 to 0.9631, Dice coefficient nearly unchanged from 0.9557 to 0.9559, and IoU slightly decreasing from 0.9447 to 0.9443.
- Architectural influence: With parallel convolutional paths at each depth level, MPNET is designed to extract a comprehensive range of

FIGURE 8. Best train dice coefficients comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

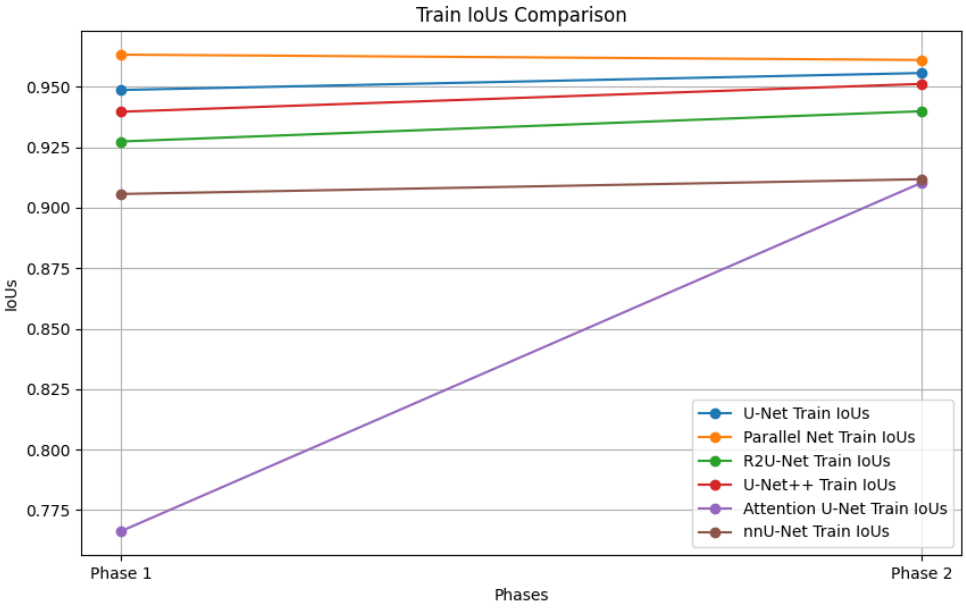


FIGURE 9. Best train IoUs comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

features by combining different kernel sizes which strengthens feature extraction. The minimal change in metrics despite the batch size reduction can be attributed to the model's capacity to maintain effective learning and generalisation.

C. QUANTITATIVE ANALYSIS OF ADIPOSE TISSUE SEGMENTATION USING MPNET ON UNSEEN DICOM DATA SETS

The MPNET model's efficacy was assessed on seven unseen datasets, containing 14 to 18 DICOM image slices taken at the L3 vertebra level. The calculations of visceral (VAT) and superficial adipose tissue (SAT) volumes are concisely presented in Table 7.

The means and medians indicated its consistent volume estimation across various anatomical structures. Standard deviations highlighted the precision of the model's measurements. The model's high accuracy, dice coefficient, and Intersection over Union (IoU) validated its robust segmentation capabilities.

D. STATISTICAL ANALYSIS

A hundred (100) test examples were divided into 10 equally sized groups (each containing 10 examples). For each group,

