CLASSIFICATION OF LIVER TUMOURS RANDOM FOREST ALGORITHM AND DNN METHOD

A JEMILA RANI
Lecturer ECE
Kamaraj Polytechnic College,
Pazhavilai.

ABSTRACT

The early stage liver disease prediction is a significant piece of health-related study, and with this type of research, one may forecast diseases and apply treatments with ease. The numerous types of liver illnesses include cirrhosis, hepatitis, fatty liver, liver cancer, and liver tumours. A critical step in the diagnosis and treatment of liver cancer is the categorization of liver tumours from computed tomography (CT) images. This research presents a convolutional neural network-based automatic approach for segmenting lesions from CT images. The CNNs is one of the deep learning models that uses convolutional filters to learn hierarchical features from data. We contrasted the DNNs model with well-known machine learning methods, Random Forests (RF) and support vector machines (SVM). Using leave-one-out cross validation, experimental evaluation was done on 30 portal phase enhanced CT images. The outcomes demonstrate that the CNNs method is promising for segmenting liver tumours and performs better than other methods.

Keywords: Computed Tomography, Liver tumours, Random forest algorithm, DNN

I INTRODUCTION

In the entire world, cancer is the second biggest killer. The World Health Organisation (WHO) reports that it was to blame for 8.8 million deaths in 2015, with liver cancer accounting for 788,000 of those [1]. According to the American Cancer Society, primary liver cancer and intrahepatic bile duct cancer will kill 28,920 persons in the United States alone in 2017 [2]. There will be around 40,710 new cases diagnosed (29,200 men and 11,510 women), and approximately 40,710 new cases (29,200 men and 11,510 women) will be diagnosed [2]. More than 600,000 people die each year from liver cancer, which is more prevalent in Southeast Asia and sub-Saharan Africa. Liver cancers are the cancers that occur in the Liver organ. The tumours that affect the liver are known as liver cancers. The source of liver cancer might come from the liver itself or from one or more other body organs. The fundamental units of the body are cells, and cancer is one particular sort of cell sickness. A mass called a tumour may develop as a result of unchecked liver cell proliferation. These tumours have a variety of causes, which leads to a variety of treatments. Information about the various forms of tumours is provided by the medical imaging of the tumour.[3] Malignant (cancerous) or benign (non-cancerous) tumours are both possible. While a malignant tumour can spread to other areas of the body, a benign tumour cannot. Liver tumours that are malignant start off there and develop either inside or outside. Deep learning has emerged as a popular area of study in the field of image processing in recent years. Numerous studies have demonstrated that CNNs are capable of achieving exceptional performance on extremely difficult tasks, including picture categorization and visual object recognition. [4] The segmenting of knee cartilage likewise used the CNNs approach. A supervised learning model made up of multi-layer neural networks, Deep

CNNs was first introduced by LeCun et al. Each intermediate layer represents an abstraction at a higher level. Highly nonlinear mappings between inputs and outputs may be captured by it. By creating high-level characteristics out of low-level ones, CNNs—which are entirely data-driven—can obtain hierarchical features.

II RELATED WORK

Deep learning has grown in popularity as a research area in the field of image processing in recent years. Numerous research have demonstrated that CNNs may achieve exceptional performance on extremely difficult tasks, such as visual object recognition and image categorization. The segmenting of knee cartilage similarly used the CNNs model [5]. A supervised learning model made up of multi-layer neural networks is called Deep CNNs. Each intermediate layer represents an increasingly more abstract level. It may be able to record extremely nonlinear input–output mappings. Due to their complete reliance on data, CNNs are capable of retrieving hierarchical features by creating high-level features from low-level ones. [6] According to the literature review, the importance of medical imaging is increasing as there is a growing need for an accurate and effective diagnosis in a short amount of time. Vascular, metabolic, secretory, and excretory functions are all supported by the liver. A pathological abnormality in the liver can be examined by doctors using the medical imaging technique known as CT. The primary problem with segmenting the liver from CT scans is the lack of contrast between the intensities of the liver and surrounding organs. Additionally, the liver may appear in multiple dimensions, which makes segmenting and identifying it even harder. [7] Segmenting liver tumours from CT images using an interactive technique. The liver was initially divided into segments using pre-processing methods, and the watershed transform was then used to divide the CT volume into the several catchment basins. By training a support vector machine (SVM) classifier using the user-selected seed points, the tumours were then retrieved from the segmented liver. [7] The watershed transform's small regions that were retrieved were used to calculate the corresponding features for training and prediction. Manual seed point selection is necessary for this strategy.

