

Article

Deep Learning Algorithms in the Automatic Segmentation of Liver Lesions in Ultrasound Investigations

Mădălin Mămuleanu ^{1,2,*}, Cristiana Marinela Urhuț ³, Larisa Daniela Săndulescu ⁴, Constantin Kamal ^{2,5}, Ana-Maria Pătrașcu ^{2,6}, Alin Gabriel Ionescu ^{2,7}, Mircea-Sebastian Șerbănescu ^{2,8} and Costin Teodor Streba ^{2,4,5}

¹ Department of Automatic Control and Electronics, University of Craiova, 200585 Craiova, Romania

² Oncometrics S.R.L., 200677 Craiova, Romania

³ Department of Gastroenterology, Emergency County Hospital of Craiova, 200642 Craiova, Romania

⁴ Department of Gastroenterology, Research Center of Gastroenterology and Hepatology, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania

⁵ Department of Pulmonology, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania

⁶ Department of Hematology, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania

⁷ Department of History of Medicine, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania

⁸ Department of Medical Informatics and Statistics, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania

* Correspondence: madalin.mamuleanu@edu.ucv.ro; Tel.: +4-0762-893-723



Citation: Mămuleanu, M.; Urhuț, C.M.; Săndulescu, L.D.; Kamal, C.; Pătrașcu, A.-M.; Ionescu, A.G.; Șerbănescu, M.-S.; Streba, C.T. Deep Learning Algorithms in the Automatic Segmentation of Liver Lesions in Ultrasound Investigations. *Life* **2022**, *12*, 1877. <https://doi.org/10.3390/life12111877>

Academic Editor: Ke-Vin Chang

Received: 1 November 2022

Accepted: 9 November 2022

Published: 14 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: The ultrasound is one of the most used medical imaging investigations worldwide. It is non-invasive and effective in assessing liver tumors or other types of parenchymal changes. Methods: The aim of the study was to build a deep learning model for image segmentation in ultrasound video investigations. The dataset used in the study was provided by the University of Medicine and Pharmacy Craiova, Romania and contained 50 video examinations from 49 patients. The mean age of the patients in the cohort was 69.57. Regarding presence of a subjacent liver disease, 36.73% had liver cirrhosis and 16.32% had chronic viral hepatitis (5 patients: chronic hepatitis C and 3 patients: chronic hepatitis B). Frames were extracted and cropped from each examination and an expert gastroenterologist labelled the lesions in each frame. After labelling, the labels were exported as binary images. A deep learning segmentation model (U-Net) was trained with focal Tversky loss as a loss function. Two models were obtained with two different sets of parameters for the loss function. The performance metrics observed were intersection over union and recall and precision. Results: Analyzing the intersection over union metric, the first segmentation model obtained performed better compared to the second model: 0.8392 (model 1) vs. 0.7990 (model 2). The inference time for both models was between 32.15 milliseconds and 77.59 milliseconds. Conclusions: Two segmentation models were obtained in the study. The models performed similarly during training and validation. However, one model was trained to focus on hard-to-predict labels. The proposed segmentation models can represent a first step in automatically extracting time-intensity curves from CEUS examinations.

Keywords: ultrasound; image segmentation; hepatocarcinoma; liver tumor

1. Introduction

The ultrasound (US) is one of the most used medical imaging investigations worldwide. It is a cheap, safe, and effective modality that can detect a large range of lesions, especially in the case of parenchymatous organs. Therefore, it is especially effective in assessing liver tumors and other types of parenchymal changes; this in turn makes it a prime investigation in the screening of malignancies [1]. This becomes even more important when dealing with at-risk populations, such as patients suffering from hepatitis or cirrhosis of either viral or noninfectious origin. Liver ultrasound is routinely performed by different

medical specialties, depending on local regulations and after completing a training course and obtaining the necessary competencies [2]. These training programs vary in length, number of required steps, or addressability. Point-of-care ultrasound (POCUS) has become increasingly utilized worldwide as a screening method, being performed in various medical settings from the emergency room to the general practitioner's office [3].

However, US has become an indispensable tool when diagnosing liver cancer, mainly due to the application of contrast agents. Safe, reliable, and minimally invasive, contrast-enhanced US (CEUS) can be applied to almost any patient due to virtually non-existent allergic reactions or organ dysfunction that may restrict its usage. The contrasting agent relies on gas microbubbles that are injected intravenously, reach liver vasculature, are degraded intravascularly under ultrasound, the produced gas being harmlessly excreted through the lungs. Liver tumors, especially hepatocellular carcinoma (HCC) produce specific filling patterns during CEUS [4]. The physician relies on following a fixed region containing the lesion over a long period of time and studying the contrast uptake and later wash-out pattern, thus indicating the possible diagnosis. The interpretation of CEUS strongly relies on the experience of the performing physician [5]. One reliable method to quantify contrast uptake and, subsequently, describe tumor vasculature, is the generation of time intensity curves (TIC) that can be later analyzed, providing a diagnosis of malignancy [6].

Multiple artifacts pertaining to breathing motion or movements of the US probe may degrade the quality of TICs generated during a normal investigation of the liver. Artificial intelligence can be employed to segment the liver parenchyma in normal B-mode US to maintain a stable area of interest (AOI) around the focal liver lesion that needs characterization. The procedure to manually adjust the AOI is time-consuming and subject to multiple sources of error [7].

Image segmentation is a computer vision technique in which each pixel from an image has a specific class associated with it. Depending on the purpose of the segmentation, the output image can have two classes (binary image) or more than two classes. In the field of medical imaging, both types of image segmentation techniques are used. Two class segmentation (binary images) is used as a mask to extract pathological or physiological information about a patient [8]. In medical imaging, multi-class image segmentation (semantic segmentation) is used to classify and extract different types of lesions [9]. In the last years there have been many solutions proposed for medical image segmentation based on artificial intelligence (AI) algorithms. These solutions can be classified as machine learning (ML) approaches and deep learning (DL) approaches. Several ML algorithms are used for image segmentation, among them are k-means and fuzzy c-means [10]. As for DL algorithms, several models based on convolutional neural networks were proposed for medical imaging segmentation. The U-Net model proposed by Olaf Ronneberger et al. [11] is a fully convolutional neural network for biomedical image segmentation. Other models were derived from the U-Net model, such as V-Net, proposed by Fausto Milletari et al. [12], are used for volumetric medical segmentation. Attention U-Net proposed by Ozan Oktay et al. [13] is used on medical image segmentation to focus on target structures. Another derived work from U-Net is U-Net++, proposed by Zongwei Zhou et al. [14], which is a nested U-Net architecture tested in medical image segmentation tasks such as nuclei segmentation, polyp segmentation, and liver segmentation. For the segmentation of the liver and liver lesions in magnetic resonance imaging (MRI) and computer tomography (CT) different authors have proposed methods based on ML or DL algorithms. Jose Denes Lima Araújo et al. [15] proposed a segmentation of liver from CT images using a U-Net model. Sebastian Nowak et al. [16] proposed a method for detecting liver cirrhosis in T2-weighted MRI scans. Their pipeline contained two DL models, one with the goal of performing liver segmentation on the T2-weighted MRI image and the second with the goal of classifying the segmented image. To date, there are not many works concerning liver lesions segmentation in ultrasound (US) investigations using DL or ML algorithms. The segmentation of the liver lesions in US is a challenging task due to interference noise

and missing boundaries of lesions [17]. In their work, Nishan Jain et al. [18] proposed a method for US liver segmentation using region difference filters. In their method, four different filters were applied to the image obtaining a region-difference image. A pixel was manually selected from the lesion. The resulting mask was computed by transforming the region-difference image into a binary image and the ROI was considered as the nearest edges which enclosed the selected pixel. Their dataset contained 56 B-mode ultrasound investigations. In terms of performance metrics, the average accuracy was measured in their study, obtaining values from 65.2 to 99.68%. Deep Gupta et. al. [19] proposed a hybrid segmentation method for US images. The method was based on the Gaussian kernel-based fuzzy c-means clustering algorithm and region based active contour model. Based on the authors conclusions, their method provided better accuracy compared with other methods such as fuzzy c-means clustering or geodesic active contours. Their dataset contained 50 US investigations. However, the US images did not contain only liver investigations.

Our aim was to build a deep learning model for automatic, real-time image segmentation during US examinations to open new opportunities in automatically computed time-intensity curves. The obtained model should have a maximum inference time of 100 milliseconds to be a baseline for creating a system to perform liver lesions classification in contrast-enhanced ultrasound investigations. Thus, besides evaluating the model in terms of performance metrics, the average inference time was also observed on two different environments. As mentioned earlier, US has become an indispensable tool when diagnosing liver cancer. It is cheap, non-invasive and can detect different types of lesions. However, US investigations have noise which can affect the interpretation of the results. A DL model which can perform image segmentation of liver lesions can be a useful tool for a gastroenterologist, especially when extracting time-intensity curves. Compared to the studies mentioned earlier, our proposed method used a DL model for the segmentation of liver lesions. More specifically, the architecture of the DL model is U-Net [11]. U-Net is an “U” shaped DL model architecture with two connected parts: the encoder and the decoder. Due to this architecture, U-Net required a single run on an image to predict the mask.

