IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# CLASSIFICATION AND DETECTION OF ULTRASOUND LIVER TUMOR USING VGG-RESNET

**[1]Anu Susan Philip, [2]Gowri S Kumar, [3]Grace Mary Abraham, [4]Rohith Balakrishnan**

[1]Assistant Professor, [2,3,4] B-Tech Student
[1]Electronics and Communication Engineering,
[1]Mar Baselios College of Engineering and Technology, Trivandrum, India

*Abstract:* Since liver cancer is the most fatal kind of cancer, it's critical to catch it early. Due to the lack of symptoms, clinical procedures make early detection hard. Reading a large number of tumor images is a perilous work for radiologists. In contemporary processes, traditional methods are employed to determine if a tumor is malignant or benign. Certain malignancies are difficult to detect visually, which leads to a high percentage of false positives and negatives. Certain tumors have comparable traits, necessitating feature extraction-based classification and identification.

Due to multiple challenges, such as low contrast between the liver and other organs and tumors, and sizes of tumors, and irregular tumor growth, the existing system has not been very good at segmenting the liver and lesions. As a result, a novel technique is required to solve these challenges.

The existing challenges are addressed using a CNN-based multiclass detection approach. Several designs are compared, including GoogLeNet, Inception-v3, ResNet, and VGG-Net, with the VGG architecture being most accurate CNN-based multiclass identification. The RCNN principle is put into practice. The features were retrieved and fed into the RCNN.

The CNN-based detection system has three stages: training, testing, and validation. Several factors such as kernel value, filter size, bias value, learning rate, and momentum can be changed to improve the accuracy of the recommended system. A novel architecture consisting of VGG-16 and ResNet-18 architecture was developed for the classification and detection for liver tumors.

***IndexTerms* - MATLAB, Convolutional Neural Network (CNN), Dataset, Learning Rate, Epoch Rate, and Minibatch Size.**

## I.INTRODUCTION

Ultrasonography is a medical examination method that is non-invasive, safe, inexpensive, and thus repeatable, suitable for patient disease monitoring. Other examination techniques, such as the Computer Tomography (CT), the Magnetic Resonance Imaging (MRI), the endoscopy, or the Contrast Enhanced Ultrasonography (CEUS) are considered irradiating or expensive. Thanks to recent developments in digital technology, it is a worldwide approach that's been widely utilized in clinical practices for years to detect or interfere with the treatment method of illnesses among those who especially suffer from severe liver disease.

However, the human eye often has difficulty distinguishing this tumor from the surrounding cirrhotic parenchyma. To accomplish automated detection and segmentation of this tumor, we create computerized, non-invasive approaches based on image processing and machine learning techniques used inside ultrasound pictures.

The Liver is the body's main organ found under the right ribcage and blows to the base of the lung. It has many functions like cleaning blood cells, transforms some of these nutrients into energy and breaks down toxic substances etc. The two principal hepatic lobes are the left and right. The quadrate and caudate lobes are two additional lobes seen from the bottom of the liver. HCC is a cancer that arises when liver cells grow out of control and spread to other places of the body. Primary hepatic tumors develop when cells act improperly, HCC frequently mimics benign tumors.

According to statistics, liver cancer is the second most prevalent cancer in men and the sixth most common disease in women. 750,000 people were diagnosed with liver cancer in 2008, with 696,000 dying as a consequence. Males are infected twice as often as females over the world. Hepatitis caused by a virus is significantly more hazardous and can lead to liver cancer. The World Health Organization (WHO) estimates that this virus kills approximately 1.45 million people per year. In 2015, Egypt was named the country with the highest rate of adult viral hepatitis C (HCV) infection, with 7% of individuals diagnosed with the virus.

Primary hepatic cancer is detected by clinical, laboratory, and imaging tests such as ultrasound scans, magnetic resonance imaging (MRI) scans, and computed tomography (CT) scans.