Proposed strategy for the two-level surgeries to segment tumours. Two techniques, global threshold with morphological operations and adaptive threshold with morphological operations, are used to segment the liver at the first level. [8] Three techniques were utilised for second level tumour segmentation at that level: fuzzy c mean clustering, region growing, and adaptive threshold with morphological operations. The suggested method makes comparisons, picks the top candidate, and then generates the final outcome. It is increasing the segmentation accuracy for different categories and quality levels of CT images. In [9], Kumar and Moni employed morphological processing and adaptive threshold to segment the liver. The fuzzy c-mean technique was used to automatically remove each suspicious tumour location from the segmented liver. Additionally, in [10], 10 liver CT scans with a resolution of 512512 are employed, and a histogram equalisation is performed to improve image quality and contrast. Neural networks are also used to distinguish between normal and pathological liver regions. Statistically based features, intensity based features, morphological based features, frequency domain based features, and wavelet domain based features are the five sets of features that are extracted in the feature extraction stage. [11][12] The feature selection process uses principal component

analysis (PCA). For the decision-making process, a feed-forward multi-layer perceptron is trained using a back propagation technique for the chosen features.

The remainder of this survey was organised as it went along. The primary deep learning methods that have been used to medical image analysis and that are mentioned throughout the survey are introduced in Section 2 of this report. Pre-processing, Segmentation, and Classification are the three aspects of medical image analysis that the Random Forest algorithm contributes to. The results are discussed in Section 4.

III PROPOSED METHOD

In this study, the input image is classified using a novel Random Forest method that is proposed. The proposed method adheres to a number of the patterns of the CNN architecture. The Random Forest technique uses the feature vectors generated by CNN to enhance classification performance. Studies show that by varying the feature vector dimension, the Random Forest approach improves classifier performance in image processing.

3.1 Image Pre-processing

Pre-processing medical images primarily serves the aim of analysing and scrutinising the image for diagnostic purposes. Many medical image processing applications start with this crucial phase, which filters the image to a higher standard. In order to accomplish this, it normalises the intensity of discrete particle rudiments, screens sections of images, and replicates. To acquire the total intensity and contrast information at this step, the normalisation technique is also used. The pre-processed image affects the segmentation and classification outcomes. In the initial data collection stage, the sliding window gathers information. In the subsequent fresh sample phase, the fresh measurement information is added to the sliding window's random matrix S rightmost column vector$x(T)$. The normalised matrix $X_n = (x_{ij})_{m,n}$ is then obtained by normalising the sliding window matrix using the formula (1). The matrix has a variance of 1, and an expected value of 0.

$$x_{ij} = [x_i(j) - \mu_{(\bar{x}_i)}] \frac{\sigma(\bar{x}_i)}{\sigma(\bar{x}_i)} + \mu_{(\bar{x}_i)}, \qquad (1)$$

The expected value and variance of each row of the normalised matrix are represented by $\mu_{(\bar{x}_i)}$ and $\sigma(\bar{x}_i)$ respectively, where $\mu_{(\bar{x}_i)}$ and $\sigma(\bar{x}_i)$ represent the expected value and variance of each row of the normalised matrix, respectively.

3.2 SEGMENTATION PROCESS OF LIVER AND LIVER TUMOR

A medical image is divided into liver parenchyma and non-liver parenchyma regions by the procedure of liver segmentation. In deep learning, every layer's nodes are taught using a different set of features from the output of the layer before them. By integrating and recombining data from earlier layers, the deep network is trained to recognise more complex features as it descends. The ability to accommodate enormously big, highly dimensional data with billions of parameters traversing the nonlinear function is known as feature hierarchy

and it is a key component of deep learning networks. Both feature categorization and automated feature extraction may be done using DNN.