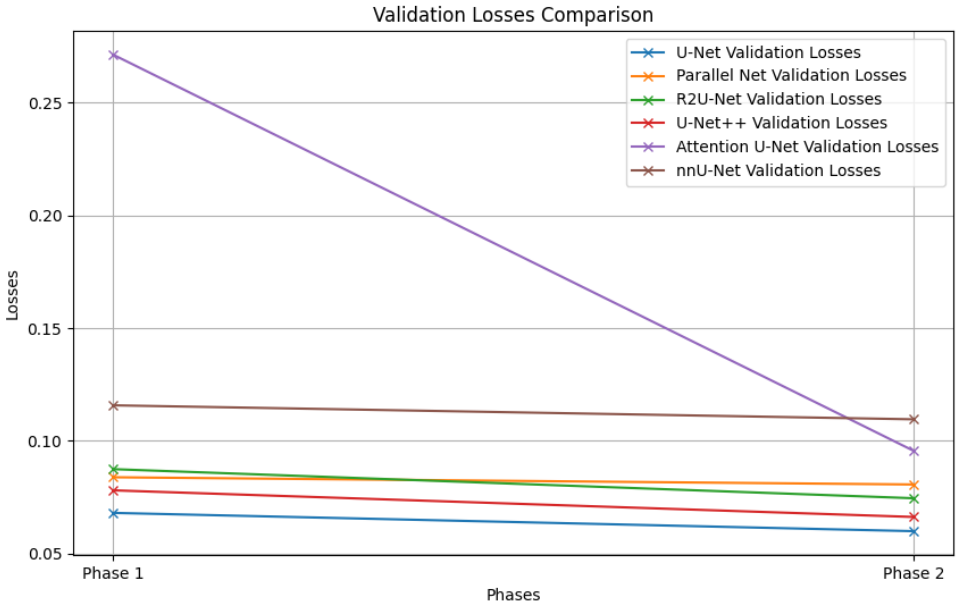


FIGURE 10. Best validation losses comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

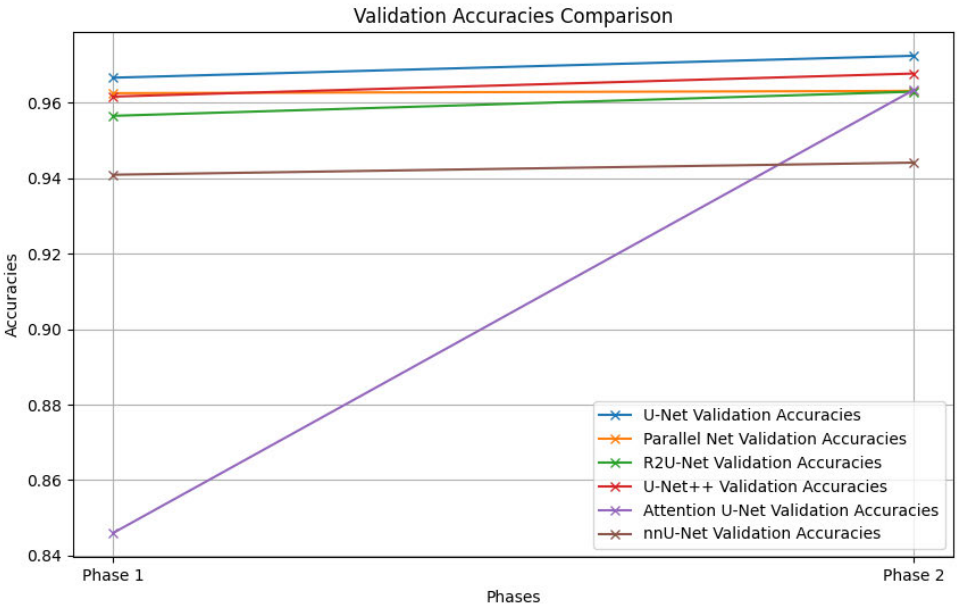


FIGURE 11. Best validation accuracies comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

we computed the Dice coefficient and IoU, then used these values to calculate 95% confidence intervals (CIs) and perform significance tests relative to MPNET. As shown in Table 5, MPNET achieves a Dice CI of (0.96299, 0.96813) and an IoU CI of (0.95159, 0.95763).

By comparison, R2U-Net reports a Dice CI of (0.94892, 0.95377) with a p-value of 6.35×10^{-7}, and nnU-Net exhibits a Dice CI of (0.92393, 0.93296) with a p-value of 9.67×10^{-11}. Attention U-Net's performance also differs significantly, with p-values on the order of 10^{-12}. All p-values remain far below the conventional 0.05 threshold, indicating MPNET's statistically superior segmentation performance.

V. DISCUSSION

The experimental results demonstrate significant insights into the performance of various CNN models in medical image segmentation. The following discussion interprets these results:

1) Performance Improvement Post-Tuning: A common trend observed across all models is the improvement in performance metrics post hyperparameter tuning. This highlighted the importance of fine-tuning in deep learning models, especially in complex tasks like medical image segmentation. The decrease in loss and increase in accuracy, Dice coefficient, and