## II.RELATED WORK.

The VGG-16 network architecture is a CNN model. The network is made up of 41 layers. There are 16 learnable weights layers, including 13 convolutional layers and three fully - connected layers. To make binary segmentation of medical pictures easier, the layer was replaced by a binary pixel classification layer.

Malignant liver tumors are considered as one of the most common cancers and a leading cause of cancer death worldwide. While using convolutional neural networks (CNNs) for feature extraction from ultrasound (US) images and tasks thereafter, most works focus on pre-trained architectures using transfer learning which can sometimes cause negative transfer and reduced performance in medical domain

In order to make binary segmentation of medical pictures easier, the usual 3D-IRCADb-01 layer was replaced with a binary pixel classification layer. With a tumor classification accuracy of over 86 percent, the recommended approach correctly detects most portions of the tumor. By manually minimizing the data's class imbalance by removing irrelevant photographs and selecting appropriate hyperparameters, the improvement was achieved. Three distinct models with equivalent architectures were employed for training. For liver segmentation, the first model was trained using abdominal CT scan images with liver annotations. To segregate the tumor from the liver, the second model was trained using liver pictures with tumor annotations. Finally, utilizing abdominal CT scan pictures with tumor annotations, the third model was trained to segment the tumor straight from the CT scan images.

Although, it frequently misclassified certain cancers as blue and misclassified other tissues as tumors. By lowering the number of filters in each convolutional block, this strategy dramatically decreases the network's complexity. The time it takes to train the network is reduced by half, from several hours to only 40 minutes.

The effectiveness of liver tumor segmentation is much improved when this segmentation approach is used. To begin with, it minimizes the network's complexity by lowering the number of filters required for each convolutional block, and hence the number of trainable parameters. As a result, the time it required to train the network was cut in half.

Regardless of the framework, the detection results of multimodal fusion are superior to those of a single modality. As a consequence, combining multi-modal information assists in object detection on medical imaging. This approach delivers a considerable 6.4 total AP gain when compared to the strong baseline Faster R-CNN with ResNet-50-FPN of two modalities. some methodologies are

MIMFNet is a multimodal liver tumor identification framework that works from start to end with multimodal semantic data. Researchers suggest an intermediate multi-modal fusion backbone and create a multi-modal EFPN to capture intermediate multi-modal information interaction and multi-modal features at different sizes. Experimental findings show that the proposed multimodal detection framework MIMFNet performs well on multimodal liver tumor pictures.

Despite improvements in liver and tumor segmentation, the system still failed to segment well in some slices. The algorithm nearly precisely segments the liver in certain slices, but it fails in others when the full liver is not caught or when the liver is concealed by other overlapping organs and appears to be divided into pieces. When it comes to tumor segmentation, the algorithm frequently fails to segment small, strangely shaped tumors. Binary cross-entropy was chosen as a loss function for the last model, which was trained to segment tumors directly from abdominal CT scan images, and Adam was chosen as an optimizer for all three models through experimentation, The proposed multi-modal fusion backbone and the multi-modal enhanced feature pyramid may be easily incorporated with existing detection or segmentation networks to efficiently employ multi-modal semantic information. The vast majority of them only snap single-modal photographs, preventing them from benefiting from the additional information supplied by other modalities. Two simple strategies for recognizing multi-modal objects are early fusion and late fusion.

To construct effective classifiers on small medical datasets, fine-tuning an existing deep convolutional neural network such as ResNeXt, ResNet18, ResNet34, ResNet50, and AlexNet can be used. A hybrid classifier is suggested, as well as a novel framework for classifying normal, cirrhosis, and hepatitis liver images. In liver scans, researchers employed a larger total patch of tissue, resulting in more reliable and less subjective results. Large datasets should be used to test the performance of the suggested strategy for applicability.

## III. METHODOLOGY

The objective of this paper is to use CNN to classify and identify multi-class liver cancers. It frequently aids the radiologist in identifying the damaged portion, which is difficult to predict by looking at the scans early on. Training, testing, and validation are the three steps of the proposed system.

