Deep Learning Based Liver Cancer Segmentation from Computed Tomography Images

Yodit Abebe Ayalew (yoditabebe9391@gmail.com)
Hawassa University Institute of Technology
Kinde Anlay Fante
Jimma University Institute of Technology
Mohammed Aliy
Jimma University Institute of Technology

Research

Keywords: Liver cancer, HCC, Segmentation, Deep learning, CNN, UNet

Posted Date: August 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-65573/v1

Deep Learning based Liver Cancer Segmentation from Computed Tomography Images

Corresponding Author: Yodit Abebe (Lecturer at Department of Biomedical Engineering, Hawassa Institute of Technology, Hawassa University, Hawassa, Ethiopia); Email: yoditabebe9391@gmail.com

Authors:

Dr. Kinde Anlay (Lecturer at Department of Electrical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia)

Mohammed Aliy (Lecturer at Department of Biomedical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia)

Abstract

Background: *Liver cancer is the sixth most common cancer worldwide. According to WHO data in 2017, the liver cancer death in Ethiopia reached 1040 (0.16%) from all cancer deaths. Hepatocellular carcinoma (HCC), primary liver cancer causes the death of around 700,000 people each year worldwide and this makes it the third leading cause of cancer death. HCC is occurred due to cirrhosis and hepatitis B or C viruses. Liver cancer mostly diagnosed with a computed tomography (CT) scan. But, the detection of the tumor from the CT scan image is difficult since tumors have similar intensity with nearby tissues and may have a different appearance depending on their type, state, and equipment setting. Nowadays deep learning methods have been used for the segmentation of liver and its tumor from the CT scan images and they are more efficient than those traditional methods. But, they are computationally expensive and need many labeled samples for training, which are difficult in the case of biomedical images.*

Results: *A deep learning-based segmentation algorithm is employed for liver and tumor segmentation from abdominal CT scan images. Three separate UNet models, one for liver segmentation and the others two for tumor segmentation from the segmented liver and directly from the abdominal CT scan image were used. A dice score of 0.96 was obtained for liver segmentation. And a dice score of 0.74 and 0.63 was obtained for segmentation of tumor from the liver and from abdominal CT scan image respectively.*

Conclusion: *The research improves the liver tumor segmentation that will help the physicians in the diagnosis and detection of liver tumors and in designing a treatment plan for the patient. And for the patient, it increases the patients' chance of getting treatment and decrease the mortality rate due to liver cancer.*

Keywords: Liver cancer, HCC, Segmentation, Deep learning, CNN, UNet

Background

Liver cancer is the sixth most common cancer worldwide. As Global Cancer Statistics, it is the second and sixth cause of cancer death for men and women respectively [1]. And according to WHO data, the percentage of liver cancer death in Ethiopia out of the total death in 2017 was 0.16% [2]. There are two types of liver cancers, primary and secondary. Among primary types of cancers, hepatocellular carcinoma (HCC) accounts for 80% of the case [3]. It is the third cause of cancer death and results in the death of around 700,000 people each year worldwide [4]. The major risk factors for primary liver cancers are cirrhosis resulted from alcohol usage, hepatitis B and C viruses, and a fatty liver disease caused by obesity [5]. It can be diagnosed and detected by using different imaging tests like ultrasound, magnetic resonance imaging (MRI) and computed tomography (CT). From these, a CT scan is a frequently used imaging test [6].

A CT scan gives detailed cross-sectional images of the abdominal region. And this abdominal CT image is further processed in order to segment the liver tumor from the image. But still, the intensity similarity between tumor and other nearby tissue in the CT images makes the detection of tumor too difficult [5]. Therefore, these images need to be processed and enhanced in order to early detect and differentiate cancerous tissue.

In a CT scan, indications of cancer presence can be identified by a difference in pixel intensity from that of the liver. It may be darker (hypodense) or brighter (hyperdense) than the surrounding healthy liver [7]. The manual segmentation of CT scan images is laborious and time-consuming for a clinical setting. Due to various factors, such as the liver typically stretching over 150 slices in a CT volume, indefinite shape of the lesions and low-intensity contrast between lesions and nearby tissues, the variation of liver shape and size among patients and the similarity with other organs of almost the same intensity [5], [8]. Considering those problems, researchers have designed different computer-aided diagnostic systems for the segmentation of liver and tumor from the abdominal CT scan images.

Nowadays, different traditional techniques are used for extracting tumor from the liver images. But these methods are not fully effective in the extraction of the tumor. Most of them are manual or semi-automatic and are dependent on edge detectors rather than analyzing the image as pixel. After hardware improvement in the 2000s, machine learning approaches came into a widely applicable system in image processing tasks like segmentation [9]. A variety of deep-learning

methods have also been developed for automatic or semi-automatic segmentation of liver tumors. Among those, convolutional neural networks (CNN) are currently the most widely used method [10].

Researchers had used CNN and its extensions, fully connected layer and UNet, for liver and tumor segmentation [10]–[15]. As our knowledge up to now, maximum dice score obtained for liver and tumor segmentation are 0.9522 and 0.63 respectively. Additionally, Christ et al and Chlebus et al had used 3D post processing methods for better segmentation result [12], [13]. But still the segmentation performance was poor.

Related Works in the Area of the Research

In 2016 Christ et al. proposed an algorithm that segments liver and its lesions using cascaded deep neural network and 3D conditional random fields. In the paper, they had used two cascaded UNet networks for liver and tumor. Then the outputs from these networks were inputted to 3D conditional random fields for better refinement. They obtained a dice similarity coefficient (DSC) of 0.943 in liver segmentation but didn't include their tumor segmentation result in their paper [12].

And after about one year Sun et al. came up with an idea of segmenting liver tumor using multichannel fully convolutional network (FCN) from contrast-enhanced multiphase CT scan image. Single-channel of FCN consists of 8 convolutional layers, 3 subsampling layers, 3 deconvolution layers, and two feature fusion layers. The convolutional layers perform convolutional operations to the output of the previous layer with different kernel size and extract features from the image by maintaining spatial correlations. The subsampling layer reduces the size of the image without affecting image resolution. And then deconvolution layer and fusion layers were used for the upsampling and fusion process respectively. For their work, they used three such FCN layers for extracting features of the image from three different phases of CT scan image. Each FCN channels were trained independently and their features were fused for better segmentation result. For evaluation, they had used different performance metrics. They had got a volumetric overlap error (VOE) of 8.1±4.5 [16]. After one year, in 2018, Chlebus et al. agreed with the idea of Christ et al. and used UNet architecture for segmenting liver tumors. But they modified the implemented UNet architecture by adding a drop out layer and additional short skip connections for parameter updates and for increasing the network speed during training. Then the

outputs from the modified UNet had been post-processed using 3D connected component, a shape-based processing method and classified again using a random forest classifier for further refinement. They got an average DSC of 0.58 in their segmentation result. But here the shape based process may not be effective as they thought since the tumors have a variable shape [13]. And then in 2019, Liu et al. modified the work of Christ et al. and Chlebus et al. and came up with an idea of segmenting liver sequence images using GIU-Net that combines the improved u net with the graph cut algorithm. In their improved u net architecture they increased the depth of the structure for better semantic segmentation result and make the skip connections to be from the pooling layers output, unlike the original UNet which concatenates the upsampling layer output with the output of the corresponding convolutional layer, in order to reduce the lost information and called this new UNet structure IU-Net. Then they combined this structure with the graph cut method to provide their new segmentation algorithm. They initially segmented a liver from a liver CT sequence using their improved version of UNet and then refine their result using graph cut. They used different metrics for evaluating the performance of their algorithm and using the DSC they got 0.9505 value. In this paper, the authors only provide an algorithm for liver segmentation they didn't forward a solution for segmenting its pathologies [10].

In the same year after few months, Li et al. disagree with Liu et al. and the work of others and used a convolutional neural network for diagnosing hepatocellular carcinoma as Christ et al. used. On the paper, they performed two main tasks, segmentation of liver and its pathology, hepatocellular carcinoma using a fully convolutional neural network and classification of hepatocellular carcinoma into diffuse, nodular and massive classes. In the segmentation step, they were dependent on the FCN-8s structure. This model contains four max-pooling layers and two skip structures to concatenate the last two outputs of the max-pooling layer with the corresponding upsampling layer. In their structure, they added two more skip connections to concatenate the remaining outputs of the max-pooling layer with the corresponding upsampling layer in order to increase the number of features used in predicting the output. The proposed model includes two parts. The first part of their model includes 13 convolution layers of the standard VGG-16 model that performed convolution operation with a kernel size of 3 x 3 and ReLU as activation function and four max-pooling layers. And In the second half of their structure upsampling using deconvolution was performed on up sampling layer and the output was fused with the corresponding output from the first half of the model on fusion layers. They could attain an accuracy of 0.994 through 100 epochs

training. But some noise spots are present in the outputs and the algorithm shows poor performance in segmenting diffuse tumors [14]. In the same year, 2019, Budak et al. developed two cascaded encoder-decoder convolutional neural networks for efficient segmentation of liver and tumor. They proposed the EDCNN algorithm that includes two symmetric encoder and decoder parts. Each part consists of ten convolutional layers with batch normalization and ReLU activation followed by max pooling layer. And in each convolutional layer, they used 64 filters and by this they can decrease the number of parameters needed. For the segmentation process they used two cascaded deep neural networks. One for liver and one for the tumor. The output of the first network becomes the input for the next network. They achieved the average DSC value of 0.9522 and 0.634 on liver and tumor segmentation respectively [15].

In this thesis, a deep learning-based segmentation algorithm is employed for liver and tumor segmentation from abdominal CT scan images. The main contributions of this work are, first it applies data augmentation tasks that solve the limitation of available data in biomedical images, second it highly reduced the time needed for training by reducing the number of filters in each convolutional block which reduces the number of trainable parameters and third it minimizes the effect of class imbalance that presents between the tumor and the background through discarding slices with no tumor information from the datasets and used only slices with full information, this improves the performance of the algorithm in detecting the tumor from the CT images.

