3

Identification of a Suitable Transfer Learning Architecture for Classification: A Case Study with Liver Tumor

B. Lakshmi Priya[1]*, K. Jayanthi[2], Biju Pottakkat[3] and G. Ramkumar[4]

[1]Department of ECE, Manakula Vinayagar Institute of Technology, Puducherry, India
[2]Department of ECE, Pondicherry Engineering College, Puducherry, India
[3]Department of Surgical Gastroenterology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India
[4]Department of Radio Diagnosis, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

Q1 ***Abstract***

Liver diseases in recent days are emerging as a predominant medico-social problem in places where there is high prevalence of alcoholism and among the subset of people with unhealthy dietary habits. The incidence and mortality rates of liver cancer and other liver related ailments witness an increasing trend year by year globally and have almost tripled in the past forty years. As no symptoms are exhibited in early days, liver cancer are diagnosed only at a later stage. Computed tomography (CT) is used as the primary imaging modality for the diagnosis of hepatocellular carcinoma (HCC), a primary liver cancer and other liver related disease. The CT image as such does not provide any clinical information pertaining to liver cancer/tumor, and hence, an intravenous iodinated contrast agent is injected prior to CT acquisition for the purpose of highlighting the tumorous tissue from the healthy liver. Accordingly, contrast enhanced computed tomography (CECT) images are acquired to make the tumorous tissue to be predominantly visible and influence well during the clinical diagnosis. In spite of the herculean visualization, CECT at times fails to provide clear picture of the abnormal parts of the liver which is otherwise called as lesions, making the diagnosis sub-optimal.

**Corresponding author*: lakshmipriyaece@mvit.edu.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (53–78) © 2021 Scrivener Publishing LLC

Additionally, different types of liver tumor exhibit similar visual appearance in CECT images, resulting in obscure diagnosis. This nudges a need for an intelligent expert system that precisely detects the presence of cancerous tissues within the organ from the CECT images.

The challenge of precise diagnosis of liver cancer can be addressed with the help of transfer learning architectures which eloquently perform well in the diagnosis of different categories of cancer. Transfer learning is a deep learning strategy, wherein the knowledge acquired by a model or network on a specific classification task is transferred to another classification task thereby leveraging in achieving a high classification accuracy. This chapter aims at exploring the classification performance of popular transfer learning architectures AlexNet, GoogLeNet, ResNet-18, and ResNet-50 in classifying the different types of liver lesions. Eventually, a suitable transfer learning architecture for liver tumor detection will be identified as a final outcome of this case study.

***Keywords*:** Liver tumour, liver CT, cancer, transfer learning, AlexNet, GoogLeNet, ResNet-18, ResNet-50

3.1 Introduction

Liver cancer is found to be the fifth and eighth leading cause of cancer related deaths among men and women, respectively. As per the survey by American Cancer Society [1], the incidence of liver cancer has almost tripled since 1980s. The statistics elucidate the substantial rise in the incidence of liver related disorders and the need for synergy between technological advancements and medical diagnostics in the remedial treatment. Apart from this, these surveys on liver cancer have some key facts in common. These are a) majority of people with liver disease are not aware that they have some sort of liver disorders. Liver diseases do not exhibit any symptoms in the early stage and are diagnosed only at a later stage; and b) alcoholic liver disease and viral hepatitis.

Against the sudden rise in liver diseases in recent years, the diagnosis of these diseases by the medical experts is highly challenging. For the diagnosis of liver disorders, imaging modalities like computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) can be used, of which liver surgeons predominantly prefer CT as the screening modality. Unfortunately, the plain CT image acquired cannot figure out the diseased portion of the liver from the healthy liver parenchyma. Henceforth, radiologists inject an iodinated contrast material into the patient's body prior to acquiring CT. The iodinated contrast does the job

of visually discriminating the unhealthy portion of the liver from the healthy ones and accordingly the CT is called contrast enhanced computed tomography (CECT). The contrast injected blood brings in transient changes in the appearance of liver tissues in the CECT images and this temporal behavior is exploited in the diagnostic procedure of liver disorders. As a consequence, liver CT is acquired at three different time intervals after contrast injection [2–4]. In the initial 20–30 seconds of contrast injection; the unhealthy portions of the organ are enhanced largely and appear to be brighter than the normal liver. This is the first phase and is referred to as hepatic arterial (HA) phase. The second phase of CT acquisition is called portal venous (PV) phase and is acquired after 40–60 seconds of contrast injection. During this time interval, the normal liver appears brighter than the abnormal regions of the organ. The third phase called delayed venous phase occurs approximately 2 minutes after injecting the contrast agent wherein there will be a gradual wash out of contrast from the entire liver except for fibrotic tissue and will appear relatively dense compared to normal tissue. The various phases of CT acquired after contrast injection are clearly shown in Figure 3.1. To facilitate the readers, the various organs observed in the abdominal CT are also marked in the plain CT image.

Lesion refers to any tissue that is not healthy and includes malignant, benign, and other abnormal tissues which are neither malignant nor benign. Tumors refer to both cancerous and non-cancerous growth. Furthermore, cancerous tissues can be categorized as primary and secondary depending upon the organ they originated. Primary liver cancer refers to the cancerous tissue originated in liver and the secondary cancer is referred as a primary cancer in another organ eventually spread to liver. This is also called as metastasis. In this case study, five types of liver abnormalities are considered as mentioned in Table 3.1.

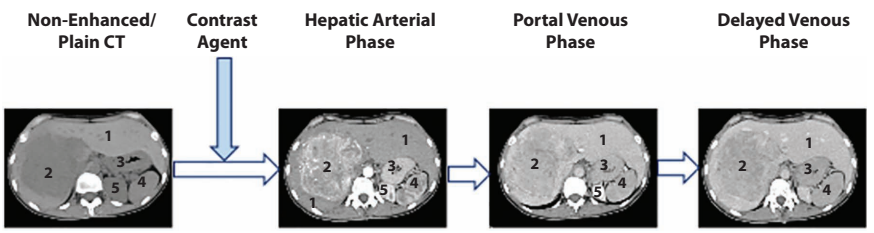

Figure 3.1 Phases of CECT images (1: normal liver; 2: tumor within liver; 3: stomach; 4: pancreas; and 5: kidney).

Table 3.1 Types of liver lesions.

Sl no	Liver disease	Type of abnormality
1	Hepatocellular carcinoma (HCC)	Primary cancer
2	Metastasis (METS)	Secondary cancer
3	Hemangioma	Tumor
4	Cyst	Tumor
5	Abscess	Neither cancer nor tumor

Though CECT images are claimed to be the preferred screening modality for liver diseases, the diagnostic accuracy is still not excellent as tissue delineation between the healthy and affected liver parenchyma in some cases is not precisely visible in CECT. Also, the differential enhancement within the areas of tumor in CECT may sometime get masked in the contrast [3], leading to misinterpretation of aggressiveness of cancer. Additionally, different lesion types exhibit similar visual appearance in CECT leading to ambiguous diagnosis. Thus, it can be said that the accuracy of diagnosis from visual inspection of CECT to be effective only when examined by a well doctor. In cases of ambiguous diagnosis, physicians rely upon painful invasive techniques like needle biopsy. These limitations enforce a need for computer aided diagnosis for liver lesion detection to facilitate doctors in their diagnosis routine by providing a second opinion to affirm their diagnosis. This work aims at providing a case study on liver lesion classification through deep learning techniques.

