Hybrid DenseNet-UNet Model for Accurate Liver Segmentation in CT images

Asmaa Sabet Anwar [a,*], Khaled Amin [b], Mohiy M. Hadhoud [b], Mina Ibrahim [c]

[a] *Department of Computer Engineering, Faculty of Engineering, May University, Cairo, 11431,Egypt*
[b] *Department of Information Technology, Faculty of Computers and Information, Menoufia University , Shebin El-Kom, 32511, Egypt*
[c] *Department of Machine Intelligence, Faculty of Artificial Intelligence, Menoufia University, Shebin Elkom, 32511, Egypt*

 * *asmaa.anwar2066@ci.menofia.edu.eg, k.amin@ci.menofia.edu.eg, mouhi.hadhood@ci.menofia.edu.eg, mina.ibrahim@ci.menofia.edu.eg*

Abstract

Liver segmentation from CT images is a critical and foundational task in medical image analysis, playing a pivotal role in accurate diagnosis, treatment planning, and patient management, particularly in liver-related diseases. The ability to precisely delineate the liver is essential for tasks ranging from assessing liver volume to planning surgical procedures and targeting radiation therapy. In this work, an advanced adaptation of the U-Net architecture, integrating DenseNet121 as its backbone is used. This combination leverages DenseNet's dense connections, ensuring efficient gradient flow and feature reuse, enhancing learning capability. Preprocessing steps, including resizing images to 256x256 pixels, histogram equalization, normalization, and binary mask conversion, are applied to ensure data consistency and enhance model performance. Two distinct datasets, 3D-IRCADb-01 and LiTS, are used. The Dice Similarity Coefficient (DSC) is used to evaluate the performance of various models. For dataset 3D-IRCADb-01, remarkable DSC scores are achieved, with the highest reaching 96.5%, and accuracy of 99.5%, indicating the effectiveness of the segmentation models. For dataset LiTS, the models excelled further, achieving DSC scores as high as 98.1% and accuracy of 99.7%. After segmentation, regions of interest (ROIs) are extracted, facilitating subsequent medical analysis and diagnosis. These results demonstrate the robustness and accuracy of the proposed model in liver segmentation tasks.

1. **Introduction**

Liver segmentation from computed tomography (CT) images is a fundamental task in medical image analysis, playing a crucial role in the diagnosis and treatment of liver diseases. Accurate segmentation is essential for various clinical applications, such as liver volume measurement, surgical planning, and the evaluation of tumor response to therapy. However, the complexity of the liver's anatomical structure, the presence of adjacent organs with similar intensities, and the variability in liver shapes and sizes among different patients pose significant challenges to this task [1].

In medical research, liver segmentation aids in developing new treatments, understanding liver pathologies, and creating educational anatomical models. Advances in machine learning and artificial intelligence have led to automated segmentation techniques, enhancing efficiency and accuracy, reducing radiologists' workload, and improving patient outcomes. [2].

Recent advancements in deep learning (DL) have significantly improved the efficiency and accuracy of segmentation approaches for various organs compared to traditional methods. U-Net-based architectures, especially 3D U-Net, are extensively used in medical image segmentation due to their ability to handle sparse volumetric data and retain more features with 3D input. The nnU-Net framework, which utilizes both 2D and

3D U-Net architectures, offers state-of-the-art performance through automatic hyperparameter tuning and data augmentation. Despite its success, 3D U-Net demands significant computational resources due to the inclusion of irrelevant features. To mitigate these issues, deep learning techniques, particularly fully convolutional neural networks (FCNs), have revolutionized medical image analysis, providing effective solutions for liver segmentation. Automated segmentation techniques are being explored to address the time-consuming and variable nature of manual segmentation by skilled operators, offering more consistent and efficient alternatives [3].

In this work, DenseNet121 is adapted as the backbone for the U-Net architecture for liver segmentation. DenseNet is chosen over other architectures, such as ResNet and VGG, for its ability to ensure efficient gradient flow through dense connectivity, where each layer has direct access to the gradients from the preceding layers. This architecture facilitates feature reuse, leading to better feature propagation and a reduction in the number of parameters compared to other architectures. ResNet, while effective at addressing vanishing gradients through skip connections, lacks the same level of feature reuse, making DenseNet more suitable for medical image analysis tasks where capturing intricate details is crucial. Similarly, VGG has a significantly larger number of parameters, making it less computationally efficient and more prone to overfitting compared to DenseNet.

This paper explores liver segmentation in CT scans using deep learning. The contributions in this challenging domain are produced in the following points:

1. Propose an improved U-Net architecture with DenseNet121 as its backbone, specifically designed for liver segmentation in CT scans. While DenseNet improves gradient flow through its dense connections, it also facilitates feature reuse, meaning that each layer can access the outputs of all previous layers. This results in a more efficient feature representation, allowing the network to capture fine-grained details in medical images that may be missed by other architectures.
2. Enhance the decoder architecture by incorporating Batch Normalization after each convolutional layer, improving training speed and stability, and employing Dropout layers with a 25% rate to prevent overfitting. Batch Normalization also stabilizes learning, allowing the model to converge faster compared to traditional architectures that do not normalize intermediate layers. This helps the model perform better across diverse datasets and reduces the risk of overfitting.
3. Optimize the model with specific filter configurations, starting with 256 filters and gradually reducing to 32 filters, effectively balancing detail capture and computational efficiency.
4. Conduct a comprehensive evaluation of the proposed method on two distinct and challenging datasets, demonstrating the model's robustness, high accuracy, and effectiveness in liver segmentation tasks.
5. Demonstrate the practical application of the model by facilitating the extraction of regions of interest (ROIs) post-segmentation, supporting subsequent medical analysis and diagnosis, thereby contributing to improved clinical outcomes and patient care.

The research is systematically organized, beginning with a comprehensive review of the current state-of-the-art in liver segmentation (Section 2). It then details the methodologies and techniques employed in the study (Section 3). Section 4 presents the results, supported by various performance metrics. The paper concludes with a discussion of the findings and future research directions in Section 5, followed by the conclusions in Section 6.

2. Related Work

The task of medical image analysis, especially the segmentation of the liver from CT scans, holds significant importance in healthcare due to its applications in accurate diagnostics and the development of personalized treatment plans. Various computational techniques have been explored over the years to enhance the precision and efficiency of liver segmentation algorithms. Notably, the U-Net architecture and its variants

have received considerable attention for their ability to effectively capture detailed patterns within medical images, contributing to their widespread adoption and success in this domain.

Hoang et al. [4], the performance of three established Convolutional Neural Networks(CNNs), DRIU, FCN-CRF and V-net was evaluated in terms of liver segmentation using diverse medical data sources. Results indicated consistent success, with all CNNs achieving mean Dice scores exceeding 0.90% on typical contrast-enhanced CT liver images. Statistical analysis revealed no significant performance differences among the networks, with DRIU demonstrating efficient processing.

Jiang et al. [5], an innovative Attention Hybrid Connection Network was introduced, combining soft and hard attention mechanisms with long and short skip connections. The research aimed to tackle liver and tumor segmentation in medical imaging using a cascade network comprising liver localization, liver segmentation, and tumor segmentation networks. The training incorporated a combined dice loss function for precise liver bounding box extraction, and fine-tuning was performed using the focal binary cross-entropy loss to enhance tumor detection and minimize false positives. The approach underwent thorough evaluation using datasets from LiTS, 3D IRCADb-01, and a clinical dataset, demonstrating rapid network convergence, precise semantic segmentation, and clinical significance. Notably, the Dice coefficient reached an impressive 95.9% in the LiTS dataset and 94.5% in the 3D-IRCADb-01 dataset.

