A Deep Learning Architecture for Classifying Medical Images of Anatomy Object

Sameer Khan* and Suet-Peng Yong†
Computer and Information Sciences Department
Universiti Teknologi PETRONAS, Malaysia.
* E-mail: sameer15khan@gmail.com
† E-mail: yongsuetpeng@utp.edu.my Tel: +605-3687414

***Abstract*—Deep learning architectures particularly Convolutional Neural Network (CNN) have shown an intrinsic ability to automatically extract the high level representations from big data. CNN has produced impressive results in natural image classification, but there is a major hurdle to their deployment in medical domain because of the relatively lack of training data as compared to general imaging benchmarks such as ImageNet. In this paper we present a comparative evaluation of the three milestone architectures i.e. LeNet, AlexNet and GoogLeNet and propose our CNN architecture for classifying medical anatomy images. Based on the experiments, it is shown that the proposed Convolutional Neural Network architecture outperforms the three milestone architectures in classifying medical images of anatomy object.**

I. Introduction

Medical images obtained from different image modalities contain vital information about various states of the patient and are an extremely important part of the diagnosis process in medical institutions [1]. Recent advances in medical imaging techniques such Computed Tomography (CT), Magnetic Resonance Imaging(MRI), X-Rays, Positron Emission Tomography(PET) have led to enormous increase in the volume of these images [2] and an increase in the stipulation for automatic methods of classifying, indexing, annotating and analyzing these medical images. From the radiology workflow perespective, following picture acquisition course of action, the images are usually archived within Picture archiving and Communication Systems (PACS) [4]. In order to make diagnosis, a radiologist retrieves an image from PACS. Retrieving similar cases from a large archive may be a daunting task and is one among the key problems within the quickly increasing domain of content-based medical image retrieval [5]. In classifying medical image anatomies, there are two main issues: intra class variability vs inter class variability [6] and data disproportion [7]. The first problem is due to the fact that images belonging to different anatomy object classes might look very similar as shown in Figure 1.

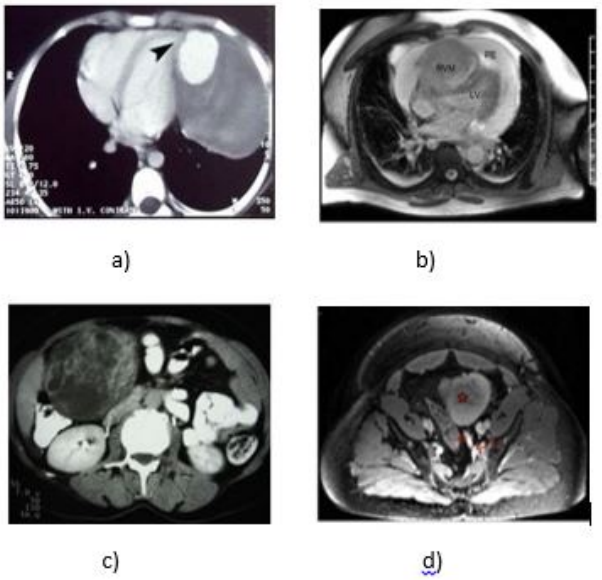

Fig. 1: Example images depicting visual variability belonging to same class i.e. heart and kidney a) CT heart, b) MRI heart, c) CT Kidney, d) MRI Kidney

The work of image classification has been conducted in a single specific domain of anatomies and modalities, such as CT lung images [8], X-ray and CT images of different body parts i.e. skull, breast, chest, hand etc [9], breast ultrasound images [10]. Although a variety of feature representation have been proposed for classifying medical images, these feature representations are domain specific, that cannot be applied to other classes keeping in mind the variability in medical images. In this study we propose a Convolutional Neural Network (CNN) architecture for automatically classifying anatomy in medical images by learning features at multiple level of abstractions from the data obtained.

The contribution of this paper is on a comprehensive evaluation of the three milestone CNN architectures, i.e. LeNet, AlexNet and GoogLeNet for classifying medical anatomy images. The findings from the performance analysis of these architectures advocates the need of a modified architecture because of their poor performance for medical image anatomy classification. Hence, a modified Convolutional Neural Network architecture for classifying anatomies in medical images

is proposed.

The rest of the paper is organized as follow: Section 2 discusses the related work, Section 3 highlights our proposed CNN while Section 4 presents the evaluation of three milestone CNN architectures and our proposed CNN architecture for classifying medical image anatomies. The paper is concluded in Section 5.

II. Related Work

Over the past decades, a number of low level feature descriptors have been proposed as an image representation ranging from global features, such as shape and texture features as reported in [11] for classification of pulmonary nodules in lung ct images, edge features [12] to the recently used local feature representations, i.e SIFT with Bag of Visual Words [13].