3.2 Related Works

Studies carried out in [5] reveals that HCC, the primary liver cancer, is characterized by more coarseness than hemangioma (HEM), a benign tumor, and exhibits uniform texture within the tumor portion. Likewise, every other lesion possesses unique texture. But, when these information are transferred to CT or MR images, the diagnosis becomes not so optimal as the differentiation between the tumor types may be not easily understood when viewed with human eye. This shortfall led to the boom of computer aided decision support system to facilitate medical experts in their precise diagnosis and sometimes to provide second opinion about the diagnostic decision. The CAD systems are usually developed using machine learning

algorithms whose performance is greatly dependent on the input features extracted from the medical images. The fact that different lesion types exhibit different degree of coarseness and roughness within the pathology of the organ makes texture features to be more proficient in the classification of lesion types of liver and other organs. A diversified set of texture features is made use of in literature for the classification of various categories of diseases from medical images.

Haralick *et al*., in [6] derived a class of texture features from gray-level co-occurrence matrix called GLCM features based on spatial inter dependency of graytones in the image. In other words, it reflects the joint probability distribution function of a pair of pixel graylevels. Many features like energy, sum average, sum entropy, inverse difference moment (IDM), homogeneity, or angular second moment (ASM), cluster shade, cluster prominence, correlation, and differential entropy can be derived from the second order co-occurrence matrix. In due course, these features revolutionized the field of pattern recognition for various image classification tasks with no exception to medical images. Apart from liver pathology, GLCM features also are being popularly used in literature for more than two decades for the task of classifying various types of medical ailments like brain tumor, breast cancer, cervical cancer, etc. [7–14]. Similar to GLCM texture features, several other texture features like fractal features and energy measures are used in literature for the classification of liver tissues. More works on liver image classification by the virtue of GLCM texture features are available in literature [15–24].

In recent times, deep learning has become a thriving force in the classification of medical images, as it solves increasingly complicated challenges with high accuracy over time. Deep convolutional neural networks (CNNs) avoids the extraction of handcrafted features and learns features at different levels very deeply by itself and therefore have manifested themselves as an effective tool for the classification of liver related disorders [25–28]. The authors of the work [25] have employed CNN-based deep learning architecture to effectively classify cysts, metastases, and hemangioma. In [26], the existing 2D CNN has been extended to three-dimensional CNN with $3 \times 3 \times 3$ kernel to differentiate secondary liver cancer from the primary one considering MRI clinical datasets. Due to the limited availability of dataset for liver lesions, the authors have synthesized high quality liver lesions using generative adversarial networks (GANs) and showed remarkable improvement in the performance. Enhanced detection capabilities for the detection of liver metastasis is achieved in [27] using fully convolution network (FCN)–based patch level analysis with superpixel sparse based classification. Watershed Gaussian–based deep learning technique is

proposed in [28] by combining the traditional feature extraction technique and deep neural network to classify three different classes of liver cancer, namely, HCC, hemangioma, and metastasis and achieved 99% accuracy with 200 epochs. Though, deep learning has recorded good performance in classification of variety of images, the training dataset requirement is huge to obtain a high degree of accuracy. This can be considered as a limitation in the availability of medical images.

Transfer learning is another paradigm shift in technology which overcomes the above mentioned challenge by using a pretrained deep neural network to train and classify any image datasets. A diverse set of such pretrained networks, namely, CifarNet [29], Alexnet [30], GoogLeNet [31, 32], VGGNet [33, 34], AggNet [35], ResNets [36, 37], and so on are developed to compete successfully in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). A deeper version of GoogLeNet-CNN has been fine-tuned and trained on thoracoabdominal lymph node datasets and interstitial lung disease dataset [38–40]. The most progressive transfer learning architecture GoogLeNet has been fine tuned to classify lung cancer images [38]. Liver fibrosis classification using transfer learning approach using VGGNet and fully connected (FC) neural network (FCNet) for US images was done [39] with an accuracy of 96.06%. The authors have claimed that performance improvement of this work was due to the presence of three FC layers used for the classification. Deep learning approach using the existing architecture of GoogLeNet was implemented by the authors in [40] for the classification of HCC and normal liver samples extracted from diagnostic information bearing phases, *viz*., HA and PV phase individually with an accuracy of 92.08%. In this chapter, the competence of popular transfer learning architectures, namely, AlexNet, GoogLeNet, ResNet-18, and ResNet-50 are analyzed for the task of liver tumor diagnosis and the competent transfer learning architecture for this case study will be identified.

3.3 Convolutional Neural Networks

CNN is a popular deep learning architecture formulated for image classification task without the need for feature engineering. A typical CNN architecture contains two modules to perform feature extraction and classification and is presented in Figure 3.2. Several convolution and max pooling layers stacked one over the other in the feature extraction module extracts the features pertinent to the input images under consideration for training and classification. Following the series of convolutional and max pooling layers, the architecture consists of FC layers analogous to feed

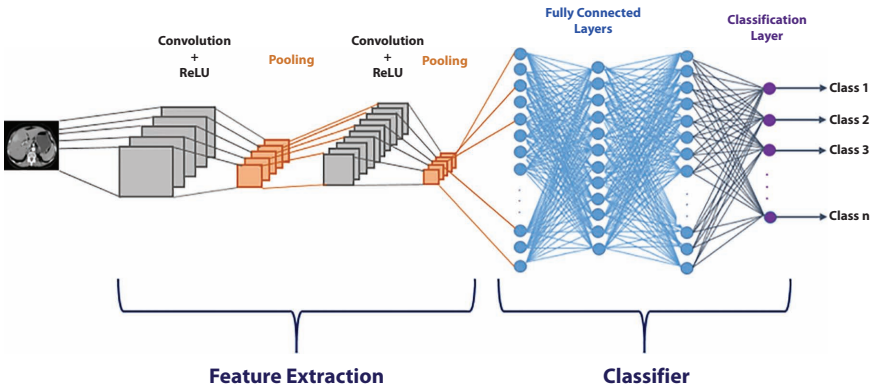

Figure 3.2 Architecture of convolutional neural network.

forward neural network (FFNN) to map the feature vectors to one of the available class.