Ahmed el al. [6] introduces a lightweight convolutional neural network (CNN), named Ga-CNN, designed to efficiently segment the liver from CT images. Ga-CNN consists of three convolutional layers and two fully connected layers, with softmax used to distinguish the liver from the background. The model uses random Gaussian distribution for weight initialization to preserve information. Experiments conducted on three benchmark datasets—MICCAI SLiver'07, 3D-IRCADb-01, and LiTS the model's effectiveness, achieving dice similarity coefficients of 92.9% on 3D-IRCADb-01, 97.3% on LiTS, and 95.0% on SLiver'07. These results indicate that Ga-CNN provides an efficient and accurate solution for liver segmentation with lower computational requirements.

Christ et al. [7], presented an automated technique for the segmentation of the liver and lesions in CT abdomen images. This approach leverages cascaded fully convolutional neural networks (CFCNs) and dense 3D conditional random fields (CRFs) to achieve effective segmentation. The process involves training and cascading two FCNs: the first segments the liver, producing Regions of Interest (ROIs) for the second FCN, which specializes in lesion segmentation within these ROIs. Enhancing segmentation accuracy involves incorporating spatial coherence and appearance information through a dense 3D conditional random field (CRF). The CFCN models undergo training using 2-fold cross-validation on the 3D-IRCADb-01 dataset. The Dice scores for liver segmentation exceed 94.0%.

Jiang et al. [8], developed the Residual Multi-scale Attention U-Net (RMAU-Net), incorporating Res-SE-Block and Multi-scale Attention Block (MAB). The Res-SEBlock mitigates gradient disappearance and enhances feature quality, while the MAB captures rich multi-scale feature information and inter-channel and inter-spatial relationships. Additionally, a hybrid loss function combining focal loss and dice loss improves segmentation accuracy and speeds up convergence. Evaluated on the LiTS dataset, the RMAU-Net achieved a dice score of 95.5% for liver segmentation and 76.1% for liver tumor segmentation.

Fern´andez et al. [9], investigates various deep learning models, including UNet and its variants, for segmenting liver tumors from CT images. The study uses the Liver Segmentation (LiTS) challenge datasets to evaluate the models' performance. The proposed methods achieved significant improvements in segmentation accuracy, with Dice similarity coefficients (DSCs) reaching up to 95.0% for liver segmentation.

Zhe Liu et al. [10], presented a GIU-Net algorithm for liver CT sequence image segmentation built upon an enhanced U-Net model in conjunction with graph-cutting techniques. To begin, initiate the segmentation process within a liver CT sequence sourced from the LiTS dataset, utilizing the improved U-Net architecture. This initial step provides a probability distribution map characterizing the liver regions. Subsequently, the algorithm identifies the starting slice for sequence segmentation and the approach integrates contextual information extracted from the liver sequence images and the liver probability distribution map to formulate a

graph cut energy function. The final segmentation is attained by minimizing this energy function, resulting in a Dice coefficient of 95.1%.

Here the state-of-the-art liver segmentation models is reviewed , each varying in terms of model complexity and computational efficiency. Models like Ga-CNN [6] offer lower complexity and high efficiency, making them suitable for resource-constrained environments, whereas architectures such as cascaded FCNs with 3D CRFs [7] and hybrid attention networks [5, 8] provide high accuracy at the cost of increased computational demands. The proposed U-Net with DenseNet121 backbone strikes a balance by leveraging dense connections for efficient feature reuse and gradient flow, resulting in a model with moderate complexity but high segmentation accuracy. A comparison of the related works is summarized in Table 1, highlighting key performance and complexity differences across the approaches.

Table 1: Overview of Liver Segmentation Techniques in related work

Study	Method	Dataset	Dice score	Model complexity	Efficiency
Hoang et al. [4]	DRIU, FCN-CRF, V-net	Contrast-enhanced CT	90.0%	Moderate	Efficient
Jiang et al. [5]	Attention Hybrid Connection Network	LiTS, 3D-IRCADb-01	95.9%, 94.5%	High	Fast convergence
Ahmed et al. [6]	Ga-CNN (Lightweight CNN)	MICCAI SLiver'07 LiTS	92.9%, 97.3%, 95.0%	Low	Highly efficient
Christ et al. [7]	CFCN + 3D CRF	3D-IRCADb-01	94.0%	High	Computationally intensive
Jiang et al. [8]	RMAU-Net (Residual Multi-scale Attention U-Net)	LiTS	95.5%	High	Moderate
Fernández et al. [9]	U-Net variants	LiTS	95%	Moderate	Standard U-Net efficiency
Zhe Liu et al. [10]	GIU-Net + Graph-cut	LiTS	95.1%	High	Graph-cut optimization

3. Methodology

This section provides an overview of the datasets utilized, details the preprocessing steps, outlines the segmentation methodology, and describes the performance metrics used for evaluation. The overall phases of the model are illustrated in Figure 1.

3.1 Dataset

The **3D-IRCADb-01** dataset [11] comprises twenty 3D CT scans from patients diagnosed with liver cancer. Each CT image contains liver density values ranging from 40 to 135 Hounsfield units (HU) and has a spatial resolution of 512 x 512 pixels per slice. The voxel spacing, which defines the physical distance between slices, varies depending on the patient and scan but typically ranges from 0.5 to 1.6 mm. The dataset is organized into discrete DICOM (Digital Imaging and Communications in Medicine) segments, where each segment includes not only the raw CT images but also corresponding segmentation masks that highlight different anatomical structures, including the liver and tumors. These segmentation masks are provided in the DICOM format, ensuring compatibility with standard medical imaging tools. Figure 2 showcases sample images from this dataset after converting the original DICOM files into PNG images, along with their corresponding segmentation masks.

The **LiTS** (Liver Tumor Segmentation) dataset [12] serves as a benchmark dataset for liver tumor segmentation, containing 200 CT scans. The CT scans in the LiTS dataset are provided in the NII