Materials

A. Datasets

Images that were used to train and test liver and liver cancer segmentation algorithm developed by this thesis were taken from two publicly available datasets, 3Dircadb01 (3D Image Reconstruction for Comparison of Algorithm Database) [17] and LITS (Liver Tumor Segmentation) Challenge [18]. The 3DIRCADb dataset is challenging to utilize since there is a high variety of data and the liver and tumor complexity [15]. Table 1 shows detail information about the two datasets.

Table 1: Liver and tumor segmentation data sets.

Dataset	**Number of patients**	**Image size**	**Pixel width and height**	**Slice thickness**	**Pixel spacing**	**Slice number**	**Tumor data out of 100%**
3D-IRCADb01	20	512 x512	0.56 -0.87 mm	1- 5 mm	0.55 - 0.95 mm	74 - 260	75%
LITS	131	512 x512	-	0.7 -5mm	-	-	63%

B. Data Preparation

Images taken from the two datasets should be prepared in order to use them for training and testing the developed algorithm. The 3D-ircadb01 dataset contains up to seven folders under each patient's data for the tumor masks depending on the anatomical position of the tumor on the liver. Therefore these tumor masks from those different folders should be added and put into one folder since the main intention is on the segmentation result not on the tumor's anatomical position.

And the images in the LITS datasets are three dimensional and there is no separate mask for the liver and its tumor. Instead, they are found on the same mask image under the segmentation folder in the dataset. Since the developed algorithm is two dimensional (2D), the data should be converted into 2D. The separate mask for the liver and tumor must also be prepared. This data preparation was done using an ImageJ tool. From both datasets, the patient data with no liver and tumor masks are discarded. And from each patient data, images or slices which are taken at the starting and ending of scanning, with no liver information were also discarded for reducing the class imbalance present between the background and foreground.

The number of images used for training and testing are included in Table 2.

Table 2: Training and testing images.

Training images	**Testing images**

2346 + Data augmentation	392

C. Image Preprocessing

The images were 512 x 512 in dimension originally. Using those images as it was is difficult due to limited GPU memory. Therefore, all images were resized with a factor of 0.25. And the images were also normalized to have a value in between 0 and 1.

D. Computing Platforms

The acquired images from both publicly available datasets were processed and analyzed on Kaggle. Kaggle is an online community of data scientists, owned by Google that provides cloud infrastructures such as built-in Python Jupyter notebook, graphical processing unit (GPU), tensor processing unit (TPU) and data storage platform for facilitating the works of data scientists.

Method

The segmentation algorithm

The algorithm is based on the UNet architecture developed by Ronneberger et al. in 2015. This algorithm includes two 2D UNet architectures, for the liver and its tumor. These architectures were designed to segment liver and tumors from the abdominal CT scan images.

A. Network Architecture

For both liver and tumor segmentation, the same U shaped network architecture is used. It consists of a contracting path, an expansive path, and a bottleneck part like the original UNet. But, here in this paper, after each convolutional layer, batch normalization is added in all the three parts of the network and a 0.5 dropout layer is added after each convolutional block of the contracting path.

The batch normalization is important for normalizing the outputs of the convolutional layers to have a mean of zero and a standard deviation of one and the dropout layer randomly deactivated some neurons in the hidden layer in order to prevent overfitting of the network. And the other modification is done on the number of filters of each convolutional block. In the first block, there

are 16 filters and it will be doubled in the consecutive three blocks and become 128 at the last convolutional block. The details of the network architecture are included in Table 3.

Contracting or Downsampling Path

The contracting path also called encoder is composed of 4 blocks. Each block is composed of:

- 3x3 Convolution Layer with ReLU activation function and batch normalization
- 3x3 Convolution Layer with ReLU activation function and batch normalization
- 2x2 Max Pooling
- Drop out layer (0.5)

The purpose of this contracting path is to capture the context or semantics of the input image to be able to do segmentation. It extracts features that contain information about what is in an image using convolutional and pooling layers. During this process, the size of the feature map gets reduced and the deep or high-level features of the image will be obtained but the network loses the spatial or location information in which those features are found.

Bottleneck

This part of the network is between the contracting and expanding paths. The bottleneck is built from two convolutional layers with batch normalization.

Expanding or Upsampling Path

The expanding path also called decoder is composed of 4 blocks. Each of these blocks is composed of

- Up convolution or Deconvolution layer with stride 2.
- Concatenation with the corresponding cropped feature map from the contracting path.
- 2 (3x3 Convolution layer with ReLU activation function and batch normalization).

The purpose of this expanding path is to recover the feature map size and to add spatial information for the segmentation image, for which it uses up-convolution layers.

The course contextual information from the contracting path will be transferred to the upsampling path by means of skip connections.

Skip Connections

There could be a loss of low-level information during the decoding process. To recover this information lost and to let the decoder access the low-level features produced by the encoder layers skip connections are used. Intermediate outputs of the encoder are concatenated with the inputs to the intermediate layers of the decoder at appropriate positions. This enables precise localization combined with contextual information from the contracting path.

Table 3: Details of the network architecture which shows layers found in each part of the model.

Encoder	2(Conv2D+ReLU+BN+Max-pooling)
	Dropout 0.5
	2(Conv2D+ReLU+BN+Max-pooling)
	Dropout 0.5
	2(Conv2D+ReLU+BN+Max-pooling)
	Dropout 0.5
	2(Conv2D+ReLU+BN+Max-pooling)
	Dropout 0.5
Bottleneck	2(Conv2D+ReLU+BN)
Decoder	Up convolution + Concat
	2(Conv2D+ReLU+BN)
	Up convolution + Concat
	2(Conv2D+ReLU+BN)
	Up convolution + Concat
	2(Conv2D+ReLU+BN)

	Up convolution + Concat
	2(Conv2D+ReLU+BN)
Output	One 1x1 Conv2D + Sigmoid

Where Conv2D is 2D convolution, BN is batch normalization and Concat is concatenation between the up convolution results with corresponding encoder output.

B. Training

The network architecture is based on the original UNet architecture. However, additional batch normalization and dropout layers are included in the network architecture of this work and the number of filters in each convolutional block is also reduced. Therefore, training the network from scratch is needed. The input images and their corresponding segmentation masks are used to train the network. Totally 2346 images from the two datasets with data augmentation were used. During training, many experiments were done by tuning the hyperparameters used in the network. Learning rate, batch size, number of epochs, and number of filters, validation split, dropout value, optimizer, loss function, and activation function had been checked for different values and assignments. After many trial and error a batch size of 8, epochs of 250, validation split of 0.30, and a dropout of 0.5 had been used.

Optimizer

Adam Optimizer is an extension for the stochastic gradient descent (SGD) and RMSprops (root mean squared). It is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirements. It finds individual adaptive learning rates for each parameter in the network. Its name is derived from adaptive moment estimation [19]. In this work, Adam optimizer with a learning rate of 0.0001 had been used.

Loss Function

Weighted dice loss and binary cross entropy were used as a loss functions to measure the variations of the predicted values from the actual values during the training of the network. The equations used for calculating weighted dice loss and binary cross entropy are given in equation 1 and 2 respectively.

$$\text{Loss} = -\,W\left(\frac{2TP}{2TP+FP+FN}\right) \quad (1)$$

Where TP is true positive, FN is false negative, FP is false positive and W is a weight factor that is introduced to balance the class frequency difference between the foreground and the background.

$$BCE = \frac{-1}{N}\sum_{i=1}^{N} y_i.\ \log(p(y_i)) + (1-y_i).\ \log(1-p(y_i)) \quad (2)$$

Where BCE is binary cross entropy, N is total number of pixels, y_i is predicted label for each pixel i and $p(y_i)$ is the predicted probability of each pixel being foreground or background.

Data Augmentation

Data augmentation is important to train the network effectively when there are small training samples available. In biomedical image segmentation tasks, there are often very little training data available. Therefore excessive data augmentation by applying affine deformations to the available training images is used. This allows the network to learn invariance to such deformations.

Data augmentation is specifically essential for biomedical image segmentation in which deformation is the basic difference in tissues. Less number of training pairs result in overfitting [20].

In the proposed work, in place or on the fly data augmentation technique had been used [21]. This type of augmentation artificially increases the size of the dataset by applying real-time data augmentation. In each epoch new randomly augmented data were given to the model.This increases the amount of data and the generalizability of the model.

Results and Discussion

Performance Metrics

For evaluating the performance of the segmentation method, the binary mask of the segmentation result is compared to the ground truth mask and their similarity is estimated. Different performance metrics like DSC, Jaccard similarity coefficient (JSC), accuracy and symmetric volume difference (SVD) are used.

A. Dice Similarity Coefficient (DSC)

It measures the overlap between two binary masks. It is the size of the overlap of the two segmentations divided by the total size of the two objects. It ranges from 0 (no overlap) to 1 (perfect overlap). It represents the overall performance of the segmentation [22], [23]. It is calculated using equation 3.

$$\text{DSC} = \left(\frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}\right) \quad (3)$$

Where TP is true positive, FN is false negative, and FP is false positive.

B. Jaccard Similarity Coefficient (JSC)

It measures the similarity between the segmented image and the binary mask. It is the ratio of the intersection of two binary masks to their union [23]. It is given by equation 4.

$$\text{JSC} = \left(\frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}}\right) \quad (4)$$

Where TP is true positive, FN is false negative, and FP is false positive.

C. Accuracy

Accuracy represents the ratio of correctly segmented samples to the total samples. It is approximately one for good segmentation results. It is calculated using equation 5 [14].

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (5)$$

Where TP is true positive, TN is true negative, FN is false negative, and FP is false positive.