On the other hand deep learning have shown promising results in image classification. Deep learning alludes to a category of machine learning techniques, where numerous layers of information processing stages in hierarchical architectures are exploited for pattern classification and feature learning. LeCun [14] adopted the deep supervised back-propagation Convolution Neural Network (CNN) for digit recognition successfully. After that, the deep Convolutional Neural Networks (CNNs) proposed in [15] turned out to be a breakthrough, that was declared first in the image classification task of ILSVRC-2012. The model was trained on more than one million images, and has achieved a successful top-5 test error rate of 15.3% over 1000 classes. Since then, more work have been done by improving CNN models to improve the image classification results. Specifically, the CNN model consists of many convolutional layers and pooling layers, that are stacked up with one on top of another. The convolutional layer shares several weights, and the pooling layer sub-samples the output of the convolutional layer and reduces the data rate from the layer below. The weight sharing in the convolutional layer, in conjunction with suitable chosen pooling schemes, subsidizes the CNN with some invariance properties e.g. invariance to translation.

On the other hand, CNNs have made a sound advancements in biomedical applications [18] too. Recent work has shown how the implementation of CNNs can significantly improve the performance of the state-of-the-art computer aided detection systems (CADe) [19–21]. However, in terms of research for classifying anatomies in medical images, there are only a few studies have been carried out using CNN [22–24].

One of the drawbacks of these studies is that they do not provide extensive evaluation of milestone deep nets [22, 23] and are just focused on single modality, such as only CT images were used in [22]. In order to overcome these limitation, an architecture that can be generalized to various anatomies with different modalities is needed which leads to the main focus of this study.

III. Proposed model

The anatomical classification problem is an important step in Computer Aided and Diagnosis Systems (CADs) [23]. Anatomical structures vary dramatically between individuals i.e normal lung structure as compared to deformed shaped due to pathological intervention, also small lumbar spine bone structure in one individual and same bone structure in other individuals appear to be elongated due to the advancement in the diseases. As a result, a robust Convolutional Neural Network (CNN) architecture is required to achieve better accuracy and that should generalize to all medical image types regardless of normal or abnormal.

Our proposed model of the CNN architecture is a modification of the basic architecture of AlexNet [15]. This architecture contains four convolutional layers (conv) followed by two fully connected layers (fc). The first convolutional layer i.e conv1 subjected to local response normalization, with kernel size 11, which depicts that each unit in each feature map is connected to 11×11 neighborhood in the input and stride of 4, which means after every four pixels perform the convolution on the input images. The output of the first convolution layer are 96 feature maps. The first layer i.e. conv1 layer is followed by pooling. The kernel size for the pooling is set to 3 with stride 2. Pooling is followed by convolution conv2 with kernel size 5 and stride 2. The pooled feature maps are again convolved in layer conv3, with parameter setting of kernel size equal to 3, stride of 2. These convolved features are again convolved in layer conv4 with parameter setting same as in layer conv3. Which is followed by fully connected layers (fc), i.e. fc5, fc6. In the layer fc6 in Alexnet two operations are applied, i.e. relu6 and drop6. While as in our proposed architecture, fully connected layer 5 (fc5) is only subjected to rectified linear unit operation. The output of of our con4 layer are 256 where as in AlexNet 384 feature maps are generated. The layer fc5 is followed by fully connected layer while fc6 which results in 4096 dimensional vector for each image.

The architecture of the proposed CNN used for medical image anatomy classification is as shown in Figure 2 while the hyperparameter specifications of the proposed CNN framework are given in Table I.

TABLE I: Hyperparameter Specifications of the proposed CNN framework in units.

HyperParameter	**Layer1**	**Layer 2**	**Layer 3**	**Layer 4**
Number of filters	96	256	384	256
kernel size	11×11	5×5	3×3	3×3
stride	4	2	2	2
Learning rate	0.01			
Momentum	0.9			
Weight Decay	0.0005			
Training epochs	30-60			
Number of units in fully connected layer				4096

In AlexNet [15], five convolutional and three fully connected layers were used, whereas our architecture contains