(Neuroimaging Informatics Technology Initiative) file format, which supports 3D image data. Each scan covers the abdominal region and has voxel resolutions that typically range from 0.6 to 1.0 mm in the x and y dimensions (in-plane resolution), with slice thickness varying between 1 and 5 mm. For model training and testing, the 3D CT scans are preprocessed into sequences of 2D image slices, stored as PNG images for ease of processing and visualization. These preprocessed slices form both the training and testing datasets. Figure 3 displays representative images segmentations from the LiTS dataset.
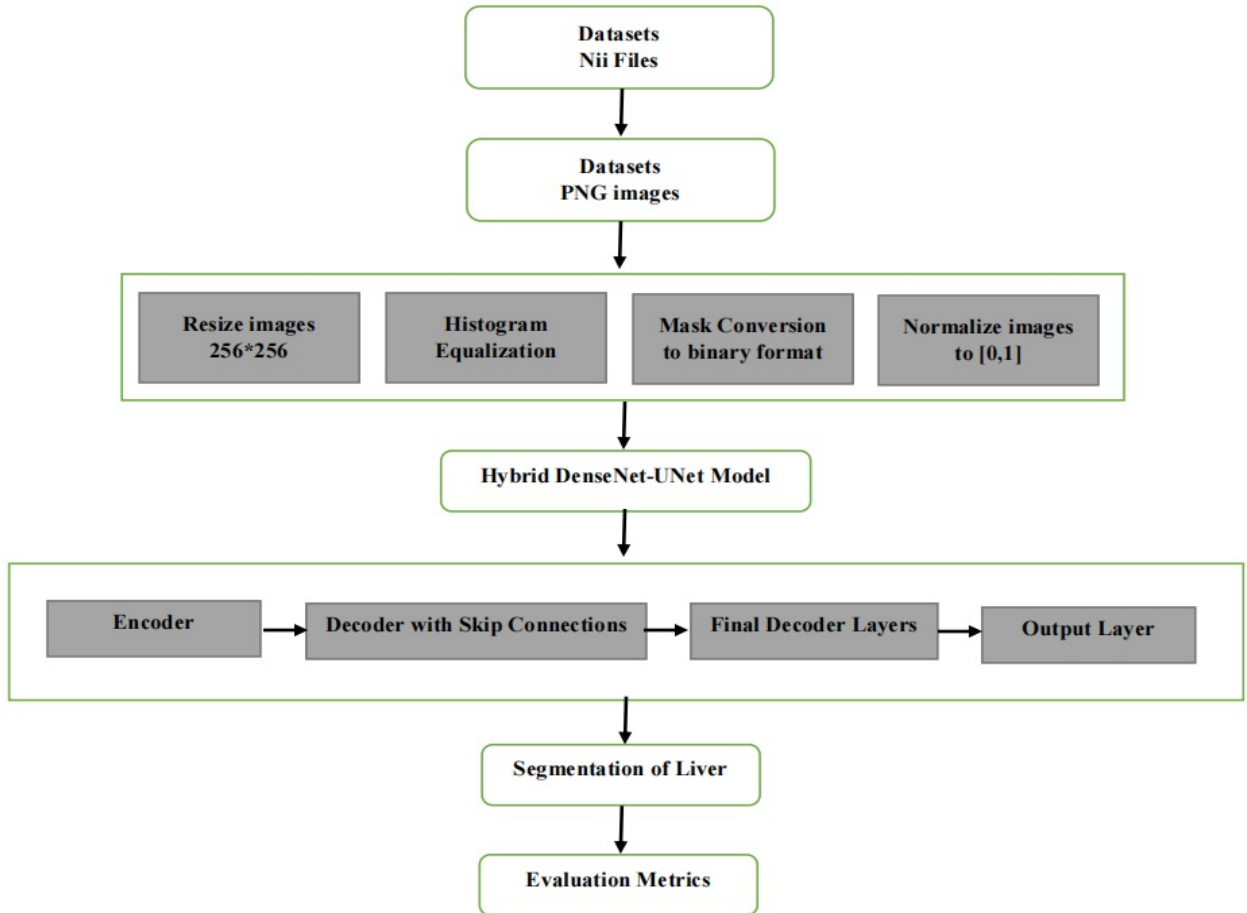

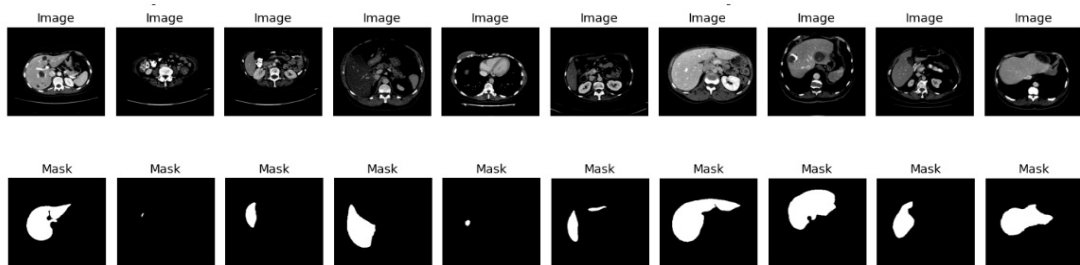Fig. 1. The workflow of proposed liver segmentation model

Fig. 2. Samples of 3D-IRCADb-01 Dataset, the first set of images represents CT image, followed by the corresponding liver masks.
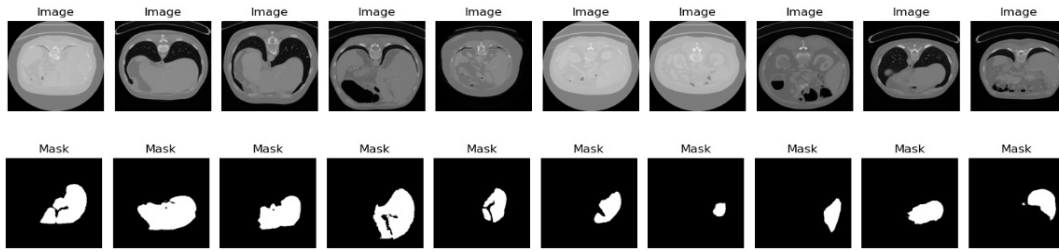
Fig. 3. Samples of LiTS Dataset, the first set of images represents CT image, followed by corresponding liver masks**.**

3.2 Pre-processing phase

In the preprocessing phase of an image analysis pipeline, a systematic series of steps is followed to prepare the dataset for subsequent analysis and model training. These steps are encapsulated in Algorithm 1, referred to as the "Image Preprocessing Algorithm". Firstly, the images from the dataset are acquired. To guarantee uniformity and consistency, all images undergo resizing to a standardized 256x256 pixel size. This specific size was chosen to strike a balance between increasing the size of the dataset, which was memory-intensive and limited by available resources, and maintaining image quality. Smaller sizes are considered, but they significantly compromised the quality of the images, which in turn had a detrimental influence on the efficacy of the outcomes, making 256x256 pixels the optimal compromise. Only the images accompanied by masks are carefully selected, as these masks provide essential ground truth information. This selection process is crucial to avoid class imbalance during training. Specifically, images without accompanying masks are excluded from the dataset because their absence would significantly skew the distribution of classes, potentially leading to biased training results and affecting the model's ability to accurately learn and generalize from the available data.

Subsequently, the contrast of the images is enhanced through histogram equalization, making subtle details more discernible. This technique is employed to address the issue of non-uniform illumination and varying contrast levels within the dataset, ensuring that important anatomical features and pathological regions are consistently and effectively highlighted across all images. Histogram equalization plays a critical role in standardizing image quality and enhancing the overall interpretability of the dataset. Pixel values are then normalized to the [0, 1] range to ensure consistent data for analysis and model training. This normalization step is crucial to mitigate the impact of varying intensity scales in the original images, thereby allowing different images to be directly comparable and preventing certain images from disproportionately influencing the training process due to their original intensity ranges. No specific noise reduction steps were applied because the dataset's quality and the preprocessing steps already in place (histogram equalization and normalization) were sufficient to handle any minor noise or artifacts.

Normalization also helps in stabilizing model convergence during training and promoting the effectiveness of subsequent image analysis and machine learning tasks. Lastly, grayscale masks are converted into a binary format, simplifying processing. This conversion is performed to streamline subsequent processing steps and facilitate the implementation of image segmentation algorithms. Binary masks, where each pixel is either classified as a relevant region or not, simplify the identification of specific anatomical structures or pathological areas in the images. This simplification reduces computational complexity, enhances the efficiency of image analysis, and improves the overall accuracy of segmentation tasks. This comprehensive preprocessing procedure ensures that the data is appropriately prepared for subsequent stages of the analysis and modeling efforts.