2. Materials and Methods

2.1. Data Acquisition

The dataset used in our study was provided by the University of Medicine and Pharmacy Craiova, Romania. The contrast-enhanced ultrasound examinations were performed using a Hitachi Aloka Arietta V70 350 (Hitachi Medical Corporation, Tokyo, Japan), equipped with the convex probe C 251 (Hitachi Medical Corporation, Tokyo, Japan). For CEUS, the contrast agent used was SonoVue (Bracco SpA, Milano, Italia). The dataset contained 50 video files from 49 consecutive patients with 59 liver lesions. Each video file was in audio video interleave (AVI) format encoded with motion jpeg and with a bit rate of 31,196 kb/s. The video files had between 7 and 12 frames per second. The video files did not contain the entire examination, they were divided into three phases: arterial, portal venous, and late phase. However, it was not an impediment as in this stage since each video file was treated individually without the need to know which file belongs to which patient and what lesion type an image contains. A more detailed analysis of the dataset is documented in [20].

2.2. Data Preprocessing

Before extracting frames from the video files, the region in B-mode was defined for each type of ultrasound device. This region was defined to automatically crop and extract frames from the video examinations in the dataset. Table 1 contains these regions for all the ultrasound devices used in this study. The values for x and y axis were determined experimentally by plotting the frame and checking the coordinates for B-mode.

Table 1. Coordinates for the edge of B-mode image.

X Min	X Max	Y Min	Y Max
0	400	78	525

Figure 1 shows how these regions were determined. After the x and y coordinates were determined for each ultrasound device, the next step was frame extraction. As mentioned, each video examination had between 7 and 12 frames per second. The changes of lesion in terms of pixels from one second to another were relatively small, therefore we have extracted the frames with a sample time of 1 s. Extracting and marking all the frames from each video examination would have resulted in marking almost the same image multiple times. Algorithm 1, used for extracting the frames from the videos, is presented below.

Algorithm 1. Extracting frames from the video examination

```

fps ← 0
ultrasound_device ← ultrasound_device_characteristics_object
fps ← ultrasound_device.getFPS()
frame_height ← read frame height of the video examination
frame_width ← read frame width of the video examination
b_mode_x_min ← ultrasound_device.b_mode_x_min
b_mode_x_max ← ultrasound_device.b_mode_x_max
b_mode_y_min ← ultrasound_device.b_mode_y_min
b_mode_y_max ← ultrasound_device.b_mode_y_max
while video examination file still has frames do:
    frame_id ← get the frame id from the video file
    frame ← get the frame from the video file
    if frame_id % fps == 0 do:
        //Process the frame and save it to disk.
        Cropped_frame ← frame[b_mode_x_min: b_mode_x_max,
                               b_mode_y_min, b_mode_y_max]
        save cropped_frame to disk
    else:
        continue //ignoring the current frame
  
```

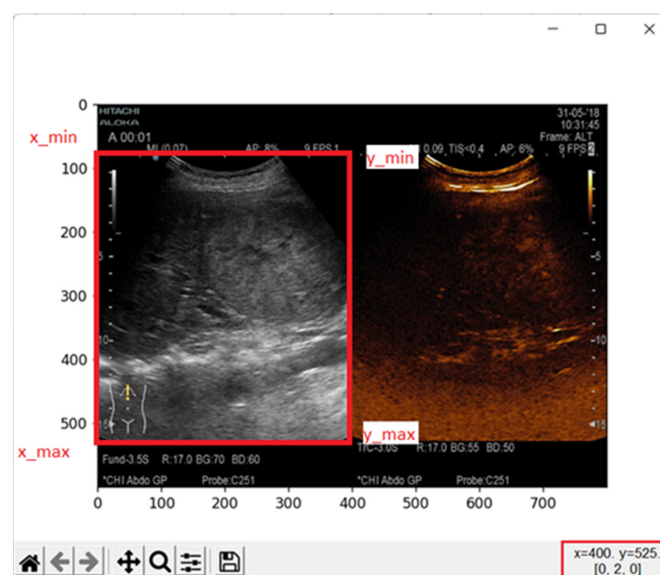


Figure 1. Plotting the first frame of the video examination and determining the edges of the B-mode image. Bottom right coordinate presented in red box (X max and Y max).

Extracting the frames from each video resulted in a total number of 6035 image files in B-mode. The region of interest (ROI) for each image was marked by a senior gastroenterologist (L.D.S.) with over 20 years of experience in abdominal ultrasound interpretation. For annotation of the region of the interest, QuPath software was used [21]. Figure 2 shows a screen capture from QuPath with a liver lesion marked by the expert gastroenterologist (L.D.S.). Examples of frames extracted by Algorithm 1 are presented in Figure 3.

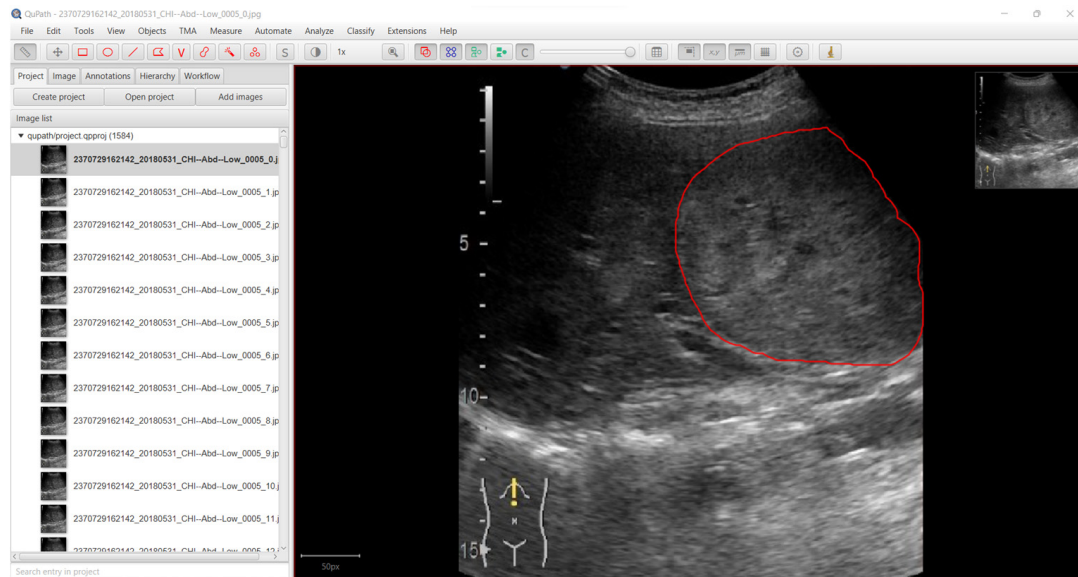


Figure 2. QuPath, the software used to mark the lesions and to apply Algorithm 2 for mask generation.

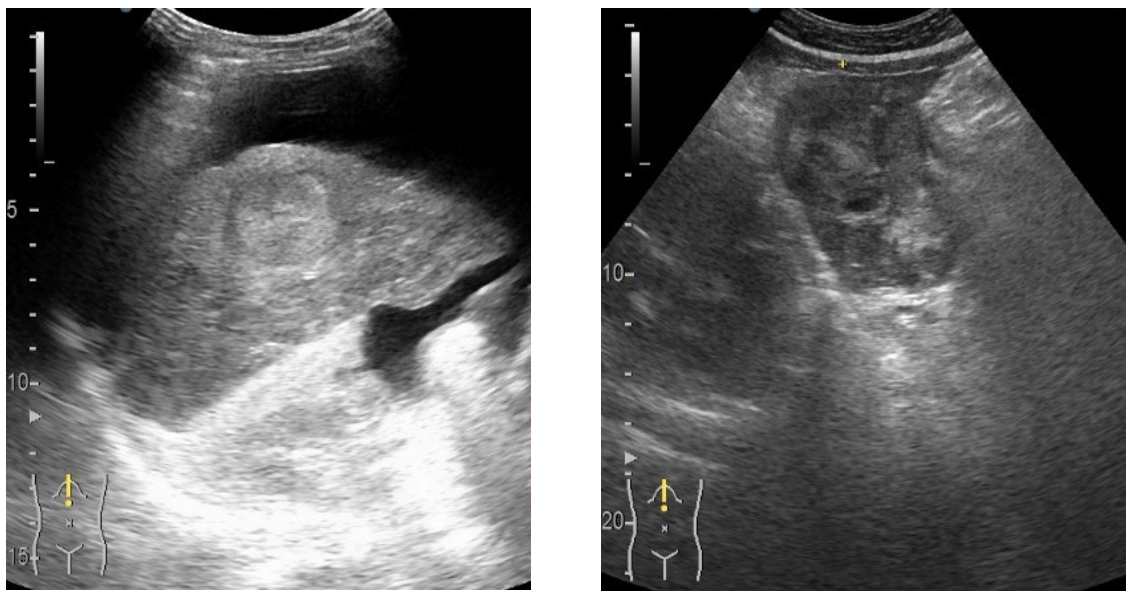


Figure 3. Example of frames extracted by Algorithm 1.

