A Novel Computer-Aided Diagnosis Framework Using Deep Learning for Classification of Fatty Liver Disease in Ultrasound Imaging

D Santhosh Reddy, R Bharath and P Rajalakshmi
Wireless Networks Laboratory, Department of EE
Indian Institute of Technology Hyderabad, Telangana, India
email: { ee15resch11005, ee13p0007, raji }@ iith.ac.in

***Abstract*— Fatty Liver Disease (FLD), if left untreated can progress into fatal chronic diseases (Eg. fibrosis, cirrhosis, liver cancer, etc.) leading to permanent liver failure. Doctors usually use ultrasound scanning as the primary modality for quantifying the amount of fat deposition in the liver tissues, to categorize the FLD into normal and abnormal. However, this quantification or diagnostic accuracy depends on the expertise and skill of the radiologist. With the advent of Health 4.0 and the Computer Aided Diagnosis (CAD) techniques, the accuracy in detection of FLD using the ultrasound by the sonographers and clinicians can be improved. Along with an accurate diagnosis, the CAD techniques will help radiologists to diagnose more patients in less time. Hence, to improve the classification accuracy of FLD using ultrasound images, we propose a novel CAD framework using convolution neural networks and transfer learning (pre-trained VGG-16 model). Performance analysis shows that the proposed framework offers an FLD classification accuracy of 90.6% in classifying normal and fatty liver images.**

I. INTRODUCTION

With the advent of Health 4.0, the way the healthcare delivered is taking a major leap and seems to be a promising solution for increasing population [1]. Both the developed and developing nations are mainly focusing on automated algorithms and remote healthcare delivery solutions for providing quality healthcare [2]. Due to degradation of human lifestyle and food habits, one of the major problem the population witnessing is Fatty Liver Disease (FLD). The development of FLD closely relates to the accumulation of excess fat in the liver cells and tissues. Although FLD in the initial stages may not cause a fatality, it progress into fatal chronic diseases (Eg. fibrosis, cirrhosis, liver cancer, etc.) if left untreated. Studies show that 20-30% of the world population is affected by the FLD and we strongly feel that addressing the challenges in accurate identification of fatty liver is the need of the hour [2]. Hence, our aim is to develop a novel computer-aided diagnosis (CAD) framework for accurate classification of FLD using ultrasound images of the liver.

Doctors treat FLD using both invasive procedures (such as blood tests and biopsies which involve complications such as infections, bile leakage, etc.), and noninvasive imaging techniques (such as ultrasound scanning, computed tomography scanning, etc.). However, multiple techniques exist ultrasound imaging is used widely for treating the FLD [4], [5]. The popular technique used by the doctors in identifying FLD is the comparison of liver echogenicity with the renal cortex [3]. Although this method is well established for the FLD diagnosis, the clinicians working in remote areas face difficulties due to their lack of skills and expertise in identifying the FLD for the ultrasound images of the liver. Computer-aided diagnosis (CAD) [6], [7] in such scenarios aid the clinicians to perform an accurate diagnosis by automating the detection and classification of the FLD. The CAD techniques also reduce the bias caused due to the skill of the radiographers. However, the methods proposed in the literature still suffer from the inaccurate classification of FLD using liver ultrasound images and hence is an active research area [6], [7].

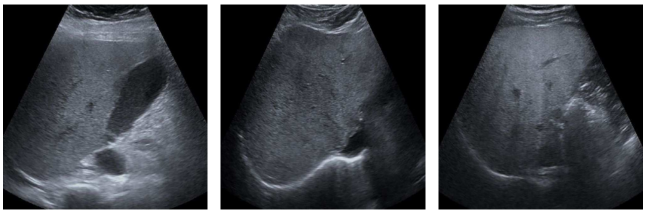
The rest of this manuscript is organized as follows. We discuss the related works in Section II and in Section III, we present the proposed CAD framework based on deep learning, transfer learning and fine-tuning along with the database description used for the experimental analysis. Finally, Section IV and Section V analyze the performance of the proposed framework and concludes the paper by discussing the future scope of the work, respectively.