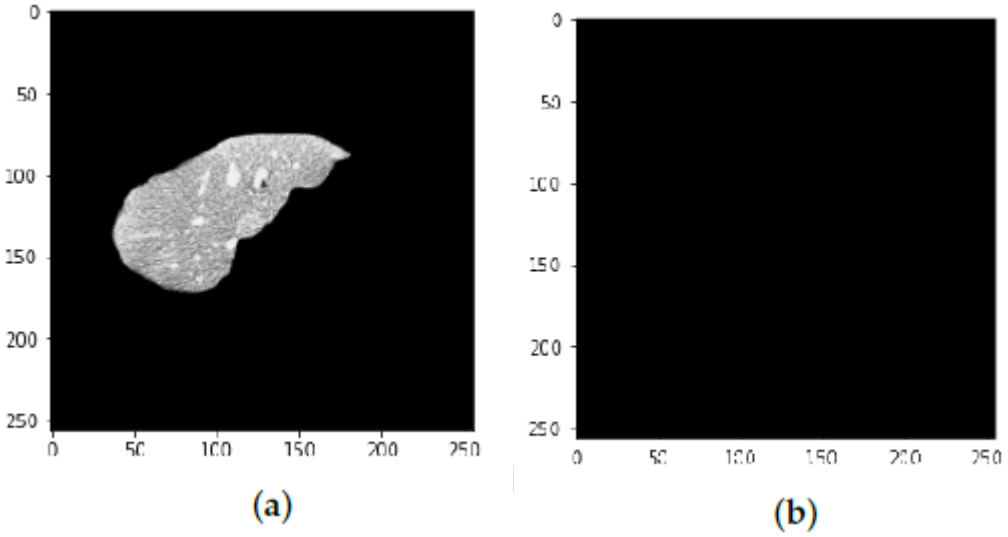

Figure 1. Results of the suggested model: (a) segmentation of the liver; (b) segmentation of the tumour.

The DNN classifier can analyse considerably larger datasets since it is trained on labelled data before using it on unlabelled, unstructured data. In addition, the methods of other researchers discussed here have only employed a single-scale, high-magnification patch with a pattern based on cell-level information. In (Figure 1) the tumor-based cells were not completely identifiable. The difficult task of distinguishing between normal and diseased liver cells required a careful analysis of the layers of 3D images.

3.3 Random Forest (RF) Classifier

A supervised learning method called random forest uses an ensemble of trees, where each tree is reliant on a collection of random variables. It combines the two categorization trees, which differ in two crucial ways. A random bootstrap sample is initially taken from the full dataset and used to fit each tree. Until our sample is the same size as the original, we randomly sample microarrays from the underlying data with replacement to create the bootstrap sample. Microarrays that were excluded from the bootstrap sample are referred to as "out-of-bag data," and these microarrays act as a natural test set for the tree that is fitted to the bootstrap sample. Because we no longer select the best viable split for every gene, the bushes differ. Instead, we choose a pattern with a handful of genes for each node and choose the most advantageous cut of the selected genes. The rectangle root of the total number of genes is often the range of genes selected at each node. We assume that the joint distribution $f_{XT}(X,T)$of the random vectors $X = (X_1, X_2, \dots, X_k)$ (denoting the real-valued predictor variables) and T (denoting the response variable) is unknown. Finding a prediction function f(x) for forecasting T is the key objective. The prediction function is obtained by minimising the loss function L(T, f(X)); the zero-one loss is frequently applied to classification.

$$L(T, f(X)) = \begin{cases} 0 & if \quad T = f(X) \\ 1 & otherwise \end{cases} \qquad (2)$$

Let $T = X = (X_1, t_1), (X_2, t_2), \dots, (X_N, t_N)$ represent the training dataset, with$i = 1,2, \dots, N$. Take an N-size $T_m$bootstrap sample from the variable$j$. Using the bootstrap sample T_m as the training data, fit a tree using binary recursive partitioning. Put each observation into its own node to start. After selecting m predictors at random from the p available predictors, calculate the ideal binary split based on the m predictors. Finally, split the node into two descendant nodes using the split, then repeat the operation until the halting requirement is met.

IV RESULTS AND DISCUSSION

The segmentation method and classifiers' implementation and simulation results are covered in this chapter. On an i7 3.6GHz processor running Windows 10 64-bit with 16.0 GB of RAM, Matlab was used to run the simulation. The 3D-IRCADb (3D Image Reconstruction for Comparison of Algorithm Database) was used to obtain all of the images used for training and testing. Each image that was captured was 512 by 512 pixels in size, with corresponding manually segmented labels offered for assessment. All of the provided photographs underwent pre-processing processes to improve their contrast. We compared each classifier's performance using random forest methods for classification.