For developing the algorithm, for each image in B-mode the corresponding label had to be created. The label was defined as a binary image with the same size as the input image. As shown in Figure 2, each lesion was marked as a shape on top of the B-mode cropped image. After every lesion was marked by the expert gastroenterologist, in order to export the binary images (masks) from QuPath [21], Algorithm 2 was applied for every image in the project. Example of masks obtained from Algorithm 2 are presented in Figure 4. The

masks presented in Figure 4 were obtained by running the frames presented as an example in Figure 3.

Algorithm 2. Mask creation (binary image)

```

current_image ← obtain current image
while current_image is not null do:
    mask ← new Image(current_image.width, current_image.height, values = 0)
    for each object in annotation_list do:
        roi ← object.getROI()
        roi.fill(values = 1)
        mask ← mask bitwise and roi
    mask_filename ← string concatenation (current_image.name, "-mask")
    save image to disk (mask, mask_filename)

```

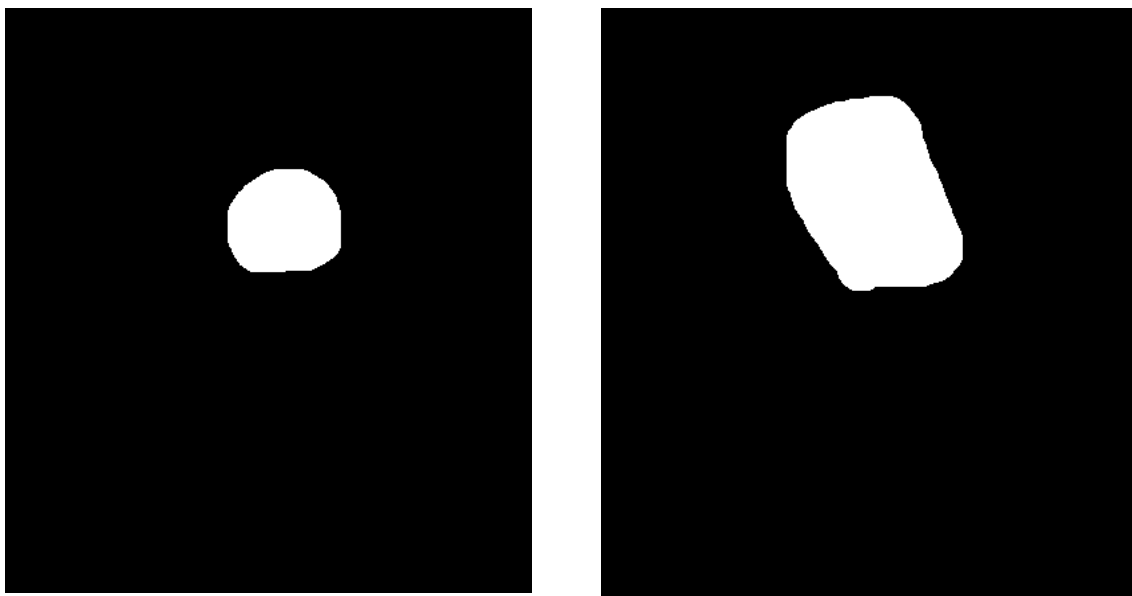


Figure 4. Example of masks obtained from Algorithm 2.

After data preprocessing, 6035 images in B-mode were obtained (with 6035 corresponding masks). The dataset preparation pipeline is presented in Figure 5. The video examination file was passed through Algorithm 1 to extract the frames. The sample time used was 1 second. The frames obtained were labelled and then each frame was passed through Algorithm 2 to obtain the mask (a binary image in which the black color represented normal tissue and the white color represented the lesion(s)). Algorithm 1 was implemented in Python programming language since the main goal of the study was to perform real-time image segmentation. Thus, Algorithm 1 will be reused in further developments. On the other hand, Algorithm 2 was developed using QuPath Automate function. The QuPath script editor (Automate) allowed us to process the images in the current project in batch. Hence, Algorithm 2 was implemented in QuPath to easily export the masks created by the expert gastroenterologist L.D.S. (Figure 2).

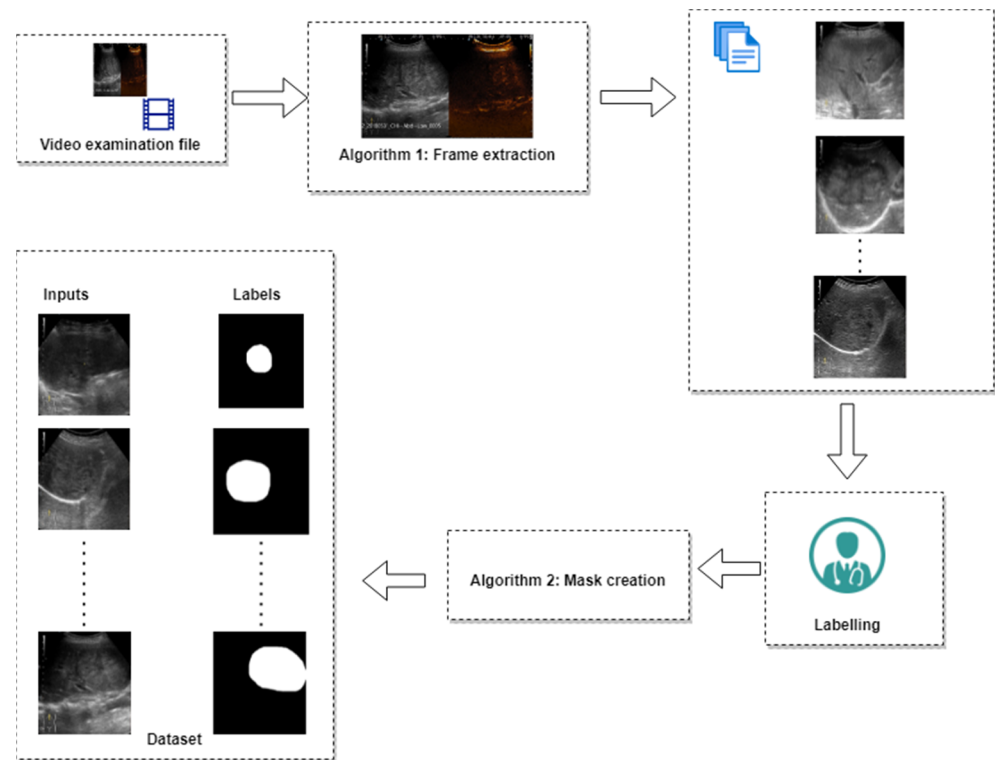


Figure 5. Dataset preparation pipeline.

2.3. Neural Network Model

In our study, the goal of the deep learning model was to perform semantic segmentation to each frame of the video examination to properly identify the lesion location. To accomplish this, we have trained a U-Net model proposed by Olaf Ronneberger et al. [11] with an input size of 256 by 256 pixels. All the images and their corresponding masks were resized (without locking aspect ratio) to fit the proposed model. U-Net is an “U” shaped fully convolutional network. The architecture has a down sampling branch (encoder) and an up-sampling branch (decoder). U-Net and other versions derived from it are used in many applications for biomedical imaging segmentation like lung ultrasound segmentation [22], cardiac magnetic resonance imaging segmentation [23], bone segmentation [24], or bones lesions segmentation [25]. Due to its architecture, U-Net can extract discriminative features from the raw images with limited training data. Segmentation of a 512 by 512 image takes less than a second [11]. The input in the U-Net architecture used in the study was a tensor with a size of 256 by 256 pixels. The encoder part of the model contained 4 blocks of convolutional layers. Each block applied two convolutional filters with a kernel size of 3 by 3 pixels and no padding. After each block of convolutional layers, a max pooling operation was performed with size 2×2 and stride 2. This operation was performed for down sampling the input tensor. The first convolutional block had 64 filters. The number of filters doubled after each max pooling operation. The scope of the encoder part of the model was to decrease resolution and increase depth to capture the context. The activation function in the encoder part for all the convolutional layers was a rectified linear unit (ReLU). The decoder part of the model contained 4 blocks of convolutional layers. Similarly to the encoder part, it applied two convolutional filters with a kernel size of 3 by 3 pixels without padding. After each block of convolutional layers, an up-sampling operation with kernel size 2×2 and nearest neighbor interpolation was performed. The resulting model had a total of 412,865 trainable parameters from a total of 414,401.

2.4. Hyperparameters and Loss Function

When training a neural network, the dataset should be split into training data, validation data, and testing data. The training and validation groups are used during training while the testing group it is used for testing the neural network. For the proposed model, we randomly divided the dataset as follows: 70% for training, 20% for validation, and 10% for testing. The model was trained for 25 epochs with a batch size of 8. The optimizer used for the model was Adam, an optimization algorithm with adaptive learning rate created specifically for training neural networks. The optimizer is computationally efficient since it can find individual learning rates for each parameter. It requires the following inputs: alpha (α)—the learning rate or step size, beta (β_1) and beta2 (β_2)—the exponential decay rates for the moment estimates and a very small number, and epsilon (ϵ), to prevent division by zero [26]. The values we have chosen for these inputs are presented in Table 2. For beta1, beta2, and epsilon the values recommended in [26] were used. The architecture of the proposed segmentation mode is presented in Figure 6.

