

Segmenting Medical Images: From UNet to Res-UNet and nnUNet

Lina Huang, Alina Miron, Kate Hone and Yongmin Li
Department of Computer Science, Brunel University London, Uxbridge, UK
{lina.huang, alina.miron, kate.hone, yongmin.li}@brunel.ac.uk

Abstract—This study provides a comparative analysis of deep learning models—UNet, Res-UNet, Attention Res-UNet, and nnUNet—evaluating their performance in brain tumour, polyp, and multi-class heart segmentation tasks. The analysis focuses on precision, accuracy, recall, Dice Similarity Coefficient (DSC), and Intersection over Union (IoU) to assess their clinical applicability. In brain tumour segmentation, Res-UNet and nnUNet significantly outperformed UNet, with Res-UNet leading in DSC and IoU scores, indicating superior accuracy in tumour delineation. Meanwhile, nnUNet excelled in recall and accuracy, which are crucial for reliable tumour detection in clinical diagnosis and planning. In polyp detection, nnUNet was the most effective, achieving the highest metrics across all categories and proving itself as a reliable diagnostic tool in endoscopy. In the complex task of heart segmentation, Res-UNet and Attention Res-UNet were outstanding in delineating the left ventricle, with Res-UNet also leading in right ventricle segmentation. nnUNet was unmatched in myocardium segmentation, achieving top scores in precision, recall, DSC, and IoU. The conclusion notes that although Res-UNet occasionally outperforms nnUNet in specific metrics, the differences are quite small. Moreover, nnUNet consistently shows superior overall performance across the experiments. Particularly noted for its high recall and accuracy, which are crucial in clinical settings to minimize misdiagnosis and ensure timely treatment, nnUNet’s robust performance in crucial metrics across all tested categories establishes it as the most effective model for these varied and complex segmentation tasks.

Index Terms—deep learning, UNet, Res-UNet, Attention Res-UNet, nnUNet, medical imaging segmentation, clinical application

I. INTRODUCTION

Medical image segmentation, the process of partitioning an image into multiple segments to identify regions of interest, plays a crucial role in various clinical applications, including disease diagnosis, treatment planning, and computer-assisted surgery [1]–[4]. However, accurate and robust segmentation is a challenging task due to the complexity and variability of medical images, often complicated by factors such as noise, intensity inhomogeneity, and low contrast [5], [6]. Traditional segmentation techniques, such as thresholding, region growing, and active contours, have limitations in handling these challenges [7], motivating the development of advanced methods. In recent years, deep learning techniques have emerged as powerful tools for medical image analysis, offering the potential to address these challenges effectively [8].

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool for medical image segmentation, demonstrating superior performance over conventional approaches [4], [9], [10]. Among the various CNN architectures, the UNet [11] and its variants, like Res-UNet [12], Attention Res-UNet [13] and nnUNet [14] have gained

significant attention due to their ability to capture multi-scale features and effectively segment complex structures.

While these models have demonstrated promising results in various applications, their relative performance on specific medical image segmentation tasks remains an area of active research. Evaluating and comparing these approaches is crucial for selecting the most suitable model for a given task and identifying potential areas for further improvement.

This study aims to comprehensively evaluate the performance of the UNet, Res-UNet, Attention Res-UNet, and nnUNet models on three medical image segmentation tasks: brain tumour segmentation, polyp segmentation, and cardiac segmentation. Through comparative analysis, this research seeks to identify the strengths and weaknesses of each model, determining which performs best across various segmentation scenarios. Such insights aim to set a benchmark in the field of medical image segmentation and provide practical guidance to future researchers on key considerations when applying UNet and its variants to medical image analysis.

The paper is structured as follows: Section 2 reviews related work in medical image segmentation and deep learning. Section 3 outlines experimental setup on brain tumour, colorectal polyp, and cardiac segmentation. Sections 4 to 6 present results. Section 7 provides the discussion, and Section 8 presents the conclusions of this study.

II. BACKGROUND

With the advent of advanced medical imaging modalities, such as X-ray (including mammography), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound, physicians have access to highly detailed and informative visual representations of the human body’s internal structures and physiological processes. Nonetheless, the pathway to becoming a radiologist involves years of specialized training and extensive clinical practice, leading to a notable shortage of professionals in this field [15], [16]. This upsurge in demand has consequentially amplified the workload for radiologists, potentially heightening the incidence of diagnostic errors and misdiagnoses due to fatigue and burnout [15]–[17].

Alternatively, the endeavor of accurately and reliably segmenting regions of interest from these varied imaging modalities presents considerable challenges. These challenges stem from the inherent complexity and variability of medical images, further complicated by factors like noise, uneven intensity distributions, low contrast, and the presence of artifacts [5], [6]. For instance, X-ray and CT images may suffer from

beam hardening and scatter artifacts, while MRI images can be affected by intensity inhomogeneities and geometric distortions. Ultrasound images, on the other hand, are frequently plagued by speckle noise and shadowing artifacts, rendering segmentation a daunting task.