3.3.1 Feature Learning in CNNs

Convolution layers basically performs convolution of the input image or matrix with a set of filters. The filters can be of size 3×3, 5×5, 7×7, and so on. Usually, convolution of an image with a known filter results in image averaging, sharpening, edge detection, and so on. Against this, in CNNs, the filter coefficients are the learnable parameters and will be optimized during the training process. The convolution operation is realized using neural networks with sparser connections in such a way that the weights associated with the neural network correspond to the filter coefficients to be learned. The implementation of convolution operation using neural networks requires an activation function and a rectified linear unit (ReLU) activation function is predominantly used in convolution layers. Following the convolution operation, max pooling is performed which significantly shrinks the input dimension (most commonly by a factor of 2). Pooling operation during dimensionality reduction replaces a portion of the matrix with the summary statistic of the local portion being dealt with. Since in feature engineering, the concern is on finding out the prominent structure of the image, max operation is performed and is accordingly referred to as max pooling. Since the max pooling layer finds the local maxima, no learning is involved in this operation. In the first stage of convolution and max pooling, one form of representation from the input image is learnt and is then forwarded to the next convolution layer on the network wherein a still

more abstract representation is learnt. This representation is then carried forward to next layers to make the feature learning a deeper one and is continued until meaningful representations are learnt. In every layer, multiple representations are being learnt using different filters and similar learning is being done in multiple such layers.

3.3.2 Classification in CNNs

At the output of final max pooling layer, feature map corresponding to an image will be present. This will now serve as an input for the classification module of the CNN. This module consists of FC layers which essentially learn the patterns of the feature map. The FC layers are simply FFNNs wherein every neuron in one layer is connected to every other neuron in the next layer. As a first step in classification process, the feature maps which are two-dimensional in nature are flatted to form a feature vector. The number of neurons at the input of first FC layer is equal to the dimension of the feature map. A number of FC layers can be added to make the classification network, a deeper one. The FC layers have dense connections unlike the convolutional layers which have sparse connections, and it is here in the FC layers, the mapping of the feature vector to one of the "n" classes will take place. The last layer of the network is called classification layer and the number of neurons in this layer is equal to the number of classes in the dataset considered for training and classification. ReLU activation function is used for all the FC layers except the final one. The activation function for the final FC layer is sigmoid for a binary classifier and softmax activation is used in the case of multiclass classifier. Both sigmoid and softmax activation functions return the confidence score for every class which is the probability that the input belongs to one particular class. The confidence scores being the probability values for all the class sum up to one and the class with maximum confidence score will be mapped to the corresponding feature vector. The weights of the filters used in convolutional layers and FC layers will be updated by means of back propagation in such a way that the classification loss is minimized. The loss function used for weight updation is cross entropy. Since the FC layers are having denser connections, the number of parameters to be learnt is very high compared to the convolutional layers. Making the classification network deeper will result in more number of learnable parameters and may lead to overfitting. Hence, an optimal trade-off between the classification accuracy and depth of the network is preferred to achieve better classification performance. In case of overfitting, dropout units which will make a portion of the neurons in the layer to be dead can be included in between convolution layers or between FC layers.

CNNs require huge set of image data, which is a daunting challenge as far as the medical images are concerned. When sufficient images are not available for training the network, pre-trained CNNs called transfer learning architectures can be used for image classification. Another confronting issue in CNN is that, it suffers from vanishing gradient problem when the network is made deeper to achieve high accuracy. Appropriate measures have to be taken while designing CNN to make it perform effectively.

3.4 Transfer Learning

Transfer learning strategy makes use of the knowledge gained by a deep learning network trained on one dataset to another related dataset. Such a deep learning network architecture is referred to as pre-trained network. Indeed, it is intelligent and efficient to use a pre-trained CNN to perform better classification on a dataset instead of training a newly built network with random weights from the scratch. While deploying a CNN trained on one data for another data, the pre-trained CNN has to be fine-tuned to the new dataset. The first task in reusing the existing pre-trained network during transfer learning is to replace the final classification layer with the number of output neurons to be equal to the number of classes in the new dataset. The fine-tuning of pre-trained CNN can be done in two ways: i) freeze the weights of all layers except the few final layers and retrain the newly added layers, retaining the weights of initial layers; ii) retrain the entire network with the existing weights of the CNN as initial weights. If the new dataset is similar to the already trained one, the first method of fine-tuning can be adopted. On the other hand, if the new dataset is different, fine-tuning of learned parameters can be done using the second method.

There exists a wide variety of pre-trained CNN architectures trained using ImageNet database [42] containing one million images with thousand classes. AlexNet [31], GoogLeNet [32], ResNet18 [37], and ResNet50 [37], the popular transfer learning architectures trained on ImageNet databases are considered in this case study of liver tumor diagnosis.

3.4.1 AlexNet

AlexNet, popular deep learning architecture, is the winner of ILSVRC, an image classification problem in the year 2012. The architecture has a depth of eight learnable layers and is presented in Figure 3.3. The network has 3 five convolutional layers and three FC layers. The input images

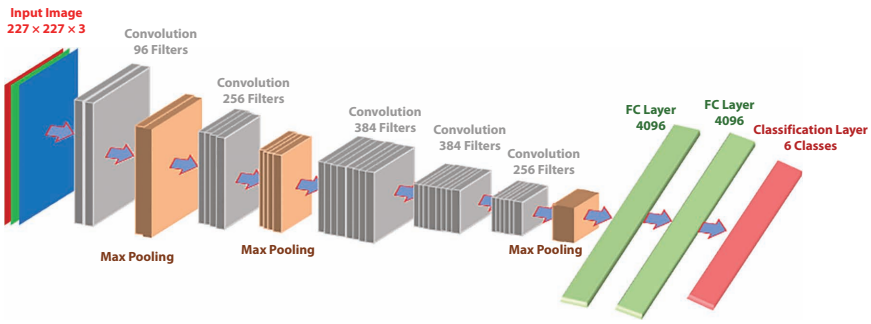

Figure 3.3 AlexNet architecture.

should be of dimension 227 × 227 × 3. The first convolutional layer consists of 96 filters with dimension 11 × 11 × 3 with a stride of 4 followed by a maxpooling layer. The output of this maxpooling layer is fed as input to the second convolution layer which has 256 filters of size 5 × 5 followed by max pooling layer. The remaining three convolutional layers are connected back to back without maxpool layer and have 384, 384, and 256 filters, respectively, with a dimension of 3 × 3 × 3. A maxpooling layer is present after the fifth convolution layer. The five convolutional layers together with the three maxpool layers serve the purpose of feature engineering. Following this, two FC layers with 4,096 neurons and a last classification layer with 1,000 class softmax activation function are present. In the proposed work of a case study on liver tumor diagnosis, there are six classes, and hence, the classification layer is replaced with six class softmax function in Figure 3.3.

