LIVER CANCER DETECTION USING ML CLASSIFIERS

Mrs.A.Selvarani M.E(Ph.D)[1], E.Babu Sajish[2], M.Pramod[3]

[1]*Associate Professor, Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai, India*

[2*, 3] *Student, Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai, India*

**Corresponding author: babusajish31@gmail.com*

---***---

Abstract - *The liver is necessary for survival and is also prone to many diseases. CT examinations can be used to plan and properly administer radiation treatments for tumors and to guide biopsies and other minimally invasive procedure. Manual segmentation and classification of CT image is a tedious task and time consuming process which is impractical for large amount of data. The liver is segmented from CT images using adaptive threshold method and morphological processing. Tumor region extraction is done by means of Fuzzy C Means (FCM) clustering from the segmented liver region. The statistical and textural information are obtained from the extracted tumor using Gabor with PCA. The features like mean, standard deviation and entropy of the obtained sub bands are calculated and stored in a feature vector (in format of mat file). The extracted features are fed as input to Extreme Machine Learning classifier to identify the presence of Liver tumor disease and to classify it as Malignant or Benign stage or not.*

Key Words: Machine Learning Classifiers, Kaggle, Medical Image Analysis,CT Scan Image, Matlab

1. INTRODUCTION:

The liver has vital has importance to human beings, so liver diseases are considered seriously. The liver is made up of various cell types, so several distinct types of tumors can develop in it. These tumors can be benign or malignant (cancerous). The benign tumors of the liver seen most frequently include cavernous hemangioma, liver cell adenomas and focal nodular hyperplasia (FNH) .The malignant tumors of the liver seen most frequently are hepatocellular carcinoma, intrahepatic cholangiocarcinoma, bile duct cystadenocarcinoma, and hepatoblastoma . Computers are used widely in medical research, where there is a vital need for better microelectronic sensors for data acquisition. Imaging modalities like Ultrasound, MRI (Magnetic Resonance Imaging), CT (Computed Tomography) and PET (Positron Emission Tomography) are widely used techniques for liver cancer tumor diagnosis. Liver cancer tumor is sixth dangerous diseases in the world. Because of the liver's vital importance to human beings, Liver diseases are considered seriously. There are two classes ofliver tumors: benign and malignant. Computer-assisted liver tumor classification which is based on the image analysis techniques provides more useful information. The conventional methods for the liver tissue classification consist of three-step process. The first step involves the segmentation of liver and tumor from CT abdominal image. The second step is the feature extraction and the third step is classification using a classifier. The Characterization of liver images based on texture analysis techniques have been developed over the years.

2. RESEARCH DISCRIPTION:

Medical image analysis is an active field of research for machine learning, partly because the data is relatively structured and labeled, and it is likely that this will be the area where patients first interact with functioning, practical artificial intelligence systems.

2.1. Machine Learning Classifiers:

Classifier in machine learning is an algorithm that automatically orders or categorizes data into one or more of a set of "classes." One of the most common examples is an email classifier that scans emails to filter them by class label: Spam or Not Spam. Machine learning algorithms are helpful to automate tasks that previously had to be done manually. They can save huge amounts of time and money and make businesses more efficient. A classifier is the algorithm itself – the rules used by machines to classify data. A classification model, on the other hand, is the end result of your classifier's machine learning. The model is trained using the classifier, so that model, ultimately, classifies your data.

CT Scan Image:

A computerized axial tomography scan (CT scan) is an xray method that combines several x-ray images with the aid of a computer to produce cross-sectional views and three-dimensional images of the interior organs and structures of the body. A CT scan is used to define normal and abnormal structures in the body and/or assist in procedures by helping to precisely guide the placement of instruments or treatments. It is a medical imaging method that employs tomography. Tomography is the method of producing a two-dimensional image of a slice or section through a 3- dimensional object (a tomogram). CT scans of the abdomen are extremely helpful in defining body organ anatomy, including visualizing the liver, gallbladder, pancreas, spleen, aorta, kidneys. To verify the existence of tumors, infection, abnormal anatomy, or changes of the body from trauma, CT scans in this area are used. The CT image is sufficient for analysis for this proposed method. Moreover MRI Scan is very costly and the tissues can't able to view clearly. But the CT is not so costly but also the tissues can be clearly visible in CT scan.