Historically, segmentation techniques such as thresholding, region growing, and active contours have seen considerable application [17]. While these methods have demonstrated utility in certain scenarios, they often falter when dealing with the intricacies of complex medical images, particularly in the presence of noise, intensity variations, and ambiguous boundaries [7]. Moreover, these approaches typically rely on hand-crafted features and prior knowledge, limiting their generalizability and adaptability to diverse imaging modalities and anatomical structures [7]. Statistical methods have also been extensively applied to medical image segmentation, including probabilistic modelling [18], [19], Markov Random Field [20], graph-cut [21]–[24], level set [25], [26].

In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have revolutionized the field of medical image analysis, offering powerful solutions for segmentation tasks across various imaging modalities [1], [27]. CNNs possess the remarkable ability to automatically learn hierarchical feature representations from raw image data, enabling accurate and robust segmentation of complex anatomical structures [28], [29]. This data-driven approach alleviates the need for hand-crafted features and prior knowledge, making it more generalizable and adaptable to various imaging modalities and anatomical structures. Automated segmentation methods based on deep learning can alleviate the workload on radiologists, improve efficiency, and reduce the risk of misdiagnosis, ultimately enhancing the clinical workflow and patient care. Among the various CNN architectures, the UNet [11] and its variants, like Res-UNet [12], Attention Res-UNet [13] and nnUNet [14] have gained significant attention due to their ability to capture multi-scale features and effectively segment complex structures.

Despite the remarkable progress made by deep learning models in medical image segmentation, several challenges persist. Existing models may struggle with accurately segmenting small or irregularly shaped lesions, preserving fine details, and maintaining robustness across diverse imaging modalities and clinical scenarios. Additionally, the successful deployment of these models often requires substantial computational resources, large-scale annotated datasets, and extensive manual tuning of hyperparameters, hindering their widespread adoption in clinical settings.

III. METHODS

This section conducts detailed experiments on various deep learning models such as UNet, Res-UNet, Attention Res-UNet, and nnUNet. It explores their network architectures, layer filter characteristics, layer interconnections, and specific functionalities.

A. UNet

UNet [11] was a pioneering CNN architecture specifically designed for biomedical image segmentation. It features an encoder-decoder structure with skip connections that allow

low-level details from the encoder to be fused with high-level semantic features from the decoder. The contracting encoder path consists of repeated convolution, batch normalization, and max-pooling layers to extract abstract representations. The expansive decoder path comprises transposed convolutions and upsampling layers to recover the original resolution. Skip connections concatenate encoder and decoder features at each level, preventing loss of spatial information and enabling precise boundary localization. In this project, the architecture's kernel size and filter count are set at 3 across all convolution layers, suited for image segmentation. The network begins with 64 filters in the first layer, increasing progressively in deeper layers to promote hierarchical learning. After 4 decoding layers, the output moves through a final fully connected convolutional layer, where the kernel size adjusts based on the number of classes in the mask to meet specific task needs. The output is then activated using a function that varies with the number of labels, ensuring customized results for each task.

B. Res-UNet

Res-UNet [12] extended the UNet by incorporating residual connections into its encoder path. Each encoder block contains residual units with identity mappings, facilitating gradient flow and enabling training of deeper networks. The decoder mirrors the encoder with upsampling and concatenation of skip connections. Residual blocks aid in reusing and refining features across layers.

C. Attention Res-UNet

Attention Res-UNet [13] further enhanced the Res-UNet by integrating attention gates after each decoder stage. These compute attention maps highlighting salient regions, which are element-wise multiplied with features from skip connections. This focuses the model on relevant areas and suppresses noise, improving segmentation of fine details. The encoder follows the Res-UNet design, while the decoder incorporates attention gates and skip connections.

D. nnUNet

nnUNet [14] is a deep learning framework designed for medical image segmentation, aiming to simplify model configuration and optimization while improving performance. nnUNet is a self-configuring framework based on the UNet architecture and robust training schemes. Compared to the original UNet model, nnUNet has undergone only minor modifications. For example, it employs instance normalization instead of batch normalization, and replaces the standard ReLU activation function with leaky ReLUs. Moreover, the framework explicitly outlines a series of steps required for model training. These steps have a significant impact on the model's performance. For instance, in the preprocessing stage, operations like resampling and regularization are included to improve data quality and model robustness. During training, aspects such as the selection of loss functions, configuration of optimizers, and application of data augmentation strategies are involved to enhance the model's learning efficacy and generalization ability.

The general settings of the UNet, Res-UNet, Attention Res-UNet, and nnUNet models in our experiments are described as follow.