Table 2. Adam optimizer parameters.

α	β_1	β_2	ϵ
0.0001	0.9	0.999	10^{-8}

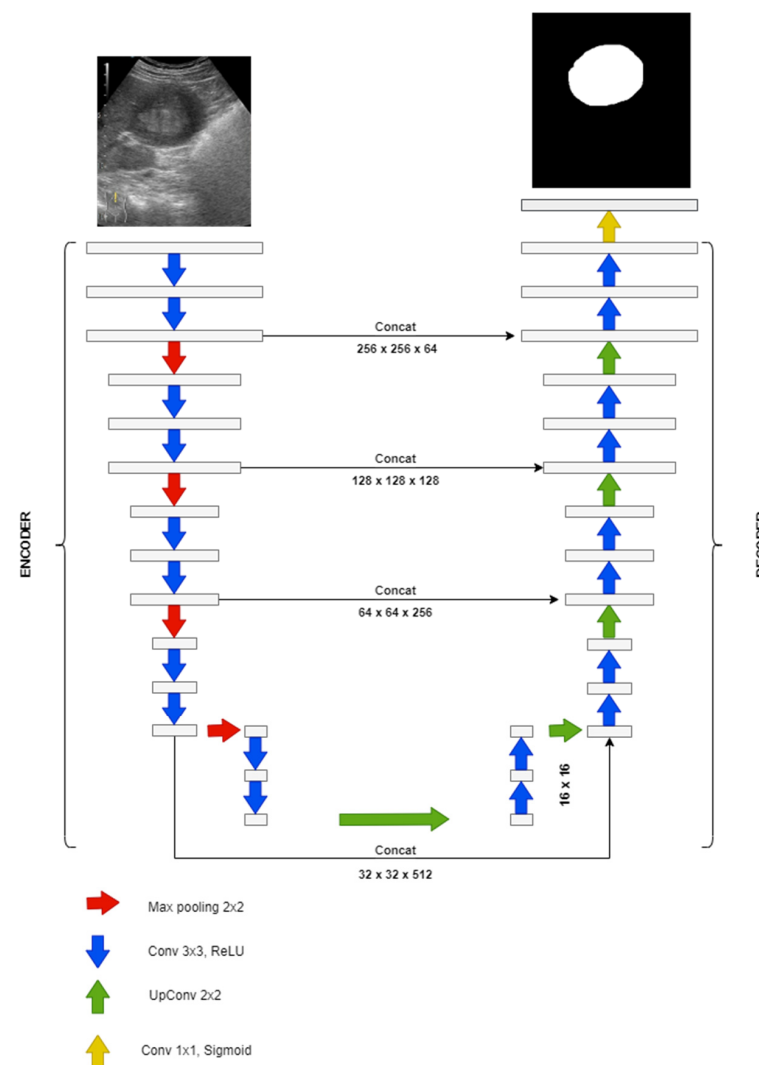


Figure 6. Architecture of the proposed segmentation model.

Since the goal of the proposed model was to predict whether a certain pixel in an ultrasound image belonged to a liver lesion or not, it can be assumed that the model was a binary classifier. Typically, in a binary classification task, the loss function used is binary cross entropy (BCE). Cross entropy is a method to measure the difference between two probability distributions and it is given by Equation (1).

$$CE = - \sum_{i=1}^C t_i \log(s_{o,i}) \quad (1)$$

where t_i is the truth label, $s_{o,i}$ is the predicted probability observation, o is of class i , and C is the total number of labels. In binary classification, the total number of classes is 2. Replacing C in Equation (1) we obtain (2). In binary cross entropy, when the predicted probability approaches 1, the loss decreases, and when the predicted value is decreasing, the loss starts to increase very fast. From Equation (2) it can be concluded that binary cross entropy penalizes equally for each class. For the proposed image segmentation model, the classes were not balanced, meaning that the pixels with lesions and the pixels with no lesions were not equally distributed into an image. Hence, for the proposed model, the loss function used was focal Tversky loss [16]. The loss function is given by Equations (3) and (4). The model was trained with two sets of α , β , and γ as described in Table 3.

$$BCE = -(y \log(p) + (1 - y)(1 - p)) \quad (2)$$

Table 3. Model performance results during training and validation.

Parameters	IoU	Recall	Precision
$\alpha = \beta = 0.5$ (Dice coefficient) $\gamma = 1$ (Training/Validation)	0.8392/0.7129	0.8911/0.8256	0.9334/0.8448
$\alpha = 0.7, \beta = 0.3$ $\gamma = 0.75$ (Training/Validation)	0.7990/0.6572	0.8171/0.7735	0.9635/0.8192

Tversky loss, also known as Tversky index (TI) [27] is an asymmetric similarity function (3) in which $X \setminus Y$ and $Y \setminus X$ represent the set difference of X and Y . α and β are two parameters of Tversky index and they must be greater than or equal to 0. If α and β both have the same value, more precisely 0.5, the Tversky index represents the dice coefficient [28]. In contrast, if α and β are set to 1, the Tversky index represents the Jaccard Index, also known as intersection over union (IoU). The Tversky index allows the proper balance of the false positives and false negatives. During training, TI can have serious problems in segmenting small liver regions since these regions do not have a major contribution to the loss. Focal loss [29] solves this problem by adding a factor to the TI to focus more on inputs that are hard to segment. Focal Tversky loss (FTL) is given by (4). γ is called a focusing parameter and it adjusts smoothly how the easy-to-segment inputs are down weighted [29].

$$TI(X, Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha |X \setminus Y| + \beta |Y \setminus X|} \quad (3)$$

$$FTL = (1 - TI)^\gamma \quad (4)$$

2.5. Experimental Setup

All the implementations were performed in Python version 3.9.0 with TensorFlow version 2.7.0 [30] on a machine equipped with Intel i7 Processor, 16 GB RAM, and Nvidia 3050Ti GPU. The operating system used was Windows 11 Pro.

2.6. Assessing the Performance of the Deep Neural Network Model

Using only pixel accuracy to measure the performance of the model can be misleading when evaluating the entire system. Assuming that a lesion in a frame from a US examination is only 10% of the entire image and the measured accuracy for that specific image is 90%, it can be concluded that the model is performing well. However, if only the non-lesion pixels were classified correctly, the model was not evaluated correctly. The issue of class imbalance in training deep learning algorithms can be corrected by applying oversampling techniques [31]. However, image segmentation scenarios, in which the liver tissue dominates the image or vice-versa, can be solved by observing more complex and precise metrics that deals with class imbalance. To properly assess the performance of the model, the metrics measured are intersection over-union (IoU or Jaccard index), area under curve (AUC), recall, and precision. Assuming that TP , TN , FP , and FN are true positives, true negatives, false positives, and false negatives, respectively, the metrics are defined by Equations (5)–(7). IoU or Jaccard index represents the overlap between model predictions and imagist segmentation divided by the union between model predictions and imagist segmentation. This metric is very useful in image segmentation to assess how much of the predicted mask overlaps with the mask created by the expert gastroenterologist.

$$IoU = \frac{TP}{TP + FP + FN} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (8)$$

$$IoU = \frac{Dice}{2 - Dice} \quad (9)$$

3. Results

We have built two models with the same architecture and hyperparameters; the only difference between these models were the α , β , and γ inputs for the FTL. The first model, for which the FTL had α equal with β equal with 0.5, did not focus on difficult-to-predict masks.

It can be observed from Table 3 that the U-Net segmentation model trained with parameters $\alpha = \beta = 0.5$ and $\gamma = 1$ performed better compared to the model trained with focus on masks which were difficult to predict, $\alpha = 0.7$ and $\beta = 0.3$ $\gamma = 0.75$. When loss function of the model was dice coefficient ($\alpha = \beta = 0.5$), both loss and IoU were converging faster compared with the model with parameters $\alpha = 0.7$, $\beta = 0.3$ $\gamma = 0.75$ for FTL (Figures 7 and 8). Furthermore, the IoU obtained for the two models were slightly different, with 0.8392 for the first model and 0.7990 for the second model. However, the first model was performing equally for both easy- and hard-to-predict inputs while the second model was performing relatively poorly on easy-to-predict inputs.

As for evaluating each model, the evolution of accuracy during training was plotted. Figure 9 shows the accuracy of both models during training and validation. It can be observed that the accuracy converged fast. While accuracy has similar values for both models, they were performing very differently, proving that accuracy is not enough to evaluate the performance of a segmentation model.

