Convolutional Neural Networks for Lung Cancer Screening in Computed Tomography (CT) Scans

Prajwal Rao*, Nishal Ancelette Pereira†, and Raghuram Srinivasan‡
Department of Electronics and Communication Engineering,
M. S. Ramaiah Institute of Technology
Bengaluru 560054, INDIA
Email: *prajwaljpj@gmail.com, †nishalpereira@gmail.com, ‡raghuram@msrit.edu

***Abstract*—Diagnosis and cure of cancer has been one of the biggest challenges faced by mankind in the last few decades. Early detection of cancer would facilitate in saving millions of lives across the globe every year. This paper presents an approach which uses a Convolutional Neural Network (CNNs) to classify tumours seen in lung cancer screening computed tomography scans as malignant or benign. CNNs have special properties such as spatial invariance, and allow for multiple feature extraction. When such layers are cascaded, leading to Deep CNNs, it has been shown widely that the accuracy of prediction increases dramatically. In this work, we have designed a CNN suitable for the analysis of CT scans with tumours, using domain knowledge from both medicine and neural networks. The results show that the accuracy of classification for our network performs better than both the traidtional neural networks, and also existing CNNs built for image classification purposes.**

I. Introduction

Trained radiologists are required in identifying cancerous tumours accurately, and due to the cost of the infrastrcture involved, are generally beyond the reach of the lower and middle classes of society. Indirectly, this leads to reduced detection of early signs of cancer, and thereby making the cure of the disease much more complicated. By developing computer aided methods for the accurate detection of malignancy, we can dramatically reduce the cost of diagnosis, thereby making the treatment and recovery process much more sucessful. In this work, a neural network based program is used to distinguish between benign and malignant cancerous tumours.

Traditionally, image processing algorithms have been used to detect specific features on images. However, such methods requires design and development of a set of image-specific hand crafted features which are tailored to the particular task at hand [1, 2]. Due to the significant complexity of the distinguishing features between benign and malignant tumours, feature engineering in this case cannot take into account all the determining factors. Deep learning architectures are capable of tackling a variety of problems such as object recognition, speech recognition and natural language processing [3]. The use of deep learning circumvents the process of manually choosing the CAD tool parameters. Manual extraction of right feature would require the designer to have domain expertise on lung cancer. Deep learning, in particular the CNN overcomes this by extracting features in a hierarchical manner by using multiple layers of convolution and max-pooling.

CNNs were introduced by LeCun et al in [4], and get their inspiration from an earlier model, the neurocognitron [12]. The neurocognitron was the first neuronal model to propose shift invariance - the relative position of the object within the image is not as important as its actual presence. In addition, CNNs allow for multiple features to be extracted at each hidden layer. Hence, multiple characteristics can be extracted, and the learning process assigns weights appropriately to the significant features, thereby automatically performing the difficult task of feature engineering. Further, when multiple hidden layers are used, leading to Deep CNNs, features are learned hierarchically [4]. For example, the learning procedure automatically determines that popcorn classification is a feature of a benign tumour. Such a feature would require a complicated algorithm, if only pure image processing methods were used. Further, since the same features are learned over the image, weight sharing drastically reduces the number of computations as compared to traditional neural networks. We propose a CNN architecture with two convolution layers, a pooling layer, a fully connected layer and a drop out layer for the classification of CT scans. Classification accuracy obtained using the proposed CNN is compared with that of an Artificial Neural network comprising of 50 neurons in the hidden layer and also with LeNet [4], which is a CNN designed for digit recognition.

II. Previous Work

Neural networks have been used extensively in the classification of cancerous tumours [13, 14]. However, for the particular case of tumour classification, CNNs are more appropriate: since cancer is primarily detected by the calcification patterns [13, 14], local connected patterns are more important for the classification process. However, if we use a fully connected neural network, where all areas of the image affect each other, excessive unnecessary weights are computed and incorrect functionalities begin to introduce into the classification procedure. Further, due to the additional property of spatial invariance, the malignant variety of classification will be detected irrespective of its position in the image. In certain works such as [2], extensive pre-processing is used, prior to using a neural network for classification. While the

978-1-5090-5256-1/16/$31.00

accuracy achieved thus is significant, multiple algorithms are used, and thereby cannot be used directly in future hardware implementation of neuromorphic circuits, such as those in [15].

Previously, CNNs have been proven to be successful for various biomedical applications such as detection of invasive ductal carcinoma in whole slide image [5], tumour tissue image classification [6], mitosis detection in breast cancer histology images [7]. In this work, we extend the existing literature to include tumour classification using a specifically designed CNN, which we will henceforth call as CanNet.