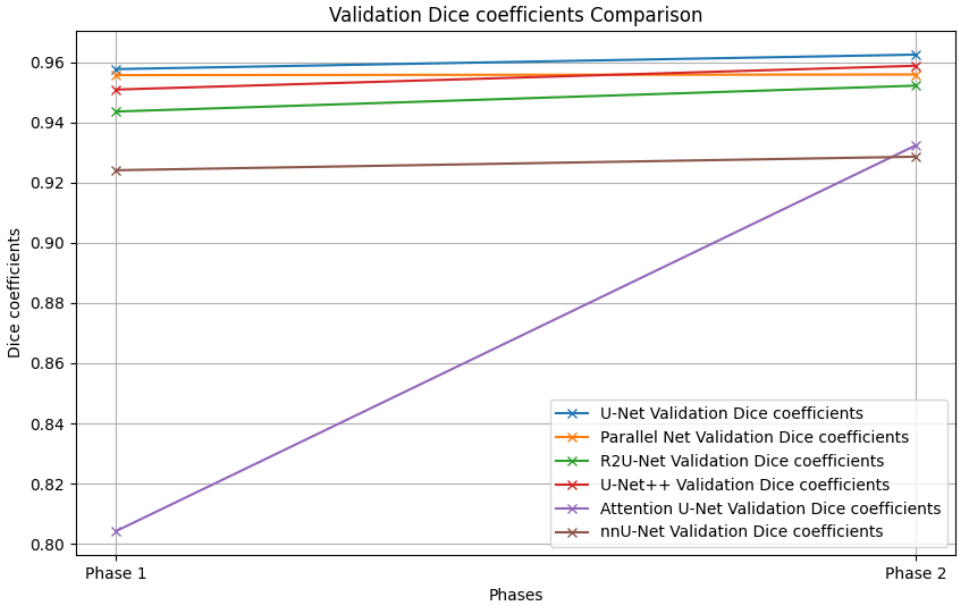


FIGURE 12. Best validation dice coefficients comparison chart for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net across two phases.

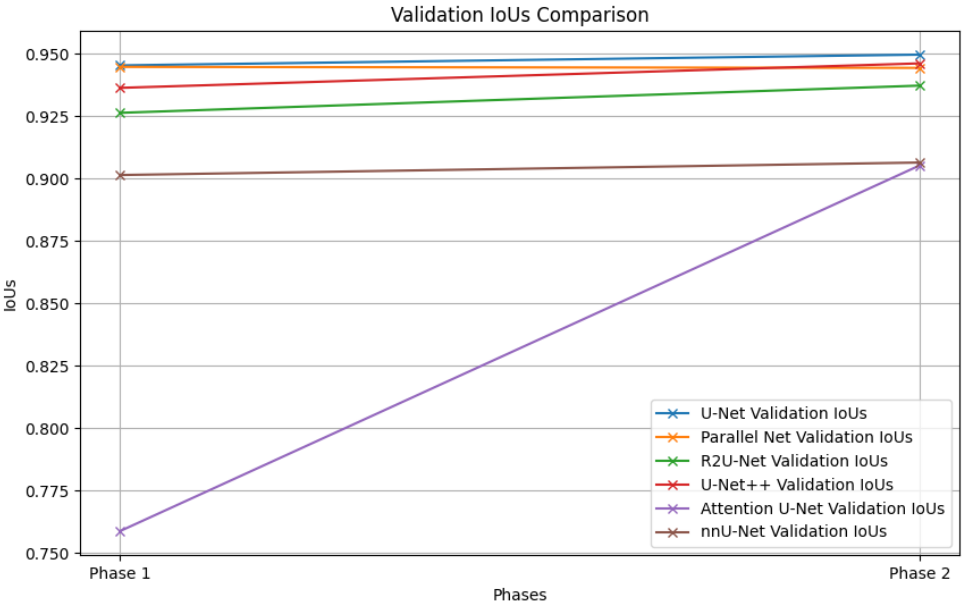


FIGURE 13. Best validation IoUs comparison graph for UNet, Parallel Net (MPNET), R2U-Net, U-Net++, Attention U-Net, and nnU-Net over two validation phases.

IoU in the post-tuning phase suggested that the models were able to generalise better and produce more accurate segmentations (see Table 4 and Table 3).

2) Technical and Clinical Generalisability of MPNET: MPNET's architecture, featuring parallel convolutional paths with varied kernel sizes, is poised for a strong performance across diverse clinical datasets. This design enables the model to effectively capture both fine and coarse anatomical details, essential for generalising to varied medical imaging scenarios. The combination of multiple feature scales and increased dropout at deeper layers suggests an enhanced ability to handle larger, more heterogeneous datasets without overfitting.
3) Comparison of Models:
   - U-Net: The U-Net model showed a substantial improvement in all metrics post-tuning, reaffirming its efficacy and robustness in medical image segmentation tasks. Its architecture, despite being the simplest among those tested, continued to deliver strong results, making it a reliable choice in the field (see Table 4 and Table 3).
   - MPNET: While MPNET shows an improvement in accuracy and Dice coefficient after tuning, its

TABLE 6. CT scans of the L3 vertebra from five patients are shown alongside segmentation outputs from the best versions of the six models: U-Net, U-Net++, Attn U-Net, nnU-Net, R2U-Net, and MPNET, highlighting differences in segmentation accuracy.