3.4.2 GoogLeNet

GoogLeNet, a powerful CNN architecture, has demonstrated its exceptional performance in image classification task and has won the ILSVRC challenge in 2014. The network is 22 layers deep and the architecture is shown in Figure 3.4. GoogLeNet does not restrict itself to one specific dimension of convolution filters like AlexNet and other series structured CNN architectures. Instead, it applies multiple kernels of different dimensions, namely, 1 × 1, 3 × 3, 5 × 5, and so on in parallel and concatenate all the feature maps. The highlight of this network is that it uses 1 × 1 convolution for dimensionality reduction across the depth of the feature maps before convolving with 3 × 3 and 5 × 5 filters and this greatly influences in reducing the computational complexity. The part of the network which

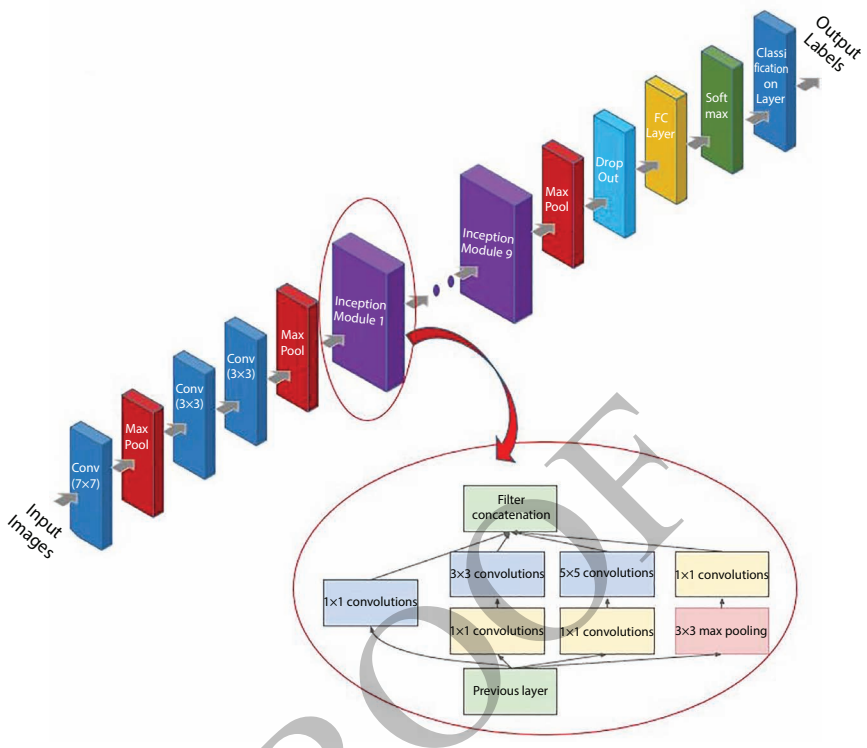

Figure 3.4 GoogLeNet architecture.

performs simultaneous convolutions of varied dimensions is called inception module and nine such inception modules are present in GoogLeNet architecture. These nine inception modules contribute more towards better performance in effectively capturing the discriminant features from the input images.

3.4.3 Residual Networks

The CNNs have started to evolve with a trend of going deeper and deeper to solve complex problems with improved classification accuracy levels as the case with above mentioned CNN architectures like AlexNet which has eight layers and subsequently GoogleNet has a depth of 22 layers. There stems two major problems when network becomes deeper. These are training them becomes very difficult and accuracy unexpectedly starts saturating and degrades too, when the deeper networks converge. These problems are solved with residual learning approach. The CNNs adopting

this approach are referred to as ResNets [40]. ResNet has won first place in the ILSVRC 2015 classification challenge with minimal error rate of 3.57%.

The conventional CNN approach is to learn several low/mid/high level features during the training phase. Unlike this in ResNets, instead of trying to learn some features, they try to learn some residuals. Residual can be simply understood as subtraction of feature learned from input of that layer. In general, in a deep CNN, several layers are stacked and are trained to the task at hand which allows the network to learn several features. In residual learning, instead of trying to learn some of the features, it tries to learn the residuals, which is the unique feature of ResNets. These residuals can be simply understood as subtraction of feature learned from input of that layer. ResNet accomplishes this residual learning using shortcut connections (directly connecting input of n^{th} layer to some $(n + x)^{th}$ layer) which is depicted in Figure 3.5. The skip connection in the Figure 3.5 is labeled "identity." It allows the network to learn the identity function, which allows it to pass the input through the block without passing through the other weight layers. Therefore, the shortcut connections simply perform identity mapping, and their outputs are added to the outputs of the stacked layers straightaway.

It has been proved that training this form of networks is easier than training simple deep CNNs and also the problem of degrading accuracy is resolved.

The basic intuitive idea behind the residue learning is that if the identity mapping is optimal, it becomes easier to push the residuals to zero ($F(x) = 0$) than to fit an identity mapping (x, input = output) by a stack of non-linear CNN layers. This function $F(x)$ is the residual function and $\boldsymbol{X}$ is the identity mapping. There are two options for exercising this shortcut

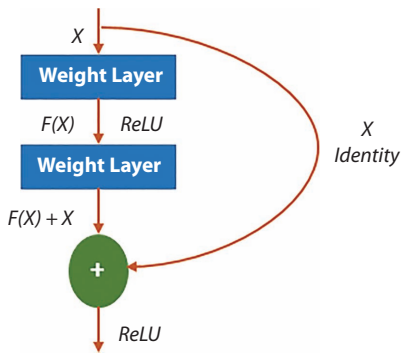

Figure 3.5 Residual learning—building block.

namely: a) The identity shortcuts ($\boldsymbol{X}$) can be directly used when the input and output are of the same dimensions. b) When the input and output are of mismatched dimensions then the shortcut still performs identity mapping, with extra zero padding. It is important to note that the Identity shortcut connections add neither extra parameter nor computational complexity. The entire network can still be trained end-to-end with back propagation, and can as well be easily implemented. There are several variants of ResNets available, namely, ResNet-18, ResNet-50, ResNet101, and ResNet152 till date.

3.4.3.1 ResNet-18

ResNet-18 is a CNN that is 18 layers deep and used mainly for image classification applications. The network accepts an image input size of 100 × 100, 224 × 224. The baseline architecture is same as the plain nets, expect that a shortcut connection is added to each pair of 3 × 3 filters as shown in Figure 3.6.

3.4.3.2 ResNet-50

ResNet-50 is a deep residual network. The "50" refers to the number of layers it has. It is a subclass of CNNs, with ResNet most popularly used for image classification. ResNets could also be treated as an ensemble of smaller networks and it as well allows several intuitive ways of training like randomly dropping its layers during training and using the full network

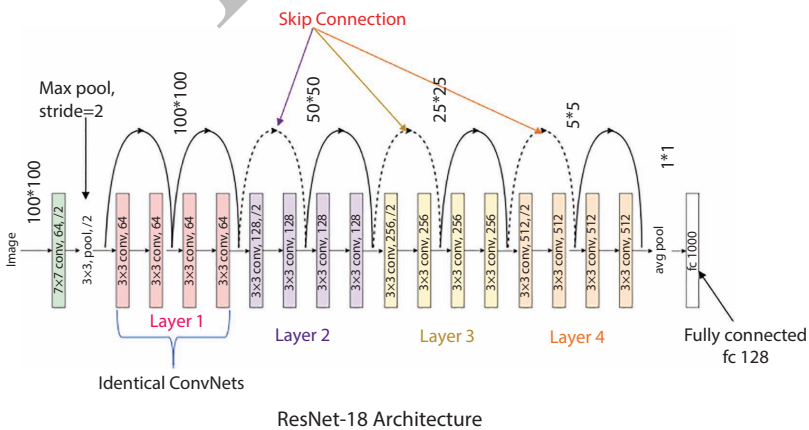

Figure 3.6 Architecture of ResNet-18.

in the testing time. This is possible in ResNets as there exist independent effective paths among the residual blocks and the majority of them remain intact even after removing a couple of layers. Thus, lot of flexibility is available in training these residual nets without compromising the performance. Because of its compelling results, ResNet quickly became one of the most popular architectures in various computer vision tasks.