Kaggle:

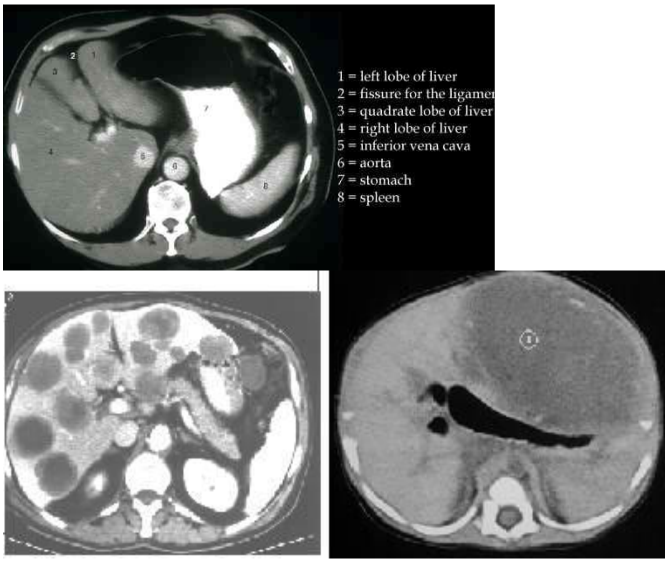
It is a subsidiary of Google LLC is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

Matlab:`

Matlab is a programming platform designed specifically for engineers and scientists to analyze and design systems and products that transform our world. The heart of MATLAB is the MATLAB language, a matrix-based language allowing the most natural expression of computational mathematics. With the help of Matlab, we can analyze data, develop algorithm to create models and applications MATLAB lets you take your ideas from research to production by deploying to enterprise applications and embedded devices, as well as integrating with Simulink® and Model-Based Design.

3.RELATED WORK:

The Liver Tumor region is segmented by using FCM. After the segmentation, the tumor region is further processed by feature extraction algorithm for extracting the features. For feature extraction we are using Multilevel Wavelet-PCA along with Laplacian mesh method. By this way we can easily identified the tumor is in which stage .

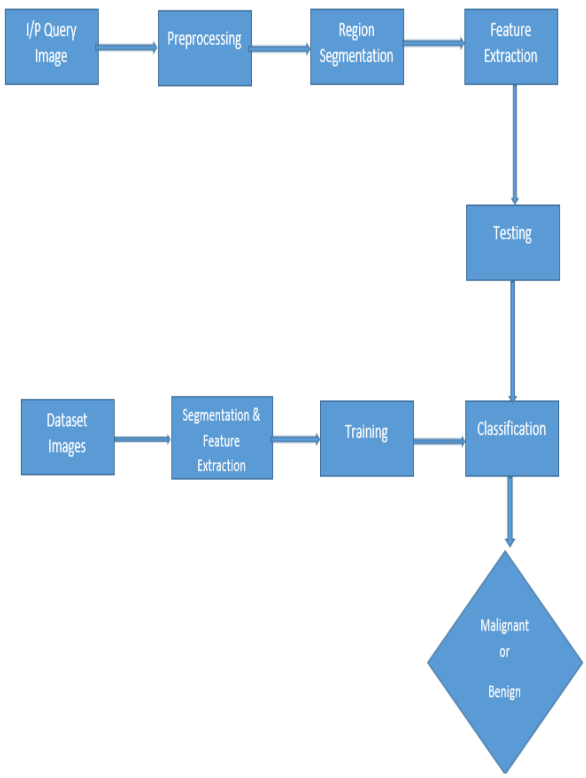


(a) CT of Normal Liver (b)CT of multiple metastases or with one large mass

The major drawbacks of the existing model are the constraints based on the Laplacian operator, the low frequencies of the mesh are preserved with the use of an attractor to the original mesh, while the iterative smoothing process removes high frequencies,the major disadvantage is the inaccuracy, as the gradient magnitude of the edges decreases. Most probably the accuracy also decreases.