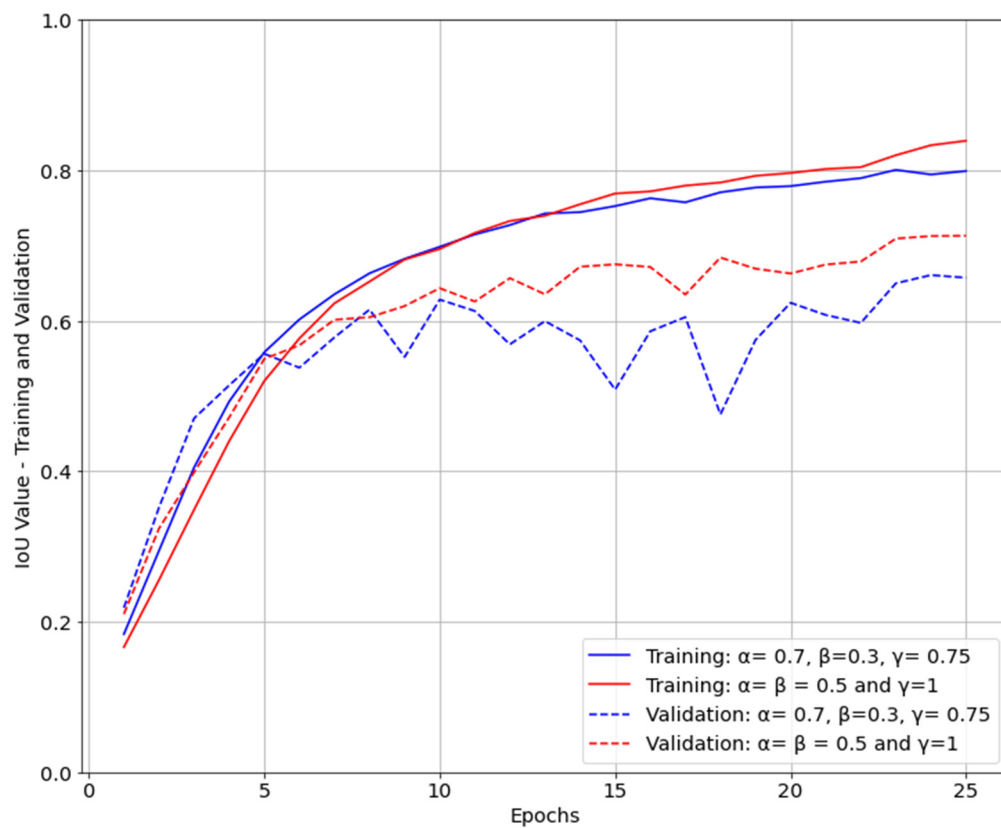


Figure 7. IoU during training and validation.

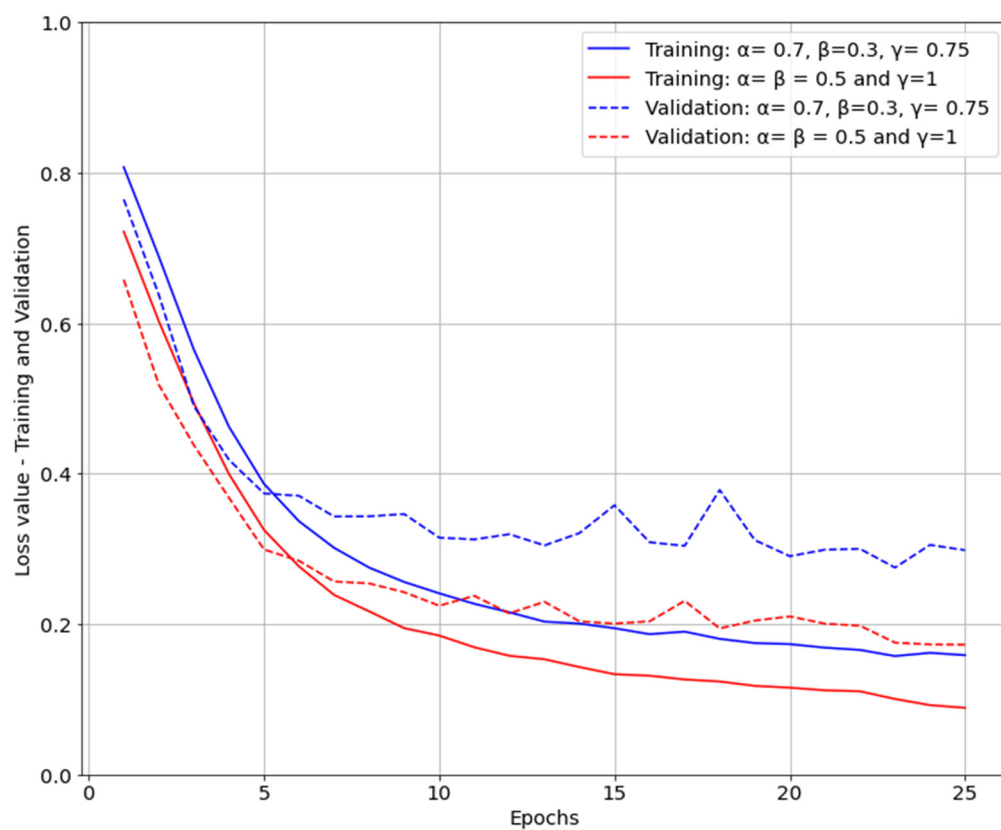


Figure 8. Loss evolution during training and validation.

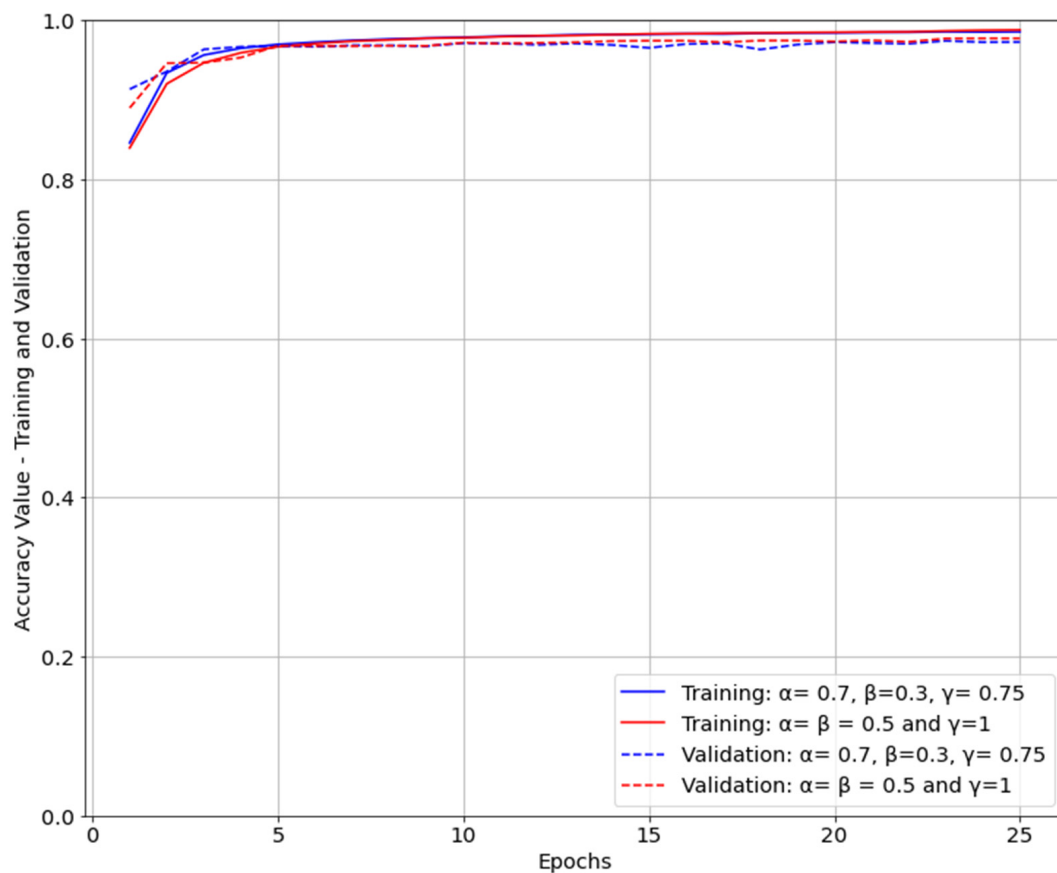


Figure 9. Accuracy evolution during training and validation.

Besides analyzing the performance of the models in terms of quality of the outputs, an analysis of the inference time, GPU memory utilization, floating point operations (FLOPs), and number of model parameters was carried out. The results were obtained on two different environments. The first test environment (ENV1) was configured on the same machine on which the segmentation models were trained: Intel i7 Processor, 16 GB RAM, and Nvidia 3050Ti GPU. The second test environment was run in the cloud using Google Colab [32]. The virtual machine provided in the Google Colab Jupyter notebook used for testing was configured as follows: Intel Xeon CPU at 2.00 GHz, 12 GB of RAM, and Nvidia Tesla T4 GPU. For testing the inference time, a clean Jupyter Kernel was used. Therefore, before running each test the runtime was restarted. From the entire dataset, 100 frames were randomly selected for testing. Each model was run on each of the selected frames in both test environments. Inference time was measured for both models for all the selected frames. Loading the model into the GPU was measured separately as this action was performed once per each test and it did not influence the inference time of the models. If the models were deployed into a system used in production, loading the model was performed only at the system startup. The metrics measured were minimum inference time, maximum inference time, and average inference time. The results are presented in Tables 4 and 5.

For analyzing the GPU memory utilization, FLOPs, and total number of model parameters, only ENV1 was used as these metrics were not GPU dependent. The results obtained are presented in Table 6.

Table 4. Inference time results for ENV1.