For brain tumour and polyp segmentation, the preprocessing includes resizing to 256×256, standardizing, and normalizing to a 0-1 range. These images are fed into the UNet, Res-UNet, and Attention Res-UNet, all using an Adam optimizer with a 1e-5 learning rate and callbacks like Early Stopping. For cardiac segmentation, due to MRI's grayscale, output masks are (128, 128, 4) with inputs at (128, 128, 1), using a Softmax for multi-class segmentation. The nnUNet model follows identical data processing steps. As previously mentioned, nnUNet can automatically specify settings for all phases of segmentation tasks, achieving task-specific optimization without the need for manual adjustments. In this project, the default nnUNet was adopted as the primary network architecture due to its outstanding performance in numerous segmentation tasks. The nnUNet architecture consists of an encoder, a decoder and transposed convolutions. Additionally, nnUNet applies normalization to the target voxels of each dataset (image) by subtracting the mean and dividing by the standard deviation, while the non-target (background) voxels are kept at 0. Aligning with nnUNet's initial configurations, downsampling is performed using strided convolutions, and upsampling is executed through transposed convolutions. The input patch size is chosen as 128×128×128, with a batch size of 2. The process includes five downsampling stages, culminating in a bottleneck feature map size of 4×4×4. The initial count of convolutional kernels starts at 32 and doubles after each downsampling stage, reaching up to a maximum of 320. The decoder's kernel count mirrors that of the encoder. Leaky ReLUs are selected for nonlinear activation, and instance normalization is employed for normalizing the feature maps. The training extends over 1000 epochs.

IV. BRAIN TUMOUR SEGMENTATION

In the medical field, brain tumour segmentation is considered one of the most challenging tasks. Accurately delineating the tumour boundaries is crucial not just for diagnosis but also plays a pivotal role in developing treatment plans and monitoring treatment progress, thereby significantly enhancing the quality of patient care [30]–[32]. This experiment aims to explore the accuracy of delineating the shape and boundaries of lower-grade gliomas (LGGs), which are slowly developing brain tumours that are generally less aggressive and have a better prognosis compared to higher-grade tumours, in brain MRI scans using models such as UNet, Res-UNet, Attention Res-UNet, and nnUNet.

A. Dataset

The brain tumour segmentation data, sourced from Kaggle (<https://www.kaggle.com/datasets/mateuszbudaj/gg-mri-segmentation>) and originally from The Cancer Imaging Archive (TCIA), features MRI brain images from 110 patients, totaling 7858 images. These images are evenly divided between brain images and their FLAIR abnormality segmentation masks. For the experiment, 1200 images and masks were randomly selected. To focus on abnormality segmentation, images and masks without low-grade gliomas

TABLE I
BRAIN TUMOUR SEGMENTATION: PERFORMANCE METRICS FOR UNET, RES-UNET, ATTENTION RES-UNET AND NNUNET ON TESTING DATASET.

| Model | Precision | Recall | Accuracy | DSC | IoU |
|--------------------|--------------|--------------|--------------|--------------|--------------|
| UNet | 0.884 | 0.767 | 0.990 | 0.821 | 0.697 |
| Res-UNet | 0.873 | 0.853 | 0.992 | 0.863 | 0.759 |
| Attention Res-UNet | 0.842 | 0.806 | 0.989 | 0.824 | 0.700 |
| nnUNet | 0.856 | 0.881 | 0.994 | 0.850 | 0.757 |

abnormalities (referred to as negative) were excluded, leaving only positive image-mask pairs. The dataset was then divided into training, testing, and validation sets in an 8:1:1 ratio, resulting in training, validation, and testing cases for the UNet, Res-UNet, and Attention Res-UNet models. The nnUNet model, employing 5-fold cross-validation, drew on the combined cases from the training and validation sets of the other models, automatically partitioning them for training and validation. In all four models, the testing set is consistent.

B. Result

In this experiment, the performance metrics include the precision, recall, accuracy, DSC and IoU. The following Table I displays information related to these performance metrics. From the performance metrics for four different neural network models, deductions can be made based on the information provided in the table.

- 1) UNet: It demonstrates high precision, effectively identifying true positives among its predictions. Yet, its recall is lower than other models, indicating it may overlook more true positives (false negatives). Additionally, both its DSC and IoU scores are the lowest, reflecting poorer overlap with the ground truth compared to other models.
- 2) Res-UNet: it has the highest DSC score (0.863) and IoU (0.759), indicating the best segmentation overlap, making it a strong model for tasks where accurate delineation of tumour boundaries is crucial.
- 3) Attention Res-UNet: It has the lowest precision at 0.842 and the lowest accuracy at 0.989. Despite this, an accuracy of 0.989 is still very high, indicating that the model performs well overall, albeit slightly less so than its counterparts in these particular metrics.
- 4) nnUNet: It stands out with the highest recall (0.881) and accuracy (0.994), suggesting it is the most reliable at identifying true positives and overall prediction correctness. This makes it a potentially valuable model for clinical applications, where missing the presence of a tumour (false negatives) could have significant consequences.
- 5) Similarity in Precision, Accuracy, and DSC Scores: Despite minor differences in precision, accuracy, and DSC scores among the models, their close performance indicates all are capable of accurately predicting true positives, identifying cases, and matching tumour boundaries well. This suggests any model could be fit for brain tumour segmentation, with selection dependent on factors like computational speed and implementation ease.