D. Symmetric Volume Difference

SVD is a measure of difference that exists between the segmented images with the ground truth. For good segmentation result, SVD approximates to zero. It is given by equation 6.

$$\text{SVD} = (1\text{-DSC}) \quad (6)$$

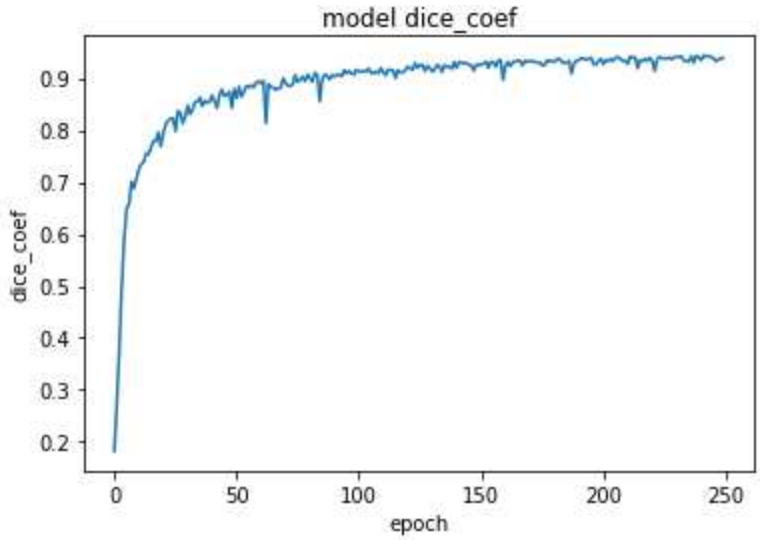
Where DSC is the Dice similarity coefficient.

- True Positive (TP): denotes all pixels belongs to the foreground and classified as foreground.

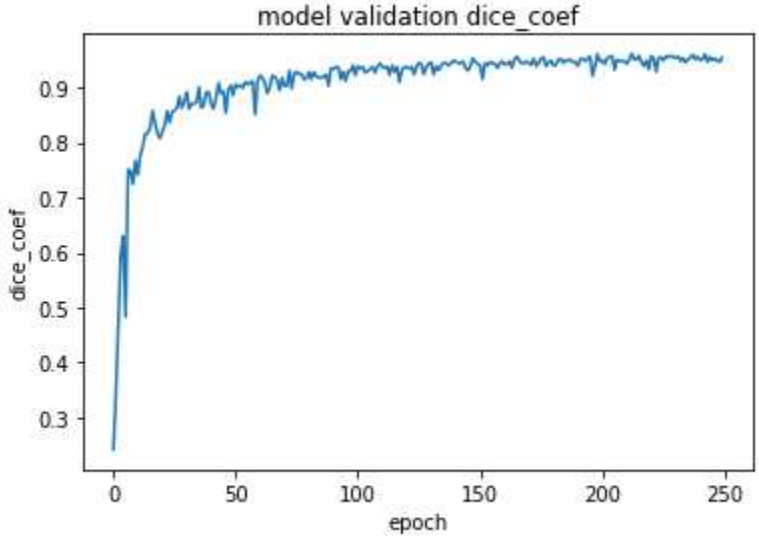
- True Negative (TN): denotes all pixels belongs to the background and classified as background.
- False Negative (FN): denotes foreground pixels that are incorrectly classified as background pixels by the classifier.
- False Positive (FP): denotes background pixels that are incorrectly classified as foreground pixels by the classifier.

During the training, three separate models with similar architectures had used. The first model was trained using abdominal CT scan images with liver annotations for liver segmentation purpose. Then the second model is trained using liver images with tumor annotations for the segmentation of tumor from the liver. And the third model was trained using abdominal CT scan images with tumor annotations for the segmentation of the tumor directly from the abdominal CT scan images.

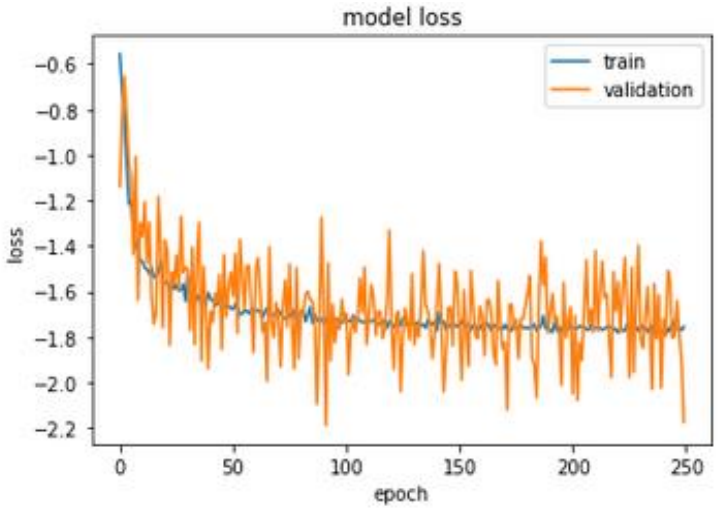
Each network was trained using 2346 images with data augmentation from scratch. Images were 512 x 512 in dimension. Since processing the whole images with this size is difficult due to limited GPU memory, the images were resized to a dimension of 128 x 128. Weighted dice loss is chosen as a loss function for the first two networks and showed better performance during training. For the last model, which is trained to segment tumor directly from abdominal CT scan images, binary cross entropy was chosen as a loss function and for all those three models Adam is selected as an optimizer through experiments. In the original UNet paper the batch size of 1 was used for the purpose of maximum usage of GPU memory without considering the time it took for training [20]. As the batch size decreases the training time will increase and the probability of using maximum GPU memory increases. Therefore the selection of batch size needs great care. Unlike [20], in this thesis batch size of 8 was used that compensates both GPU memory problem and training time after many trials. That means the network was trained using eight images at a time. The network's model DSC and model loss for liver segmentation, tumor segmentation from the liver and tumor segmentation from the abdominal CT scan images were plotted from Figure 1 up to Figure 3.


model dice_coef
dice_coef
epoch
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
50
100
150
200
250

a)


model validation dice_coef
dice_coef
epoch
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
50
100
150
200
250

b)



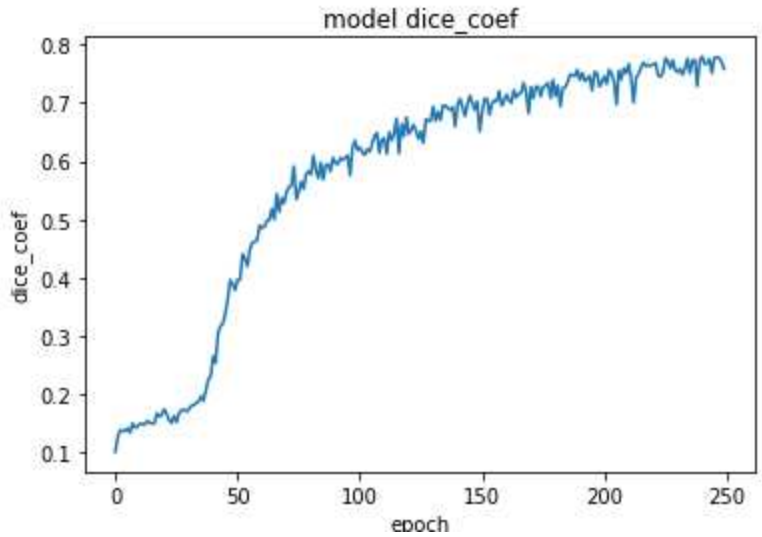
c)

Figure 1: DSC and loss plot of the model during training for liver segmentation. (a) and (b) are model DSC for training and validation data respectively and (c) is model loss.

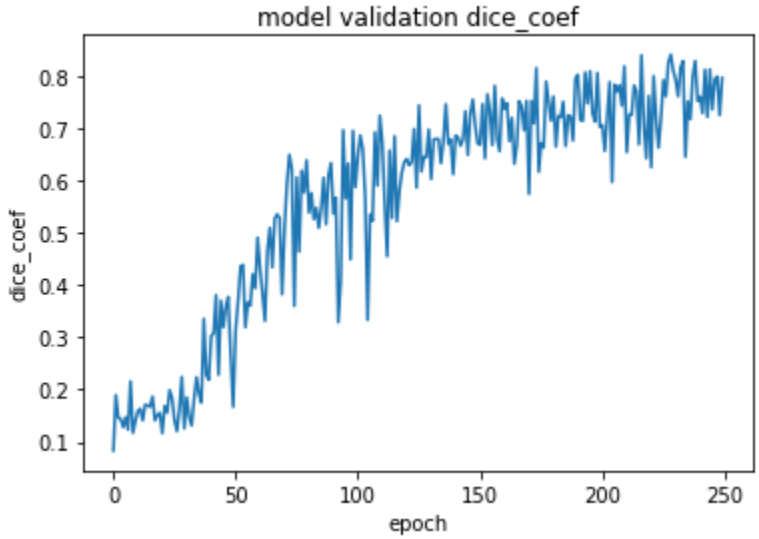
The first two plots (a and b) in Figure 1, show the model DSC for training and validation data during the training of the model for liver segmentation. And as considered from the two graphs, the model has good performance for both training and validation data. The DSC for both graphs increases greatly around the first 100 epochs and its increment becomes gradual and becomes nearly constant. Finally, the DSC becomes 0.9511 and 0.9633 for training and validation data respectively.

And the third plot in Figure 1 shows the model loss for training and validation data. As observed the loss for both data decreases greatly up to around the first 100 epochs and after that it becomes nearly constant. The final loss for training and validation was -1.7567 and -2.1753 respectively.