Model	Minimum Inference (Milliseconds)	Maximum Inference (Milliseconds)	Average Inference (Milliseconds)	Loading Time (Seconds)
$\alpha = \beta = 0.5$ (Dice coefficient)	32.50	56.48	41.76	294.29
$\gamma = 1$ $\alpha = 0.7, \beta = 0.3$ $\gamma = 0.75$	32.15	59.70	43.04	373.16

Table 5. Inference time results for ENV2.

Model	Minimum Inference (Milliseconds)	Maximum Inference (Milliseconds)	Average Inference (Milliseconds)	Loading Time (Seconds)
$\alpha = \beta = 0.5$ (Dice coefficient)	48.76	77.59	59.68	5.86
$\gamma = 1$ $\alpha = 0.7, \beta = 0.3$ $\gamma = 0.75$	51.90	76.43	61.15	7.89

Table 6. The complexity of the system.

Metric	Value	Unit
FLOPs	43.2433	MFLOPs
Memory requirement (GPU)	0.9291	GB
Total number of parameters	414,401	N/A

4. Discussion

The goal of the proposed study was to build a deep learning model which can perform image segmentation from CEUS video examinations. To train and validate the DL segmentation model, we used 59 liver tumors corresponding to 49 patients evaluated through ultrasound and contrast-enhanced ultrasound in the Gastroenterology Department of the Emergency Clinical County Hospital of Craiova between 7 January 2018 and 12 December 2020. Clinical data regarding age, gender, underlying liver disease, history of previous malignancy, the final diagnosis, and the standard method used for confirmation of the diagnosis were collected from the medical records of the patients found in the hospital system. Ultrasound investigations were performed using Hitachi Arietta V 70, convex probe C250. The contrast agent used for the examination was SonoVue (Bracco, Milan, Italy). CEUS examinations were stored as videos corresponding to arterial, portal, and delayed phases and were performed according to the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) guidelines. The majority of patients were male, 31 in total (63.26%). The mean age of the patients was 69.57 ± 10.65 . In what concerns the presence of a subjacent liver disease, 36.73% had liver cirrhosis and 16.32% had chronic viral hepatitis (5 patients: chronic hepatitis C and 3 patients: chronic hepatitis B). History of previous neoplasia was detected in 11 patients (22.44%). The mean size of the tumors was 51.63 mm. The definitive diagnosis of liver tumors was hepatic hemangioma ($n = 6$; 10.16%), focal nodular hyperplasia ($n = 1$; 1.69%), liver cysts ($n = 5$; 8.47%), liver abscess ($n = 1$; 1.69%), liver adenoma ($n = 1$; 1.69%); hepatocellular carcinoma ($n = 24$; 40.67%), cholangiocarcinoma ($n = 4$; 6.77%), liver metastases ($n = 15$; 25.42%), and malignant liver adenoma ($n = 1$; 1.69%). The final diagnosis was established by contrast-enhanced CT, MRI, or both. CEUS alone was used for typical liver hemangioma or hepatic metastases developed in a known clinical context of neoplasia. The reference diagnosis for uncertain

cases was obtained through pathology. A statistical overview of the liver lesions and patients involved in the study is presented in Table 7.

Table 7. Patient cohort involved in the study.

Variable	n ¹ (%)
Gender	M-63.26% F-36.74%
Age (mean value $\pm$ SD)	69.57 $\pm$ 10.65
Age Wise Classification of Samples	
Age group	Number of patients
<40	2
40–49	2
50–59	7
60–69	17
70+	21
Underlying liver disease	
1. Liver cirrhosis	36.73%
2. Chronic viral hepatitis	HBV-6.12% HCV-10.20%
History of previous malignancy	22.44%
Tumor size (mm), mean value	51.65
Final diagnosis	
Hepatic hemangioma	10.16%
Liver cysts	8.47%
Focal nodular hyperplasia	1.69%
Liver adenoma	1.69%
Liver abscess	1.69%
Hepatocellular carcinoma	40.67%
Liver metastases	25.42%
Cholangiocarcinoma	6.77%
Malignant liver adenoma	1.69%

¹ n = 49.

In Figure 10, an example of images from the validation batch is presented. The same video frame was run through both models. Figure 10c presents the predicted output for the model in which the values for the loss function were $\alpha = \beta = 0.5$ and $\gamma = 1$. Figure 10d presents the predicted output for model in which the values for the loss function were $\alpha = 0.7$, $\beta = 0.3$, and $\gamma = 0.75$. The image contains multiple liver lesions with different shapes and sizes. It can be observed that when the lesions are multiple and close to each other, both models predict a much larger region without separation. Moreover, in the predicted mask (Figure 10c) the multiple lesions are not the same shape as marked by L.D.S. and, if a binary operation was performed on the B-mode image with this mask, parenchyma would also be extracted from the image. However, the second model ($\alpha = 0.7$, $\beta = 0.3$, $\gamma = 0.75$) is closer in obtaining a separation of the lesions. This shows the fact that during training and validation this model focused on hard-to-predict masks while the other model ($\alpha = \beta = 0.5$ and $\gamma = 1$ for FTL) treated all the labels from the dataset equally.

In Figure 11, another image from the validation batch is presented. The lesion is large with no other small lesions in the surrounding tissue. It can be observed that the location of the lesion was predicted correctly by both models, however one model predicted a mask very close to the expected output while the other model predicted two separated lesions. Even though the mean tumor size was 51.65 mm, both obtained segmentation models performed very well on small lesions (Figures 12 and 13).

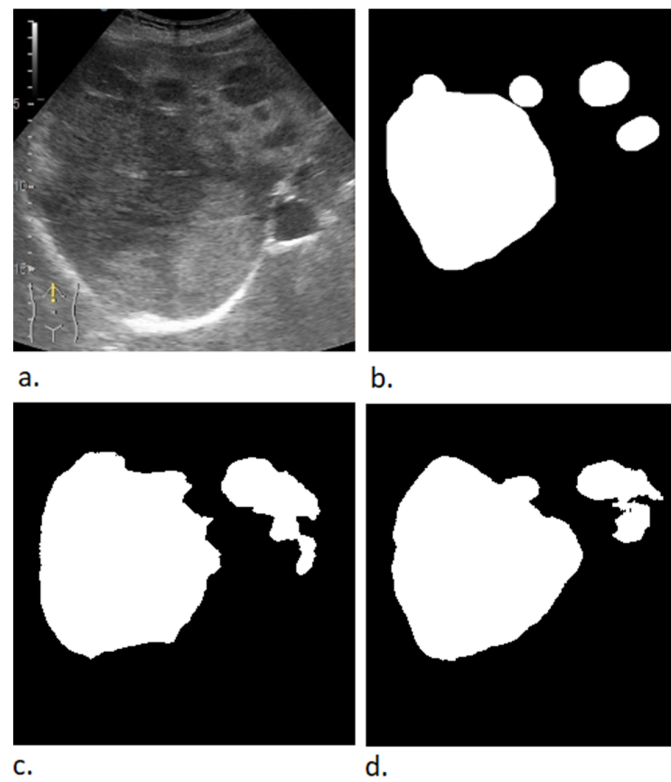


Figure 10. Results from the validation batch. (a) B-mode frame from the video examination file. (b) Label (expected output). (c) Predicted output for model trained with $\alpha = \beta = 0.5$ and $\gamma = 1$. (d) Predicted output for model trained with $\alpha = 0.7$, $\beta = 0.3$, $\gamma = 0.75$.

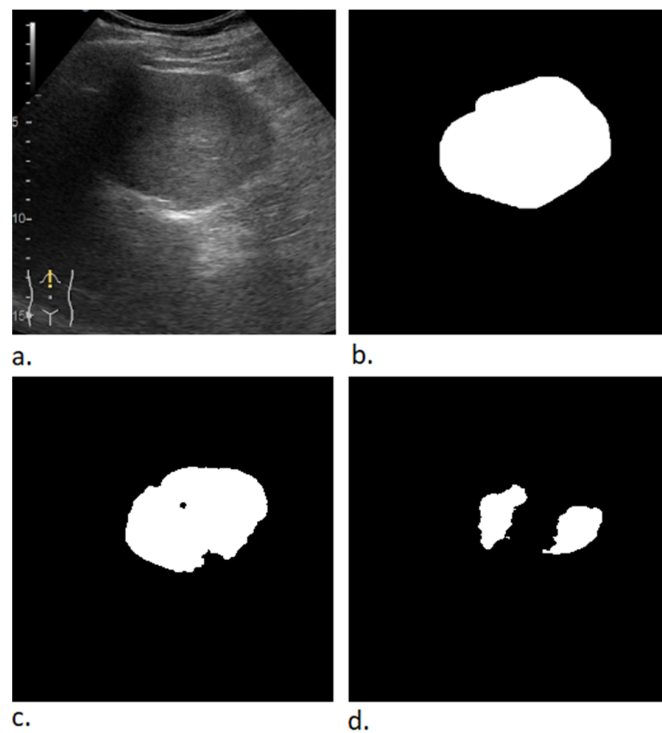


Figure 11. Results from the validation batch. (a) B-mode frame from the video examination file. (b) Label (expected output). (c) Predicted output for model trained with $\alpha = \beta = 0.5$ and $\gamma = 1$. (d) Predicted output for model trained with $\alpha = 0.7$, $\beta = 0.3$, $\gamma = 0.75$.