II. RELATED WORKS

In ultrasound images, the texture appears as granular patterns at the parenchyma of a liver, whereas in the case of FLD the texture appears finer and smoother as shown in Fig. 1. Recently several studies proposed novel classification methods of the normal and fatty liver using ultrasound images. Andrade *et al.* in [19] developed an FLD classification model using Support Vector Machines (SVM) classifier based on the following features: First-Order Gray Level Parameters (FOGLP), Gray-Level Co-occurrence Matrix (GLCM), Law Texture Energy (LTE), Fractal Dimension (FD) with an accuracy of 79.7%. Singh *et al.* in [20] performed FLD classification using Fishers linear discriminant analysis based on features such as Spatial Gray-Level Co-occurrence Matrix (SGLCM), Statistical Feature Matrix (SFM), LTE, Fourier Power Spectrum (FPS) and fractals, with 92% accuracy. Minhas *et al.* in [21] used SVM with features extracted using Wavelet Packet Transform (WPT) and achieved 95% accuracy. Acharya *et al.* in [22] detected the FLD using features obtained from Wavelet, Higher-Order Spectra (HOS) and achieved 93.3% accuracy when used along with a decision tree classifier. Dan *et al.* [23] used random forest

978-1-5386-4294-8/18/$31.00

(a) Normal Liver



(b) Fatty Liver

Fig. 1: Ultrasound images of liver with and without FLD.

and SVM Classifiers using attenuation gray level, a variance of the skewness, and the kurtosis measure of the texture as features and obtained an accuracy of 90.8% and 87.7%, respectively. In [24], the authors proposed a classification framework of FLD using the handcrafted texture features with an accuracy of 95%. Acharya *et al.* [25] achieved 98% accuracy using Probabilistic Neural Network (PNN) along with GIST descriptors. In literature, researchers are mainly focused on handcrafted features. Hence, in this study, we propose a convolution neural network (CNN) and transfer learning (using pre-trained VGG-16 model) based CAD framework to improve the FLD classification accuracy using ultrasound images.

Many deep learning techniques have gained significant attention and are emerging in all the domains of engineering. Particularly, CNN's have been proven to be an efficient and reliable method for many of the computer vision tasks [8]–[10], which automatically learns mid-level and high-level abstractions from the database [11]. Recently, deep learning approaches appear promising in the domain of medical image analysis. However, in the medical imaging, acquiring the required amount of data for training the deep learning models remains to be a significant challenge when compared to natural images. When deep learning models are trained with the limited medical data available, deep CNN tends to suffer from over-fitting, and there arise the convergence problems. Also, training deep CNNs usually require large computational resources. To address these issues, a new transfer learning approaches have been applied to deep CNN's thereby enabling their use in medical imaging applications with smaller datasets [18]. Transfer learning enables us to use the pre-trained CNN models (models trained using natural image datasets or any other medical image datasets) for the classification task. The pre-trained CNN models can be used for generation of features from the input images which then can be then to train a new classifier [12]. For example, in [13], the authors used a pre-trained CNN using adequate natural images and implemented the convolutional layers of pre-trained CNN in a new CNN framework for locating the fetal abdominal plane in the ultrasound video sequence. Tajbakhsh *et al.* [12] used an already CNN along with layer-wise fine-tuning to improve the performance in locating the fetal abdominal standard plane using a small dataset. Using the available already trained CNN models in conjunction with fine-tuning for achieving better classification accuracy is already being explored at a rapid pace and also found to be successful in diversified applications [14]–[16]. Also, we would like to advise the readers to refer [17] for more such applications in the medical imaging field. Motivated by the success of transfer learning, in this paper, we adhere to transfer learning and fine-tuning an already trained CNN layer-wise for classifying the normal and fatty liver ultrasound image accurately.

III. Database description and the Proposed CAD algorithm for classification of Fatty Liver

In this section, we discuss the database considered for experimental analysis and the proposed CAD framework the classification of FLD using CNN, transfer learning and fine-tuning.