3.5 System Model

The framework of the proposed system model for liver tumor diagnosis is presented in Figure 3.7. The HA phase CT images, which contains most of the diagnostic information are considered as for this case study.

The liver CT images as presented in Figure 3.1 contain many adjacent organs in it and hence for the classification task of liver tumor, liver, and the lesion regions have to be segmented prior to classification. The segmentation of liver from the CT images is highly challenging as liver and the adjacent organs exhibit similar intensity patterns in the CT image. Also, the boundary of the liver will not be precisely defined and the size and shape of the organ varies from slice to slice. The liver and liver lesions from the CT images are segmented using bidirectional region growing algorithm implemented on edge enhanced CT images proposed by the authors in [42]. The output of segmentation algorithm carried out is depicted in Figure 3.8. The edge components of the input image are enhanced by means of unsharp masking in NSCT domain. Following the enhancement, liver is segmented using bidirectional region growing algorithm. Consequently, from the segmented liver, lesion region is segmented using the same bidirectional region growing algorithm.

After the segmentation of lesions, classification of six different classes of lesion is carried out using four different types of transfer learning

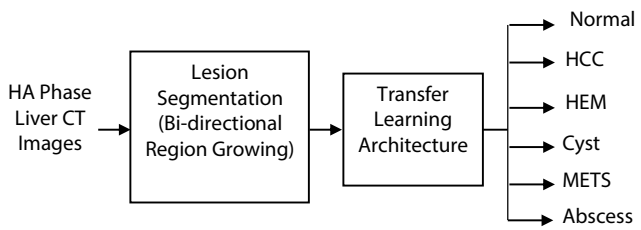

Figure 3.7 System model for case study on liver tumor diagnosis.

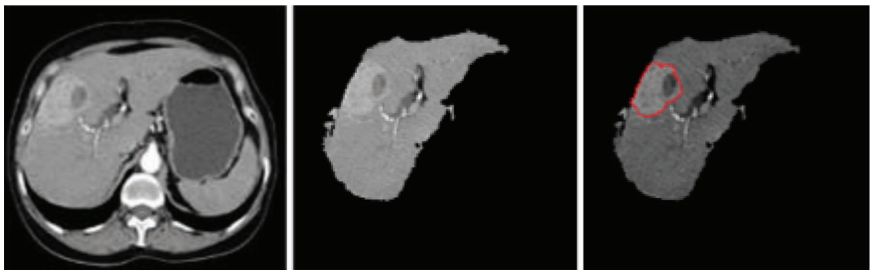

Figure 3.8 Output of bidirectional region growing segmentation algorithm: (a) input image; (b) segmented liver; (c) segmented liver with tumor boundary marked.

architectures, *viz*., Alexnet, GoogLeNet, ResNet-18, and ResNet-50 and their performances are compared.

3.6 Results and Discussions

3.6.1 Dataset

The images used in this case study are obtained from Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Puducherry, India with ethical clearance. Siemens Sensation 64 detector scanner was used to acquire the CT images via a standard four phase contrast enhanced imaging protocol. All the CT images were in DICOM (Digital Imaging and Communications in Medicine) format with a resolution of size 512 × 512 pixels and are then converted to JPG format using RadiAnt DICOM viewer software for this case study. A total of 661 image samples were used in study. The count of number of images in each class is presented in Table 3.2. The HA phase CT images for all six classes considered for this work are presented in Figure 3.8. The tumor part is highlighted using a red coloured box.

3.6.2 Assessment of Transfer Learning Architectures

In this section, the performance of four different transfer learning architectures on the classification of liver tumor is presented. From the visual inspection of the images presented in Figure 3.9, it is seen that the texture pattern of hemangioma, cyst, METS, and abscess seems to be similar. This obviously makes the classification of liver tumor from the CECT images a challenging task. Four popular transfer learning architectures which have manifested

Table 3.2 Dataset count.

Liver lesion type	Number of CT images
Normal	107
HCC	145
Hemangioma	118
Cyst	126
Metastasis	66
Abscess	99
Total	**661**

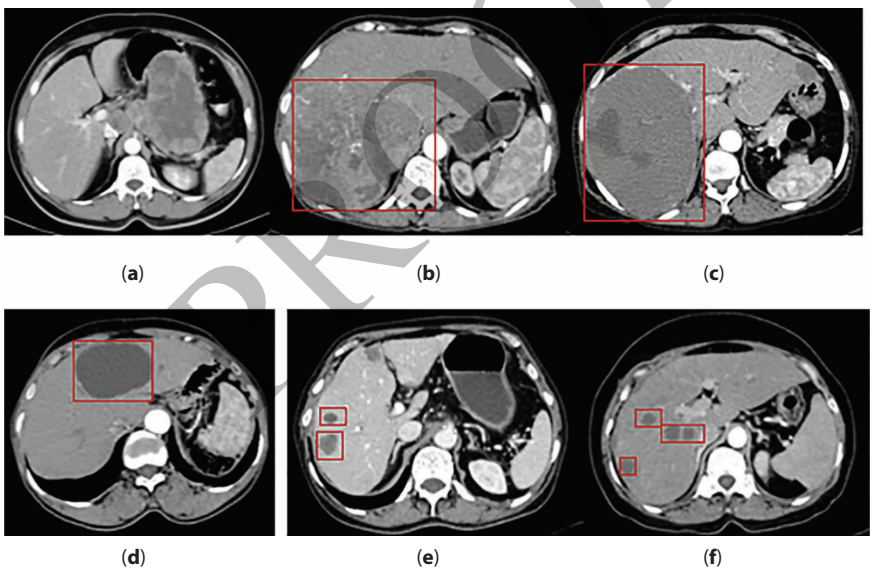

Figure 3.9 HA Phase Liver CT images: (a) normal liver; (b) HCC; (c) hemangioma; (d) cyst; (e) metastasis; (f) abscess.

their proficiency in various image classification tasks are deployed in this case study of liver tumor diagnosis and the best performing architecture will be identified. The training and classification of the deep neural networks were executed using deep learning toolbox of MATLAB 2020a software. A total of 661 image samples as given in Table 3.2 were considered in this

case study. Among the entire dataset, 70% of the images were considered for training and 30% for testing. To make the performance comparison of all four transfer learning architectures to be fair, hyperparameter settings of the network are maintained the same and are tabulated in Table 3.3.