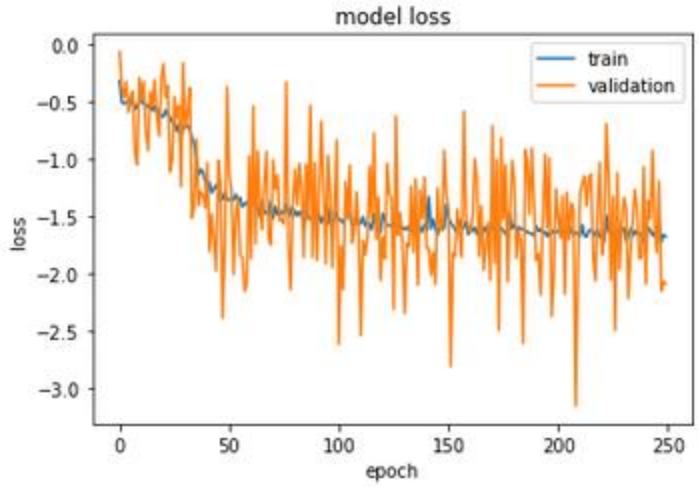
As considered from the three graphs in Figure 1, the developed algorithm has good performance in segmenting the liver from the abdominal CT scan images.


model dice_coef
dice_coef
epoch
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250

a)


model validation dice_coef
dice_coef
epoch
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250

b)

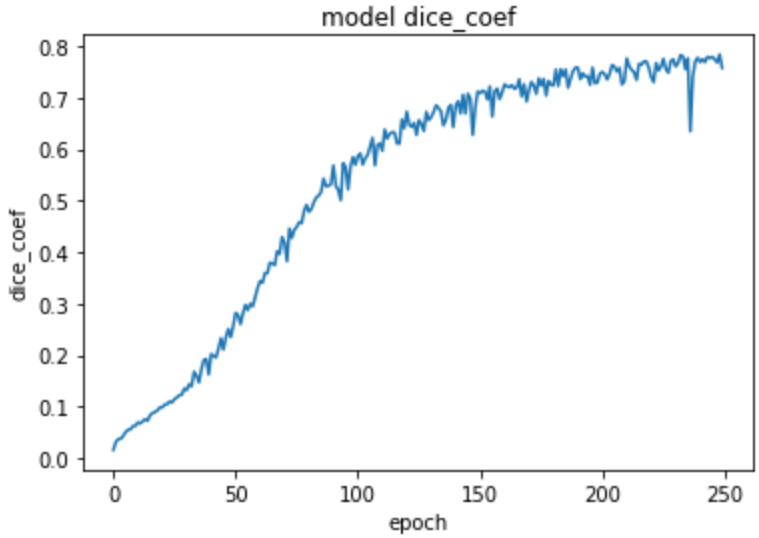


c)

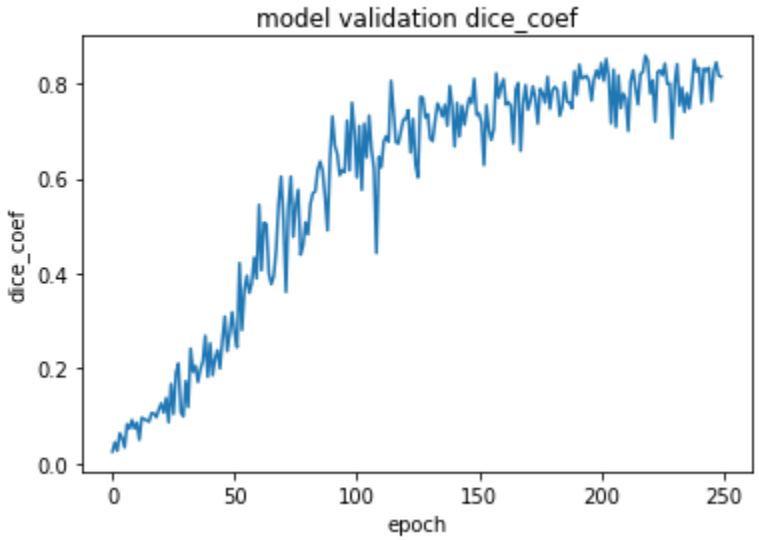
Figure 2: DSC and loss plot of the model for tumor segmentation from the liver. (a) and (b) are model DSC for training and validation data respectively where (c) is the model loss.

As Figure 2 (a) shows the model DSC for the training data increases up to some point and becomes nearly constant. This shows the network was good during training. In the second plot (b), the model DSC for validation data is also plotted and some fluctuations are observed. At last epoch, a DSC of 0.7769 and 0.8375 was obtained for training and validation data respectively.

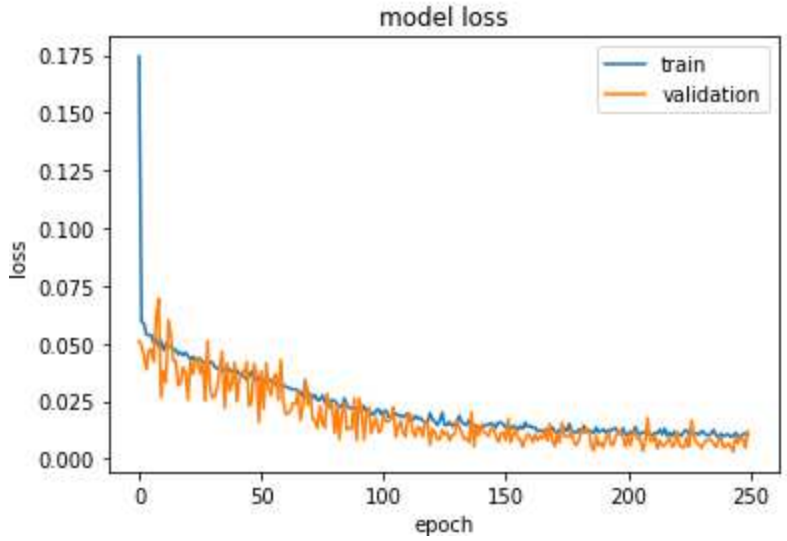
In Figure 2 (c) the model loss is plotted. The loss for the training and validation data decreases as expected and becomes nearly constant. And finally, a loss of -1.6291 and -2.0278 was obtained for training and validation data respectively.


model dice_coef
dice_coef
epoch
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
50
100
150
200
250

a)


model validation dice_coef
dice_coef
epoch
0.0
0.2
0.4
0.6
0.8
0
50
100
150
200
250

b)



c)

Figure 3: DSC and loss plot of the model for tumor from the abdominal CT scan images during training. (a) And (b) are model DSC for training and validation data respectively and (c) is a model loss.

As shown in Figure 3, model DSC and model loss is plotted for tumor segmentation from the abdominal CT scan image. The first two plots are model DSC for training and validation data. At last epoch, a DSC of 0.7734 and 0.8240 was obtained for training and validation data respectively.

And in the third plot Figure 3, the model loss for the two data are plotted. Here also some fluctuation in validation loss is observed. But the training loss decreased almost constantly. And obtained a loss of 0.0093 and 0.001 for training and validation respectively.

A. Test Results for Liver Segmentation

To test the liver segmentation performance of the developed network, 392 images were used. And those images were preprocessed using the same preprocessing technique that was implemented on the training data. The result of the network was evaluated using the respective ground truths of the images and the comparison result of this algorithm with works of Christ et al. who had used a cascaded deep neural network with a 3D conditional random fields to segment liver and its lesions [12], Liu et al. who came up with GIU-Net that combines the improved UNet with the graph cut

algorithm for segmenting liver sequence images [10] and lastly with Budak et al. who developed two cascaded encoder-decoder convolutional neural network for the segmentation of liver and its tumor [15], were also included.

Objective Comparison

The performance of the liver segmentation algorithm was evaluated using different performance metrics and the result is included in Table 4. The segmentation result of the algorithm with the respective ground truth images is included in Figure 4.

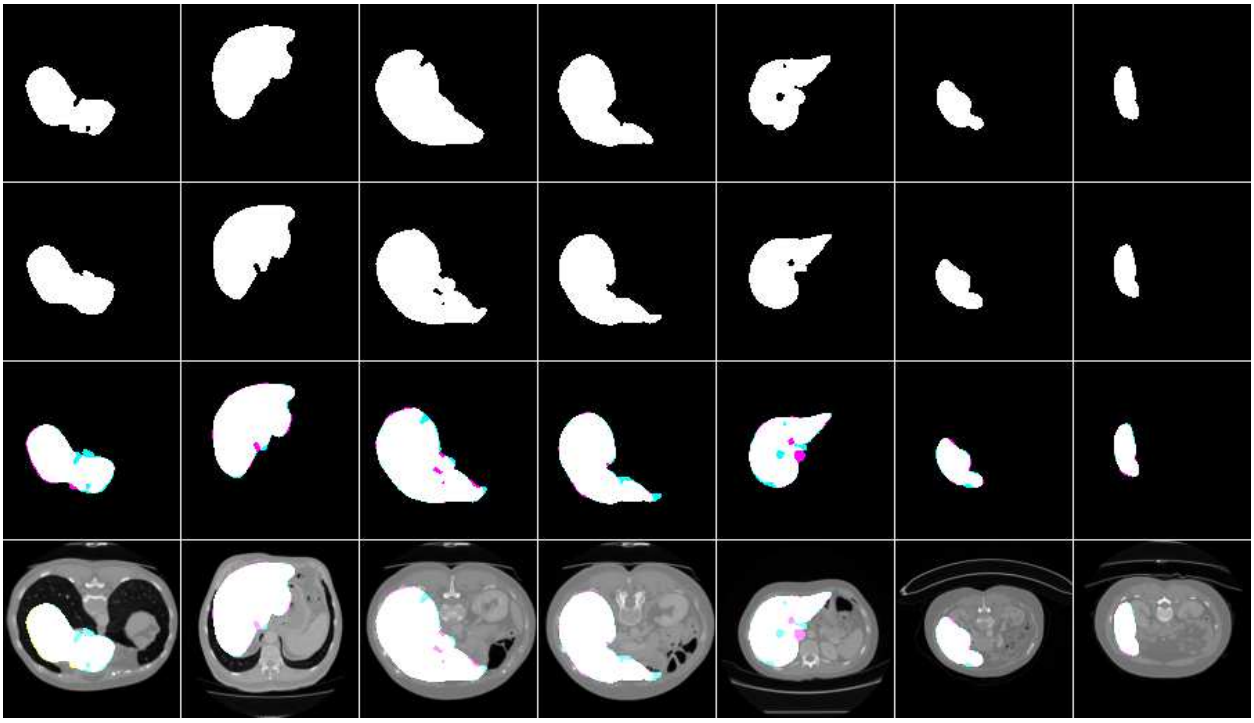


Figure 4: Liver segmentation results, ground truths, and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the original CT scan image.