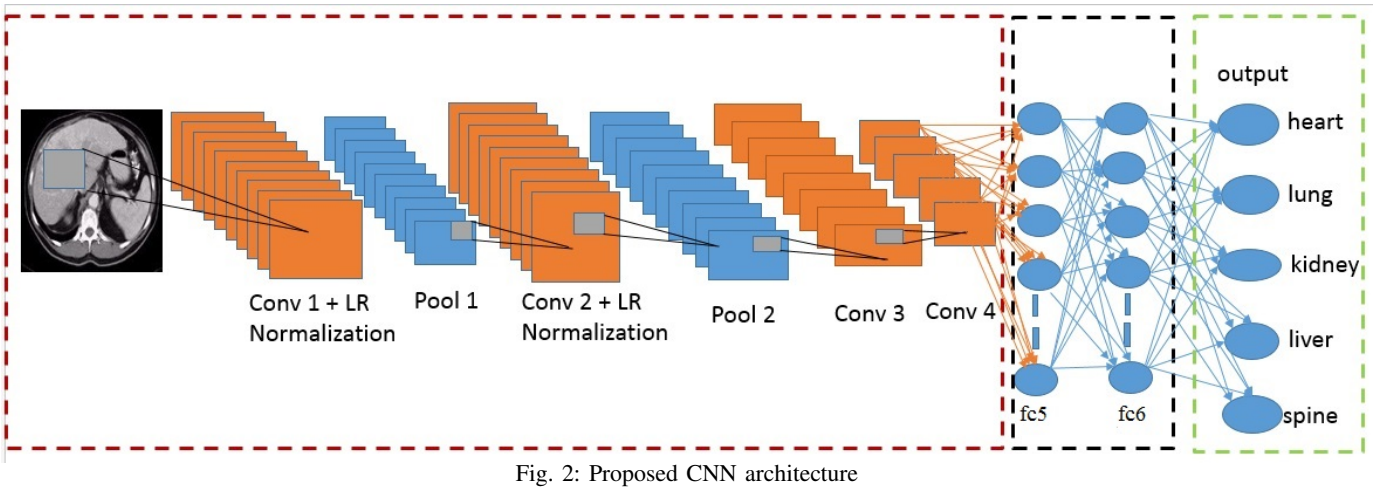

Fig. 2: Proposed CNN architecture

only four layers with two fully connected layers (fc): fc5 and fc6. We did not use the dropout layers that have been used with fc6 and fc7 layers in AlexNet, because looking at the visualization of the feature maps most of the activations are dumped out in higher layers. The result of which it does not control any overfitting but rather adds complexity to the network. Outputs from the convolution layer 4 are calculated as:

$$Y^{l}_{i,j} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} W_{ab} * X^{l-1}_{(i+a)(j+b)} \tag{1}$$

The features maps resulted from convolution are subjected to rectified linear unit operation as follows:

$$y_{ij} = max\{0, Y_{ij}\} \tag{2}$$

In AlexNet, layers fc6 and fc7 are subjected to dropout for regularization. Dropout prevents co-adaptation of hidden units by randomly dropping out i.e., setting to zero a proportion p of the hidden units during forward back-propagation. That is, given the penultimate layer $l = [l_1, ..., lm]$, where m represents the filters. The dropout is formulated as :

$$y = w(l \odot r) + b, \tag{3}$$

where $\odot$ is the element-wise multiplication operator and r $\in \mathbb{R}_m$ is a masking vector of Bernoulli random variables with probability p of being 1.

Gradients are backpropagated only through the unmasked units. So if the Drop out masks the maximum unit it will cause the weights to update in such a way that the neuron will never activate on any data point again. If this happens, then the gradient flowing through the unit will forever be zero from that point on. So these activated units will ultimately vanished during training process.

In our modified architecture, this does not subject to fully connected layer, fc5 to drop out operation, rather feed it with the output of the conv4 layer, as shown in a simplified expression,

$$y = w.l + b \tag{4}$$

IV. Experiment and Results Discussion

Experiments were conducted with a machine incorporated with NVIDIA GeForce GTX 980M, using a data set acquired from the U.S. National Library of Medicine, National Institutes of Health, Department of Health and Human Services[23]. The open accessed medical image database contains thousands of anonymous annotated medical imaging data. Anatomical images that are used in this experimentation consist of CT, MRI, PET, Ultrasound and X-ray modalities. This database contains images with various pathologies. For our experimental evaluation, we adopted 37198 images of five anatomies to train the CNN models. For testing, we used 500 images other than that in the training set, i.e. 100 images per anatomy. So a total of 37698 images were used in the experiments. The anatomies considered in our experiments were lung, liver, heart, kidney and lumbar spine. Sample images are shown in Figure 3.

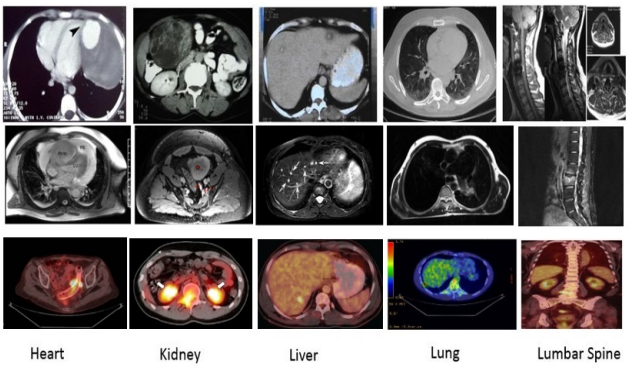

Fig. 3: Example images of five anatomies from various modalities. First row corresponds to CT modality, Second row corresponds to MRI modality and third row corresponds to PET modality.

The normal and pathological images were used, so that these frameworks should be generalized to classify any image of the same organ if it varies in shape or contrast.