- 6) **Significant Variability in Recall and IoU:** The models show notable differences in recall and IoU, indicating varied effectiveness in capturing all relevant tumour cases and accurately outlining tumour areas against ground truth. High recall is essential in clinical settings to avoid missing tumours, essential for treatment planning, pointing to some models (e.g., nnUNet) as more appropriate for this task. IoU variations reveal differences in segmentation precision, critical for precise treatment or surgical planning, with models like Res-UNet and nnUNet showing superior accuracy in delineating tumour boundaries.
- 7) **Generally Lower IoU Scores:** IoU scores being not quite high across all models could be due to several factors inherent to the task of brain tumour segmentation: complexity of tumour shapes (irregular and complex shapes); variability in tumour appearance; image quality and resolution; overlap with non-tumour tissues (tumours might have similar intensity values to surrounding tissues).
- 8) **DSC Coefficient vs. IoU:** Minor variations in DSC scores compared to IoU suggest models capture tumours well but differ more in precisely outlining tumour boundaries, as IoU indicates. This is crucial for treatment planning and monitoring tumour growth, highlighting IoU's importance in model evaluation. In essence, despite DSC and IoU assessing overlap, IoU's calculation highlights more significant boundary delineation differences among models, essential for clinical application insights. Improving IoU scores might involve incorporating more sophisticated image processing techniques, enhancing the models with additional contextual information, or using higher-quality imaging data. It may also benefit from more advanced training strategies that specifically target the improvement of segmentation boundary accuracy, such as utilizing more refined loss functions or incorporating post-processing techniques to refine the segmentation outputs.

Four examples were selected to demonstrate the predictive results of the four models throughout the experiment, including the provided images and their corresponding ground truth masks (see Fig. 1).

V. POLYP SEGMENTATION

Polyps are a leading cause of colorectal cancer, with colonoscopy being the preferred detection and removal method. Accurately identifying polyps is key to diagnosing and treating colorectal cancer, yet segmenting them during colonoscopy is difficult due to their varied shapes, sizes, and color contrasts [33], [34]. This study focuses on precisely defining intestinal polyps' contours and boundaries using models like UNet, Res-UNet, Attention Res-UNet, and nnUNet.

A. Dataset

CVC-ClinicDB [35] is a dataset of frames from colonoscopy videos, containing original images and corresponding polyp masks (ground truth). It comprises 612 original images paired with 612 masks. In this experiment, the data is split into training, validation, and test sets in an 8:1:1 ratio. Similarly to the previous brain segmentation experiment, all models (UNet,

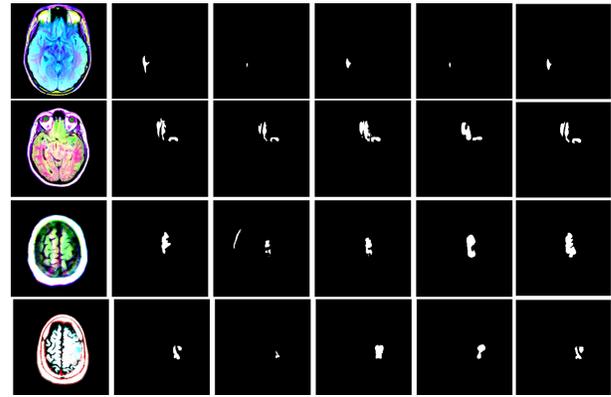


Fig. 1. Brain tumour segmentation: Four examples showing segmentation by four models. From left to right: input images, ground-truth, and segmentation by UNet, Res-UNet, Attention Res-UNet, nnUNet, in order.

Res-UNet, Attention Res-UNet, and nnUNet) use the same dataset. During nnUNet training, it receives the combined training and validation data from the other models. In all four models, the testing set is consistent.

B. Result

The following Table II displays performance metrics including precision, recall, accuracy, DSC, and IoU for four different models. Deductions can be made based on the information provided in the table.