The classifier performance using the four transfer learning architectures are assessed in terms of classification accuracy. The confusion matrix gives the information about prediction results in a more precise manner. The rows of the confusion matrix correspond to actual class and every column represents the predicted class. The diagonal entries of the matrix correspond to correct prediction and are referred to as true positive (TP) and true negative (TN). The entries in the non-diagonal cell represent the wrong prediction and they represent the false positive (FP) and false negative (FN). Zero values in the non-diagonal cells of the confusion matrix illustrate better classifier performance. The confusion matrices corresponding to the classification task of all four networks are presented in Tables 3.4 to 3.7.

Table 3.3 Hyperparameter settings for training.

Hyperparameter	**Value**
No. of epochs	40
Optimization	Adaptive moments
Learning rate	0.0001
Mini batch size	27

Table 3.4 Confusion matrix for AlexNet.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	38	5	0	0	0
	HEM	0	3	26	0	3	4
	Cyst	0	0	3	27	3	5
	METS	0	0	8	0	7	5
	Abscess	0	0	0	8	0	22

Table 3.5 Confusion matrix for GoogLeNet.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	4	27	0	3	1
	Cyst	0	0	1	27	4	6
	METS	0	0	3	0	15	2
	Abscess	0	0	1	5	3	21

Table 3.6 Confusion matrix for ResNet-18.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	4	28	1	2	0
	Cyst	0	0	0	28	5	5
	METS	0	0	3	1	16	2
	Abscess	0	0	0	6	3	21

Table 3.7 Confusion matrix for ResNet-50.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	3	28	0	2	2
	Cyst	0	0	0	28	4	6
	METS	0	1	2	2	16	0
	Abscess	0	0	1	3	2	24

Classification accuracy, which is defined as the percentage of correct predictions, is considered to be the most important parameter in quantifying the performance of a classifier. It is computed from the confusion matrix as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Table 3.8 Comparison of classification accuracies.

Transfer learning architecture	Classification accuracy
ResNet-50	***86.43%***
ResNet-18	84%
GoogLeNet	83.33%
AlexNet	76.38%

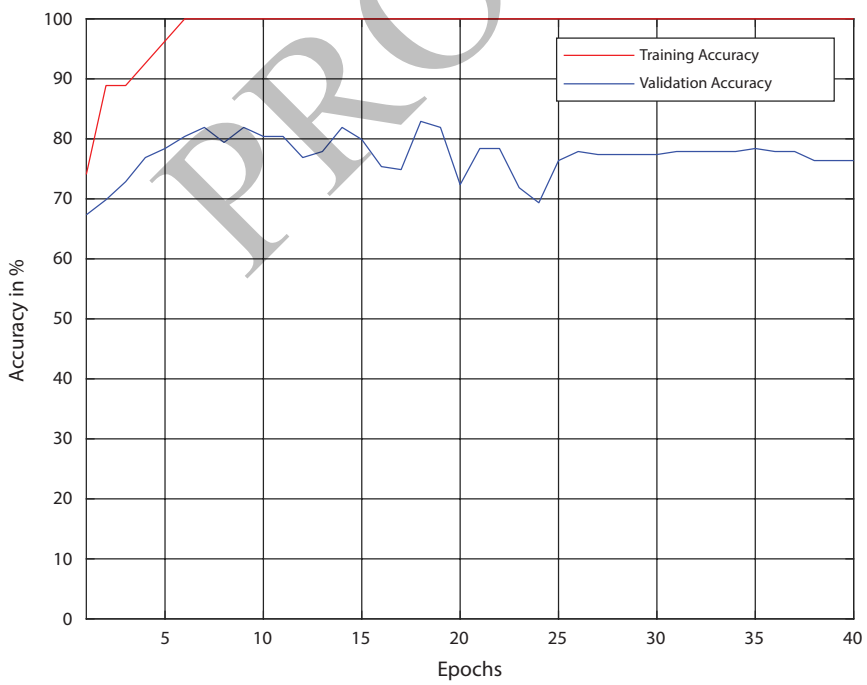

Figure 3.10 Training progress for AlexNet.

The classification accuracy obtained in this case study of liver tumor classification is computed for all four architectures under consideration and are tabulated in Table 3.8.

The variation of training and validation accuracies with increase in epochs for all the four architectures is shown in Figures 3.10 to 3.13.

From the comparison of classification accuracy of different networks from Table 3.8 and Figures 3.10 to 3.13, it is evident that ResNet-50 architecture records a better performance for the case study of liver tumor diagnosis. ResNet-18 and GoogLeNet architectures also perform well comparatively and produced a classification accuracy closer to ResNet-50. The fact that different lesion samples exhibit highly similar texture pattern justifies why classification accuracy obtained is less than 90%. Deeper the network, better is the classification accuracy and this justifies the better performance of GoogLeNet, ResNet-18, and ResNet-50 over AlexNet. Increasing the depth of the network will cause vanishing gradient problem wherein the back propagation of the gradient of loss function becomes negligible in initial layers of the network and they seldom update their weights. This ultimately results in degradation in the classification performance.

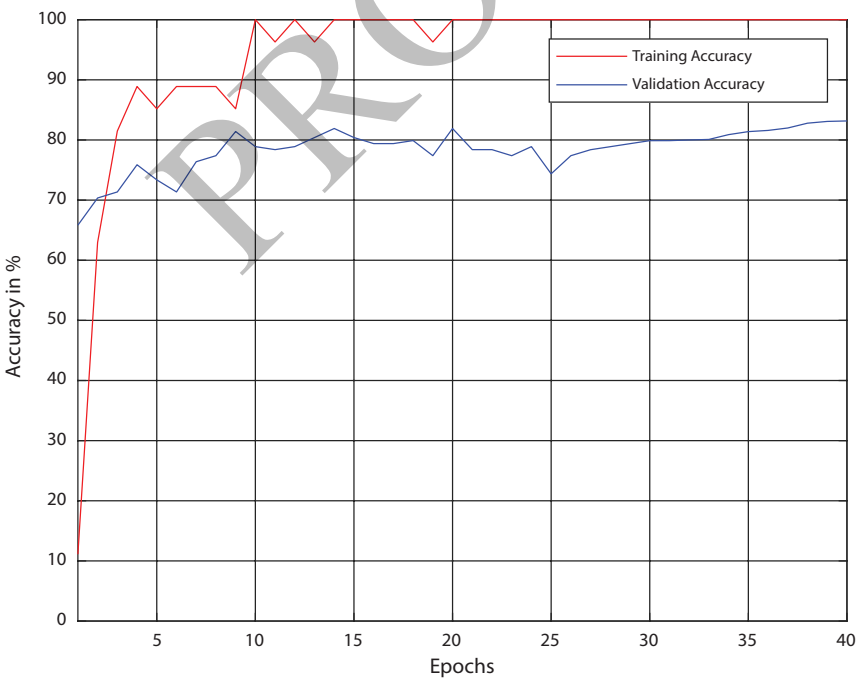

Figure 3.11 Training progress for GoogLeNet.