4.1 PERFORMANCE MATRIX

We used Accuracy (AC), Sensitivity (SN), and Specificity (SP) to assess the performance of the classifiers for the liver and tumour pixel classes (Tables 1). The following section goes into further depth about the parameters for performance evaluation of the liver and tumour segmentation algorithms.

Accuracy (AC): The AC is determined by dividing the total number of accurate predictions by the total dataset. The formula for AC can be written as follows: The total number of all right predictions is equal to the sum of the TN and TP value.

$$Accuracy = \frac{TP}{(TP + TN + FP + FN)} \quad (3)$$

Sensitivity (SN): SN is determined by dividing the total number of correctly predicted positive outcomes by the total number of positive outcomes. Recall or True Positive Rate (TPR) are other names for SN. SN may be written as:

$$Sensitivity = \frac{TP}{(\mathrm{TP} + \mathrm{FN})} \quad (4)$$

Specificity: Specificity (SP) is determined by dividing the total number of correctly predicted negative outcomes by the total number of negatives. True Negative Rate (TNR) is another name for it, and the following formula is used to calculate it:

$$\mathrm{Specificity} = \frac{TN}{(TN + FP)} \quad (5)$$

The values of AC, SN, and SP are noticeably high at 97%, as shown in Tables 1 above. A high AC value does not always imply that the classifier or the entire segmentation process is effective. The value of accuracy alone is typically unreliable in the situation of unbalanced data, similar to the scenario of liver and tumour segmentation when a big percentage of pixels or cells in an image aren't liver or tumour and a smaller percentage are.

Classifiers	Accuracy	Sensitivity	Specificity
RF	0.9952	0.9994	0.9936
SVM	0.9807	0.9996	0.9802
DNN	0.9845	0.9959	0.9842

Table 1: Accuracy, Sensitivity and Specificity of ML classier for Tumor segmentation

We have therefore examined sensitivity and specificity in order to confirm that high accuracy values indicate higher performing classifier. Both positive and negative classes are classified using the SN or TPR, and vice versa. As a result of our high SN and SN values, the classifier is effective at separating the positive and negative classes, respectively, as seen by the high values for each class. The classifier appears to be effective at differentiating both positive and negative classes, as indicated by these numbers, which also support high accuracy values.