TABLE II
POLYP SEGMENTATION: PERFORMANCE METRICS FOR UNET, RES-UNET, ATTENTION RES-UNET AND nnUNET ON TESTING DATASET.

| Model | Precision | Recall | Accuracy | DSC | IoU |
|--------------------|--------------|--------------|--------------|--------------|--------------|
| UNet | 0.868 | 0.764 | 0.969 | 0.813 | 0.685 |
| Res-UNet | 0.907 | 0.751 | 0.971 | 0.821 | 0.697 |
| Attention Res-UNet | 0.902 | 0.754 | 0.971 | 0.821 | 0.697 |
| nnUNet | 0.940 | 0.949 | 0.993 | 0.941 | 0.895 |

- 1) **Precision and Accuracy:** All four models excel in precision and accuracy, although UNet has the lowest scores (0.868 precision and 0.969 accuracy, but the values are still quite high). Yet, the minor differences in these metrics among the models imply a consistent strength in identifying true positives across tested datasets.
- 2) **Recall, DSC, and IoU:** The UNet, Res-UNet, and Attention Res-UNet models show similar recall, DSC, and IoU scores, evaluating their ability to detect all relevant cases (recall) and their segmentation accuracy against the ground truth (DSC and IoU). Despite being lower than nnUNet's scores, they remain acceptable. Nonetheless, in critical fields like medical imaging where overlooking a true positive can be crucial, even minor discrepancies matter.
- 3) **nnUNet Performance:** The nnUNet excels, showing the highest precision (0.940), recall (0.949), accuracy (0.993), DSC (0.941) and IoU (0.895) scores, markedly surpassing other models. Its recall indicates a stronger capability in identifying actual positives, minimizing missed polyp cases. The nnUNet's DSC and IoU, significantly higher than its counterparts, suggests a closer

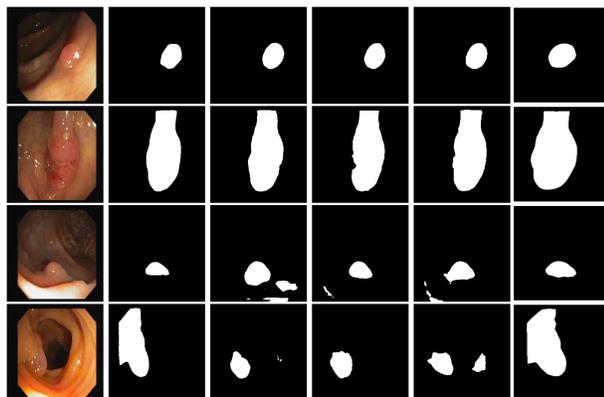


Fig. 2. Polyp segmentation: Four examples showing segmentation by four models. From left to right: input images, ground-truth, and segmentation by UNet, Res-UNet, Attention Res-UNet, nnUNet, in order.

match to the ground truth in predictions. This accuracy is critical in medical tasks for precise area segmentation. In summary, nnUNet outperforms in all aspects, especially in recall, DSC and IoU, crucial for accurate detection and segmentation. It emerges as the top choice for polyp segmentation, provided its computational demands are manageable. Other models, while effective, serve as alternatives under nnUNet’s resource or speed constraints, or specific clinical needs. Based on the results of the above polyp segmentation experiment, four examples were selected to demonstrate the performance of the four models in polyp segmentation. From these four examples, it is clear that nnUNet has the best performance (see Fig. 2).

VI. HEART SEGMENTATION

Medical imaging of the heart is essential for diagnosing and treating cardiovascular diseases, offering key insights into cardiac structure and function. Precise segmentation is critical for accurately measuring features like chambers, wall thickness, and blood flow dynamics, essential for diagnosing heart failure, arrhythmias, and other cardiac conditions [36], [37]. Yet, this process is challenging due to the heart’s complex anatomy and motion artifacts from dynamic imaging. Manual segmentation, while precise, is time-consuming and subject to variability [38], [39]. Addressing this, our study evaluates deep learning models—UNet, Res-UNet, Attention Res-UNet, and nnUNet—for automating heart image segmentation.

A. Dataset

This study utilized the ACDC dataset [40] from the University Hospital of Dijon, featuring cardiac MRI of 150 patients for cardiac segmentation. It spans diastole and systole phases, with 100 and 50 patients in the training and testing sets, respectively. Sets include original images and masks for the diastole phase, categorizing heart parts from 0 to 3 for background, RV, Myo, and LV. The goal was multi-class segmentation during the diastole phase. After preprocessing, the dataset comprised 951 images and masks, divided into training, validation, and testing sets in an 8:1:1 ratio. The UNet, Res-UNet, Attention Res-UNet, and nnUNet models were trained on these sets, with nnUNet utilizing combined

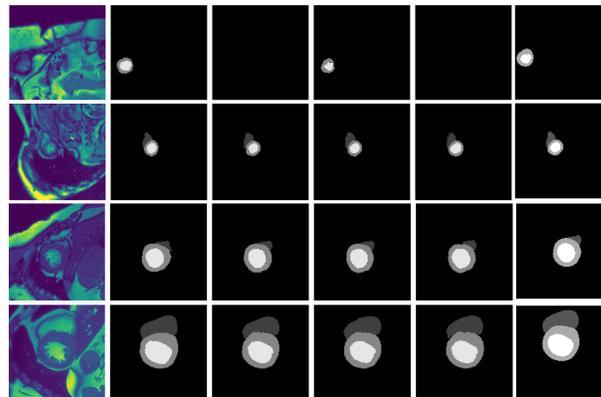


Fig. 3. Heart segmentation: Four examples showing segmentation by four models. From left to right: input images, ground-truth, and segmentation by UNet, Res-UNet, Attention Res-UNet, nnUNet, in order.

training and validation sets. All models were tested using a consistent set of images and masks.