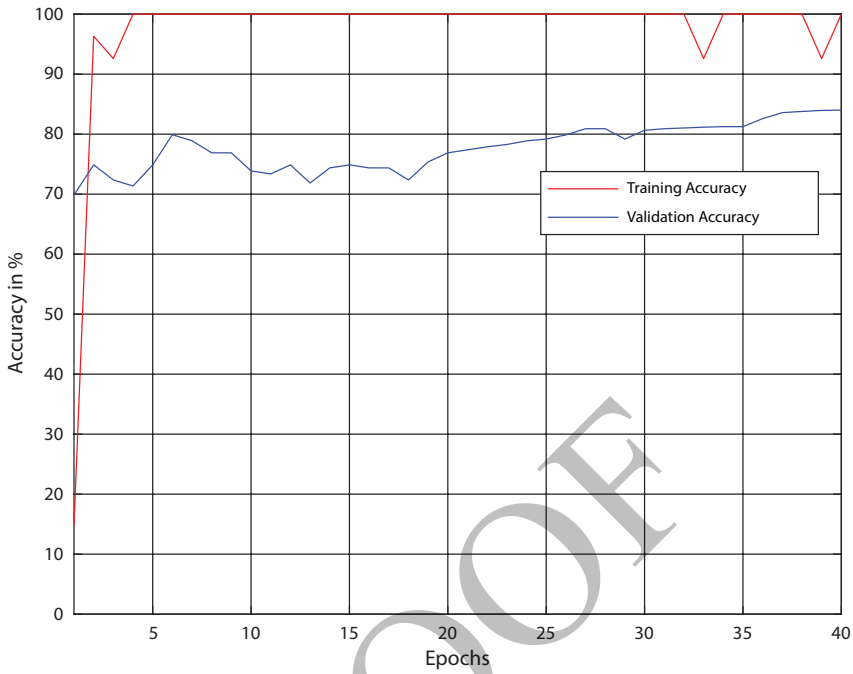

Figure 3.12 Training progress for ResNet-18.

Since this problem of vanishing gradient is addressed in ResNet family of networks through residual connections, ResNet-18 and ResNet-50 register better performance than GoogLeNet. Comparing the two variants of ResNets dealt with in this chapter, owing to deep layers with residual connections, ResNet-50 reports higher classification accuracy.

3.7 Conclusion

In this chapter, a performance study of popular transfer learning architectures for the case of liver tumor classification is proposed. The classification of five different types of liver lesions along with normal liver is performed using the HA phase liver CT images. Since, the lesion types exhibit similar textural pattern in the CT images this study attempts to investigate the proficiency of the transfer learning architectures in performing a challenging classification task. The result presented in Table 3.8 ascertains that ResNet-50 architecture recorded a remarkable performance of 86.43%. To further improve the classification performance, various other approaches

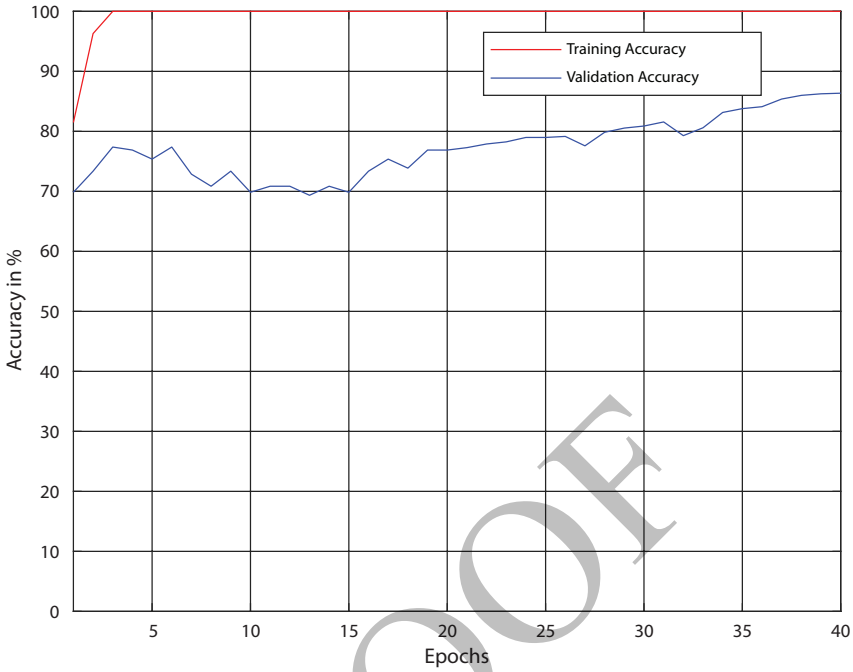

Figure 3.13 Training progress for ResNet-50.

can be followed. These are i) the raw CT images can be enhanced to discriminate the textural patterns of the lesions and ii) the pre-trained CNN model can be improvised to compensate for vanishing gradient problem. These methodologies are adopted by the authors in [43] and an exceptional improvement in classification accuracy is achieved. Additionally, the performance of the classifier can be greatly improved by considering more number of images for training the CNN. Since the dataset considered has limited number of images, a large number of liver CT images can be artificially generated either using GANs or Variational Auto Encoders (VAEs) and consequently the classification accuracy can be increased.

References

1. https://cancerstatisticscenter.cancer.org
2. Pandharipande, P.V., Krinsky, G.A., Rusinek, H. *et al.*, 'Perfusion imaging of the liver: current challenges and future goals. *Radiology*, 234, 3, 661–673, 2005.