The dataset was tested with the three milestone architectures, i.e LeNet [14], AlexNet [15] and GoogLeNet [17].

The comparative results after applying these CNN architectures on our dataset are shown in Figure 4, 5 and 6 respectively. The validation accuracy and validation losses are computed from the last layers of each architecture i.e., LeNet, AlexNet and GoogLeNet respectively. There are three different accuracies in GoogLeNet, i.e., loss1/accuracy(val), loss2/accuracy(val) and accuracy(val) that correspond to three different classifiers that this GoogLeNet network uses during training. In this network, loss1/accuracy(val) is evaluated after inception layer 4a and loss2/accuracy(val) is evaluated after inception layer 4d. This is the naming convention in GoogLeNet architecture and the final accuracy (accuracy(val)) is evaluated at the end of the net which has been used in this study.

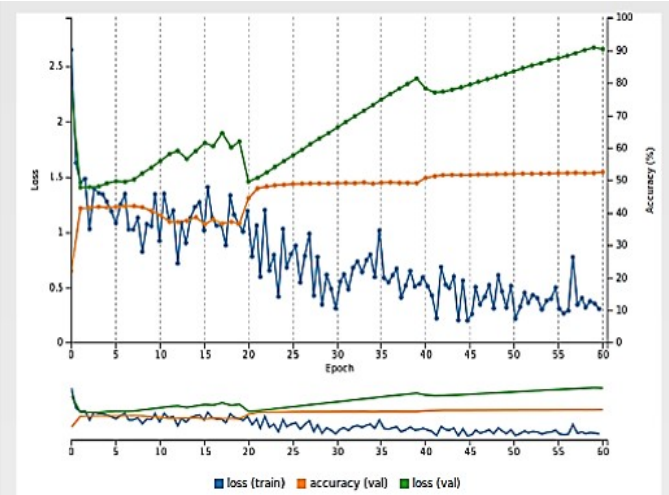

Fig. 4: Training and Validation error with each epoch for LeNet[14]

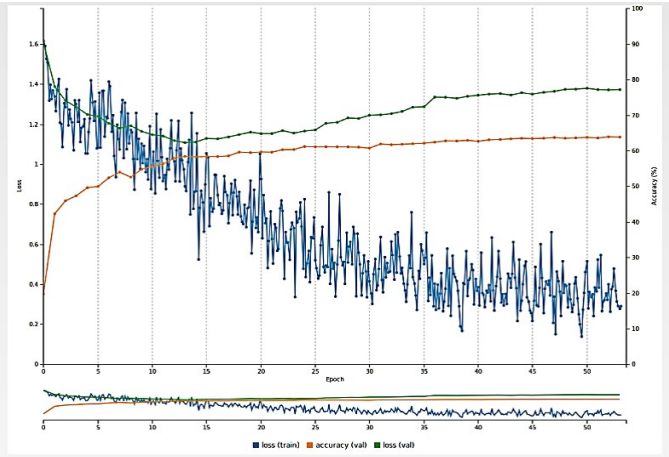

Fig. 5: Training and Validation error with each epoch for AlexNet [15]

All these CNNs have been trained with Stochastic Gradient Descent algorithm. After evaluating these three milestone architectures, that clearly depicts from the above results that these CNNs over fits for the task of medical image classification. To figure out what is the reason for over fitting, we visualized the filters and feature maps of these three

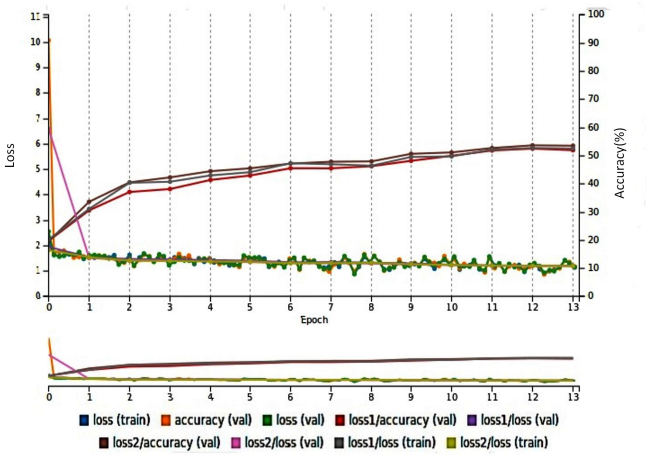

Fig. 6: Training and Validation error with each epoch for GoogLeNet [17]

architectures as shown in Figure 7, Figure 8 and Figure 9 respectively.

Analyzing these visualizations clearly depict that the filters learned by LeNet and GoogLeNet are not distinguishable enough to depict the edge like features, that are supposed to be learned by the first convolutional layer as there is lot of noise in filters as shown in Figure 7 and Figure 8. After progressing through the convolutional layers, the visualization shows that most of the feature maps in LeNet does not clearly figure out the structure representation of the anatomical structure. Where as most of the features maps are dumped out in GoogLeNet convolutional layers as shown in Figure 8. These visualizations clearly depict that these architectures are not learning the representations effectively.