4.PROPOSED METHODOLOGY:

In our proposed scheme, we are going to use the FCM clustering for tumor region finding. The same segmented region is completely analyzed by using the Gabor features on it. After the features are extracted, the processed dataset is completely trained and classified by using either KNN or SVM classifier of machine learning concept.



Segmentation of liver and tumor from abdominal Computed Tomography (CT) is important for proper planning and treatment of liver disease. Variable size, intensity overlap, and complexity of CT images probe a problem for a radiologist. These issues make accurate and reliable delineation of liver and tumor very difficult and time-consuming. So, an automatic method is desired and beneficial. In this paper, we propose a fully automatic method to segment both liver and tumor using an array of Gabor Filter (Gabor Bank (GB)) and Machine Learning (ML) classifiers: Support Vector Machine (SVM), Probabilistic Neural Network (PNN) and K Nearest Neighbor (KNN). First, GB extract pixel level Gabor features from CT images [13]. Secondly, the liver is segmented using ML classifiers trained on Gabor features. Finally, tumor segmentation is done on the segmented liver image using the same approach as in liver segmentation as shown in the Figure 4.

4.1 Data Acquisition: The dataset collected from various European hospitals with different CT scanners. For our study, we selected 31 CT slices with hepatic tumors from a CT volume of 118 CT slices. The analyzed CT volume had a varying liver shape across the slices with tumor being visible only on some slices.

4.2 Data preprocessing: Preprocessing was carried out in a slice-wise fashion. First, each image slice was windowed in the range of [-100, 400] Hounsfield Unit (HU) to exclude irrelevant organs and objects. Secondly, Histogram Equalization (HE) was performed to increase the contrast of the windowed image. Finally, the original image of 512 by 512 was resized to 256 by 256. The entire proposed method is depicted in Fig. 1 and has been discussed in following subsections in details.

4.3 Gabor Bank for feature extraction:A Gabor Bank (GB) is an array of Gabor Filters (GFs). GF is a transform based method for extracting the texture information. The basic function of GF is to analyze the presence of any specific frequency content of an image in a specific direction within a localized region of interest [14] and uses a GB made up of 24 2-D GFs for extraction of liver and tumor features from abdominal CT image in a spatial domain. A GF is a sinusoid function modulated by a Gaussian and a 2-D GF in a x-y co-ordinate system is given by

$$g(x, y) = \exp - x\ 2\ c + \gamma\ 2y\ 2\ c\ 2\sigma\ 2\ \mathrm{expi}\ 2\pi xc\ \lambda + \psi\ , (1)$$

$$xc = x\cos\theta + y\sin\theta, (2)$$

$$yc = -x\sin\theta + y\cos\theta. (3)$$

The input parameter σ is the standard deviation of Gaussian function, λ is the wavelength of harmonic function, θ is the orientation, γ is the spatial aspect ratio, and ψ is the phase shift of harmonic function giving description of how much the pattern needs to be shifted with respect to the center, as in (1) .

The output value is the weight of the filter at the (x, y) location. The orientations and frequencies used are θ= 0 ∘, 45 ∘, 90 ∘ , 135 ∘ , and λ= 2.48, 5.69, 11.317, 20.204, 45.678, 90.597 respectively as shown in Fig. 2.

Each of 24 filters designed on these orientations and frequencies was moved through the entire image to extract features. The size of each filter was 7 by 7. To determine feature at each pixel, one of these filters were convolved with a 7 by 7 block of neighborhood pixels with selected pixel as a central pixel. Using this process, we obtained 24 Gabor features. We also incorporated the spatial location of each pixel to existing Gabor features. For each pixel, this resulted in 26 features thereby producing a feature matrix of size 256*256 by 26 for a single CT image slice. D. Classification using DNN Deep

Neural Networks (DNNs) are the advanced or modified form of Neural Network (NN) with more than one hidden layer or more than three layers (including input and output), as in Fig. 3.