As shown in Figure 4 the liver segmentation result is satisfactory and the algorithm can almost segment liver from the abdominal CT scan images fully. It has an average dice score of 0.96 which is greater than the others by around 0.009. But in some cases, it loses some portion of the liver as it is shown with cyan and segments nearby tissues as a liver as it is shown as magenta in row 3.

Comparison with Previous Works

This algorithm highly reduces the complexity of the network by reducing the number of filters in each convolutional block. This decreases the time needed to train the network from a few hours to 40 minutes. In this thesis, 2346 images with data augmentation were used to train the network which is very small when it is compared with other works that had used more than 20,000 images. And here the class frequency difference between the liver and the background was minimized through removing the CT slices with no liver that affects the segmentation performance in addition to introducing a weight vector to the loss function. The result of this algorithm was compared with other works in order to show how this algorithm improves liver segmentation performance. Table 4 shows the result of this algorithm and other works.

Table 4: Test results of others and our work for liver segmentation

Papers	**Dice score**	**Jaccard**	**SVD**	**Accuracy**
Christ et al. [12]	0.9430	_	_	_
Liu et al. [10]	0.9505	_	_	_
Budak et al. [15]	0.9522	_	_	_
This work(U net)	**0.9612**	**0.9999**	**0.0388**	**0.9931**

B. Test Results for Tumor Segmentation

For testing the segmentation ability of the developed algorithm on segmenting the liver tumor a total of 392 images with their respective ground truths were used. The tumor was segmented in two ways. The first is the segmentation of the tumor directly from the abdominal CT scan image and the other is from the liver after segmenting it first. The result of the network was evaluated using the respective ground truths of the images and the comparison result of this algorithm with works of Chlebus et al. who used UNet by modifying it with object-based post-processing to segment liver tumor [13], and Budak et al. who implemented an encoder-decoder convolutional neural network for liver tumor segmentation [15], were also included.

Objective Comparison

The segmentation result of this network on segmenting liver tumors from the liver and directly from the abdominal CT scan images is evaluated using different performance metrics and the result is included in Table 5. The segmentation result of the algorithm with the respective ground truth images is included in Figure 5 and Figure 6.

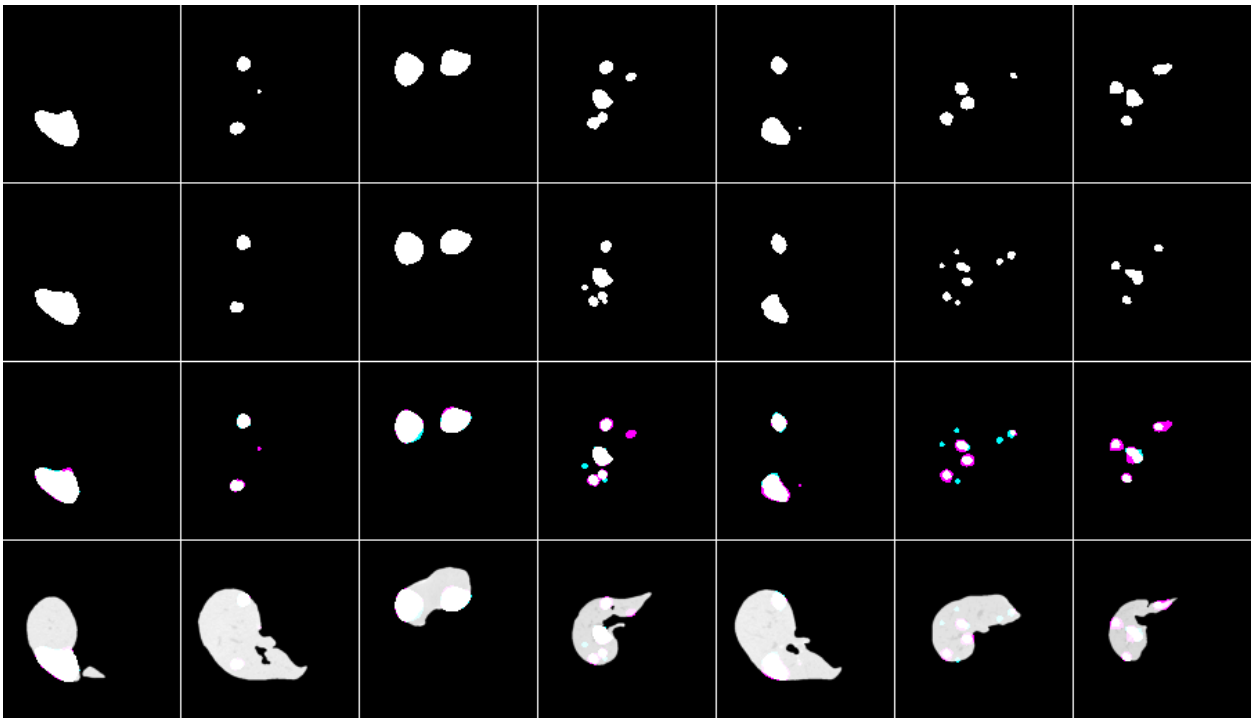


Figure 5: Tumor segmentation results from the liver with the respective masks and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the liver.

As shown in Figure 6, the algorithm has good segmentation ability on circular tumors and can also detect the distributed tumors from the same liver slice. It has an average dice score of 0.74, which is greater than the others by around 0.11. But in some cases, it fails to segment some tumors as it is seen as cyan and segments other tissues as tumor as it is seen as magenta in row 3.

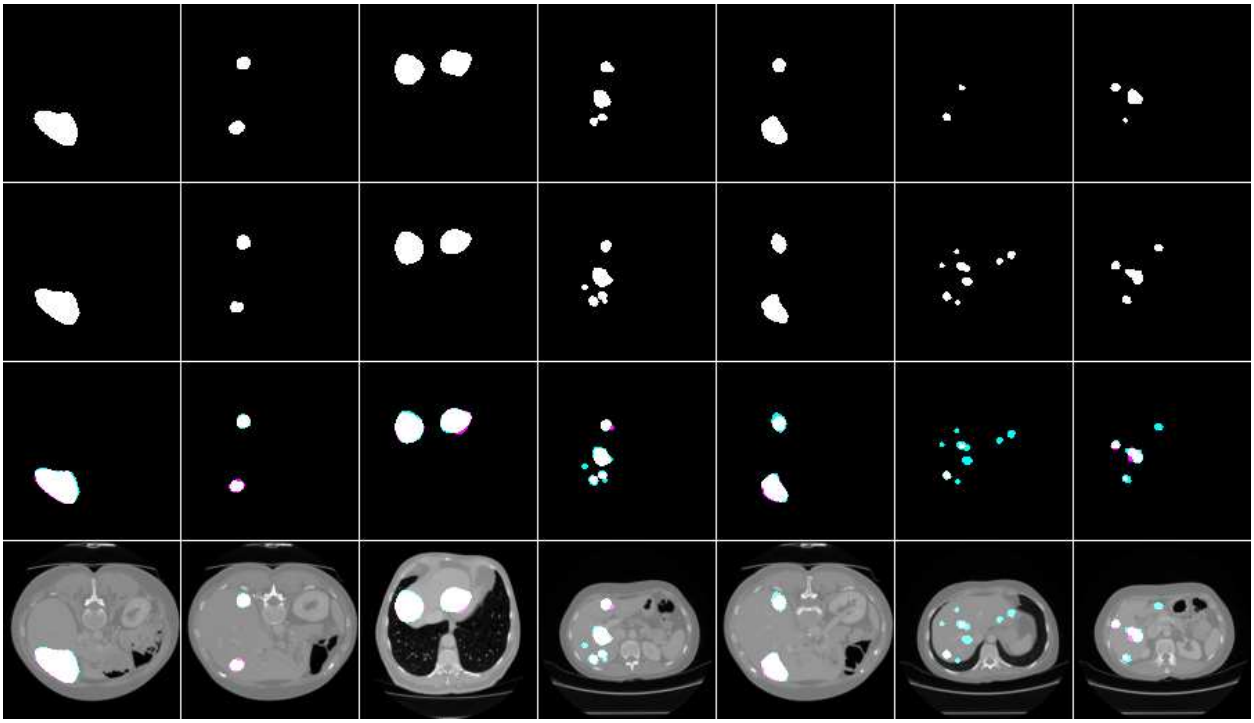


Figure 6: Result of tumor segmentation from abdominal CT images with the respective masks and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the abdominal CT image.

As shown in Figure 6, the tumor segmentation directly from the abdominal CT scan image shows good performance relative to works done by other researchers. It has a relatively similar performance with those works that segment the tumor with a two-way process. It has an average dice score of 0.63. But it fails to segment some tumors as it is seen as cyan in row 3 and it also segments other nearby tissues as a tumor as it is shown with magenta in row 3.

Comparison with Previous Works

As it is already discussed, this work decreases the complexity and the time needed for training. And the work also shows a new way for liver tumor segmentation. It can segment the tumor directly from the abdominal CT scan images, unlike the others which followed two steps to segment it. In other work, to segment liver tumors, the liver has to be segmented first and then the tumor segmentation precedes next to that.

But this work came up with segmenting of the tumor directly without liver segmentation and by this, a comparable segmentation result of 0.63 DSC was obtained. In addition to this, the tumor segmentation was also done using the previous way. That means by following a two-step process like others and obtained a DSC of 0.74 that differs by an average of 0.11 from the previous works. Chlebus and his colleagues had used the post-processing method, which includes 3D connected components and random forest classifiers. However, the segmentation result obtained from this algorithm is greater than them by 0.16. This improvement is due to the class balancing that the work implemented. As it is discussed above, the class balancing was done by removing slices with no tumor. The difference between the numbers of tumor pixels to background pixels largely affects the segmentation result. Therefore this work tried to decrease this class imbalance by removing those slices with no tumor from the whole dataset in addition to the weight factor added to the loss function and observed a performance improvement. The segmentation result is compared with other works in order to show the improvements in liver tumor segmentation. Table 5 shows the tumor segmentation result of two papers and the current work.