B. Result

From performance metrics of precision, recall, DSC, and IoU for four different models, deductions can be made based on the information provided in the Table III:

- 1) Background (Class 0): Not surprisingly, all models excel at identifying the background class due to its larger area and simpler texture, making it easier to learn and segment compared to anatomical structures.
- 2) Right Ventricle (RV, Class 1): The nnUNet’s lower scores in precision, recall, DSC, and IoU for RV segmentation suggest it is less suitable for this task compared to others. Res-UNet and Attention Res-UNet both excel in heart segmentation, with Res-UNet slightly outperforming Attention Res-UNet in certain metrics.
- 3) Myocardium (Myo, Class 2): Segmenting the Myocardium (Myo) proves more difficult for the models, showing lower metrics than the background, RV, and LV classes, likely due to Myo’s smaller size and complex shape. However, the nnUNet stands out for its superior Myo segmentation, indicating its adaptive architecture and training strategy successfully tackle Myo’s unique features.
- 4) Left Ventricle (LV, Class 3): Res-UNet and Attention Res-UNet indeed have similar, high performance metrics for LV segmentation. UNet and nnUNet also perform well, with only slight differences compared to the other two models. This implies that the LV, while segmented reasonably well by all models, benefits from the features of the Res-UNet and Attention Res-UNet architectures.

Based on the results of the above heart segmentation experiment, four examples were selected to demonstrate the performance of the four models in heart segmentation (see Fig. 3).

VII. DISCUSSION

This study provides a thorough comparison of four deep learning models—UNet, Res-UNet, Attention Res-UNet, and

TABLE III
HEART SEGMENTATION: PRECISION, RECALL, DSC AND IOU SCORE FOR EACH CLASS BY FOUR MODELS.

| Model | Precision | | | | Recall | | | | DSC | | | | IoU | | | |
|--------------------|-----------|-------------|-------------|-------------|--------|-------------|-------------|-------------|-------|--------------|--------------|--------------|-------|--------------|--------------|--------------|
| | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| UNet | 0.99 | 0.93 | 0.87 | 0.97 | 0.99 | 0.92 | 0.86 | 0.94 | 0.993 | 0.924 | 0.869 | 0.956 | 0.987 | 0.860 | 0.768 | 0.916 |
| Res-UNet | 0.99 | 0.95 | 0.89 | 0.98 | 1.00 | 0.92 | 0.85 | 0.94 | 0.994 | 0.936 | 0.869 | 0.961 | 0.988 | 0.879 | 0.769 | 0.924 |
| Attention Res-UNet | 0.99 | 0.94 | 0.89 | 0.96 | 0.99 | 0.92 | 0.86 | 0.96 | 0.994 | 0.930 | 0.874 | 0.960 | 0.988 | 0.869 | 0.777 | 0.924 |
| nnUNet | 1.00 | 0.82 | 0.89 | 0.95 | 1.00 | 0.81 | 0.90 | 0.95 | 0.996 | 0.812 | 0.894 | 0.953 | 0.992 | 0.776 | 0.818 | 0.921 |

nnUNet—on three medical imaging tasks: brain tumour, polyp, and multi-class heart segmentation.

- 1) Brain Tumour Segmentation: In brain tumour segmentation, Res-UNet and nnUNet excelled, showing superior performance in critical metrics. Res-UNet achieved the highest DSC and IoU scores, indicating exceptional accuracy in tumour boundary delineation. nnUNet was notable for its high recall and accuracy, proving to be highly reliable in identifying tumour presence, which is vital for effective clinical diagnosis and treatment planning. These models' advanced architectures allow for superior capture of complex image details, crucial for accurate medical image analysis. Despite nnUNet's intensive computational demands, its outstanding overall performance highlights its significant potential in medical applications.
- 2) Polyp Segmentation: The nnUNet model outperformed others, exhibiting exemplary precision and recall, indicating its potential as a reliable tool for endoscopic analysis, provided the computational resources for its extensive training are available.
- 3) Heart Multi-Class Segmentation: In the segmentation of cardiac structures, Res-UNet and Attention Res-UNet showed promising results, especially in LV segmentation, where precision in anatomical delineation is paramount. As for segmenting the RV, Res-UNet stands out as the best performer in this class, suggesting that the residual connections in its architecture are beneficial for capturing the complex features of the RV. The nnUNet's performance, particularly in segmenting the myocardium with high DSC and IoU scores, suggests its utility in scenarios where the most detailed tissue differentiation is required.