Algorithm 1 Image Preprocessing Algorithm

Input: Images from the dataset
Output: Preprocessed images with masks
procedure ImagePreprocessing
    for each image in the dataset **do**
        Image Acquisition
        Resizing to 256x256 pixels
        if image has a mask **then**
            Histogram Equalization
            Normalization to [0, 1] range
            Mask Conversion to binary format
        else
            Exclude image (no mask available)
        end if
    end for
end procedure

3.3 Segmentation Approach

3.3.1   Hybrid DenseNet-UNet Model

The proposed model for liver segmentation represents an advanced adaptation of the U-Net architecture, integrating the powerful DenseNet121 [13] as its backbone which summarized the steps in detailes in Table 2. This combination leverages DenseNet's dense connections, which ensure efficient gradient flow and feature reuse, enhancing the overall learning capability of the model. Key enhancements in the proposed model include the strategic use of Batch Normalization, Dropout layers, and specific configurations of filters. This model is structured into three distinct steps: Encoder, Decoder, and Output Layer, as explained below.

- **Encoder**

The encoder leverages DenseNet's dense connections, which ensure efficient gradient flow and feature reuse, enhancing the overall learning capability of the model. By incorporating DenseNet121 up to the conv4_block24_concat layer, the model benefits from DenseNet's densely connected blocks, which enhance feature propagation and utilization through direct connections between layers. This deep feature extraction captures complex patterns and textures in the input images, crucial for accurate segmentation. The pre-trained weights ensure that the model starts with a strong foundation of visual knowledge, significantly improving convergence speed and performance on the liver segmentation task.

- **Decoder**

The decoder component systematically upsamples the feature maps to the original image dimensions while integrating encoder features through skip connections. Each Conv2DTranspose layer performs upsampling, effectively increasing the spatial resolution of the feature maps. The concatenation with corresponding encoder layers (such as conv3 block12 concat and conv2 block6 concat) reinstates lost spatial information, enhancing the detail and accuracy of the segmentation. Subsequent Conv2D layers with ReLU activation and batch normalization refine these features, promoting stability and efficient training. Batch Normalization is employed after each convolutional layer and skip connections, standardizing the inputs to a layer, improving training speed and stability. This technique mitigates issues related to internal covariate shift, allowing for higher learning rates and faster convergence. Dropout layers, particularly after Batch Normalization layers, serve as a regularization technique to prevent overfitting. Dropout works by randomly deactivating a fraction of neurons during training, forcing the model to learn more robust features that generalize better to unseen data. In the proposed model, a dropout rate of 25% is used, which was empirically determined through experimentation to provide the best balance between regularization and maintaining model performance. Comparative trials with lower and higher dropout rates

(10% and 50%) demonstrated that 25% offered the optimal trade-off, preventing overfitting while preserving learning capacity. The decoder layers start with 256 filters, gradually reducing to 32 filters as the spatial resolution increases. This gradual reduction aligns with the increasing spatial dimensions, ensuring that the model can capture fine details necessary for accurate segmentation while managing computational complexity. The filter size of (3,3) is chosen for its effectiveness in capturing spatial features at various scales.

- **Output Layer**

The output layer comprises a Conv2D layer with a single filter and a 1x1 filter size, followed by a sigmoid activation function. This layer translates the refined feature maps from the decoder into a final segmentation map, where each pixel's value represents the probability of it being part of the liver. The 1x1 convolution ensures that spatial dimensions are maintained while reducing the depth to one channel, suitable for binary segmentation tasks. The sigmoid activation function normalizes the output to a range between 0 and 1, facilitating a probabilistic interpretation that is essential for segmentation tasks, as it clearly distinguishes between liver and non-liver regions in the image.

In summary, the proposed model enhances the original U-Net by incorporating DenseNet121 for its encoder, and introduces Batch Normalization and Dropout layers to improve training stability and prevent overfitting. The carefully chosen filter sizes and number of filters at each layer ensure a balanced approach to capturing detailed features and managing computational efficiency, making the model robust and effective for liver segmentation tasks.

Table 2: Architecture Details of Modified Unet using DenseNet model

Component	Details
Encoder(DenseNet121)	Pre-trained with weights from ImageNet, up to layer conv4_block24_concat
Decoder	Conv2DTranspose(Num_filter=256, filter_size=3x3, stride=2x2, padding='same') Concatenate with encoder's skip connection (conv3_block12_concat) Conv2D(Num_filter=256, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization Dropout (rate=0.25) Conv2D(Num_filter=256, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization
	Conv2DTranspose(Num_filter=128, filter_size=3x3, stride=2x2, padding='same') Concatenate with encoder's skip connection (conv2_block6_concat) Conv2D(Num_filter=128, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization Dropout (rate=0.25) Conv2D(Num_filter=128, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization
	Conv2DTranspose(Num_filter=64, filter_size=3x3, stride=2x2, padding='same') Conv2D(Num_filter=64, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization Dropout(rate=0.25) Conv2D(Num_filter=64, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization
	Conv2DTranspose(Num_filter=32, filter_size=3x3, stride=2x2, padding='same') Conv2D(Num_filter=32, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization Dropout(rate=0.25) Conv2D(Num_filter=32, filter_size=3x3, activation=ReLU, padding='same') BatchNormalization
Output	Conv2D (Num_filter=1, filter_size=1x1, activation=sigmoid)

3.4 Performance Metrics

To assess the effectiveness of the segmentation technique, the comparison between the binary mask derived from the segmentation output and the reference ground truth mask is quantified. Various performance metrics, including but not limited to DSC, JSC, and accuracy, are employed for this evaluation.

Several common metrics are utilized to evaluate the performance of segmentation models:

- **True Positives (TP):** The number of correctly predicted foreground pixels or regions.
- **True Negatives (TN):** The number of correctly predicted background pixels or regions.
- **False Positives (FP):** The number of background pixels or regions incorrectly predicted as foreground.
- **False Negatives (FN):** The number of foreground pixels or regions incorrectly predicted as background.

Dice Similarity Coefficient (DSC): Also recognized as the F1-score for binary segmentation, DSC quantifies the spatial overlap between the predicted and ground truth segmentations.

$$DSC = 2 * TP / (2 * TP + FP + FN) \tag{1}$$

Jaccard Similarity Coefficient (JSC): Also known as the Intersection over Union (IoU), JSC measures the ratio of the intersection to the union of the predicted and ground truth regions.

$$JSC = TP / (TP + FP + FN) \tag{2}$$

Accuracy: Measures the overall correctness of the segmentation.

$$ACC = (TP + TN) / (TP + TN + FP + FN \tag{3}$$

Recall (Sensitivity): Assesses the model's ability to correctly identify all positive instances.

$$TP / (TP + FN) \tag{4}$$

Surface Volume Difference (SVD): Measures the difference in volume between the predicted and ground truth surfaces.

$$SVD = |Vp - Vg| / Vg \tag{5}$$

Where Vp is the volume of the predicted surface, and Vg is the volume of the ground truth surface.

Generally, higher values for DSC, JSC, Accuracy, Precision, Recall, and Specificity, and lower values for SVD indicate better segmentation performance.

4. Experiments and Results

4.1. Experimental Set-Up

In the experimental setup, a laptop with an Intel Core i7 processor, 16.0 GB of RAM, and Windows 11 Pro is utilized. For model training, the cloud-based Kaggle platform is employed, leveraging dual Nvidia 2 x T4 GPUs. This GPU configuration provided accelerated training and inference, facilitating efficient development and evaluation of the machine learning models for image analysis tasks. The chosen hardware and GPU setup were sufficient to handle the computational demands of the DenseNet-UNet architecture, ensuring efficient convergence without excessive resource usage.