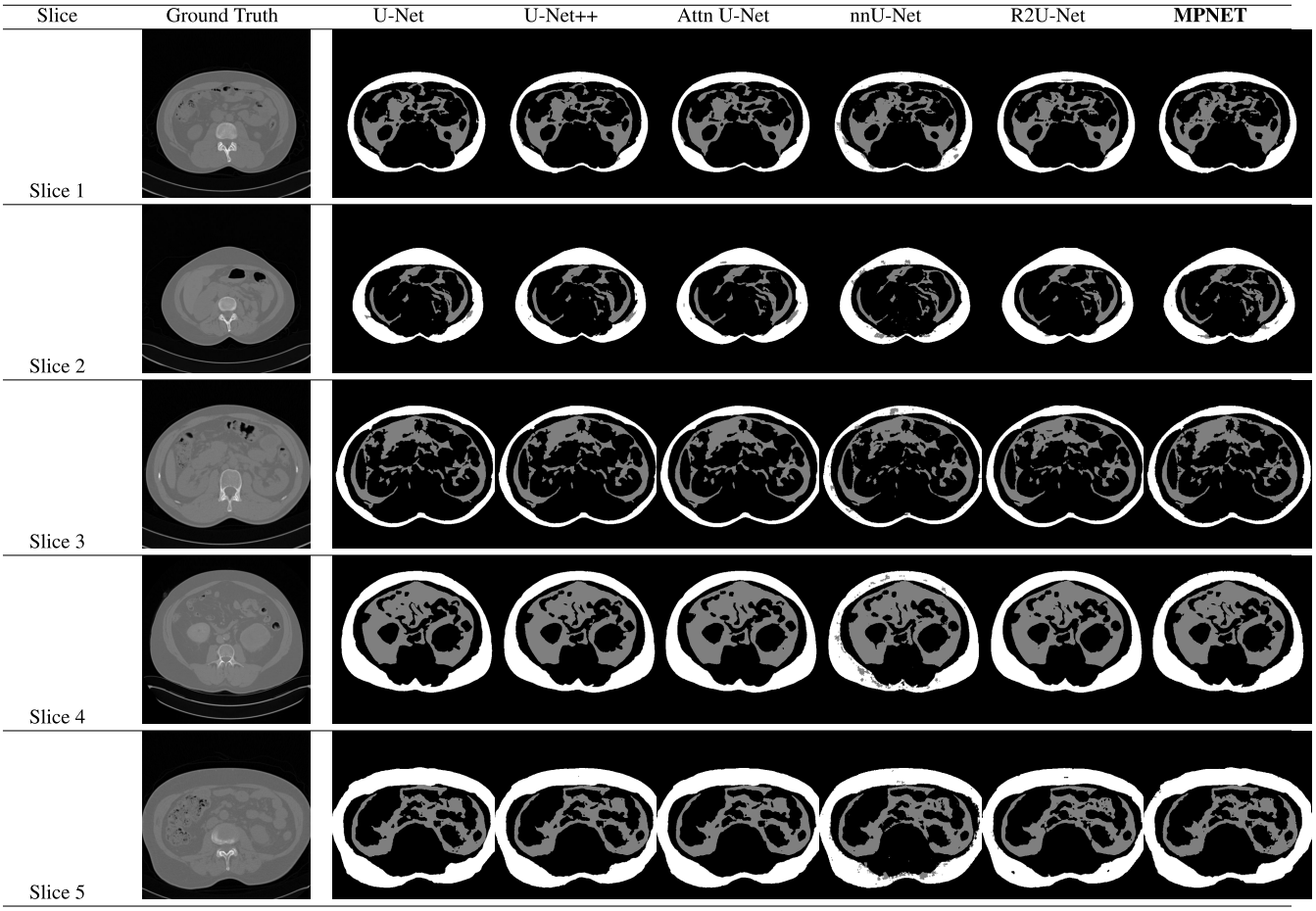


TABLE 7. Summary of MPNET's volume analysis including the sum, mean, median and standard deviation for VAT and SAT across seven datasets with 14-18 slices each.