A. *Description of the Developed Dataset*

The open datasets corresponding to FLD is not available for research activities. Hence, in this study, the database is collected with the help of two well-trained sonographers. The database collection involved both male and female patients with an age group ranging from 20 to 55 years. We have collected 81 normal liver images and 76 were suffering from FLD. All the images used for this study are collected from a Siemens Acuson S1000 ultrasound scanner with curved array transducer. Every image in the database is labeled by the radiographers. The excessive blank/black space on either side of the liver organ does not convey any diagnostic information and hence it is cropped before training the network. The cropped images are then resized to a fixed size of 224×224 pixels and are used in our experiment and also for training the model.

B. *Proposed automated CAD framework for the classification of fatty liver in ultrasound images*

Fig. 2 shows the proposed CAD architecture comprising of the pre-trained VGG-16 model along with the transfer learning and fine-tuning. A VGGNet using CNN is designed markedly with different layer depths for image recognition. The VGGNet gave an accuracy of 92.7% when validated using the ImagNet dataset consisting of 14 million images from 1000 classes [26]. To reduce the complexity of computing weight parameters, a small (3×3) convolution filter with a stride size of 1 is utilized for all the convolutional layers. The max-pooling considers a 2×2 pixel window, with a stride of 2 pixels for the last convolutional layer of each block as shown in Fig. 2.

From Fig. 2, one can observe that three fully-connected (FC) layers follow the stack of convolutional layers. The

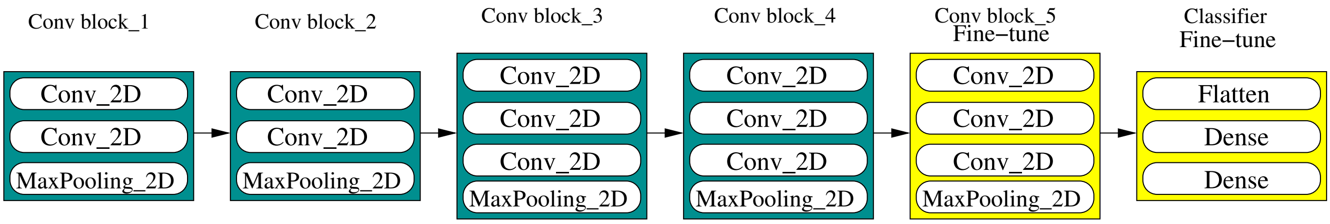


Fig. 2: The proposed VGG-16 architecture using CNN with transfer learning and fine-tuning. We fine tuned the last two blocks in the proposed architecture for getting optimal classification.

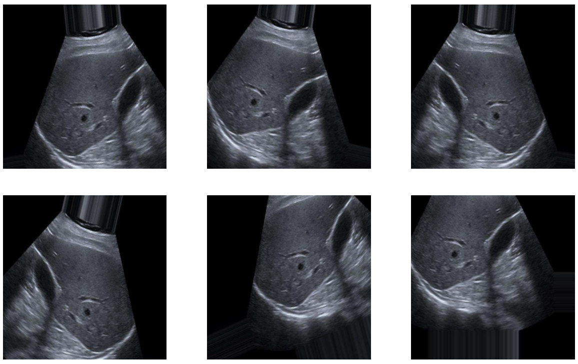


Fig. 3: Fatty liver ultrasound images generated using data augmentation technique.

TABLE I: Performance analysis of the proposed algorithm for detection and classification of FLD for 100 epochs

Class	Precision	Recall	$F1_{score}$	Support
Normal	0.92	0.85	0.88	13
Abnormal	0.90	0.95	0.92	19
Avg/total	0.91	0.91	0.91	32