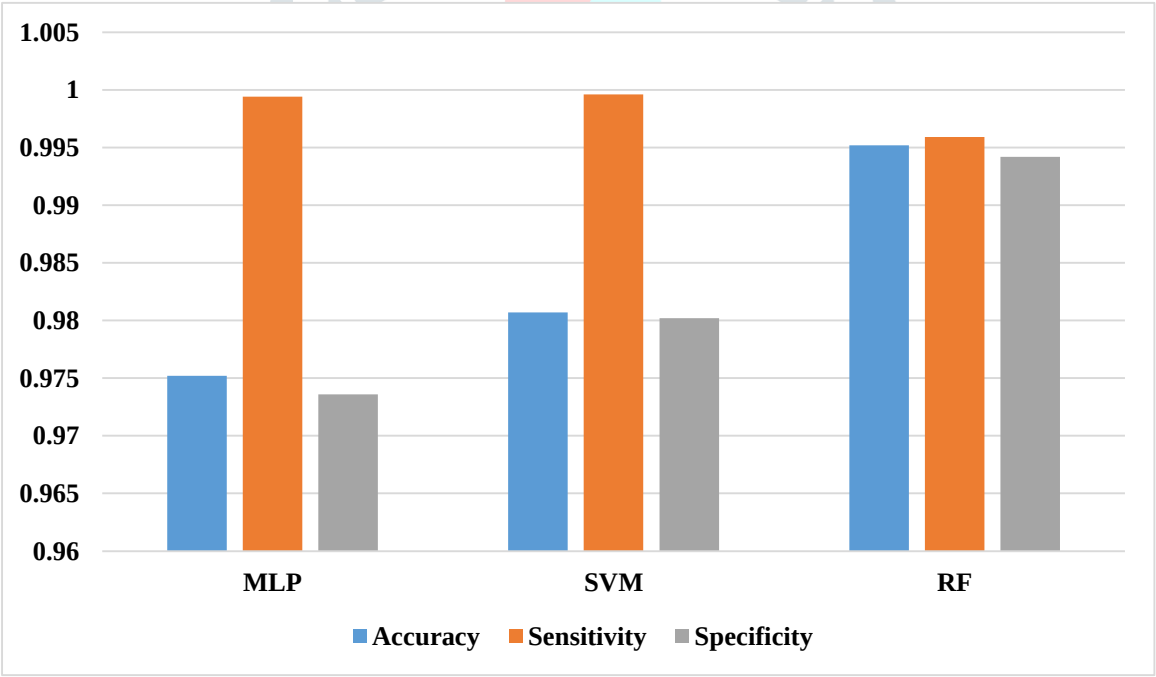

Figure 1: Comparison Graph

The accuracy, sensitivity, and specificity comparison values for the methods MLP, support vector machine (SVM) algorithm-based Convolutional Neural Network (CNN), and the suggested Random forest algorithm are each shown in figure 1 above. This comparison graph unmistakably demonstrates that when compared to the current approaches, the proposed method had the greatest values.

CONCLUSION

The research discussed in this article suggests that a CT scan of the liver can be utilised to distinguish between the healthy liver and the area of the liver that contains tumours. The Random Forest algorithm was employed in this work to categorise liver cancers. The major objective is to use CT scan pictures of the liver to categorise cancer as benign or malignant. Finally, it is recommended that computed tomography (CT) pictures of livers containing malignancies be categorised using a Random forest algorithm technique. Four machine learning classifiers were utilised with this upgraded hybrid feature dataset: MLP, SVM, and RF. It was shown that after applying the Random forest method classifier, a 99.6% overall accuracy was obtained in these liver tumours.

REFERENCES

1. [1] Anter A, Azar A, Hassanien A, El-Bendary N, ElSoud M. Automatic computer aided segmentation for liver and hepatic lesions using hybrid segmentations techniques. IEEE proceedings of federated conference on computer science and information systems 2013:193–8.
2. Priyadarsini S, Selvathi D. Survey on segmentation of liver from CT images. 2012 IEEE international conference on advanced communication control and computing technologies (ICACCCT). 2012. p. 234–8.
3. Hassan, Elmogy, Sallam. Medical image segmentation for liver diseases: a survey. Int J Comput Appl 2015;118(19).
4. Tustison, N.J., Shrinidhi, K.L., Wintermark, M., Durst, C.R., Kandel, B.M., Gee, J.C., et al. (2015) Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR. Neuroinformatics, 13, 209-225. http://dx.doi.org/10.1007/s12021-014-9245-2 [Citation Time(s):1]
5. Ciresan, D.C., Giusti, A., Gambardella, L.M. and Schmidhuber, J. (2013) Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. MICCAI 2013, LNCS 8150, 411-418. http://dx.doi.org/10.1007/978-3-642-40763-5_51 [Citation Time(s):1]
6. Selvathi, D.; Priyadarsini, S.; Malini, C.; Shanmugavalli, P. Performance analysis of multi resolution transforms with kernel classifiers for liver tumor detection using CT images. *Int. J. Appl. Eng. Res.* **2014**, *9*, 30935–30952. [**Google Scholar**]
7. Li, D.; Liu, L.; Chen, J.; Li, H.; Yin, Y. A multistep liver segmentation strategy by combining level set based method with texture analysis for CT images. In Proceedings of the 2014 International Conference on Orange Technologies, Xi'an, China, 20–23 September 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 109–112.
8. Xu, Z.; Burke, R.P.; Lee, C.P.; Baucom, R.B.; Poulose, B.K.; Abramson, R.G.; Landman, B.A. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. Med. Image Anal. 2015, 24, 18–27. [CrossRef]

9. Song, X.; Cheng, M.;Wang, B.; Huang, S.; Huang, X.; Yang, J. Adaptive fast marching method for automatic liver segmentation from CT images. Med. Phys. 2013, 40, 091917. [CrossRef]
10. Maklad, A.S.; Matsuhiro, M.; Suzuki, H.; Kawata, Y.; Niki, N.; Satake, M.; Moriyama, N.; Utsunomiya, T.; Shimada, M. Blood vessel-based liver segmentation using the portal phase of an abdominal CT dataset. Med. Phys. 2013, 40, 113501. [CrossRef]
11. Peng, J.; Dong, F.; Chen, Y.; Kong, D. A region-appearance-based adaptive variational model for 3D liver segmentation. Med. Phys. 2014, 41, 043502. [CrossRef]
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012, 25. [CrossRef]