S/N	Sum (cm3)			Mean (cm3)			Median (cm3)			Standard Deviation (cm3)		
	VAT	SAT	Total	VAT	SAT	Total	VAT	SAT	Total	VAT	SAT	Total
1.	246.62	1423.78	1670.4	16.44	94.92	111.36	15.9	95.43	112.38	2.01	5.72	5.17
2.	1080.44	963.39	2043.83	77.17	68.81	145.99	77.23	66.64	145.86	2.88	6.16	6.48
3.	581.94	891.36	1473.3	36.37	55.71	92.08	36.39	55.94	92.7	0.82	1.51	2.05
4.	325.44	371.87	697.31	20.34	23.24	43.58	20.49	23.4	43.78	0.96	1.21	0.71
5.	340.79	784.45	1125.24	20.05	46.14	66.19	19.68	46.61	66.48	1.84	1.83	1.38
6.	340.02	770.32	1110.34	21.25	48.15	69.4	21.4	48.69	69.3	1.29	1.5	1.11
7.	538.82	881.47	1420.29	29.93	48.97	78.91	30.4	50.36	78.72	2.47	2.93	2.96

loss increased slightly. This could indicate a more nuanced balance between precision and recall, and suggests areas for further optimization in its architecture. Also, from the early epochs, MPNET displayed higher validation and train loss figures compared to other models. This suggests that its architecture might be particularly effective, hinting at potentially significant results if tested on a larger scale.

- R2U-Net: Exhibits con1''sistent improvements across all metrics, which may be attributed to its recurrent convolutional layers that enhance feature extraction capabilities.
- U-Net++: Showed notable improvements post-tuning, particularly in accuracy and Dice coefficient. The nested architecture of U-Net++ seems to benefit significantly from the fine-tuning process, possibly due to its enhanced feature extraction through dense skip pathways.
- Attention U-Net: Despite its lower starting point, Attention U-Net showed remarkable improvement after tuning, particularly in terms of accuracy and

the Dice coefficient. This underscored the utility of attention mechanisms in improving model focus and accuracy in segmentation tasks.
- nnU-Net: Demonstrated consistent performance improvements, supporting the idea that a self-adapting framework can effectively tailor itself to the specifics of the segmentation task.

4) Radiological Assessment: An experienced consultant radiologist reviewed images produced by various models to ensure that they included only SAT or VAT. The review also verified that the models omitted abdominal organs, vascular structures, muscles, and bones. Any models not meeting these criteria were subsequently refined.
5) Implications:
   - The results reinforce the notion that no single model universally outperforms others in all aspects. Each model has its unique strengths and weaknesses, which become evident in different performance metrics (see Table 4, Table 3, and Table 6).
   - The improvement in IoU and Dice coefficient across all models post-tuning suggests that these models are particularly sensitive to hyperparameter adjustments. This is crucial for practitioners to consider when employing these models in real-world scenarios (see Table 3).
   - The variation in model performance also highlights the importance of selecting the right model based on specific requirements of the segmentation task, such as the level of detail required, the presence of imbalanced classes, or the computational resources available.

The findings from this study underscore the intricate dynamics and performance characteristics of various deep learning models in medical image segmentation. These insights not only reinforce the criticality of tailored model selection and parameter optimization but also hint at promising directions for future research to further refine and exploit these advanced modeling techniques.

VI. CONCLUSION

This study embarked on a comparative analysis of deep learning models for the segmentation and quantification of adipose tissue in CT imaging, culminating in the evaluation of traditional models and the introduction of the novel MPNET. The quantitative results demonstrated that while conventional models like U-Net and U-Net++ performed admirably, the novel MPNET showed promising qualitative results that could translate into better clinical applicability.

The utilization of Dice loss combined with categorical cross-entropy proved effective across models, facilitating a balance between shape similarity and pixel-wise classification accuracy. Data augmentation techniques successfully expanded the limited dataset, underscoring the potential of deep learning in scenarios where data scarcity is a challenge.

Furthermore, the post-processing step employed by MPNET highlighted its potential to produce clinically realistic segmentations, despite some quantitative metrics not fully capturing its performance. This suggests a possible divergence between numerical evaluation and practical utility, advocating for a more nuanced approach to model assessment in future research.