first two FC layers have 4096 nodes each, while the third layer performs classification using 1000 nodes (one for each class). In addition Rectified Linear Unit (ReLU) is used as activation function in all the hidden layers, while the softmax activation function is employed in the final layer of the FC layers. In this paper, we make use of the VGG-16 comprising 16 layers along with convolution blocks (comprises of convolutional layers and max-pooling layers) and a fully-connected classifier. The fine-tuning is carried out using the pre-trained VGG-16 model in Keras. In the model shown in Fig. 2, the convolution 2D layers of the $conv\ block\_5$ consists of 512 nodes for convolution layers, in our experiment we fine-tuned the network from using 512 nodes to 256 nodes. Also, we fine-tuned fully connected layers (Fc_1 and Fc_2) having 4096 nodes to 256 nodes, and the output layer comprises two neurons whose output corresponds to the two classes (normal and abnormal) in this study. The weight parameters of the output layer are initialized randomly which follows a Gaussian distribution, and the training is performed using 100 epochs to prevent the over-fitting. Because the usage of fine-tuning also requires moderate size database and considering the limitation of having a small database, we used augmentation techniques based on transformations to increase the size of the database. The image transformation techniques like vertical flipping, horizontal flipping, rotation, zoom range, shear range with different orientations are used to generate the new training images. The new training images resulted from the data augmentation is shown in Fig. 3.

IV. Results

From a total of 157 ultrasound liver images, 125 images are used for training and 32 images are used for testing the trained model. Using the 125 images of the training set, a total of 2500 images are generated with different transformations using data augmentation techniques. This training set of 2500 images are used for training the model described in Section III. The performance of the proposed model is analyzed using classification accuracy, confusion matrix, F_{score}, $Precision$, and $Recall$ is shown in Table. I. The description of the considered performance metrics are given below:

$$F_{score} = 2\frac{recall * precision}{(recall + precision)}, \tag{1}$$

$$Recall = \frac{N(TP)}{N(TP) + N(FN)}, \tag{2}$$

$$Precision = \frac{N(TP)}{N(TP) + N(FP)}, \tag{3}$$

$$Accuracy = \frac{N(TP) + N(TN)}{N(TP) + N(TN) + N(FP) + N(FN)}, \tag{4}$$

Where $N(TP)$ indicates the total true positives, $N(FP)$ indicates total false positives, $N(TN)$ indicates total true negatives and $N(FN)$ indicates the false negatives. All these measures are computed for each class, and an overall measure of the algorithm is computed by taking the average of all these measures across the two classes.

The proposed network is trained for 100 epochs. Fig. 4 shows the training and testing accuracy of the proposed model at various epochs. From the Fig. 4, it can also be observed that the proposed framework achieves an overall classification accuracy of 90.6%. Also, the performance comparison of the proposed method is with basic CNN, VGG-16 transfer learning and fine-tuning the FLD is shown in Table II. The architecture of basic CNN consists of six convolutional layers followed by max-pooling layers

TABLE II: Performance analysis of the proposed classification algorithms

Classifier	Sensitivity(%)	Specificity(%)	Accuracy(%)
CNN	0.89	0.85	84.3
VGG16 + Transfer Learning	0.95	0.76	87.5
VGG16 Transfer Learning + Fine Tuning	**0.95**	**0.85**	**90.6**

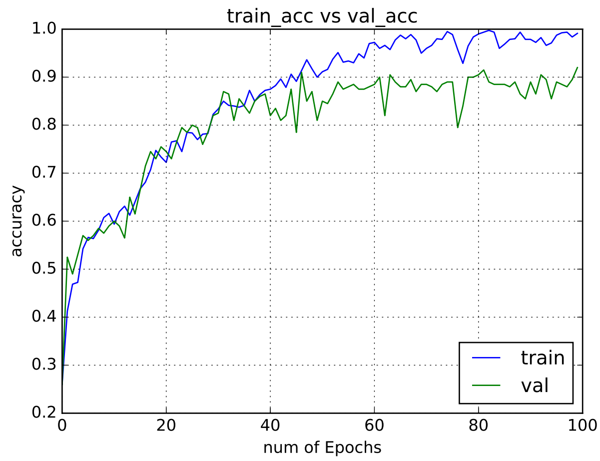


Fig. 4: Training and validation accuracy of the proposed CNN concerning the number of epochs (blue curve indicates training accuracy and green curve indicates the validation accuracy).