The model architecture, as summarized in Table 3, has been thoughtfully designed and meticulously optimized for two distinct medical image datasets: 3D-IRCADb- 01 and LiTS. For both the 3D-IRCADb-01 and LiTS datasets, the entire dataset is partitioned into a training set and a test set, with 0.8 allocated for training and the remaining 0.2 for testing. This split is chosen to ensure that the model could be effectively trained and rigorously evaluated within the available memory constraints and resource limitations.

Table 3: Hybrid DenseNet-UNet model Parameter

Dataset	No images	Train images	Test images	No Epochs	Batch size	Shuffle	Optimizer	Loss	Early Stopping
LiTS	6500	5200	1300	50	8	True	Adam	B.C.E	Val Loss
3D-IRCADb-01	2823	2258	565	50	8	True	Adam	B.C.E	Val Loss

where B.C.S is binary cross-entropy

For the 3D-IRCADb-01 dataset, which contains a total of 2,823 images, training is extended over 50 epochs (as depicted in Figure 4), as this epoch count consistently yielded the most reliable results. Training for 50 epochs is chosen to ensure the model effectively captured the nuances of the images and reached the desired level of performance. By the 50th epoch, the training and validation loss and accuracy had stabilized, indicating that the model had converged and learned the dataset's characteristics effectively.

Similarly, for the LiTS dataset, given the memory limitations, a total of 6,500 images are randomly selected. Training for the LiTS dataset is conducted over 50 epochs (as shown in Figure 5), a choice made to strike the optimal balance between model convergence and stability. The decision to use 50 epochs is based on empirical observations and experimentation, which indicated that the model effectively learned the dataset's characteristics within this timeframe.
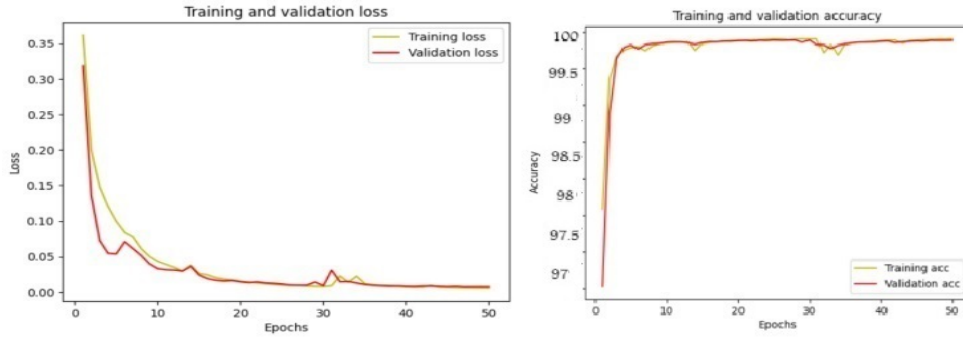

Fig. 4. (a) Training and validation loss ,(b) Training and validation accuracy via Epochs in 3DIRCADb-01 dataset
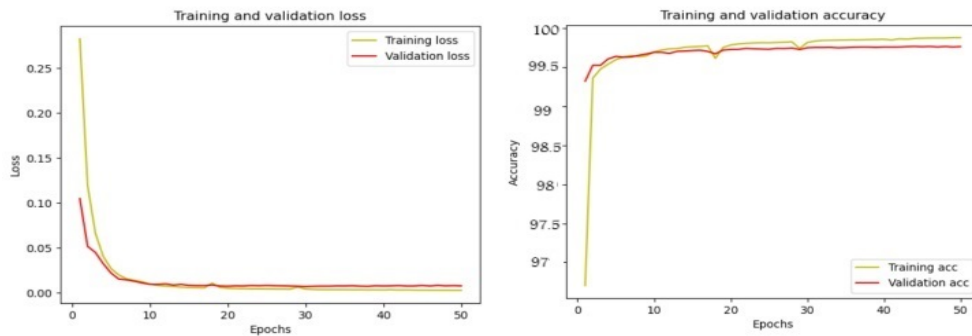

Fig. 5. (a) Training and validation loss ,(b) Training and validation accuracy via Epochs in LiTS dataset

Several key hyperparameters were carefully chosen to optimize the model's performance. The Adam optimizer was selected for its efficiency in handling sparse gradients, with binary cross-entropy (BCE) as the loss function, which is suitable for segmentation tasks. The initial learning rate was set to 1e-4, which was empirically determined to provide stable training without causing rapid oscillations in the loss function. A learning rate scheduler was employed to reduce the learning rate by a factor of 0.1 if no improvement in validation loss was observed for a predefined number of epochs.

The batch size was set to 8 across all training configurations to manage memory constraints while ensuring the model could generalize effectively. Early stopping was used to monitor validation loss, with training halting if no improvements were observed for 10 consecutive epochs, preventing overfitting. The model was trained for a maximum of 50 epochs to ensure convergence, with early stopping usually occurring before this limit was reached.

To prevent overfitting, a dropout rate of 0.25 was applied during training, alongside batch normalization after each convolutional layer. Shuffling was enabled during training to introduce randomness in the sample order, improving generalization. The model used ReLU activation functions for all intermediate layers.

4.2. Preprocessing Results

The preprocessing phase aims to ensure data consistency and enhance image quality for analysis. It begins by resizing all images to a uniform 256x256 pixel size, excluding those without masks. Histogram equalization is then applied to improve contrast, followed by normalization to scale pixel values to the [0, 1] range. Additionally, grayscale masks are converted into binary format to simplify processing. Figure 6 illustrates the results of these preprocessing steps for the 3D-IRCADb-01 dataset, while Figure 7 showcases the outcomes for the LiTS dataset. These processed images and masks serve as the foundation for the subsequent analysis and model development, ensuring high-quality and standardized data.
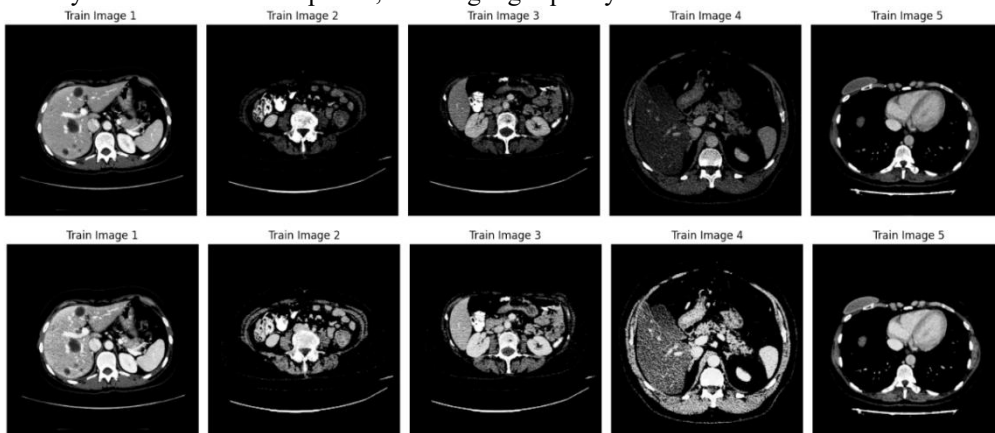