In light of these findings, it is evident that model selection cannot be guided by quantitative metrics alone, especially when considering the end application in clinical settings. The MPNET, with its unique architecture and promising clinical performance, warrants further investigation. Subsequent research could refine its architecture and explore its applicability across a broader range of medical imaging tasks.

VII. RECOMMENDATIONS AND FUTURE WORK

The outcomes of this study, while shedding light on the capabilities of CNN architectures in adipose tissue segmentation from CT images, also open several avenues for future research. The following recommendations and areas of future work are proposed to extend the findings of this research:

1) Architectural Refinement: Future studies could focus on refining the architecture of MPNET, perhaps by experimenting with different convolutional layer configurations or by introducing new forms of regularization to improve generalisation. The introduction of advanced techniques such as attention mechanisms or capsule networks could also be explored to enhance segmentation quality.
2) Dataset Expansion: While data augmentation techniques were effective in this study, acquiring larger and more diverse datasets could provide richer information for model training. Collaborating with medical institutions to gather more data would be a significant step forward. This study serves as a pilot, laying the groundwork for a forthcoming validation study that will include a larger and more ethnically diverse dataset, enhancing the generalisability and strength of the findings.
3) Computational Efficiency: Investigating the computational efficiency of the models is crucial, especially for deployment in clinical settings where resources may be limited. Studies could focus on model compression techniques or efficient network designs that do not compromise on performance.
4) Explainability and Trust: As AI models become more prevalent in healthcare, their explainability becomes paramount. Future work should aim to make these models more interpretable, which is vital for gaining trust from healthcare practitioners.
5) Cross-Validation: This study did not utilize cross-validation due to limited computing resources. Future research should incorporate cross-validation to better assess model performance and generalisability.

ACKNOWLEDGMENT

The authors would like to thank Birmingham City University for providing access to the necessary GPUs and also would like to thank Dr. Roger Trait for granting access to the computational clusters, facilitating this research.

REFERENCES

[1] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/s12065-020-00540-3.

[2] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[3] A. J. Weisman, D. T. Huff, R. M. Govindan, S. Chen, and T. G. Perk, "Multi-organ segmentation of CT via convolutional neural network: Impact of training setting and scanner manufacturer," *Biomed. Phys. Eng. Exp.*, vol. 9, no. 6, Nov. 2023, Art. no. 065021, doi: 10.1088/2057-1976/acfb06.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, Jan. 2015, pp. 234–241.

[5] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[6] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.

[7] Z. Zhou, M. Mahfuzur Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," 2019, *arXiv:1912.05074*.

[8] S. Pan, C. Chang, T. Wang, J. Wynne, M. Hu, Y. Lei, T. Liu, P. Patel, J. Roper, and X. Yang, "Abdomen CT multi-organ segmentation using token-based MLP-mixer," *Med. Phys.*, vol. 50, no. 5, pp. 3027–3038, May 2023.

[9] Z. Hong, M. Chen, W. Hu, S. Yan, A. Qu, L. Chen, and J. Chen, "Dual encoder network with transformer-CNN for multi-organ segmentation," *Med. Biol. Eng. Comput.*, vol. 61, no. 3, pp. 661–671, Mar. 2023.

[10] X. Zhang and H. Huang, "P-UNet: Parallel attention based UNet for crack detection," in *Proc. 7th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2022, pp. 311–315.

[11] J. H. Lee, S. H. Choi, K. J. Jung, J. M. Goo, and S. H. Yoon, "High visceral fat attenuation and long-term mortality in a health check-up population," *J. Cachexia, Sarcopenia Muscle*, vol. 14, no. 3, pp. 1495–1507, Jun. 2023.

[12] P. Finch, "Intra-abdominal fat: Comparison of computed tomography fat segmentation and bioimpedance spectroscopy," *Malawi Med. J.*, vol. 29, no. 2, pp. 155–159, Aug. 2017.

[13] Z. Xu, Y. Liu, C. Yan, R. Yang, L. Xu, Z. Guo, A. Yu, X. Cheng, L. Ma, C. Hu, G. Guglielmi, and K. Hind, "Measurement of visceral fat and abdominal obesity by single-frequency bioelectrical impedance and CT: A cross-sectional study," *BMJ Open*, vol. 11, no. 10, Oct. 2021, Art. no. e048221.

[14] Y. Tong, J. K. Udupa, and D. A. Torigian, "Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space: A novel approach," *Med. Phys.*, vol. 41, no. 6, Jun. 2014, Art. no. 063501.