On the other hand looking at the visualizations from AlexNet. It is shown that the weights learned by AlexNet as shown in Figure 9 are comparatively distinguishable than LeNet and GoogLenet as it able to capture the edge like features and also the most of the feature maps are retained through various convolutional layers. But looking at its training and validation process this network is also over fits because of its large number of parameters as they progress through higher layers.

Based on the comparative performance of AlexNet with other architectures and to overcome its overfitting problem, we modified the basic architecture of AlexNet. The visualizations from proposed framework are shown in Figure 10 and Figure 11 .

The visualizations clearly shows that the modified architecture learns better representations than other architectures and also training and validation graphs depicts the same behavior. The training and validation results from the proposed CNN architecture is shown in Figure 12.

It is evident from Figure 12 that the proposed CNN gives good validation accuracy with low training loss and validation loss.

For evaluating the performance of proposed CNN and three

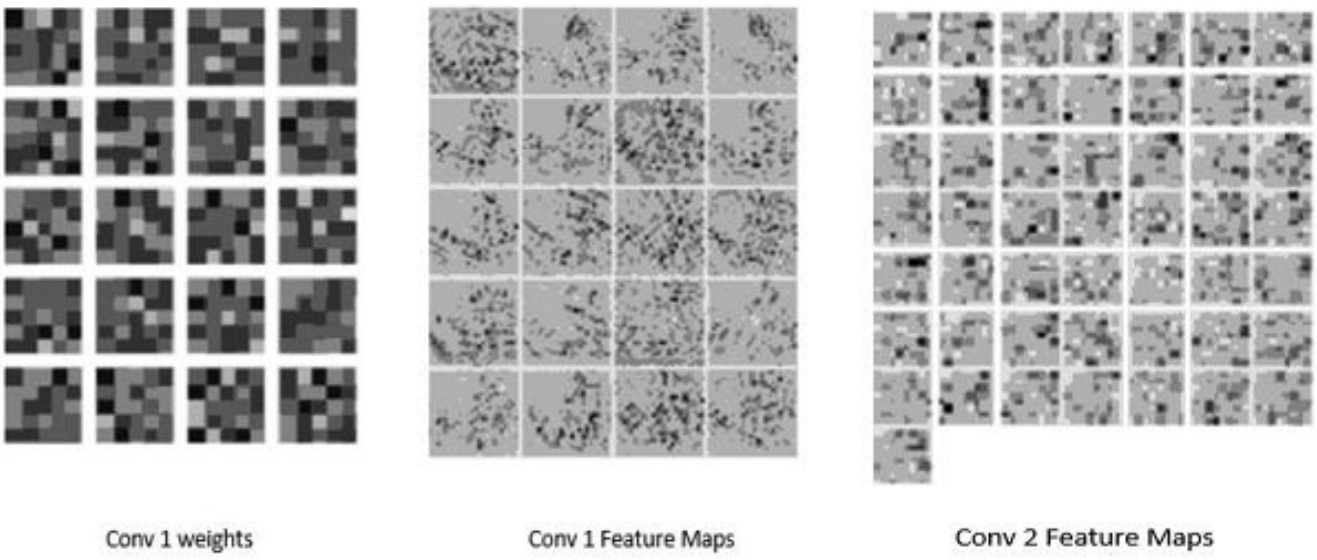

Fig. 7: Filter and feature map visualization of LeNet for analyzing the training and validation process