Table 5: Test results of others and our work for tumor segmentation

Papers	**Dice score**	**Jaccard**	**SVD**	**Accuracy**
Chlebus et al. [13]	0.58	_	_	_
Budak et al. [15]	0.63			
This work (U net) from liver	**0.74 ±0.02**	**0.9999**	**0.26 ±0.02**	**0.9954**
This work (U net) from abdominal CT image	**0.63±0.02**	**0.9999**	**0.37 ± 0.02**	**0.9950**

C. General Results of the Architecture

This segmentation algorithm highly improves the efficiency of liver tumor segmentation. First, it reduces the complexity of the network by reducing the number of filters needed on each convolutional block that decreases the number of trainable parameters. Due to this the time needed for training the network greatly reduced. The total time needed to train the network for 250 epochs is about 40 minutes on Kaggle kernel. This is a great achievement in deep learning-based segmentation in which time and complexity of the network matter a lot. And the other burning issue in deep learning-based segmentation was the absence of enough training samples to train the network. And this also solved by the developed algorithm. It only needs small training samples and used excessive data augmentation. By this, it can increase the number of training samples present. Data augmentation applied affine deformations on those available images that helped the network to learn invariance to those deformations hence, deformation is the most common variation in biomedical images.

And the other important thing that should be considered during liver tumor segmentation or other biomedical image segmentation is the class imbalance between the two classes to be segmented. There is a large difference in size between the tissue to be segmented and the background. This highly affects the segmentation performance. For example, in Figure 7 the number of white pixels to black pixels shows a high difference.

a) b)

Figure 7: Class imbalance between liver and tumor with the background. (a) shows the liver and (b) shows the tumor.

The ratio of white pixels to black pixels can be calculated using equation 7. It is 1: 9 and 1: 85 for liver and tumor masks respectively. Due to this, the network gets more black pixels than white pixels and learns from that during training. Its probability of learning from white pixels is very small when it is compared with the black ones. This results in poor performance of the network.

$$\text{Ratio} = \frac{\text{Number of white pixels}}{\text{Number of black pixels}} \quad (7)$$

On the original UNet paper, the authors included a weight map that pre-computed from the ground truth images for balancing the class frequencies. In addition to this, in this paper, it is reduced using the removal of slices with no tumor information. During data preparation, the first step was checking all patients' data with tumor from both datasets. Then remove data that is obtained from healthy ones from the dataset and next search and remove for slices with no tumor. Lastly, the data with the tumor only was arranged and saved sequentially. The network had been trained using those data and it's observed that the network's performance shows a great difference. The network performance increases with 0.009 and 0.11 for liver and tumor respectively.

This work also introduces a new way for tumor segmentation. Before this work, tumor segmentation has been done from the liver after segmenting it first from the abdominal CT scan image. The segmentation was a two-way process. But here liver tumors can be detected and segmented directly from the abdominal CT scan images with relatively comparable performance. This decreases the time and the effort needed during the segmentation of the tumor.

Comparison of Modified UNet with the Orginal UNet model

Experiments were done to show the effect of filter reduction and the application of data augmentation on the overall model performance. Table 6 demonstrates the results of the models with original and reduced number of filters and their performance before and after applying a data augmentation.

Table 6: Experimental results for liver and tumor segmentation with filter size reduction and application of data augmentation.

	Liver (In DSC)	**Tumor from liver (In DSC)**	**Tumor from abdominal CT image (In DSC)**
UNet (with original filter size)	0.9529	0.7384	0.6743
Modified UNet without data augmentation	0.9027	0.0992	0.0287
Modified UNet with data augmentation.	0.9612	0.74	0.63

As Table 6 shows, reducing the filter size didn't reduce the models performance, rather it shows small improvements in both liver and tumor segmentation and the training time is also reduced by about 1/3. The model performance is checked with and without data augmentation. Without data augmentation, it shows overfitting. It was good during the training but it is worse at testing time especially for tumor segmentation since most of tumors are very small.

Conclusion

This thesis focused on segmenting the liver and its tumor using a deep learning method. The method consists of two modified UNet models for the liver and the tumor segmentation. Using this algorithm a DSC of 0.96 and 0.74 for the segmentation of liver and tumor respectively were attained which shows an improvement of around 0.009 and 0.11 for liver and tumor segmentation respectively. This improvement was obtained due to the reduction of the class imbalance occurred in the data manually by removing unnecessary images and the selection of good hyperparameters through many trials.

List of abbreviations

2D	Two dimensional
3D	Three dimensional
BN	Batch normalization
CT	Computed tomography
DSC	Dice similarity coefficient
FCN	Fully convolutional network
GPU	Graphical processing unit
HCC	Hepatocellular carcinoma
IRCAD	Image Reconstruction for Comparison of Algorithm Database
JSC	Jaccard similarity coefficient
LITS	Liver Tumor Segmentation
MRI	Magnetic resonance imaging
ReLU	Rectified linear unit
SVD	Symmetric volume difference
TPU	Tensor processing unit

Declaration

Authors' contribution

The study was written, drafted and done by YA. MA had a great role during coding and editing the document. KA had read and gave his valuable comments. All authors read and approved the final manuscript.

Funding

No specific grant from any funding agency.

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Acknowledgment

Not applicable

References

[1] A. M. Anter, A. E. Hassanien, and G. Schaefer, "Automatic Segmentation and Classification of Liver Abnormalities Using Fractal Dimension," *2013 2nd IAPR Asian Conf. Pattern Recognit.*, pp. 937–941, Nov. 2013.

[2] F. Chung Choong Lon, "Regional appearance modeling for deformable model-based image segmentation Modélisation de l'apparence de régions pour la segmentation d'images basée modèle," Paris Institute of Technology, France, 2011.

[3] J. C. P. D, "Physiology of the liver," vol. 4, no. 8, pp. 13–24, 2017.

[4] J. Chiang, *Liver Physiology: MetaboLism and Detoxification*. Elsevier Inc., 2014.

[5] H. Alahmer and A. Ahmed, "Computer-aided Classification of Liver Lesions from CT Images Based on Multiple ROI," in *Procedia Computer Science*, 2016, vol. 90, pp. 80–86.

[6] WORLD HEALTH RANKING, "Liver Cancer in Ethiopia," 2017. [Online]. Available: http://www.worldlifeexpectancy.com/ethiopia-liver-cancer. [Accessed: 24-May-2018].

[7] "Liver Cancer - Types, Causes, Symptoms, Diagnosis, Treatment, Prevention, Progosis." [Online]. Available: https://www.medindia.net/patients/patientinfo/liver-cancer.htm. [Accessed: 10-Feb-2020].

[8] G. Chlebus, A. Schenk, J. H. Moltz, H. K. Hahn, and H. Meine, "Deep learning based automatic liver tumor segmentation in CT with shape-based post-processing," *Int. Conf. Med. Imaging with Deep Learn.*, pp. 1–9, 2018.

[9] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J. Digit. Imaging*, vol. 3, pp. 582–596, 2019.

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

[11] C. Sun *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.

[12] Patrick Ferdinand Christ *et al.*, "Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields," *Springer*, pp. 415–423, 2016.

[13] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," *Sci. Rep.*, vol. 8, no. 1, p. 15497, Dec. 2018.

[14] J. Li *et al.*, "A fully automatic computer-aided diagnosis system for hepatocellular carcinoma using convolutional neural networks," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 238–248, 2019.

[15] Ü. Budak, Y. Guo, E. Tanyildizi, and A. Şengür, "Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation," *Med. Hypotheses*, vol. 134, no. 109431, Jan. 2019.

[16] C. Sun *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, 2017.

[17] "3Dircadb | IRCAD France." [Online]. Available: https://www.ircad.fr/research/3dircadb/. [Accessed: 13-Feb-2020].

[18] "LiTS - challenge and data," *CodaLab - Competition*, 2017. [Online]. Available: https://competitions.codalab.org/competitions/17094. [Accessed: 13-Feb-2020].

[19] A. Raj and M. Jayasree, "Automated Liver Tumor Detection Using Markov Random Field Segmentation," *Procedia Technol.*, vol. 24, pp. 1305–1310, Jan. 2016.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[21] L. Meng, Y. Tian, and S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale," *J. Appl. Clin. Med. Phys.*, vol. 21, no. 1, pp. 144–157, Jan. 2020.

[22] V. Bevilacqua *et al.*, “A novel approach for Hepatocellular Carcinoma detection and classification based on triphasic CT Protocol,” in *A Novel Approach for Hepatocellular Carcinoma Detection and Classification Based on Triphasic CT Protocol*, 2017, pp. 1856–1863.

[23] M. Ahmad *et al.*, “Deep Belief Network Modeling for Automatic Liver Segmentation,” *IEEE Access*, vol. 7, pp. 20585–20595, 2019.

Figures

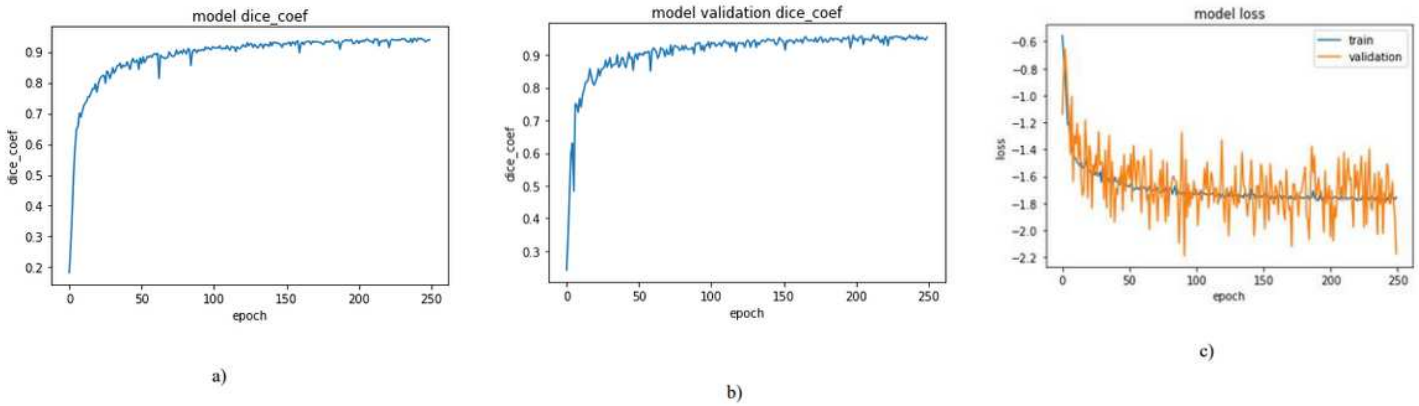


Figure 1

DSC and loss plot of the model during training for liver segmentation. (a) and (b) are model DSC for training and validation data respectively and (c) is model loss.