Fig. 6. The 3D-IRCADb-01dataset images before and after the preprocessing phase. The first set of images represents the original data, followed by the corresponding images after undergoing preprocessing.
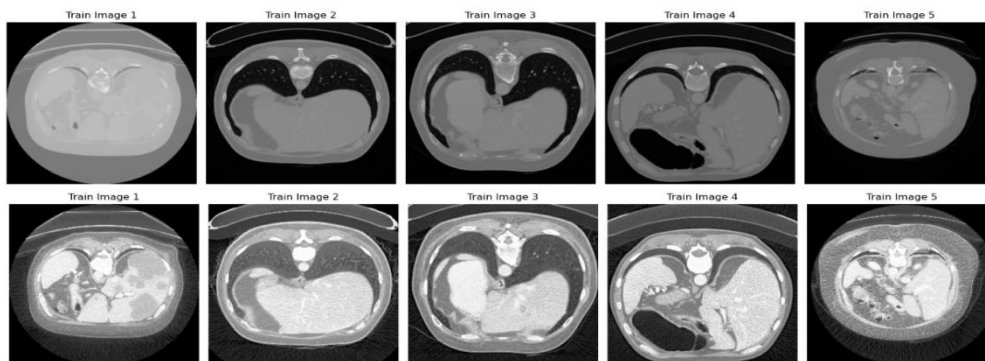

Fig. 7. The LiTS dataset images before and after the preprocessing phase. The first set of images represents the original data, followed by the corresponding images after undergoing preprocessing.

4.3. Segmentation Results

The proposed model for liver segmentation has been evaluated on two distinct and challenging datasets, LiTS and 3D-IRCADb-01, with the results presented in Table 4. The performance metrics clearly demonstrate the robustness and effectiveness of the model across different evaluation criteria.

For the LiTS dataset, the model achieved an exceptional accuracy of 99.7%, indicating a high level of correctness in the segmentation process. The Dice Similarity Coefficient (DSC) further supports this, with the model attaining a score of 98.1%, reflecting a substantial overlap between the predicted segmentation and the ground truth. This high DSC value is complemented by a Mean IoU of 96.5%, showcasing the model's capability to accurately segment liver regions even in complex scenarios. The recall rate of 98.4% highlights the model's proficiency in identifying true positive liver regions, minimizing the instances of missed segments. Additionally, the Surface Volume Difference (SVD) is impressively low at 0.003, underscoring the precision of the volume estimations compared to the ground truth.

Similarly, for the 3D-IRCADb-01 dataset, the model maintained a high level of performance with an accuracy of 99.5%. The DSC value of 96.5% indicates strong agreement between the predicted and actual liver segments, while the Mean IoU of 93.0% confirms the model's effectiveness in achieving accurate segmentation. The recall rate matches the DSC at 96.5%, further illustrating the model's reliability in detecting liver regions. The SVD value of 0.007, though slightly higher than that for LiTS, still represents a minimal deviation, demonstrating the model's consistency in volume estimation across different datasets.

Overall, the results in Table 4 validate the proposed model's superior performance in liver segmentation tasks, providing a reliable tool for clinical applications and further medical analysis. The high accuracy, DSC, and recall rates, coupled with low SVD values, indicate that the model not only performs well in terms of segmentation precision but also ensures minimal volumetric errors, making it highly suitable for practical deployment in medical imaging scenarios.

Table 4: Performance Metrics for Liver Segmentation

Metrics	Liver Segmentation LiTS	Liver Segmentation 3D-IRCADb-01
Accuracy (%)	99.7	99.5
Dice Coefficient (%) (DSC)	98.1	96.5
Mean IoU (%)	96.5	93.0
Recall (%)	98.4	96.5
SVD	0.003	0.007

In Figure 8 and Figure 9 the comprehensive results of the model's performance are presented on the 3D-IRCADb-01 dataset and LiTS respectively. The first row of the figure showcases the output of the preprocessing phase, where images are uniformly resized, contrast is enhanced through histogram equalization, and grayscale masks are converted into binary format for simplicity. In the second row, a visual representation of the ground truth for the images is provided. The third row reveals the predictions generated by the model, demonstrating its ability to classify and interpret the images effectively. Lastly, the fourth row exhibits the overlap of the model's predictions onto the original images, offering a visual insight into the alignment of the model's output with the actual data. These four rows collectively offer a comprehensive overview of the model's performance and its alignment with the ground truth.
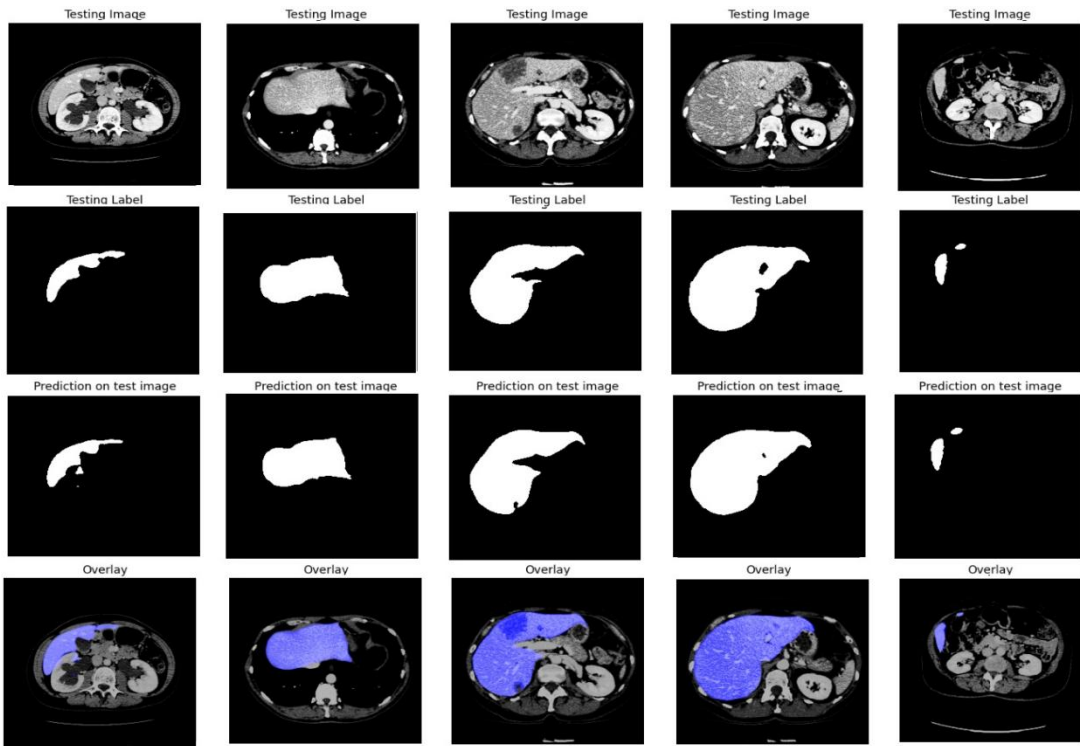
Fig. 8. Segmentation results on the 3D-IRCADb-01 dataset using the Hybrid DenseNet-UNet model. The images start with the original CT scans, followed by the ground truth masks. Then, the predicted segmentation masks are presented, and finally, the overlap between the predicted masks and the CT images is shown.
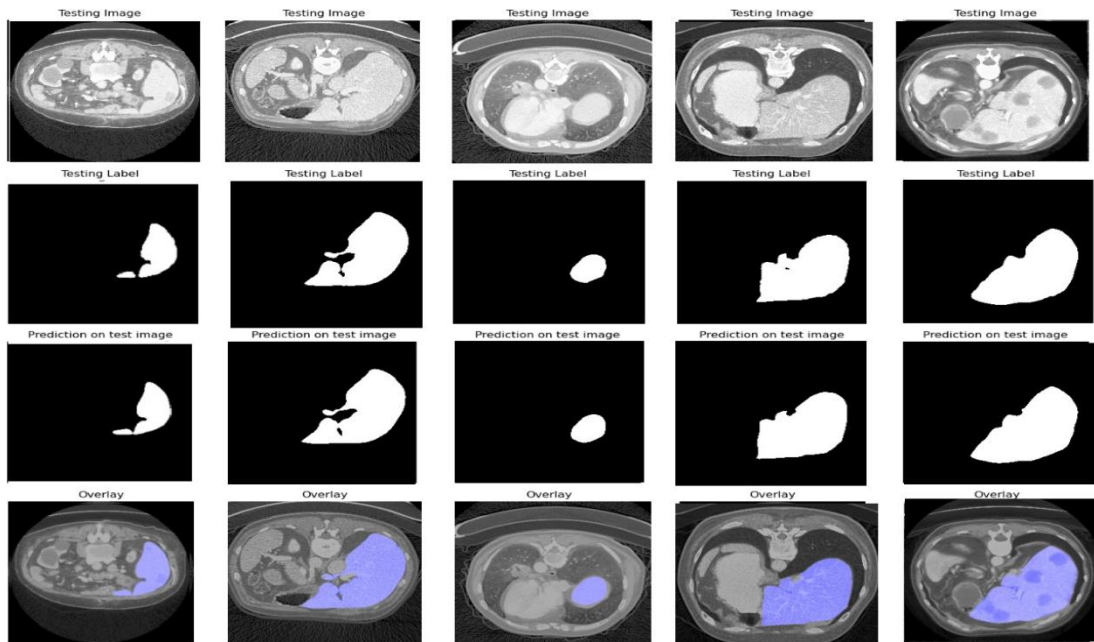