[15] F. M. Mauad, F. A. Chagas-Neto, A. C. G. S. Benedeti, M. H. Nogueira-Barbosa, V. F. Muglia, A. A. O. Carneiro, E. M. M'uller, and J. Elias Jr., "Reproducibility of abdominal fat assessment by ultrasound and computed tomography," *Radiol. Brasileira*, vol. 50, no. 3, pp. 141–147, Jun. 2017.

[16] D. Sottier, J.-M. Petit, S. Guiu, S. Hamza, H. Benhamiche, P. Hillon, J.-P. Cercueil, D. Krausé, and B. Guiu, "Quantification of the visceral and subcutaneous fat by computed tomography: Interobserver correlation of a single slice technique," *Diagnostic Interventional Imag.*, vol. 94, no. 9, pp. 879–884, Sep. 2013.

[17] N. Tajbakhsh, "CNNs and their impact on medical image analysis," *J. Med. Imag.*, vol. 1, no. 1, pp. 10–20, 2022.

[18] H. Kim, J. Jung, J. Kim, B. Cho, J. Kwak, J. Y. Jang, S.-W. Lee, J.-G. Lee, and S. M. Yoon, "Abdominal multi-organ auto-segmentation using 3D-patch-based deep convolutional neural network," *Sci. Rep.*, vol. 10, no. 1, p. 6204, Apr. 2020.

[19] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, Dec. 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26263899/

[20] I. Naseer, S. Akram, T. Masood, A. Jaffar, M. A. Khan, and A. Mosavi, "Performance analysis of state-of-the-art CNN architectures for LUNA16," *Sensors*, vol. 22, no. 12, p. 4426, Jun. 2022, doi: 10.3390/s22124426.

[21] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. Maier-Hein, "NnU-Net: Self-adapting framework for U-Net-based medical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, Jan. 2018.

[22] M. Zahangir 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*.

[23] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," 2022, *arXiv:2202.05273*.

[24] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, Jan. 2020, pp. 465–476, doi: 10.18653/v1/2020.acl-main.45.

[25] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019, *arXiv:1902.09630*.

[26] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," 2021, *arXiv:2102.04525*.

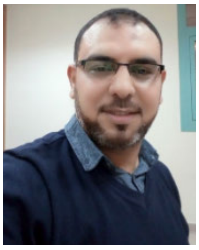
[27] A. M. Parikh, A. M. Coletta, Z. H. Yu, G. M. Rauch, J. P. Cheung, L. E. Court, and A. H. Klopp, "Development and validation of a rapid and robust method to determine visceral adipose tissue volume using computed tomography images," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183515, doi: 10.1371/journal.pone.0183515.

[28] N. Ahmad, R. Strand, B. Sparresäter, S. Tarai, E. Lundström, G. Bergström, H. Ahlström, and J. Kullberg, "Automatic segmentation of large-scale CT image datasets for detailed body composition analysis," *BMC Bioinf.*, vol. 24, no. 1, p. 346, Sep. 2023, doi: 10.1186/s12859-023-05462-2.

[29] Z. Ma, C. Li, T. Du, L. Zhang, D. Tang, D. Ma, S. Huang, Y. Liu, Y. Sun, Z. Chen, J. Yuan, Q. Nie, M. Grzegorzek, and H. Sun, "AATCT-IDS: A benchmark abdominal adipose tissue CT image dataset for image denoising, semantic segmentation, and radiomics evaluation," 2023, *arXiv:2308.08172*.

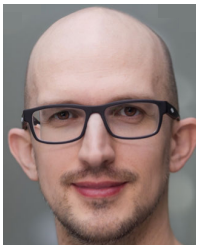
JOSTEVE ADEKANBI is a Staff Software Engineer at Bridge and a Machine Learning Researcher at the UK's National Health Service (NHS). He holds a First-Class Bachelor's degree in Computer Science from Bowen University and a Master's degree from Birmingham City University, where he graduated as the overall best student. His expertise spans software engineering and artificial intelligence, with a research focus on applying machine learning to healthcare, particularly in imaging and diagnostics. Additionally, Josteve is committed to mentoring emerging talents in Human-Computer Interaction (HCI), fostering innovation in user-centered design and interactive systems.