### 3.1. DATA PREPARATION AND AUGMENTATION

For Liver tumor detection a large Dataset is required. So a multi class Liver tumor. Dataset is being collected for the Process from ultasoundcases.info website. The collected Dataset has 4 classes which are hemangiomas, hepatic adenoma, carcinoma and metastases and cirrhosis and fatty liver. The image size of the dataset is (225,225,3).

Image transformations and pre-processing techniques are being used to enhance datasets. The first step in data augmentation in this research was to rotate the raw image into 180 degree. The concept that image dimensions may not be kept after rotation is a crucial component of this procedure. Since the dataset's images are rectangular, rotating them in 180 degrees would preserve their shape. Followed by this step, the raw images and the rotated images are to be mirrored in y direction.

Denoising algorithms such as mean filtering for AWGN removal and median filtering for speckle/multi-noise removal are used for further augmentation. In different layers of CNN, low level features like edges, intermediate level features such as texture, and high level features such as form and structures have been trained and investigated.

### 3.2 PROPOSED NOVEL ARCHITECTURE

VGG-16 is widely regarded as one of the most advanced vision model architectures currently available. Instead of having a large number of hyper-parameters, VGG16 focused on having 3x3 filter convolution layers with stride 1 and always used the same padding and max pool layer of 2x2 filter stride 2. The number 16 in VGG16 refers to the fact that there are 16 layers with different weights. With roughly 138 million (estimated) parameters, this network is quite large. The fact that VGG has a comprehensive linear sequential structure allows it to analyze more properties than other designs like AlexNet, GoogLeNet, and Inception is one of its primary advantages.


Fig.1 VGG-ResNet Architecture

Additional layers in Deep Neural Networks improve accuracy and performance, usually in order to solve a challenging problem. The idea behind adding more layers is that the layers would gradually learn more complicated features. But it has been found that there is a maximum threshold for depth with the traditional Convolutional neural network model and there is a significant increase in error percentage as the layer increases due to vanishing gradient. Skip connections are used in the network to solve vanishing gradient issues by bypassing several convolution stages and connecting straight to the output. The suggested architecture for categorizing the images combines VGG-16 and ResNet18 incorporating the advantages of both architectures. The network's input is a two-dimensional image (225,300,3). The first layer has 64 channels with a 1*1 filter size and the same padding. Then, after a stride (2,2) max pool layer, two convolution layers with 64 filters and a filter size of (3,3). In the layers, the ReLu activation function is utilized. This is followed by a stride (2,2) max-pooling layer, which is identical to the preceding layer.

After that, the picture is processed through two sets of convolution layers of filter size (3,3) and 64 filters, followed by a max pooling layer of stride (2,2). Then there are four more sets of convolution layers with filter sizes of (5, 5) and 64 filters. After that, the image is processed through two sets of convolution layers, each with a filter size of (7,7) and 64 filters. The output is then flattened, resulting in a single completely linked layer with four classes. The activation function for all hidden layers is ReLU. ReLU is more computationally efficient since it speeds up learning and reduces the chances of vanishing gradient issues. With a learning rate of 0.0001, Adam is chosen as the optimization method. The system achieved an accuracy of 83.78 % after training the dataset in the suggested architecture, which is suitable for biological pictures with restricted characteristics and datasets.