Fig. 9. The segmentation results on the LiTS dataset using the Hybrid DenseNet-UNet model. The first set of images shows the original CT scans, followed by the ground truth segmentation masks. Next, the predicted segmentation masks are displayed, and finally, the overlap between the predicted masks and the original CT scans is visualized in the last row.

4.4. *Comparative Analysis*

The segmentation performance results are presented in two distinct tables, providing a comprehensive comparison of various models and their respective DSC. In Table 5 and Table 6, the performance of established models from previous studies is observed, including U-Net, Improved U-Net and Graph Cut, GANs and the Mask R CNN, CDNN-II, Cascaded ResNet (with Multi-scale Fusion), Enhanced M-RCNN, H-DenseUNet, Attention U-Net++, and AHCNET. These models exhibit a wide range of DSC values, demonstrating their varying degrees of effectiveness in the task. Notably, the models introduced in this paper, leveraging DenseNet, emerged as top performers, consistently achieving the highest DSC values and demonstrating superior segmentation capabilities. This underscores the significant advancements and contributions made by the models presented in this study, setting a new benchmark for medical image segmentation

In Table 5 and 6, The proposed model is introduced, which leveraged Hybrid DenseNet-UNet Model. This model demonstrates a remarkable DSC score of 98.1% for LiTS dataset and 96.5% for 3D-IRCADb-01 dataset, indicating its proficiency in accurately segmenting medical images. By comparing the proposed model's performance with these established benchmarks, Its superior segmentation capabilities are highlighted, offering promising results for medical image analysis tasks. These two tables collectively offer an informative overview of the segmentation performance landscape, with the proposed model standing out as a promising advancement in the field.

Table 5: Segmentation Performance Comparison based on 3D-IRCADb-01 dataset

Paper	Model	DSC
Jiang in [5]	AHCNET	94.5%
Christ [7]	CFCN	94.3%
Ümit Budak in [15]	EDCNN	95.2%
Rafiei in [16]	3D-2D-FCN	93.5%
Ayalew in [17]	Modified U-Net	96.1%
Alirr in [18]	FCN and level-set	95.2%
Proposed	Hybrid DenseNet-UNet Model	96.5%

Table 6: Segmentation Performance Comparison based on LiTS dataset

Paper	Model	DSC
Ayalew in [19]	U-Net	96.1%
Jiang in [5]	AHCNET	95.9%
Balasubramanian in [20]	Enhanced M-RCNN	95.7%
Wei in [21]	GANs and the Mask R CNN	95.3%
Zhe Liu in [10]	Improved U-Net and Graph Cut	95.1%
Bi Lei in [22]	Cascaded ResNet (w Multi-scale Fusion)	95.5%
Li Xiaomeng in [23]	H-DenseUNet	96.5%
Yuan in [24]	CDNN-II	0.967
Elnakib in [25]	CNN with Majority voting	94.2%
Proposed	Hybrid DenseNet-UNet Model	98.1%

5. **Discussion**

The proposed model, which integrates DenseNet121 into the U-Net architecture, demonstrates significant advancements in liver segmentation from CT images. Leveraging DenseNet's dense connections in the encoder component enhances gradient flow and feature reuse, capturing complex patterns essential for accurate segmentation. The use of pre-trained weights accelerates convergence and improves performance.

The decoder's design, incorporating skip connections, reinstates lost spatial information and refines features through Batch Normalization and Dropout layers. This approach enhances training stability, mitigates internal covariate shift, and prevents overfitting. A dropout rate of 25% strikes a balance between

regularization and learning capacity. The progressive reduction of filters from 256 to 32, with a (3,3) filter size, ensures detailed feature capture and computational efficiency.

The output layer, featuring a Conv2D layer with a 1x1 filter and sigmoid activation, translates feature maps into a segmentation map with clear distinctions between liver and non-liver regions.

Evaluations on the 3D-IRCADb-01 and LiTS datasets demonstrate the model's robustness and accuracy. Achieving DSC scores of up to 96.5% and 98.1%, and accuracy rates of 99.5% and 99.7%, respectively, underscores the model's effectiveness in liver segmentation tasks. These high-performance metrics facilitate the extraction of regions of interest (ROIs) for subsequent medical analysis and diagnosis.

In summary, the integration of DenseNet121, along with strategic enhancements like Batch Normalization and Dropout layers, significantly improves the U-Net architecture's capability in liver segmentation tasks. The proposed model's robustness and accuracy establishes it as a valuable tool in medical image analysis.

6. Conclusion

In this paper, a sophisticated and highly effective approach is presented to liver segmentation from CT images by integrating DenseNet121 into the U-Net architecture. The proposed model capitalizes on DenseNet's dense connections, ensuring efficient gradient flow and feature reuse, which significantly enhances the model's learning capability and performance. The strategic incorporation of Batch Normalization and Dropout layers within the decoder has proven to improve training stability and prevent overfitting, while the meticulously designed filter configurations ensure a balanced approach to capturing detailed features and managing computational complexity. The comprehensive evaluations on the LiTS and 3D-IRCADb-01 datasets have demonstrated the model's robustness and superior performance, achieving Dice Similarity Coefficients of 98.1% and 96.5%, and accuracies of 99.7% and 99.5%, respectively. These results underscore the effectiveness of the proposed approach in delivering precise and reliable liver segmentation. Furthermore, the ability of the model to facilitate the extraction of regions of interest (ROIs) post-segmentation holds significant promise for subsequent medical analysis and diagnosis, thereby contributing to improved clinical outcomes and patient care. Overall, this work sets a new benchmark in medical image segmentation, offering a valuable tool for healthcare professionals and paving the way for future advancements in the field.