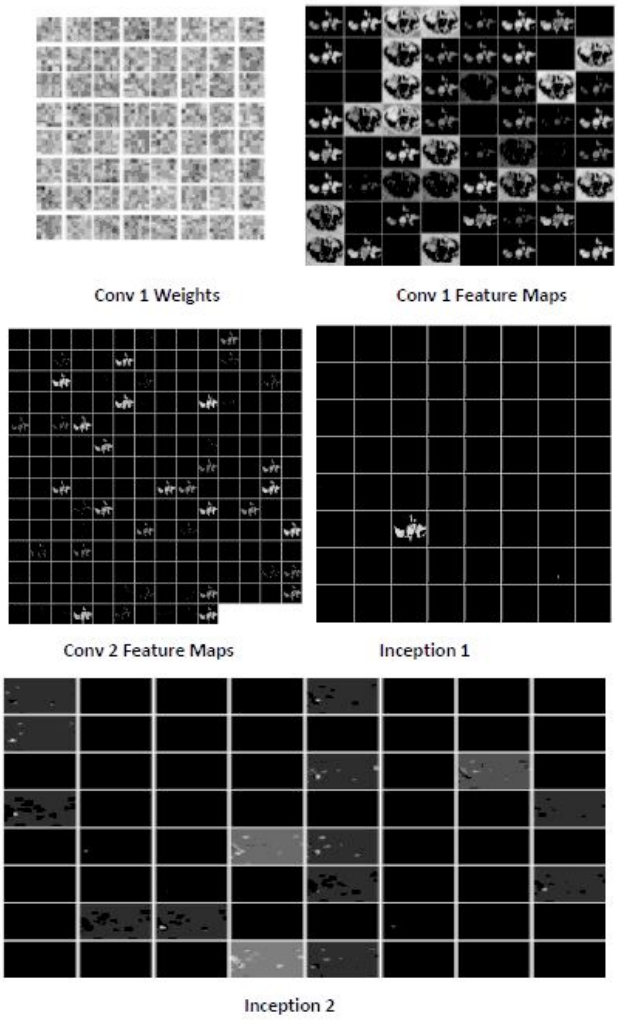

Fig. 8: Filter and feature map visualization of GoogLeNet for analyzing the training and validation process

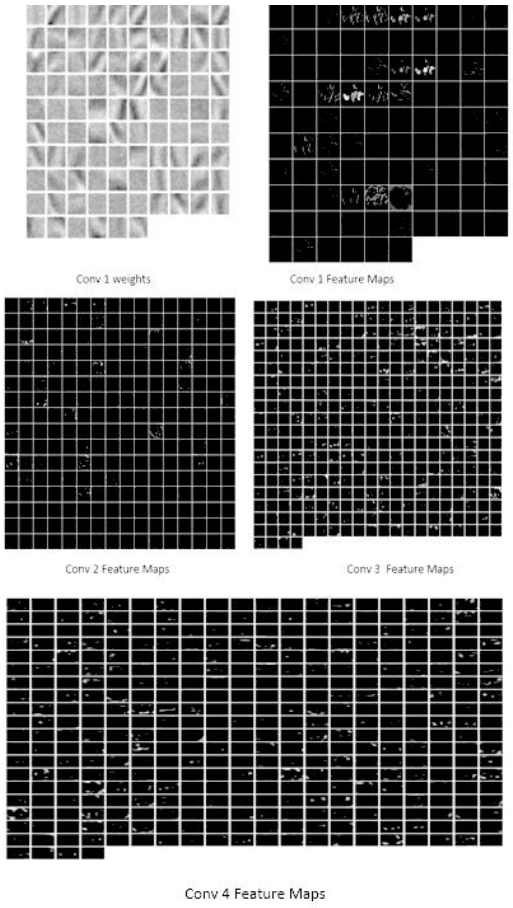

Fig. 9: Filter and feature map visualization of AlexNet for analyzing the training and validation process