## IV. RESULTS AND DISCUSSION

Table 4.1: Comparison of different architectures

| Architecture | Validation Accuracy | Time |
|---|---|---|
| VGG-16 | 95.30% | 284 min 17 sec |
| RESNET -18 | 91.28% | 3 min 15 sec |
| INCEPTION V-3 | 92.62% | 200 min 39 sec |
| GOOGLENET | 95.97% | 5 min 25 sec |
| VGG -RESNET | 83.78% | 12 min 12 sec |

In this experiment,VGG-16 and ResNet18 structures were combined and are used for comparison, mainly by the method of skip connection and batch normalization.VGG-16 is one of the best vision model designs. ResNet18 are combined together for classifying the images. In VGG -16 the input to any of the network configurations is considered to be a fixed size 224 x 224 image with three channels – R, G, and B. The only pre-processing done is normalizing the RGB values for every pixel. This is achieved by subtracting the mean value from every pixel.
One of the key advantages of VGG is that it has a complex linear sequential structure that allows it to analyze more characteristics than other designs such as AlexNet, GoogLeNet, and Inception. VGG-16 and ResNet18 are coupled in the suggested architecture for picture classification.

We use the input network as an image of dimension (225,300,3). The first layers have 64 channels of 1*1 filter size with the same padding. Then after a max pool layer of stride (2,2), two layers have convolution layers of 64 filters and a filter size of (3,3). Here we used the ReLu activation function in the layers. Then followed a max-pooling layer of stride (2,2) which is the same as the previous layer. Then there are two further convolution layers with filter sizes of (3,3) and 64 filters, followed by a stride max-pooling layer (2,2). There are four sets of convolution layers with filter sizes of (5,5) and 64 filters. After that, the picture is processed through two sets of convolution layers, each with a filter size of (7,7) and 64 filters. The output is then flattened, resulting in a single completely linked layer with four classes. The activation function for all hidden layers we used is ReLU.

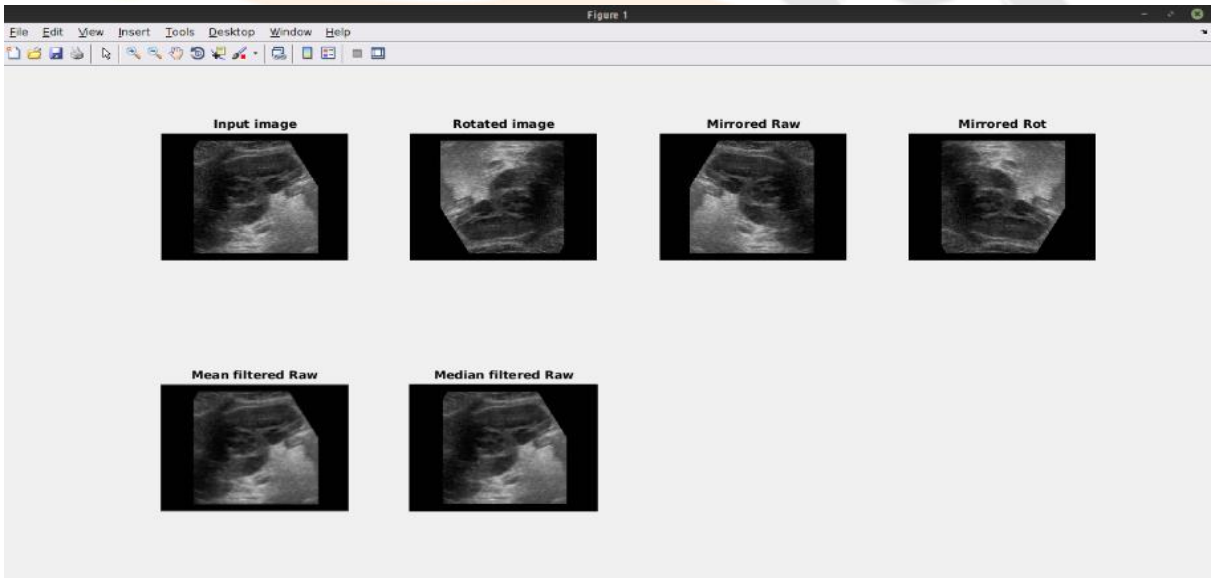


Fig.2 Data Augmentation

Research Through Innovation

ReLU is more computationally efficient since it speeds up learning and reduces the chances of vanishing gradient issues. Skip connections are used in the network to solve vanishing gradient problems by skipping several convolutional layers and connecting directly to the output. ReLu changes negative numbers to zero while maintaining positive values, allowing for more efficient and successful training. This is commonly referred to as activation since only the active traits are carried on to the next layer.
The optimization method Adam is used, with a learning rate of 0.0001. We achieved an accuracy of 83.78 percent after training the dataset in our proposed architecture, which is suitable for biological pictures however due to dataset limitations for our architecture we only got this much accuracy in the line of work.