7. References

[1] J. Li *et al.*, "Literature survey on deep learning methods for liver segmentation from CT images: A comprehensive review," *Multimedia Tools and Applications*, vol. 83, pp. 14539–14565, 2024.

[2] R. Loomba, S. L. Friedman, and G. I. Shulman, "Mechanisms and disease consequences of nonalcoholic fatty liver disease," *Cell*, vol. 184, pp. 2537–2564, 2021.

[3] A. Gupta, G. Cazoulat, M. Al Taie et al., "Fully automated deep learning based auto-contouring of liver segments and spleen on contrast-enhanced CT images," Scientific Reports, vol. 14, article no. 4678, 2024.

[4] H. S. Hoang, C. Phuong Pham, D. Franklin, T. van Walsum, and M. Ha Luu, "An evaluation of CNN-based liver segmentation methods using multi-types of CT abdominal images from multiple medical centers," in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 20–25, 2019.

[5] H. Jiang, T. Shi, Z. Bai, and L. Huang, "AHCNet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes," IEEE Access, vol. 7, pp. 24898–24909, 2019.

[6] M. Ahmad et al., "A lightweight convolutional neural network model for liver segmentation in medical diagnosis," Computational Intelligence and Neuroscience, article ID 7954333, 2022.

[7] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty et al., "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016), Lecture Notes in Computer Science, vol. 9901, Springer, Cham, pp. 415–423, 2016.

[8] L. Jiang et al., "RMAU-Net: Residual multi-scale attention U-Net for liver and tumor segmentation in CT images," Computers in Biology and Medicine, vol. 158, article no. 106838, 2023.

[9] J. G. Fernandez, V. Fortunati, and S. Mehrkanoon, "Exploring automatic liver tumor segmentation using deep learning," in Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1–8, 2021.

[10] Z. Liu et al., "Liver CT sequence segmentation based on improved U-Net and graph cut," Expert Systems with Applications, vol. 126, pp. 54–63, 2019.

[11] L. Soler et al., "3D image reconstruction for comparison of algorithm database: A patient-specific anatomical and medical image database," Tech. Rep., vol. 1, no. 1, IRCAD, Strasbourg, France, 2010.

[12] P. Bilic et al., "The liver tumor segmentation benchmark (LiTS)," Medical Image Analysis, vol. 84, article no. 102680, 2023.

[13] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, Los Alamitos, CA, USA, 2017. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.243.

[14] H. Rahman et al., "A deep learning approach for liver and tumor segmentation in CT images using ResUNet," Bioengineering (Basel), vol. 9, p. 368, 2022.

[15] U. Budak, Y. Guo, E. Tanyildizi, and A. Şengür, "Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation," Medical Hypotheses, vol. 134, p. 109431, 2020.

[16] S. Rafiei et al., "Liver segmentation in CT images using three dimensional to two dimensional fully convolutional network," in Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2067–2071, 2018.

[17] Y. A. Ayalew, K. A. Fante, and M. Mohammed, "Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method," BMC Biomedical Engineering, vol. 3, p. 4, 2021.

[18] O. I. Alirr, "Deep learning and level set approach for liver and tumor segmentation from CT scans," Journal of Applied Clinical Medical Physics, vol. 21, pp. 200–209, 2020.

[19] M. Rela, N. R. Suryakari, and P. R. Reddy, "Liver tumor segmentation and classification: A systematic review," in Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1–6, 2020.

[20] P. Balasubramanian, W.-C. Lai, H. Gan, K. C., and J. Selvaraj, "APESTNet with mask R-CNN for liver tumor segmentation and classification," Cancers, vol. 15, p. 330, 2023.

[21] X. Wei et al., "Automatic liver segmentation in CT images with enhanced GAN and mask region-based CNN architectures," Biomed Res Int, vol. 2021, article ID 9956983, 2021.

[22] L. Bi, J. Kim, A. Kumar, and D. D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," ArXiv, abs/1704.02703, 2017.

[23] X. Li et al., "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," IEEE Transactions on Medical Imaging, vol. 37, pp. 2663–2674, 2018.

[24] Y. Yuan, "Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation," ArXiv, abs/1710.04540, 2017.

[25] A. Elnakib, H. E.-D. Moustafa, and N. Elmenabawy, "Automated deep system for joint liver and tumor segmentation using majority voting (Dept.E)," MEJ. Mansoura Engineering Journal, vol. 45, pp. 30–36, 2020.

101

النموذج الهجين DenseNet-UNet للتجزئة الدقيقة للكبد في صور الأشعة المقطعية

اسماء ثابت انور*¹ ,خالد محمد امين ² , محي محمد هدهود²,مينا ابراهيم³

١. قسم هندسه الحاسب , كليه الهندسه ,جامعه مايو , القاهره, ١١٤٣١ , مصر

٢. قسم تكنولوجيا المعلومات, كليه الحاسبات والمعلومات, جامعه المنوفيه, شبين الكوم,٣٢٥١١, مصر

٣. قسم ذكاء الاله, كليه الذكاء الاصطناعي, جامعه المنوفيه, شبين الكوم, ٣٢٥١١, مصر

* *asmaa.anwar2066@ci.menofia.edu.eg, k.amin@ci.menofia.edu.eg, mouhi.hadhood@ci.menofia.edu.eg,*
mina.ibrahim@ci.menofia.edu.eg

الملخص

تُعتبر تجزئة الكبد من صور الأشعة المقطعية مهمه حاسمه واساسيه في تحليل الصور الطبية، حيث تلعب دورًا محوريًا في التشخيص الدقيق، تخطيط العلاج، وإدارة المرضى، لا سيما في الأمراض المتعلقة بالكبد .القدرة على تحديد الكبد بدقة ضرورية في المهام التي تتراوح بين تقييم حجم الكبد وتخطيط الإجراءات الجراحية واستهداف العلاج الإشعاعي .في هذا العمل، يتم استخدام تعديل متقدم لبنية U-Net، مع دمج DenseNet121 كعمودها الفقري .يستفيد هذا الدمج من الروابط الكثيفة في DenseNet مما يضمن تدفقاً فعالاً للتدرجات وإعادة استخدام الميزات، مما يعزز القدرة على التعلم .يتم تطبيق خطوات معالجة مسبقة، بما في ذلك تغيير حجم الصور إلى 256x256 بكسل، وتعديل التباين التاريخي ، والتطبيع، وتحويل القناع إلى ثنائي، لضمان اتساق البيانات ولتعزيز أداء النموذج تم استخدام مجموعتي بيانات مختلفتين وهما 3D-IRCADb-01 و LiTS تم استخدام معامل تشابه Dice (DSC) لتقييم أداء النماذج المختلفة .بالنسبة لمجموعة بيانات 3D-IRCADb-01 تم تحقيق نتائج ممتازة في معامل DSC حيث بلغت أعلى قيمة 96.5% وبلغت الدقة 99.5% مما يشير إلى فعالية نماذج التجزئة .أما بالنسبة لمجموعة بيانات LiTS فقد تفوقت النماذج أكثر حيث حققت درجات DSC تصل إلى 98.1% ودقة بلغت 99.7% بعد التجزئة، يتم استخراج مناطق الاهتمام مما يسهل التحليل الطبي اللاحق والتشخيص .تظهر هذه النتائج قوة ودقة النموذج المقترح في مهام تجزئة الكبد.