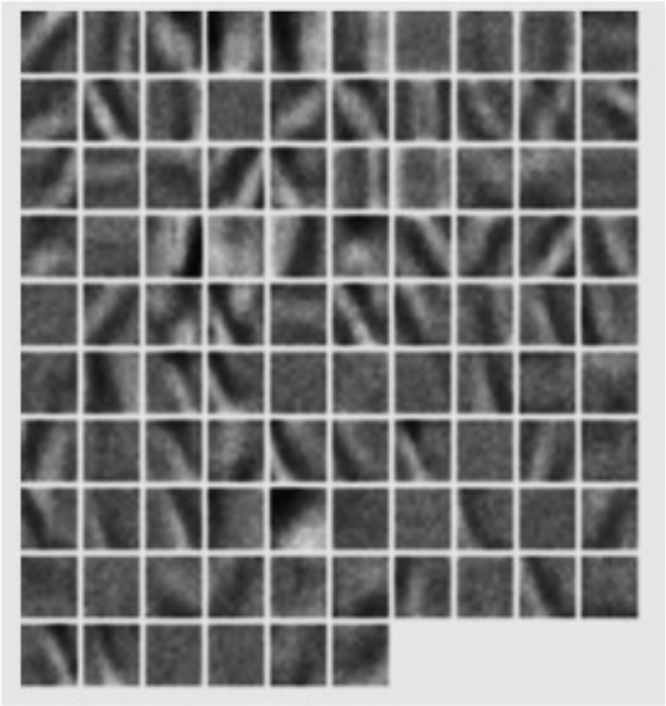

Fig. 10: Filter visualization of proposed framework

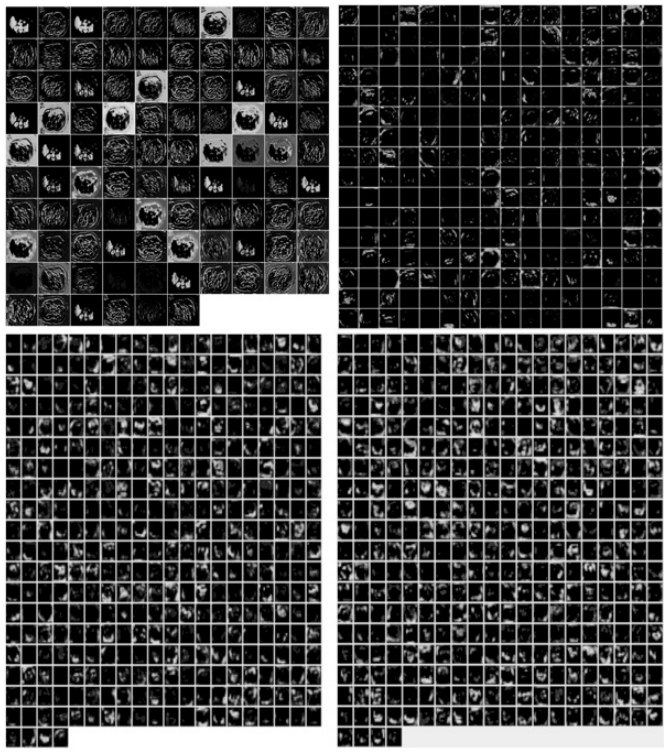

Fig. 11: Filter and feature map visualization of proposed framework for analyzing the training and validation process

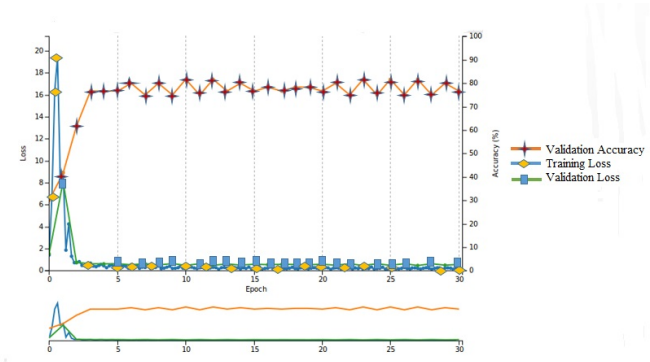

Fig. 12: Training and Validation error with each epoch for Proposed CNN architecture for classifying medical image anatomies

milestone architectures, we conducted 7 different experiments using CNN, by using varying sizes of training sets and recorded the result classification accuracy. Each experiment was validated using the randomly selected 20% data from the training data set. The comparative performance of these three milestone architectures with increasing number of datasets is shown in Figure 13. It clearly showed that larger sets of training data leads to increased accuracy in classification which supports our claim that CNN are data intensive architectures and relatively lack of training data in medical imaging needs the modification of the milestone architectures. Since these three milestone architectures have been used for the natural image classification, a subtle treatment to the parameter tuning and layer formulation is needed. Therefore, we came up with the modified CNN architecture, in which the modifications have been carried out on [15]. In [15], five convolutional and three fully connected layers were used. However, our architecture contains only four convolutional layers with two fully connected layers.

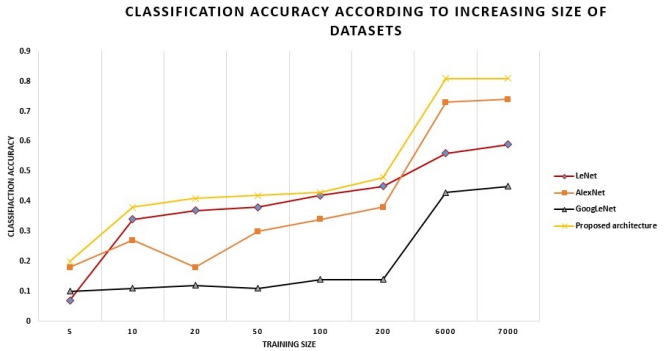

Fig. 13: Classification Accuracy According to the Increasing Size of Datasets

The summary of the comparative performance of the proposed CNN and the three milestone architectures in terms of runtime, training loss, validation accuracy and test accuracy is given in Table II. It can be seen that our proposed CNN outperforms other three milestone CNN architectures by having 81% accuracy while AlexNet achieved only 74%, followed by LeNet 59% and GoogleLeNet 45%.

TABLE II: Comparative performance of proposed CNN with LeNet, AlexNet and GoogLenet in terms of runtime, training loss, validation accuracy and test accuracy

Model	**Runtime in seconds**	**Validation Loss**	**Validation Accuracy (%)**	**Test Accuracy (%)**
LeNet[14]	1655	1.3	58	59
AlexNet[15]	33466	1.39	65	74
GoogLeNet[17]	52470	1.2	55	45
Proposed CNN	16728	**0.67**	**76.6**	**81**

The results in Table II shows that the architectures used for natural image classification cannot be generalized on medical images of anatomies. It is evident from the above results, that modification of the basic CNN architecture in terms of number of layers, the normalization function and subtle tuning of hyper parameters shall yield better results for the task of medical image anatomy classification.