3. Baron, R.L., Understanding and optimizing use of contrast material for CT of the liver. *Am. J. Roentgenol.*, 163, 2, 323–331, 1994.
4. Roy, S., Chi, Y., Liu, J. *et al.*, 'Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions'. *IEEE Trans. Biomed. Eng.*, 61, 11, 2768–2778, 2014.
5. Wegene, O.H., *Whole body computerized tomography*, vol. 10, no. 20, English translated by J.H. Long, Schering AG, Berlin, West Germany, 1983.
6. Haralick, R.M., Shanmugam, K., Dinstein, I., "Textural features for image classification". *IEEE Trans. Syst. Man Cybern.*, 3, 6, 610–621, Nov. 1973.
7. Sabino, D.M.U., Costa, L.F., Rizzatti, E.G., Zago, M.A., "A texture approach to leukocyte recognition". *Real-Time Imaging*, 10, 4, 205–16, Aug. 2004.
8. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., "A methodological approach to the classification of dermoscopy images". *Comput. Med. Imaging Graph*, 31, 6, 362–373, Sep. 2007.
9. Desir, C., Petitjean, C., Heutte, L., Thiberville, L., Salaün, M., "An SVM-based distal lung image classification using texture descriptors". *Comput. Med. Imaging Graph.*, 36, 4, 264–270, Jun.2012.
10. Gomez, W., Pereira, W.C.A., Infantosi, A.F.C., Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Trans. Med. Imaging*, 31, 10, 1889–1899, Oct. 2012.
11. Nagarajan, M.B., Coan, P., Huber, M.B., Diemoz, P.C., Glaser, C., Wismüller, A., "Computer-Aided Diagnosis in Phase Contrast Imaging X-Ray Computed Tomography for Quantitative Characterization of ex vivo Human Patellar Cartilage". *IEEE Trans. Biomed. Eng.*, 60, 10, 2896–2903, Oct. 2013.
12. Torheim, T. *et al.*, Classification of Dynamic Contrast Enhanced MR Images of Cervical Cancers Using Texture Analysis and Support Vector Machines. *IEEE Trans. Med. Imaging*, 33, 8, 1648–1656, Aug. 2014.
13. Angel Arul Jothi, J. and Mary Anita Rajam, V., "Effective segmentation and classification of thyroid histopathology images". *Appl. Soft Comput. J.*, 46, 652–664, Mar. 2016.
14. Arya, M., Mittal, N., Singh, G., Texture-based feature extraction of smear images for the detection of cervical cancer. *IET Comput. Vision*, 12, 8, 1049–1059, Dec. 2018.
15. Kadah, Y.M., Farag, A.A., Zurada, J.M., Badawi, A.M., Youssef, A.M., Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Trans. Med. Imaging*, 15, 4, 466–478, Aug. 1996.
16. Chen, E.L., Chung, P.C., Chen, C.L., "An automatic diagnostic system for ct liver image classification". *IEEE Trans. Biomed. Eng.*, 45, 6, 783–794, Jun. 1998.
17. Gletsos, M., Mougiakakou, S.G., Matsopoulos, G.K. *et al.*, A computer aided diagnostic system to characterize ct focal liver lesions: design and optimization of a neural network classifier". *IEEE Trans. Inf. Technol. Biomed.*, 7, 3, 153–162, Sep. 2003.

18. Bilello, M., Gokturk, S.B., Desser, T., "Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous phase CT". *Med. Phys.*, 31, 9, 2584–2593, Sep. 2004.
19. Poonguzhali, S. and Ravindran, G., "Evaluation of feature extraction methods for classification of liver abnormalities in ultrasound images". *Int. J. Bio Med. Eng. Technol.*, 1, 2, 134–143, 2007.
20. Laws, K., II, Rapid texture identification, in: *Proc. of SPIE Image Processing for Missile Guidance*, Dec. 1980, vol. 238, pp. 376–380.
21. Diamant, I., Goldberger, J., Klang, E., Amitai, M., Greenspan, H., Multi-phase liver lesions classification using relevant visual words based on mutual information, in: *Proc. IEEE Int. Sym. on Biomedical Imaging (ISBI)*, New York, NY, Jul. 2015, pp. 407–410.
22. Ragesh, K.K. and Radhakrishnan, S., Focal and diffused liver disease classification from ultrasound images based on iso-contour segmentation. *IET Image Process.*, 9, 4, 261–270, Apr. 2015.
23. Chang, C.C., Chen, H.H., Chang, Y.C., Lo, C.M., Ko, W.C., Lee, Y.F., Liu, K.L., Chang, R.F., Computer-aided diagnosis of liver tumours on computed tomography images. *Comput. Methods Programs Biomed.*, 145, 45–51, Jul. 2017.
24. Huang, Y.L., Chen, J.H., Shen, W.C., Diagnosis of hepatic tumors with texture analysis in non enhanced computed tomography images. *Acad. Radiol.*, 13, 6, 713–720, Jun. 2006.
25. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification". *Neurocomputing*, 321, 321–331, Dec. 2018.
26. Trivizakis, E., Manikis, G.C., Nikiforaki, K., Extending 2-d convolutional neural networks to 3-D for advancing deep learning cancer classification with application to mri liver tumor differentiation. *IEEE J. Biomed. Health Inform.*, 23, 3, 923–930, May 2019.
27. Ben-Cohen, A., Klang, E., Kerpal, A., Konen, E., Amitai, M.M., "Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations". *Neurocomputing*, 275, 1585–1594, Jan. 2018.
28. Das, A., Acharya, U.R., Panda, S.S., Sabut, S., "Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques". *Cognit. Syst. Res.*, 54, 165–175, May 2019.
29. Krizhevsky, A., *Learning multiple layers of features from tiny images*, Master's Thesis, Dept. of Comp. Science, University of Toronto, 2009.
30. Krizhevsky, A., Sutskever, I., Hinton, G.E., ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25, 2, 1097-1105, Jan. 2012.
31. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Going deeper with convolutions, in: *Proc. CVPR '2015*, 9 pages, 2015.

32. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., Rethinking the inception architecture for computer vision, in: *Proc. CVPR '16*, Las Vegas, NV, pp. 2818–2826, 2016.
33. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., Overfeat: integrated recognition, localization and detection using convolutional networks, in: *Proc. CVPR '14*, Feb. 2014.
34. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, in: *Proc. of ICLR*, Apr 2015.
35. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., Navab, N., AggNet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Trans. Med. Imaging*, 35, 5, 1313–1321, May 2016.
36. He, K., Zhang, X., Ren, S., Sun, J., Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
37. He, K., Zhang, X., Ren, S., Sun, J., Identity mappings in deep residual networks, in: *European conference on computer vision*, 2016, October, Springer, Cham, pp. 630–645.
38. Fang, T., a novel computer-aided lung cancer detection method based on transfer learning from GoogLeNet and median intensity projections, in: *Proc. IEEE Int. Conf. on Computer and Communication Engineering Technology*, Beijing, pp. 286–290, Aug. 2018.
39. Meng, D., Zhang, L., Cao, G., Cao, W., Zhang, G., Hu, B., Liver fibrosis classification based on transfer learning and fcnet for ultrasound images. *IEEE Access*, 5, 5804–5810, Mar. 2017.
40. Brunetti, A., Carnimeo, L., Trotta, G.F., Bevilacqua, V., "Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: a survey based on medical images". *Neurocomputing*, 335, 274–298, Mar. 2019.
41. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L., Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, IEEE, pp. 248–255, June 2009.
42. Lakshmi Priya, B., Jayanthi, K., Pottakkat, B., Ramkumar, G., "Liver segmentation using bidirectional region growing with edge enhancement in nsct domain". *IEEE International Conference on Systems, Computation, Automation and Networking (ICSCAN 2018)*, IEEE, pp. 248–255, June 2009.
43. Balagourouchetty, L., Pragatheeswaran, J.K., Pottakkat, B., R.G., GoogLeNet-Based Ensemble FCNet Classifier for Focal Liver Lesion Diagnosis. *IEEE J. Biomed. Health Inform.*, 24, 6, 1686–1694, June 2020.

Author Query

Q1 Please condense abstract to a maximum of 200 words, as required by Scrivener guidelines.