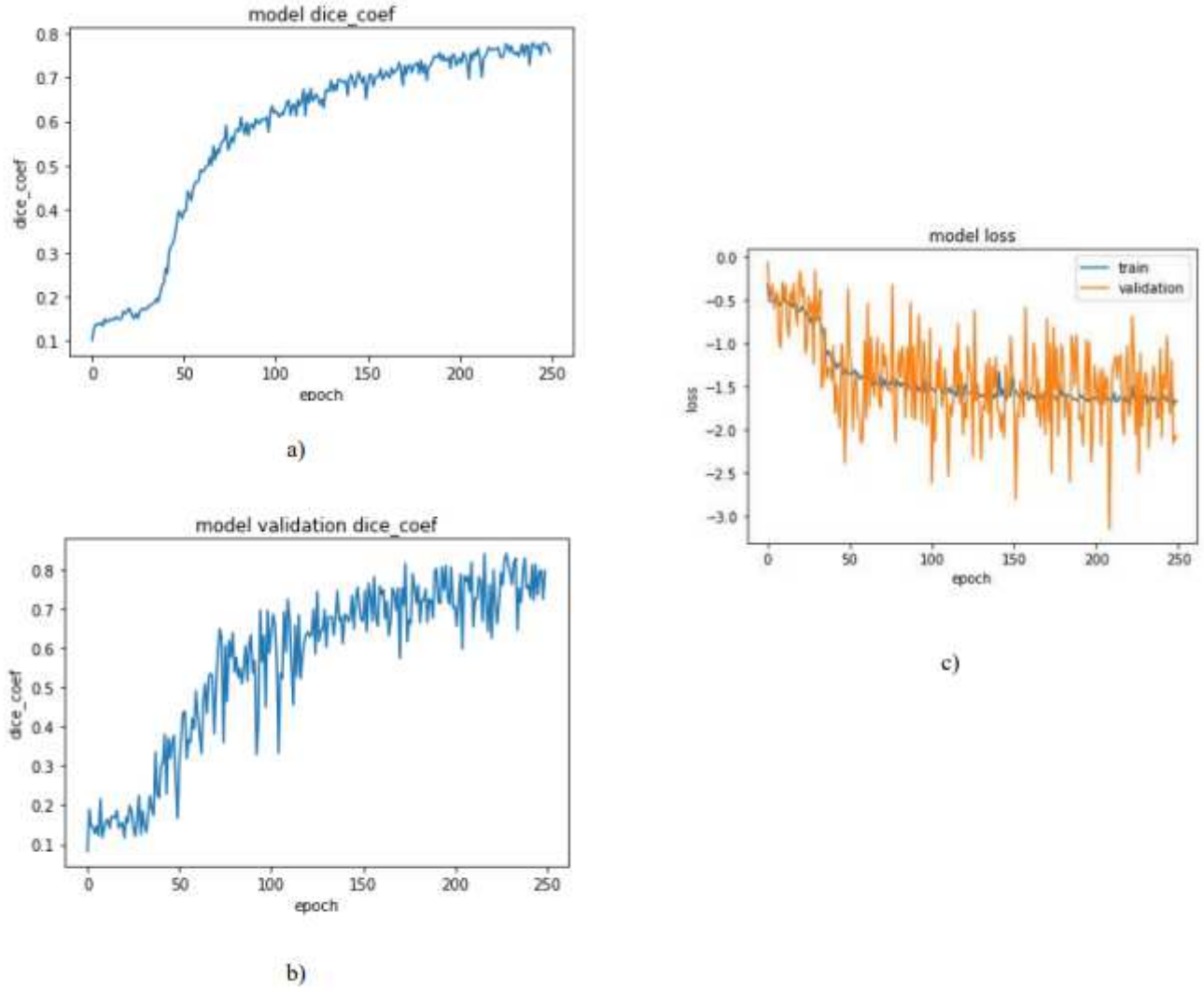
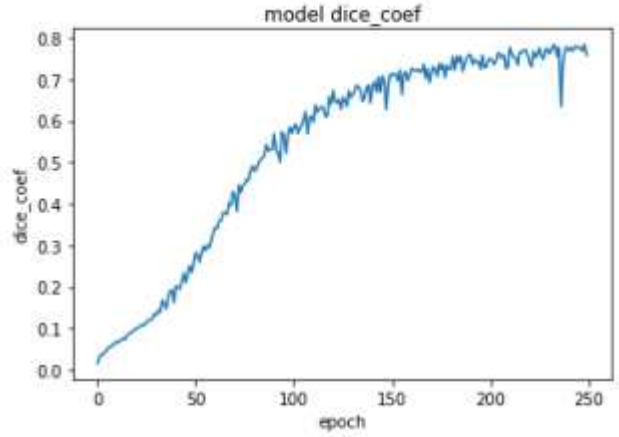
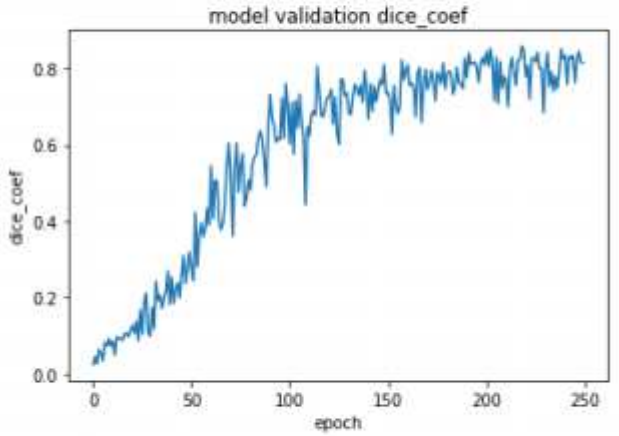


Figure 2

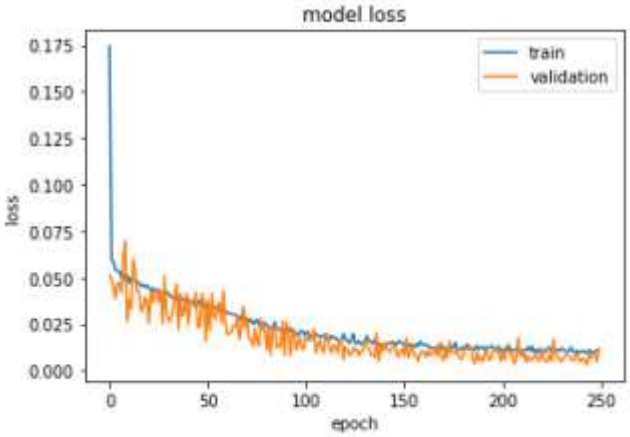
DSC and loss plot of the model for tumor segmentation from the liver. (a) and (b) are model DSC for training and validation data respectively where (c) is the model loss.



a)



b)



c)

Figure 3

DSC and loss plot of the model for tumor from the abdominal CT scan images during training. (a) And (b) are model DSC for training and validation data respectively and (c) is a model loss.