V. Conclusion

In this paper, we proposed a modified CNN architecture that combines multiple convolution and pooling layers for higher level feature learning. The experiments for medical image anatomy classification has been carried out and it shows that the proposed CNN feature representation outperforms the three baseline architectures for classifying medical image anatomies. The modification of CNN has been done on the basis of experimentation, that is carried out with the three milestone architectures. These models overfit due to the number of layers and the hyper-parameters used in these architectures have been used for large set of natural images. However, medical image datasets are different in terms of their acquisition medium and less availability because of privacy and security policies as compared to natural images. In this paper, We also provide an insight into the deep features that have been learned through training, that will help in analyzing various abstraction of features ranging from low level to high level and their role in final classification.

Our future work will extend to recognition and classification of pathological structures from these classified anatomies, leading to a fully automated medical image classification system.

References

[1] Xin Zhou, Adrien Depeursinge, and Henning Müller. Hierarchical classification using a frequency-based weighting and simple visual features. *Pattern Recognition Letters*, 29(15):2011–2017, 2008.

[2] Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. Discriminative cue integration for medical image annotation. *Pattern Recognition Letters*, 29(15):1996–2002, 2008.

[3] Jayashree Kalpathy-Cramer and William Hersh. Effectiveness of global features for automatic medical image classification and retrieval–the experiences of ohsu at imageclefmed. *Pattern recognition letters*, 29(15):2032–2038, 2008.

[4] Uri Avni, Hayit Greenspan, Eli Konen, Michal Sharon, and Jacob Goldberger. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *Medical Imaging, IEEE Transactions on*, 30(3): 733–746, 2011.

[5] Adrien Depeursinge, Alejandro Vargas, Alexandra Platon, Antoine Geissbuhler, Pierre-Alexandre Poletti, and Henning Müller. 3d case–based retrieval for interstitial lung diseases. In *Medical Content-Based Retrieval for Clinical Decision Support*, pages 39–48. Springer, 2010.

[6] Yang Song, Weidong Cai, Heng Huang, Yun Zhou, Yue Wang, and David Dagan Feng. Locality-constrained subcluster representation ensemble for lung image classification. *Medical image analysis*, 22(1):102–113, 2015.

[7] M Mostafizur Rahman and DN Davis. Addressing the class imbalance problem in medical datasets. *International Journal of Machine Learning and Computing*, 3(2):224–228, 2013.

[8] Jennifer G Dy, Carla E Brodley, Avi Kak, Lynn S Broderick, and Alex M Aisen. Unsupervised feature selection applied to content-based retrieval of lung images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(3):373–378, 2003.

[9] M Srinivas, R Ramu Naidu, CS Sastry, and C Krishna Mohan. Content based medical image retrieval using dictionary learning. *Neurocomputing*, 168:880–895, 2015.

[10] Dar-Ren Chen, Ruey-Feng Chang, Chii-Jen Chen, Ming-Feng Ho, Shou-Jen Kuo, Shou-Tung Chen, Shin-Jer Hung, and Woo Kyung Moon. Classification of breast ultrasound images using fractal feature. *Clinical imaging*, 29(4):235–245, 2005.

[11] Ashis Kumar Dhara, Sudipta Mukhopadhyay, Anirvan Dutta, Mandeep Garg, and Niranjan Khandelwal. A combination of shape and texture features for classification of pulmonary nodules in lung ct images. *Journal of digital imaging*, 29(4):466–475, 2016.

[12] Mohammad Reza Zare, Ahmed Mueen, and Woo Chaw Seng. Automatic medical x-ray image classification using annotation. *Journal of digital imaging*, 27(1):77–89, 2014.

[13] Wei Yang, Zhentai Lu, Mei Yu, Meiyan Huang, Qianjin Feng, and Wufan Chen. Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced ct images. *Journal of digital imaging*, 25(6):708–719, 2012.

[14] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[16] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition,localization and detection using convolutional networks. In *International Conference on*

Learning Representations. ICLR 2014, 2014.

[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[18] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 411–418. Springer, 2013.

[19] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 246–253. Springer, 2013.

[20] Holger R Roth, Le Lu, Ari Seff, Kevin M Cherry, Joanne Hoffman, Shijun Wang, Jiamin Liu, Evrim Turkbey, and Ronald M Summers. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 520–527. Springer, 2014.

[21] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pages 844–848. IEEE, 2014.

[22] Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. Medical image deep learning with hospital PACS dataset. *CoRR*, abs/1511.06348, 2015.

[23] Holger R Roth, Christopher T Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, and Ronald M Summers. Anatomy-specific classification of medical images using deep convolutional nets. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 101–104. IEEE, 2015.

[24] David Lyndon, Ashnil Kumar, Jinman Kim, Philip HW Leong, and Dagan Feng. Convolutional neural networks for medical clustering. volume 1391. CEUR Workshop Proceedings, 2015.