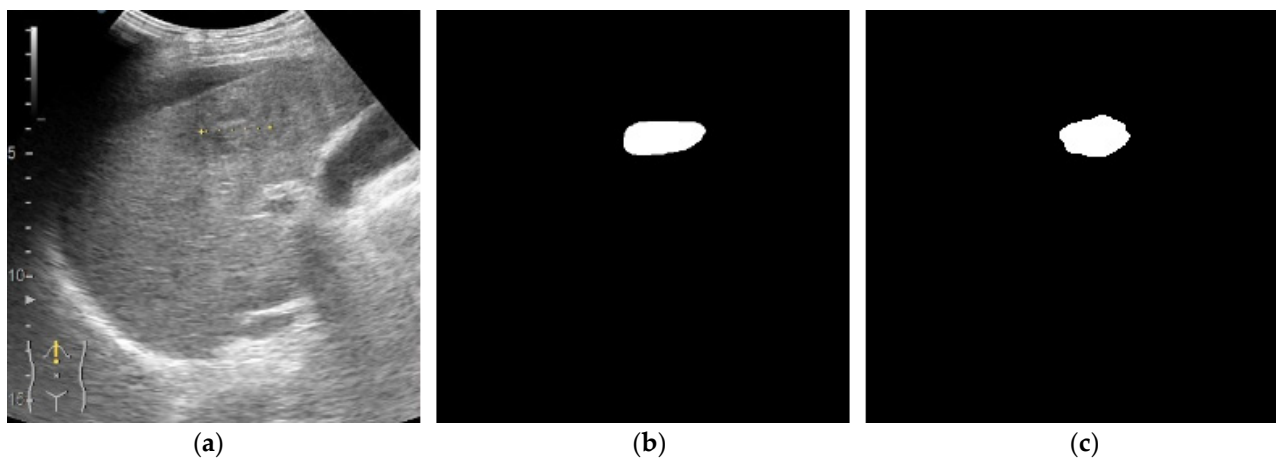


Figure 12. Small tumor size. Results from the validation batch for model with $\alpha = \beta = 0.5$ and $\gamma = 1$. (a) B-mode frame from the video examination file. (b) Label. (c) Predicted output.

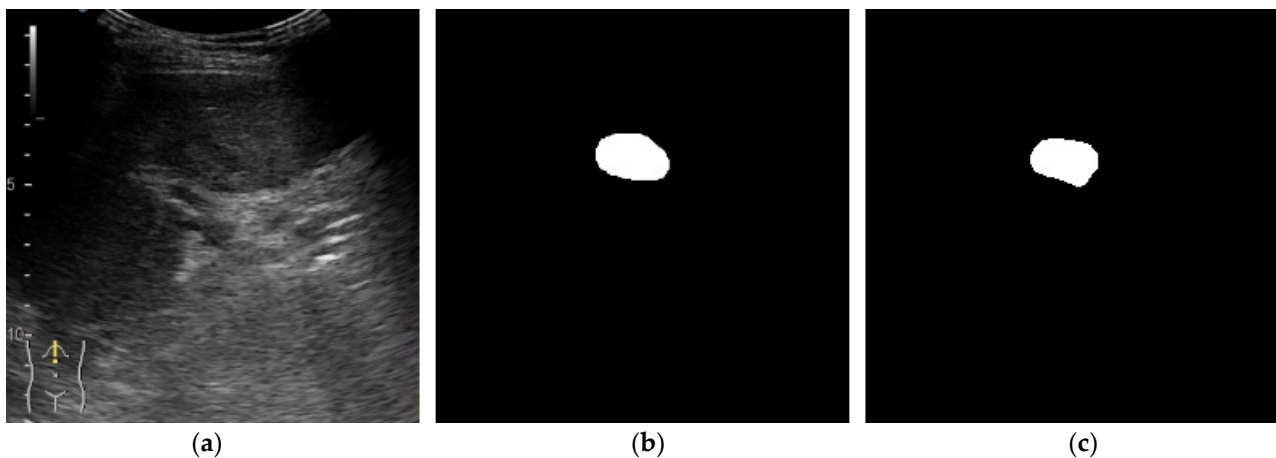


Figure 13. Small tumor size. Results from the validation batch for model with $\alpha = 0.7$, $\beta = 0.3$, $\gamma = 0.75$. (a) B-mode frame from the video examination file. (b) Label. (c) Predicted output.

Other studies [33,34] use dice coefficient [28] for evaluation of image segmentation models. Dice coefficient can be expressed as Equation (8). However, Jaccard index (IoU) can be expressed as related to dice coefficient as in Equation (9) hence observing both metrics when evaluating a segmentation model does not add additional information [35]. There are not many studies which have explored liver tumor segmentation in US. Compared with other existing studies [17,18] in which the authors performed segmentation using region difference filters or fuzzy clustering algorithms, our proposed method used a DL model for the image segmentation of US liver lesions. The architecture used by our study is U-Net [11], a DL model architecture containing two parts: an encoder and a decoder. Due to this architecture, the model requires a single run on an image to predict the corresponding mask. Another advantage of the U-Net model is that it can extract discriminative features from the raw images with limited training data [11]. The proposed method is uses a different approach in terms of loss function and performance metrics observed. Therefore, FTL function was used. As mentioned earlier, observing only accuracy is not enough to assess the performance of a segmentation model. The evolution of accuracy during training and validation for both models is observed in Figure 7. Thus, besides recall and precision, intersection over union was observed for both models. A solution for image segmentation in US investigations based on DL models can be more flexible as transfer learning can be applied on the pre-existing weights of the model. Using transfer learning, the model can be

retrained when new datasets are available. Furthermore, a DL model can be trained in a federated learning approach to preserve the privacy of the dataset [36].

As presented in Tables 4 and 5, the model was tested on two different environments and the results obtained show that the minimum inference time was 32.15 milliseconds and maximum inference time was 77.59 milliseconds. While there is a significant difference between the minimum and maximum inference time, these values are still in the accepted range of a 7 to 12 frames per second video investigation. The memory requirement for the model, presented in Table 6, is 0.9291 GB which is an acceptable size for any modern GPU.

Automatic tumor segmentation in US images is of great importance when using TICs as a method to quantify tumor perfusion in cases of possible liver hepatocellular carcinoma [7]. Keeping the region of interest in focus sometimes proves to be a difficult task, eminently operator-dependent and greatly influenced by comorbidities of the patient (breathing difficulties due to either concomitant respiratory or cardiac diseases, increased abdominal pressure due to underlying cirrhosis, etc.) [37]. In addition, the medical professional tries to keep contrast bubble destruction to a minimum, to maximize the efficiency of the procedure, therefore intentionally moving the probe and thus shifting focus from the liver tumor. The main strength of our method is therefore the capacity to identify the lesion in the regular B-mode window of an US machine, thus providing an accurate TIC measurement for the tumor zone, that can be compared with the one generated by the parenchyma. A possible limitation of our study is the limited number of patients, CEUS examinations and the relatively low variety of tumor types. Furthermore, as presented in Table 4, a possible drawback of the study is the loading time of the model which is 294.29 seconds on ENV1. In image segmentation, the model searches for patterns in B-mode images, and in some cases the lesion shape is more important than the lesion type. If the segmentation model is used for time-intensity curve extraction, a filtering of the values needs to be carried out if sudden changes are observed.

5. Conclusions

The aim of the study was to build a DL model for automatic, real-time image segmentation during US examinations. Besides the DL model, two algorithms were defined to achieve the goal. Algorithm 2 was used only for the creation of the dataset while Algorithm 1 was defined to serve two different scopes: dataset preparation and real-time frame extraction and cropping. The DL model used for image segmentation was U-Net with an input of 256 by 256 pixels. The dataset contained 50 video examinations from 49 patients. Two different models were obtained during the study. One model was trained to focus on hard-to-predict labels while the other one was trained to treat all the labels equally. The results presented show that the models built in this study can be useful for gastroenterologists to keep track of lesion movement during examinations. Usually, the extraction of the TIC(s) is performed manually. The operator defines the two regions, lesion and parenchyma, and updates these regions during the entire investigation as the image is moving due to patient's breathing or if the probe is moving. The inference time was tested for both models in two different environments. The results show that the proposed method can be used for real-time segmentation of US examinations if the video investigations are in a range of 7 to 12 frames per second.

The proposed segmentation models can represent a first step in automatically extracting time-intensity curves from CEUS examinations.

Author Contributions: Conceptualization, M.M., C.T.S., C.M.U. and M.-S.S.; methodology, M.M., C.T.S., L.D.S. and A.G.I.; validation, C.T.S., C.M.U., L.D.S. and A.-M.P.; formal analysis, M.M., C.M.U., M.-S.S. and C.K.; data curation, M.M., L.D.S., C.M.U. and A.G.I.; writing—original draft preparation, M.M., C.T.S. and C.M.U.; writing—review and editing, M.M. and C.T.S.; project administration, C.T.S. All authors have read and agreed to the published version of the manuscript.

Funding: The article processing charges were funded by the University of Medicine and Pharmacy of Craiova. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethical Committee of the University of Medicine and Pharmacy of Craiova (initial approval for the dataset used in this study: 36/22.04.2016).