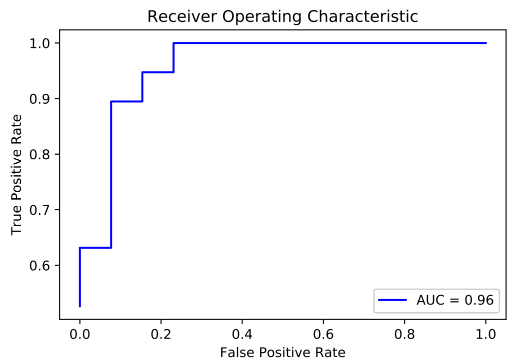


Fig. 5: The ROC curve for the proposed architecture

and four FC layers. In all hidden layers, ReLu is used as an activation function and softmax activation function is employed for the FC layer. A 3X3 convolutional filter size is employed for convolutional layers with the stride 1 and a size of 2X2 is used for max-pooling with the stride 2. The convolutional layers future map 32 is employed for the first two layers followed by 128 to the next two convolutional layers and the last two layers with the 256. All the FC layers with the 256 nodes gave an accuracy of 84.3%. VGG-16 with transfer learning without changing the intermediate layers parameters, last output classification layer is changed to two nodes achieved an accuracy of 87.5% and VGG-16 model with fine tuning last layers of the network achieved an accuracy of 90.6%. VGG-16 fine tuning

TABLE III: Confusion matrix obtained using the proposed CAD framework

True class	**Predicted class**	
	Normal	**Abnormal**
Normal (13)	11	2
Abnormal (19)	1	18

performed better compared to transfer learning and basic CNN model. The performance of the proposed algorithms is assessed with the sensitivity, specificity and accuracy. Sensitivity is defined as the ratio of the true positives that are correctly identified and specificity is the proportion of true negatives that are correctly identified. Table. III provides the confusion matrix obtained using the proposed algorithm and the receiver operating characteristics (ROC) curve for the proposed algorithm is shown in Fig. 5. The ROC represents the performance of the proposed algorithm concerning true positive and false positive rate, the proposed algorithm gave an ROC of 0.96.

The proposed CAD architecture is implemented using Tensorflow and Keras frameworks in Python. The simulations are performed on Intel Xeon(R) E5-2650 V2 CPU running at 2.6 GHz with 16 parallel cores and an NVIDIA GeForce GTX 1080 GPU. The model classifies the FLD using ultrasound images with better accuracy and we strongly feel that this study will have a significant impact in driving the future research in this area.

V. Conclusion

In this paper, we proposed a novel CAD architecture to accurately detect the fatty liver disease using ultrasound images. In many of the developed and developing nations, due to the rapid growth of population, the healthcare delivery is becoming a challenging problem due to the unavailability of skilled clinicians. Especially, in the cases of fatty liver diseases are increasing, and it is estimated that about 15-20% of the world population is suffering from FLD. While ultrasound is the primary modality used for treating the FLD, due to unavailability of the skilled sonographers, the quality of diagnosis offered is being affected severely. To address this problem, we proposed deep learning, transfer learning and fine-tuning for classifying the fatty liver in ultrasound images. The performance analysis of the proposed framework shows that the FLD in ultrasound images can be detected with an accuracy of 90.6%. As a future extension of this work, we try to port the developed algorithm on the hardware platform to make it more translational.

Acknowledgment

The authors would like to thank Dr. Abdul Mateen Mohammad and his team of radiologists at Hyderabad, Telangana, India, for their valuable support in the collection of the database throughout this research. This work is partly funded by Visvesvaraya Ph.D. Scheme, Media Lab Asia, MEITY, Govt. of India and partly funded by Indian Institute of Technology Hyderabad.

REFERENCES

[1] C. T huemmler and C. Bai, *Health 4.0: How virtualization and big data are revolutionizing healthcare,* Springer International Publishing, 2017.

[2] Bellentani, Stefano, et al. “Epidemiology of non-alcoholic fatty liver disease.” *Digestive diseases,* 28.1, 155-161, 2010.