VIII. CONCLUSIONS

The discussion above separately addresses brain segmentation, polyp segmentation, and heart multi-class segmentation. From the three experiments, it is evident that Res-UNet performs better than UNet and Attention Res-UNet, achieving higher scores in certain performance metrics than nnUNet. Nevertheless, when combining the results from all three experiments, nnUNet emerges as the best model for the following reasons:

- 1) Although nnUNet requires considerable computational resources, in clinical applications, high diagnostic accuracy and low misdiagnosis rates are paramount. For example, missing a tumour in a patient can lead to delayed treatment with serious consequences.
- 2) In brain segmentation, the DSC score (0.863) and IoU (0.759) for Res-UNet are higher than those for nnUNet

(0.850 and 0.757, respectively), though the margins are very small. However, the values of recall and accuracy for nnUNet are higher than those for Res-UNet.

- 3) In polyp detection, all performance metrics for nnUNet are higher than the other three models, especially recall, DSC, and IoU, where nnUNet's scores (0.949, 0.941, and 0.895, respectively) far surpass those of the others.
- 4) In heart multi-class segmentation, nnUNet's metrics for RV are the lowest at 0.82, 0.81, 0.812, and 0.776, but still high and acceptable. nnUNet leads in myocardial segmentation with the highest scores, including a notable recall of 0.90 and an IoU of 0.818. In LV segmentation, while Res-UNet and Attention Res-UNet perform slightly better, nnUNet's results are closely comparable, nearly matching those models.

This study sets a benchmark in the field of medical image segmentation and offers insights to future researchers on key considerations when applying UNet and its variants to medical image analysis. By systematically comparing models such as UNet, Res-UNet, Attention Res-UNet, and nnUNet, the research highlights their respective strengths and potential applications in segmenting MRI and endoscopic images, as well as in binary and multi-class tasks.

However, despite notable progress in two-dimensional image segmentation, there are limitations in the types of data handled. In real medical scenarios, many critical datasets typically exist in three-dimensional forms, yet all experiments conducted in this study are limited to two-dimensional data. This limitation could restrict the models' ability to address real medical imaging challenges effectively. Future research could also explore more extended models similar to nnUNet, and develop or fine-tune large AI models, to enhance their capability to process and analyze three-dimensional medical images. Additionally, this experiment faces a data leakage issue due to the division of data at the image level rather than the patient level. To prevent such problems in future work, it is recommended: 1. Strict data segmentation ensures that all images from the same patient are contained within a single dataset (training, validation, or test) before training begins. 2. Proper data handling before augmentation to ensure that enhancements are only applied to the training set, thus avoiding indirect data leaks into validation or test sets.

REFERENCES

- [1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [2] L. Liu, J. Cheng, Q. Quan, F.-X. Wu, Y.-P. Wang, and J. Wang, "A survey on u-shaped networks in medical image segmentations," *Neurocomputing*, vol. 409, pp. 244–258, 2020.