DEBASHISH DAS received the Ph.D. degree in computer science from Universiti Malaysia Pahang, Malaysia. He is currently a Faculty Member with the Faculty of Computing, Engineering and the Built Environment, Birmingham City University (BCU), U.K. Overall, he has 22 years of teaching and research experience at various leading universities in U.K., Malta, Malaysia, and Bangladesh. He is the author or the co-author of many articles in international journals and conferences. His research interests include but not limited to artificial intelligence, optimization, classification, prediction, data science, data analytics, data mining and machine learning algorithms, biomedical, artificial intelligent applications, and programming languages. He has received several recognitions for quality teaching and research, including the Top Performing Academic, the UMP Research Grant, the XMUM Research Fund, the Doctoral Scholarship Scheme (DSS), the Gold Medal in Three Minutes Thesis Completion, and the Merit Award (GOLD) for SCI Indexed Journal Publication.



NOUH ELMITWALLY received the M.Sc. degree in computer science from Cairo University and the Ph.D. degree in computer science from the University of Surrey, U.K., in 2014. He is currently a Ph.D. Scholar specializing in data science. With extensive experience in both academia and industry, he has a strong background in applied AI within medical imaging and natural language processing (NLP). His research interests include data science, big data, and artificial intelligence, with a particular focus on machine learning, deep learning, large language models, and generative AI. He has conducted numerous research studies in various healthcare areas and has published his findings in prestigious international journals and conferences.

ALIYUDA ALI is currently a Lecturer in computer science and software engineering with Sheffield Hallam University, U.K. His research interests include responsible and trustworthy AI, novel deep learning architectures, and data-driven methodologies to tackle complex real-world challenges in healthcare, energy systems, and environmental monitoring. He has published extensively in prestigious peer-reviewed journals and conferences, contributing to the advancement of knowledge in his field. He combines his passion for research with his dedication to teaching, aiming to inspire, and mentor the upcoming generation of computer scientists and engineers.



VINCE I. MADAI is currently the Lead of the "Responsible Algorithms" Team and a Principal Investigator with the QUEST Centre for Responsible Research with Berlin Institute of Health at Charité, Berlin, and a Visiting Professor in medical informatics with Birmingham City University. His research interests include trustworthy AI in healthcare, meta-research in AI, AI ethics, robustness, transparency, xAI, bias, reproducibility, the translation of AI in healthcare solutions, and improving healthcare product development.

BAHADAR BHATIA is currently the Chair of the Medical Physics Group, Institute of Physics. He is also an experienced Medical Physicist and Specialty Research Lead with the Diagnostic Radiology, Imaging Group, Sandwell and West Birmingham NHS Trust. Supplementing these roles, he has developed local clinical academic partnerships in artificial intelligence. In addition, he is also a part-time Honorary Visiting Fellow with the Space Research Centre, University of Leicester, where he develops his interests in photon counting detectors, X-ray gamma photon medical imaging, and Monte Carlo modeling.

JOHN MORLESE is currently a Consultant Radiologist with Sandwell and West Birmingham NHS Trust. His work focuses on improving the outcomes of radiological systems using advanced technologies. His aims are to improve the diagnostic accuracy and efficiency of radiological reports. He has published 30 peer-reviewed articles, one book, three book chapters, and 20 posters. His research interest includes chest and abdominal AI solutions.

MUHAMMED AFSAL received the European Diploma in Radiology (EDiR), Diplomate of National Board (DNB), and the M.D. degrees in diagnostic radiology from the Maulana Azad Medical College, Delhi.

He is currently a Consultant Radiologist with extensive experience in diagnostic radiology and subspecialty expertise in neuroradiology and cardiac imaging. He is also a Consultant Radiologist and the subspecialty Lead of U.S. and MRI with Sandwell and West Birmingham Hospitals NHS Trust. He holds a Fellowship of the Royal College of Radiologists (FRCR). He is an active member of several professional organizations, including the Royal College of Radiologists, the European Society of Radiology, Society for Cardiovascular Magnetic Resonance, and British Society of Cardiovascular Imaging (BSCI). He is committed to advancing radiological practices and has actively participated in research and teaching and has publications in international journals.

TANAY SHAH is currently a Consultant Radiologist working with SWBH NHS Trust, U.K. Following this, he went on to complete his FRCR from the Royal College of Radiologists, U.K., in 2020. He has an experience of working as a consultant radiologist for about six years now. His subspecialty research interests include MSK and neuroradiology along with general radiology and also have interest in AI in radiology. He was part of AI meetings in MSK imaging. His dissertation topic was MRI spectroscopy findings in prostate tumors. He also has a couple of publications to his name.