In deep learning, each layer of nodes is trained on a distinct set of the features from previous layer output. As the network goes deep, the deep network is trained to recognize for more complex features by aggregating and recombining features from preceding layers. In this paper, we have used DNN in a binary supervised mode. We have implemented the 3 hidden layers with Rectified Linear units (ReLU) as an activation function. In case of ReLU, if X is the input neuron matrix, Bl is the bias and Wl is the corresponding weight matrix for layer l, then the output at each layer from nodes is given by

Zl = ReLU (XWl + Bl) = max (0, XWl + Bl), (4)

where l is a positive value ranging from 1 to a sum of total number of hidden layers and output layer. The output layer is then configured as a logistic function to obtain output value between 0 and 1 which can be represented by

pk = 1 1 + exp (−zl) , (5)

where hidden unit maps input X to a class probability pk. As a loss function for back propagating gradients in training stage, we used softmax cross entropy function given by

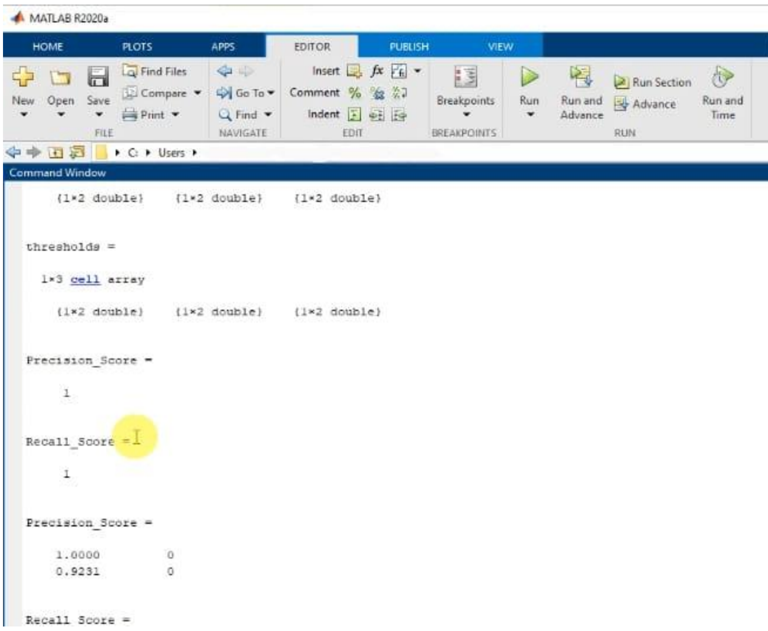
C = − 1 N X N n=1 [ynlogpn+ (1−yn)log(1−pn)]+ λ 2 kWk 2 2 , (6)

If a seed point is selected outside the region of interest (ROI), then the final segmentation would be incorrect. Most of the region growing methods require the seed point to be selected manually in advance. In order to make the region growing segmentation fully automatic, it is necessary to develop an automatic and accurate seed point selection method. For this, a novel method of finding the seed point is proposed in which centroid of the largest connected area in the third eroded image is determined and the coordinates of the centroid point acts as the initial seed point for region growing. The region growing process for liver segmentation starts from the seed point thus selected. Subsequently, the 4- connected neighborhoods are determined. The similarity measure of each neighborhood pixel with the seed point region is calculated. The similarity measure b(X) is the difference between the pixel's gray level value and the mean intensity of the region grown as in equation (l).

Let L be the set of pixels labeled as part of grown region, and I be the set of pixels x in the ROI which are not part of L but are neighbors to L [14].

$\delta(x) = |A(x) - mean(A(y))|$ Where A (x) is the gray level value of the current pixel and mean (A(y)) is the mean of the already grown region. The pixel with the minimum difference measured by this way was added to the grown region. This process continues iteratively by comparing all unallocated adjacent neighboring pixels of the grown region using the similarity measure.