## V. CONCLUSION

As liver cancer is the most lethal type of cancer, early detection is crucial, as it can save the lives of many people. Clinical techniques make early detection difficult due to the lack of symptoms. For radiologists, reading a huge number of tumor images is a dangerous task. Visual detection of some cancers is challenging, resulting in a high percentage of false positives and negatives. Feature extraction-based classification and detection are essential since different tumors have similar features. The most challenging problem in medicine is still detecting liver tumors. Computerized diagnosis now depends only on conventional interpretation and

prediction methods. To produce a superior learning-based multiclass prediction solution, the proposed technique uses CNN. Due to high false positive and false negative rates, the presence of some visually challenging tumors might make identification difficult. As a result, there is a considerable demand for an adaptive multiclass tumor detection system.
Prior to training, noise reduction and enhancement are performed during data augmentation. Mean filtering and median filtering, in addition to standard transformation, are used to reduce AWGN and speckle, respectively. As a consequence, for better augmentation, the upgraded denoised images are connected to the dataset. Malignancies can be detected early with multiclass prediction, allowing patients to get more effective therapy.
Several state-of-the-art pretrained designs are examined, including GoogLeNet, Inception-v3, ResNet, and VGG-Net, with accuracies of 95.97%, 92.62%, 91.28%, and 95.30%, respectively. Using Adam as the optimization method, a novel design consisting of VGG-16 and ResNet-18 architecture was constructed for the classification and detection of liver cancers, with an accuracy of 83.78 percent.

## VI. ACKNOWLEDGMENT

Words cannot express our gratitude to Dr Jayakumari (HOD, Electronics and communication) for providing all necessary facilities for the completion of our project and her invaluable patience and feedback. We express our profound gratitude seminar coordinator to Dr. Swapna (Associate professor) for her support and guidance during the course of the project. Without the tremendous knowledge and experience supplied by our project guide, Ms. Anu Susan Philip (Assistant professor), we would not have been able to embark on this journey.
Additionally, this endeavor would not have been possible without the generous support from Ms.Revathy Sivanandan. We also appreciate the moral support received from our family and friends throughout the project.

## REFERENCES

[1] Almotairi, Sultan, Ghada Kareem, Mohamed Aouf, Badr Almutairi, and Mohammed A-M. Salem. "Liver tumor segmentation in CT scans using modified SegNet." *Sensors* 20, no. 5 (2020): 1516.

[2] Ayalew, Yodit Abebe, Kinde Anlay Fante, and Mohammed Aliy Mohammed. "Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method." *BMC Biomedical Engineering* 3, no. 1 (2021): 1-13.

[3] Pan, Chao, Peiyun Zhou, Jingru Tan, Baoye Sun, Ruoyu Guan, Zhutao Wang, Ye Luo, and Jianwei Lu. "Liver Tumor Detection Via A Multi-Scale Intermediate Multi-Modal Fusion Network on MRI Images." In *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 299-303. IEEE, 2021.

[4] Pasyar, Pezhman, Tahereh Mahmoudi, Seyedeh-Zahra Mousavi Kouzehkanan, Alireza Ahmadian, Hossein Arabalibeik, Nahid Soltanian, and Amir Reza Radmard. "Hybrid classification of diffuse liver diseases in ultrasound images using deep convolutional neural networks." *Informatics in Medicine Unlocked* 22 (2021): 100496.

[5] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." In *2017 international conference on engineering and technology (ICET)*, pp. 1-6. Ieee, 2017