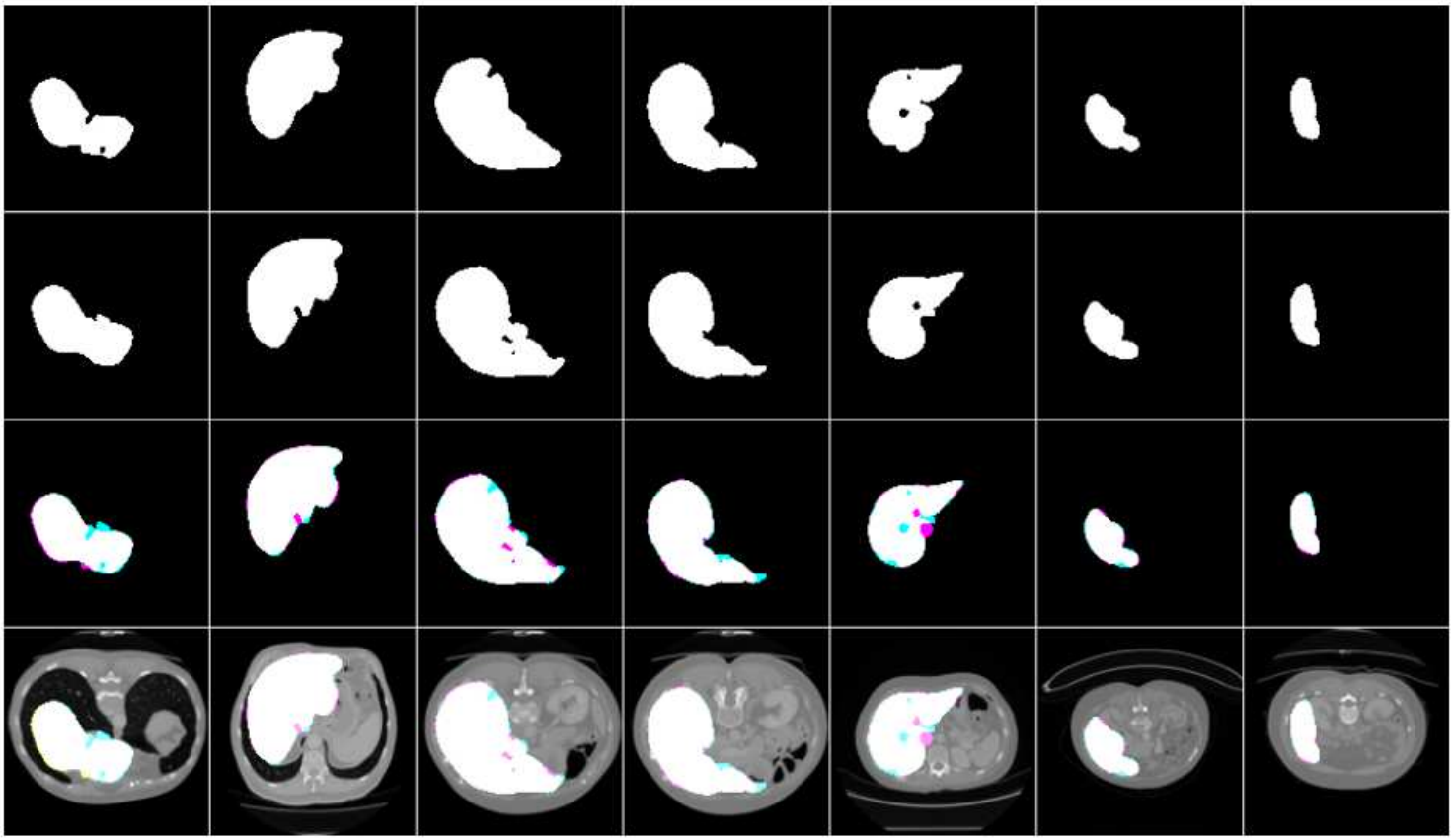


Figure 4

Liver segmentation results, ground truths, and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the original CT scan image.

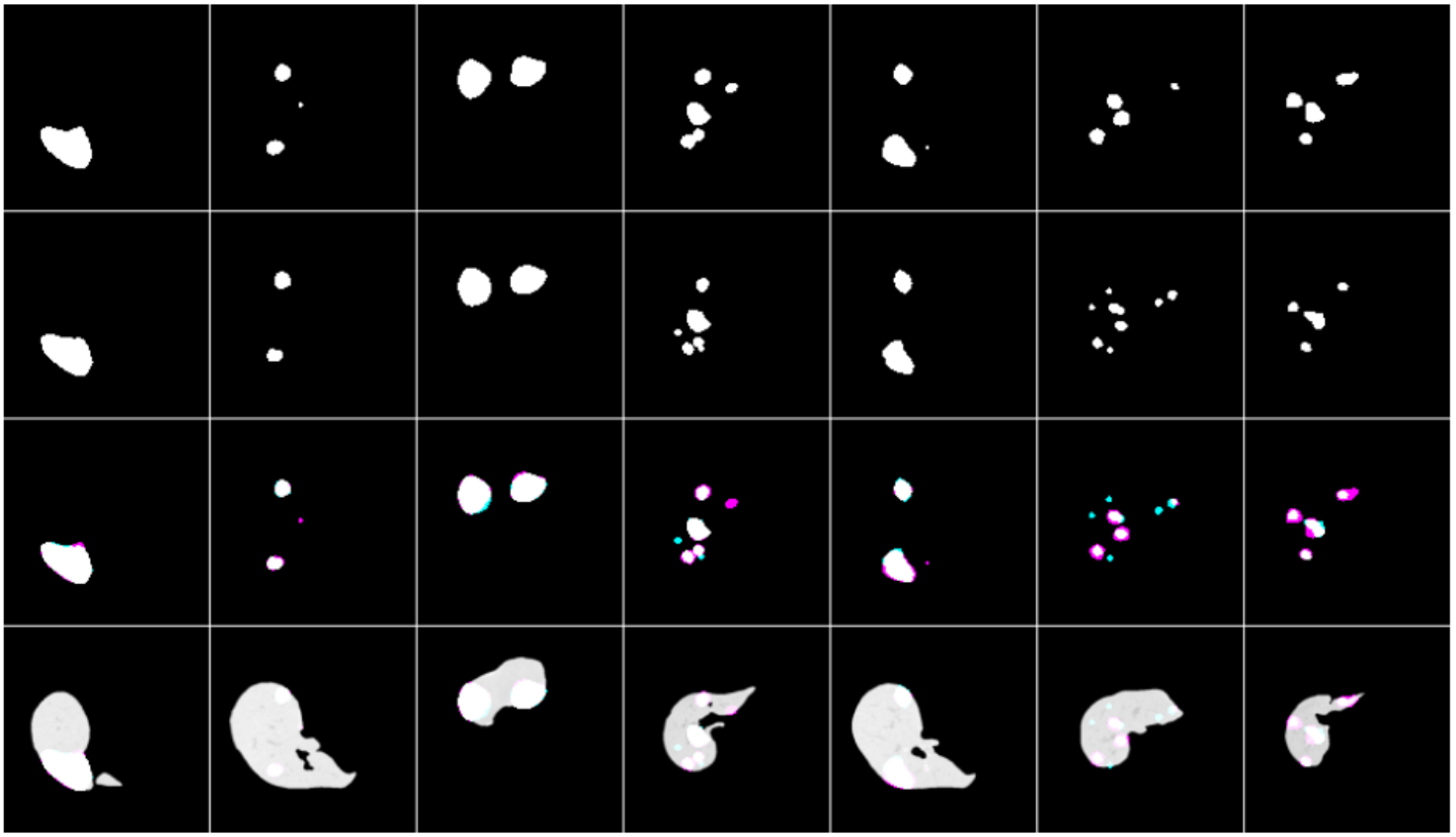


Figure 5

Tumor segmentation results from the liver with the respective masks and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the liver.

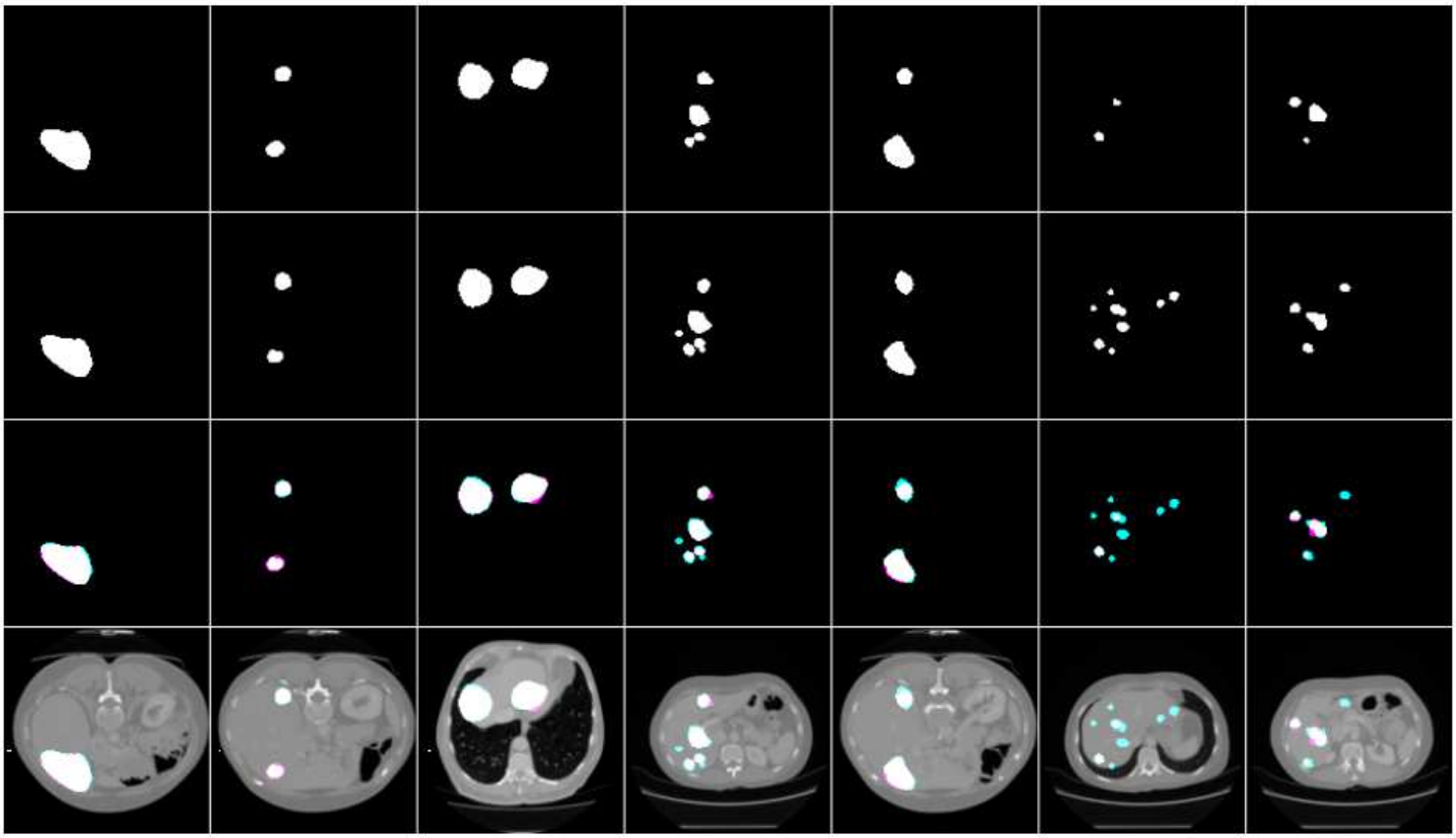


Figure 6

Result of tumor segmentation from abdominal CT images with the respective masks and overlap images. Row 1 shows the result of the model, row 2 shows the respective masks, row 3 shows overlap images of result with a mask. And row 4 shows both results and mask on the abdominal CT image.

Figure 7

Class imbalance between liver and tumor with the background. (a) shows the liver and (b) shows the tumor.