5.OUTPUT:



6.CONCLUSION:

This paper presents the machine learning classifiers method for liver cancer and lesion identification and segmentation. Various layers in the neural network are utilized to extract features of medical images to improve the accuracy of the detection of medical images. 2D feature maps are combined with several slices in the feature extraction process. The algorithm showed very accurate liver volume measurements of above 95 percent. The study showed a high accuracy of the segmentation method had an average coefficient of 0.92. The results show that the PNN produces the best results with data changes, adjacent slices, and appropriate class weights. Note that a limited dataset and testing have been done 3 fold cross-validation. Probabilistic Neural Network is used to generate the follow-up segmentation of tumors as knn classifier. The segmentation leaks are then eliminated in the resulting segmentation so that the final results can be obtained. The proposed machine learning classifiers method has high accuracy in terms of identifying liver tumors.

7.REFERENCES:

1. Guo, L. H., Wang, D., Qian, Y. Y., Zheng, X., Zhao, C. K., Li, X. L., ... & Xu, H. X. (2018). A two-stage multi-view learning framework based on computeraided diagnosis of liver tumors with contrast-enhanced ultrasound images. Clinical hemorheology and microcirculation, 69(3), 343-354. 2. Vorontsov, E., Cerny, M., Régnier, P., Di Jorio, L., Pal, C. J., Lapointe, R., ... & Tang, A. (2019). Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases. Radiology: Artificial Intelligence, 1(2), 180014. 3. Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2), 574-582. 4. Yan, K., Wang, X., Lu, L., & Summers, R. M. (2018). DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of medical imaging, 5(3), 036501. 5. Selvarani.A and Suresh.G.R(2020),"Preclinical Diagnosis of diabetes with tongue infrared thermography and PSO algorithm",Medico-Legal Update Vol.20 No.1 pp.191-197,April,ISSN:0971-720 DOI:10.37506/v20/I1/2020/mlu/194322. 6.VJ Ramya,Navarajan.J,Prathipa.R,LA Kumar.Detection of melanoma skin cancer using digital camera images.ARPN Journal of Engineering and Applied Sciences 10(7),3082-3085,2015. 7.Selvarani.A and Suresh.G.R (2016),"Texture Analysis of Diabetic tongue using binary pattern "International Journal of Control Theory and Applications,Vol.9. No.5,pp.2205-2212, January,ISSN:09745572 8.Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., ... & Sommer, W. H. (2016, October). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 415-423). Springer, Cham. 9. Hu, P., Wu, F., Peng, J., Liang, P of a solitary metastatic liver tum., & Kong, D. (2016). Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Physics in Medicine & Biology, 61(24), 8676. [2] Kenei Furukawa and Katsuhiko Yanag," Successful resection or from prostate cancer 15 years after radical prostatectomy: a case report," Springer, DOI: 10.1186/s40792-017- 0292-4, Jan 2017. 10. Anisha P R, Kishor Kumar Reddy C, Narasimha Prasad L V "A Pragmatic approach for Detecting Liver Cancer using Image Processing and Data Mining Techniques", SPACES-2015, Dept. of ECE, K L UNIVERSITY 2015. 11. 11. M V Sudhamani, G T Raju,"Segmentation and Classification of Tumor in Computed Tomography Liver Images for Detection, Analysis and Preoperative Planning", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-1 Issue-14 March-2014. 12. L. Ali, A. Hussain, J. Li, U. Zakir, X. Yan, A. Shah, U. Sudhakar, B. Luo" Intelligent Image Processing Techniques for Cancer Progression Detection, Recognition and Prediction in the Human Liver", 978-1-4799-5375-2/14 ©2014 IEEE. 13. Yu Masuda, Amir Hossein Foruzan, Tomoko Tateyama and Yen Wei Chen., "Automatic Liver Tumor Detection Using EM/MPM Algorithm and Shape Information", 13th International Conference on Control, Automation, Robotics & Vision Marina Bay Sands, Singapore, 10-12th December 2014 (ICARCV 2014). 14. El Masry and et al., "Automatic liver CT image clustering based on invasive weed optimization algorithm", Engineering and Technology, 1-5, IEEE 2014. 15. Abdalla Zidan, "Level Set based CT liver image segmentation with watershed and artificial neural networks, International conference on hybrid intelligence systems, 96-102, IEEE 2012. 16.Pedro Rodrigues, Jaime Fonseca, João L. Vilaça, "An Image Processing Application for Liver Tumour Segmentation", 1st Portuguese Meeting in Bioengineering, February 2011