Informed Consent Statement: Patient consent was waived as no patient personal information was retrieved or processed off-site.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The images were made available from the University's repository as an anonymized dataset.

Acknowledgments: This work was conducted within the project “Innovative expert computer network-based system neuronal for classification and prognosis of liver tumors”, MYSMIS ID 109722 within the National Competitivity Program, POC/62/1/3/.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hu, J.; Zhou, Z.-Y.; Ran, H.-L.; Yuan, X.-C.; Zeng, X.; Zhang, Z.-Y. Diagnosis of Liver Tumors by Multimodal Ultrasound Imaging. *Medicine* **2020**, *99*, e21652. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Birch, M.S.; Marin, J.R.; Liu, R.B.; Hall, J.; Hall, M.K. Trends in Diagnostic Point-of-Care Ultrasonography Reimbursement for Medicare Beneficiaries Among the US Emergency Medicine Workforce, 2012 to 2016. *Ann. Emerg. Med.* **2020**, *76*, 609–614. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Hata, J. Point-of-Care Abdominal Ultrasound. *Masui* **2017**, *66*, 503–507. [\[PubMed\]](#)
4. Lencioni, R.; Piscaglia, F.; Bolondi, L. Contrast-Enhanced Ultrasound in the Diagnosis of Hepatocellular Carcinoma. *J. Hepatol.* **2008**, *48*, 848–857. [\[CrossRef\]](#)
5. Jacobsen, N.; Nolsøe, C.P.; Konge, L.; Graumann, O.; Dietrich, C.F.; Sidhu, P.S.; Piscaglia, F.; Gilja, O.H.; Laursen, C.B. Contrast-Enhanced Ultrasound: Development of Syllabus for Core Theoretical and Practical Competencies. *Ultrasound Med. Biol.* **2020**, *46*, 2287–2292. [\[CrossRef\]](#)
6. Dietrich, C.F.; Nolsøe, C.P.; Barr, R.G.; Berzigotti, A.; Burns, P.N.; Cantisani, V.; Chammas, M.C.; Chaubal, N.; Choi, B.I.; Clevert, D.-A.; et al. Guidelines and Good Clinical Practice Recommendations for Contrast-Enhanced Ultrasound (CEUS) in the Liver—Update 2020 WFUMB in Cooperation with EFSUMB, AFSUMB, AIUM, and FLAUS. *Ultrasound Med. Biol.* **2020**, *46*, 2579–2604. [\[CrossRef\]](#)
7. Streba, C.T. Contrast-Enhanced Ultrasonography Parameters in Neural Network Diagnosis of Liver Tumors. *World J. Gastroenterol.* **2012**, *18*, 4427. [\[CrossRef\]](#)
8. Zhang, Q.; Du, Y.; Wei, Z.; Liu, H.; Yang, X.; Zhao, D. Spine Medical Image Segmentation Based on Deep Learning. *J. Healthc. Eng.* **2021**, *2021*, 1917946. [\[CrossRef\]](#)
9. Yu, F.; Zhu, Y.; Qin, X.; Xin, Y.; Yang, D.; Xu, T. A Multi-Class COVID-19 Segmentation Network with Pyramid Attention and Edge Loss in CT Images. *IET Image Process.* **2021**, *15*, 2604–2613. [\[CrossRef\]](#)
10. Sammouda, R.; El-Zaart, A. An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method. *Comput. Intell. Neurosci.* **2021**, *2021*, 4553832. [\[CrossRef\]](#)
11. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
12. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016.
13. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
14. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018.
15. Araújo, J.D.L.; da Cruz, L.B.; Diniz, J.O.B.; Ferreira, J.L.; Silva, A.C.; de Paiva, A.C.; Gattass, M. Liver Segmentation from Computed Tomography Images Using Cascade Deep Learning. *Comput. Biol. Med.* **2022**, *140*, 105095. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Nowak, S.; Mesropyan, N.; Faron, A.; Block, W.; Reuter, M.; Attenberger, U.I.; Luetkens, J.A.; Sprinkart, A.M. Detection of Liver Cirrhosis in Standard T2-Weighted MRI Using Deep Transfer Learning. *Eur. Radiol.* **2021**, *31*, 8807–8815. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Hiransakolwong, N.; Hua, K.A.; Khanh, V.; Windyga, P.S. Segmentation of Ultrasound Liver Images: An Automatic Approach. In Proceedings of the 2003 International Conference on Multimedia and Expo—ICME '03 (Cat. No.03TH8698), Baltimore, MD, USA, 6–9 July 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 1–573.

18. Jain, N.; Kumar, V. Liver Ultrasound Image Segmentation Using Region-Difference Filters. *J. Digit. Imaging* **2017**, *30*, 376–390. [\[CrossRef\]](#)
19. Gupta, D.; Anand, R.S.; Tyagi, B. A Hybrid Segmentation Method Based on Gaussian Kernel Fuzzy Clustering and Region Based Active Contour Model for Ultrasound Medical Images. *Biomed. Signal. Process. Control* **2015**, *16*, 98–112. [\[CrossRef\]](#)
20. Ciocalteu, A.; Iordache, S.; Cazacu, S.M.; Urhut, C.M.; Sandulescu, S.M.; Ciurea, A.-M.; Saftoiu, A.; Sandulescu, L.D. Role of Contrast-Enhanced Ultrasonography in Hepatocellular Carcinoma by Using LI-RADS and Ancillary Features: A Single Tertiary Centre Experience. *Diagnostics* **2021**, *11*, 2232. [\[CrossRef\]](#)
21. Bankhead, P.; Loughrey, M.B.; Fernández, J.A.; Dombrowski, Y.; McArt, D.G.; Dunne, P.D.; McQuaid, S.; Gray, R.T.; Murray, L.J.; Coleman, H.G.; et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci. Rep.* **2017**, *7*, 16878. [\[CrossRef\]](#)
22. Cheng, D.; Lam, E.Y. Transfer Learning U-Net Deep Learning for Lung Ultrasound Segmentation. *arXiv* **2021**, arXiv:2110.02196.
23. Jia, S.; Despinasse, A.; Wang, Z.; Delingette, H.; Pennec, X.; Jais, P.; Cochet, H.; Sermesant, M. Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*; Springer: Cham, Switzerland, 2018.
24. González Sánchez, J.C.; Magnusson, M.; Sandborg, M.; Carlsson Tedgren, Å.; Malusek, A. Segmentation of Bones in Medical Dual-Energy Computed Tomography Volumes Using the 3D U-Net. *Phys. Med.* **2020**, *69*, 241–247. [\[CrossRef\]](#)
25. Wu, J.; Yang, S.; Gou, F.; Zhou, Z.; Xie, P.; Xu, N.; Dai, Z. Intelligent Segmentation Medical Assistance System for MRI Images of Osteosarcoma in Developing Countries. *Comput. Math. Methods Med.* **2022**, *2022*, 7703583. [\[CrossRef\]](#)
26. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
27. Tversky, A. Features of Similarity. *Psychol. Rev.* **1977**, *84*, 327–352. [\[CrossRef\]](#)
28. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [\[CrossRef\]](#)
29. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
30. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.04467.
31. Butaru, A.E.; Mămuleanu, M.; Streba, C.T.; Doica, I.P.; Diculescu, M.M.; Gheonea, D.I.; Oancea, C.N. Resource Management through Artificial Intelligence in Screening Programs—Key for the Successful Elimination of Hepatitis C. *Diagnostics* **2022**, *12*, 346. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Google Colab. Available online: <https://colab.research.google.com/> (accessed on 3 October 2022).
33. Milletari, F.; Ahmadi, S.-A.; Kroll, C.; Plate, A.; Rozanski, V.; Maiostre, J.; Levin, J.; Dietrich, O.; Ertl-Wagner, B.; Bötzel, K.; et al. Hough-CNN: Deep Learning for Segmentation of Deep Brain Regions in MRI and Ultrasound. *Comput. Vis. Image Underst.* **2017**, *164*, 92–102. [\[CrossRef\]](#)
34. Milletari, F.; Ahmadi, S.-A.; Kroll, C.; Hennemersperger, C.; Tombari, F.; Shah, A.; Plate, A.; Boetzel, K.; Navab, N. Robust Segmentation of Various Anatomies in 3D Ultrasound Using Hough Forests and Learned Data Representations. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 111–118.
35. Taha, A.A.; Hanbury, A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Med. Imaging* **2015**, *15*, 29. [\[CrossRef\]](#)
36. Florescu, L.M.; Streba, C.T.; Șerbănescu, M.-S.; Mămuleanu, M.; Florescu, D.N.; Teică, R.V.; Nica, R.E.; Gheonea, I.A. Federated Learning Approach with Pre-Trained Deep Learning Models for COVID-19 Detection from Unsegmented CT Images. *Life* **2022**, *12*, 958. [\[CrossRef\]](#)
37. Dietrich, C.; Ignee, A.; Hocke, M.; Schreiber-Dietrich, D.; Greis, C. Pitfalls and Artefacts Using Contrast Enhanced Ultrasound. *Z. Gastroenterol.* **2011**, *49*, 350–356. [\[CrossRef\]](#)