- [3] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.
- [4] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Deep learning in medical image analysis: challenges and applications*, pp. 3–21, 2020.
- [5] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.
- [6] E. K. Wang, C.-M. Chen, M. M. Hassan, and A. Almgren, "A deep learning based medical image segmentation technique in internet-of-medical-things domain," *Future Generation Computer Systems*, vol. 108, pp. 135–144, 2020.
- [7] I. Despotović, B. Goossens, and W. Philips, "Mri segmentation of the human brain: challenges, methods, and applications," *Computational and mathematical methods in medicine*, vol. 2015, 2015.
- [8] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, 2021.
- [9] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, pp. 611–629, 2018.
- [10] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [13] D. Maji, P. Sigedgar, and M. Singh, "Attention res-unet with guided decoder for semantic segmentation of brain tumors," *Biomedical Signal Processing and Control*, vol. 71, p. 103077, 2022.
- [14] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [15] T. C. Kwee and R. M. Kwee, "Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence," *Insights into imaging*, vol. 12, pp. 1–12, 2021.
- [16] D. Yang, C. Martinez, L. Visaña, H. Khandhar, C. Bhatt, and J. Carretero, "Detection and analysis of covid-19 in medical images using deep learning techniques," *Scientific Reports*, vol. 11, no. 1, p. 19638, 2021.
- [17] E. Michael, H. Ma, H. Li, F. Kulwa, and J. Li, "Breast cancer segmentation methods: current status and future potentials," *BioMed research international*, vol. 2021, pp. 1–29, 2021.
- [18] D. Kaba, C. Wang, Y. Li, A. Salazar-Gonzalez, X. Liu, and A. Serag, "Retinal blood vessels extraction using probabilistic modelling," *Health Information Science and Systems*, vol. 2, pp. 1–10, 2014.
- [19] D. Kaba, A. G. Salazar-Gonzalez, Y. Li, X. Liu, and A. Serag, "Segmentation of retinal blood vessels using gaussian mixture models and expectation maximisation," in *Health Information Science: Second International Conference, HIS 2013, London, UK, March 25-27, 2013. Proceedings 2*, pp. 105–112, Springer Berlin Heidelberg, 2013.
- [20] A. Salazar-Gonzalez, Y. Li, and D. Kaba, "Mrf reconstruction of retinal images for the optic disc segmentation," in *Health Information Science: First International Conference, HIS 2012, Beijing, China, April 8-10, 2012. Proceedings 1*, pp. 88–99, Springer Berlin Heidelberg, 2012.
- [21] A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu, "Segmentation of the blood vessels and optic disk in retinal images," *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, pp. 1874–1886, 2014.
- [22] A. G. Salazar-Gonzalez, Y. Li, and X. Liu, "Optic disc segmentation by incorporating blood vessel compensation," in *2011 IEEE Third International Workshop On Computational Intelligence In Medical Imaging*, pp. 1–8, IEEE, 2011.
- [23] A. G. Salazar-Gonzalez, Y. Li, and X. Liu, "Retinal blood vessel segmentation via graph cut," in *2010 11th International Conference on Control Automation Robotics & Vision*, pp. 225–230, IEEE, 2010.
- [24] B. Dodo, Y. Li, K. Eltayef, and X. Liu, "Graph-cut segmentation of retinal layers from oct images," in *International Conference on Bioimaging*, 2018.
- [25] C. Wang, D. Kaba, and Y. Li, "Level set segmentation of optic discs from retinal images," *Journal of Medical Systems*, vol. 4, no. 3, pp. 213–220, 2015.
- [26] B. I. Dodo, Y. Li, D. Kaba, and X. Liu, "Retinal layer segmentation in optical coherence tomography images," *IEEE Access*, vol. 7, pp. 152388–152398, 2019.
- [27] L. Liu, F.-X. Wu, Y.-P. Wang, and J. Wang, "Multi-receptive-field cnn for semantic segmentation of medical images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 3215–3225, 2020.
- [28] M. Arabahmadi, R. Farahbakhsh, and J. Rezazadeh, "Deep learning for smart healthcare—a survey on brain tumor detection from medical imaging," *Sensors*, vol. 22, no. 5, p. 1960, 2022.
- [29] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
- [30] P. Kickingeder, F. Isensee, I. Tursunova, J. Petersen, U. Neuberger, D. Bonekamp, G. Brugnara, M. Schell, T. Kessler, M. Foltyn, et al., "Automated quantitative tumour response assessment of mri in neuro-oncology with artificial neural networks: a multicentre, retrospective study," *The Lancet Oncology*, vol. 20, no. 5, pp. 728–740, 2019.
- [31] P. Kickingeder, U. Neuberger, D. Bonekamp, P. L. Piechotta, M. Götz, A. Wick, M. Sill, A. Kratz, R. T. Shinohara, D. T. Jones, et al., "Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma," *Neuro-oncology*, vol. 20, no. 6, pp. 848–857, 2018.
- [32] P. Kickingeder, M. Götz, J. Muschelli, A. Wick, U. Neuberger, R. T. Shinohara, M. Sill, M. Nowosielski, H.-P. Schlemmer, A. Radbruch, et al., "Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response," *Clinical Cancer Research*, vol. 22, no. 23, pp. 5765–5771, 2016.
- [33] M. Akbari, M. Mohrekehsh, E. Nasr-Esfahani, S. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Polyp segmentation in colonoscopy images using fully convolutional network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 69–72, IEEE, 2018.
- [34] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu, "Adaptive context selection for polyp segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*, pp. 253–262, Springer, 2020.
- [35] J. Bernal, N. Tajikbaksh, F. J. Sanchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, et al., "Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge," *IEEE transactions on medical imaging*, vol. 36, no. 6, pp. 1231–1249, 2017.
- [36] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: a review," *Frontiers in cardiovascular medicine*, vol. 7, p. 25, 2020.
- [37] T. D. Karamitsos, J. M. Francis, S. Myerson, J. B. Selvanayagam, and S. Neubauer, "The role of cardiovascular magnetic resonance imaging in heart failure," *Journal of the American College of Cardiology*, vol. 54, no. 15, pp. 1407–1424, 2009.
- [38] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. von Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanian, G. Lavi, et al., "Automatic model-based segmentation of the heart in ct images," *IEEE transactions on medical imaging*, vol. 27, no. 9, pp. 1189–1201, 2008.
- [39] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac mri," *IEEE transactions on medical imaging*, vol. 29, no. 9, pp. 1612–1625, 2010.
- [40] O. Bernard, A. Lalonde, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, et al., "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.