[3] Palmentieri B, de Sio I, La Mura V, *et al.* The role of bright liver echo pattern on ultrasound B-mode examination in the diagnosis of liver steatosis. *Dig Liver Dis 2006*; 38: 485-489

[4] Roldan-Valadez E, Favila R, Martnez-Lpez M, Uribe M, MndezSnchez N. “Imaging techniques for assessing hepatic fat content in nonalcoholic fatty liver disease.” *Ann Hepatol* 2008; 7: 212-220

[5] Saverymuttu SH, Joseph AE, Maxwell JD. “Ultrasound scanning in the detection of hepatic fibrosis and steatosis.” *Br Med J* 1986; 292: 13-15

[6] Goceri, Evgin, et al. “Quantification of liver fat: A comprehensive review.” *Computers in biology and medicine,* 71, 174-189, 2016

[7] P. Bharti, D. Mittal, R. Ananthasivan, “Computer-Aided characterization and Diagnosis of Diffused Liver Diseases Based on Ultrasound Imaging A Review,” *Ultrasonic Imaging,* doi: 10.1177/0161734616639875, 2016.

[8] C. Szegedy *et al.*, “Going deeper with convolutions,” *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 1-9.

[9] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Representations,* 2015.

[10] M. D. Zeiler and R. Fergus. “Visualizing and understanding convolutional networks.” In *Computer VisionECCV 2014*, pages 818833. Springer, 2014

[11] Y. LeCun, K. Kavukcuoglu and C. Farabet, “Convolutional networks and applications in vision,” *Proceedings of 2010 IEEE International Symposium on Circuits and Systems,* Paris, 2010, pp. 253-256.

[12] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” in *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1299-1312, May 2016.

[13] H. Chen et al., “Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks,” in *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 5, pp. 1627-1636, Sept. 2015.

[14] Hayaru Shouno, Satoshi Suzuki, and Shoji Kido. “A transfer learning method with deep convolutional neural network for diffuse lung disease classification”, In *International Conference on Neural Information Processing,* pages 199207, 2015.

[15] R. Girshick, J. Donahue, T. Darrell and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” *2014 IEEE Conference on Computer Vision and Pattern Recognition,* Columbus, OH, 2014, pp. 580-587.

[16] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki and S. Carlsson, “From generic to specific deep representations for visual recognition,” *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Boston, MA, 2015, pp. 36-45.

[17] G. Litjens *et al.*, (2017). “A survey on deep learning in medical image analysis.” [Online]. Available: https://arxiv.org/abs/1702.05747

[18] O. Hadad, R. Bakalo, R. Ben-Ari, S. Hashoul and G. Amit, “Classification of breast lesions using cross-modal deep learning,” *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, 2017, pp. 109-112.

[19] A. Andrade, J. S. Silva, J. Santos, and P. Belo-Soares, “Classifier Approaches for Liver Steatosis using Ultrasound Images,” *Procedia Technology,* vol. 5, pp. 763-770, 2012

[20] M. Singh, S. Singh, and S. Gupta, “A new quantitative metric for liver classification from ultrasound images,” *International Journal of Applied Physics and Mathematics,* vol. 2, no. 6, 2012.

[21] D. Sabih and M. Hussain, “Automated classification of liver disorders using ultrasound images,” *Journal of medical systems,* vol. 36, pp. 3163-3172, 2012.

[22] U. R. Acharya, S. V. Sree, R. Ribeiro et al., “Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm,” *Medical Physics,* vol. 39,no. 7, pp. 42554264, 2012.

[23] D. M. Mihilescu, V. Gui, C. I. Toma, A. Popescu, and I. Sporea. “Computer aided diagnosis method for steatosis rating in ultrasound images using random forests.” *Medical Ultrason,* 15(3):184190, 2013.

[24] M. Singh, S. Singh and S. Gupta, “An information fusion based method for liver classification using texture analysis of ultrasound images”, *Information Fusion*, vol. 19, pp 91-96, 2014.

[25] Acharya, U. Rajendra, et al. “Decision support system for fatty liver disease using GIST descriptors extracted from ultrasound images.” *Information Fusion,* 29, 32-39, 2016.

[26] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition.” *Computer Science,* 2014.