III. Methodology

In this section, we discuss the data and the architecture of the network that we have designed for tumour classification. We begin this section with details about the data, and the steps in converting it into a suitable form for the CNN.

A. Data

The database used is obtained from Lung Image Database Consortium (LIDC) [10]. This is a lung nodule classification database containing the scans of a total of 1018 patients. Each patients' CT scan in turn is comprising of around 150 to 550 dicom format images. The database provides four classifications namely-(i)Unknown, (ii)Benign, (iii)Malignant, and (iv)Metastatic. Out of which, we consider only the benign and malignant labelled data. This is done because of two reasons: Classification of an Unknown category is futile and a Metastatic category will require the origin of the cancer cells from other CT scans which is not provided. Overall, segregating the classes will result in 71 patients' data, which is being used. All dicom format images are, at first, obtained as 512x512 grayscale images. Since the overall size of the CT scan becomes very large, this is scaled down to a 128x128 grayscale image. Each image is then normalized by dividing this grayscale image pixel value by 255 (this is done so that the computations stay within finite values, otherwise error terms, etc. very quickly become unbounded). Next, each of the grayscale images are concatenated linearly to create a 3D volume in HDF5 format. Further, this 3D array is zero padded to equalize the third dimension, i.e., the number of images in each CT scan provided for one patient. For conversion to HDF5 format, the 3D array is reshaped into an array of (1, 500, 128, 128). The conversion process is done using a python script. HDF5 contains two inputs: one being the data and the other being the label. The data are the 3D array and the label is another dimension to specify the class of the CT scan which in our case is 0 for benign and 1 for malignant. The sequence of operations for the preparation of the data are shown in Figure 1.

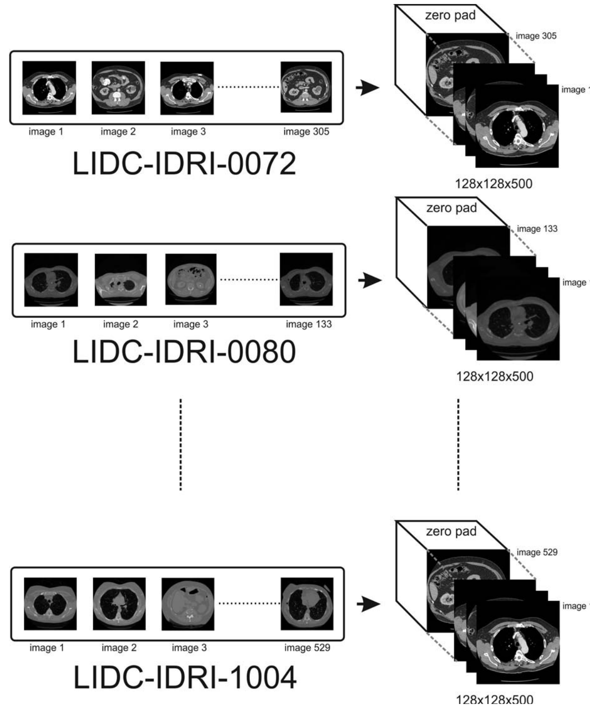

Fig. 1. Representation of input to CNN

B. Convolutional Neural Network

A CNN is a neural network architecture that efficiently exploits the spatial correlation of the input data. Moreover, weight sharing in CNN facilitates in learning a feature regardless of its position in the image, along with having the added advantage of reduced computations as compared to a fully connected ANN. The convolution layer of a CNN produces a feature map by convolving different sub regions of the image with a learned kernel (learned during the training process). Further, non-linear activation functions such as a sigmoid, tanh or rectified linear (ReLu) can also be applied. The ReLu layer is also known to improve the convergence properties when the error is low, leading to stagnation in the traditional sigmoid activation function [4]. Another method for reducing computations is the pooling layer, where a region of the image/feature map is chosen and the maximum among them is chosen as the representative pixel. Hence, a 2x2 or 3x3 grid can be reduced to a single scalar value. Average values can also be used for pooling, but in either case, the net effect is a large reduction in the sample size. More details about kernel size, feature maps, stride and other parameters related to a convolutional layer can be found in [4]. Additionally, traditional fully connected layer can also be used in conjunction with the convolutional layers, and are usually used towards the output stage, as in [4].

C. Proposed CanNet Architecture

We are proposing CanNet which is based on the layers that are used in a CNN, along with appropriate values for network parameters. CanNet contains two convolution layers immediately after the input layer, followed by a pooling layer, a dropout layer and a fully connected layer. Each convolution layer is in turn followed by a rectified linear output layer and the pooling layer is followed by a Dropout layer. The architecture is shown in Figure 2, and the layers are explained briefly:

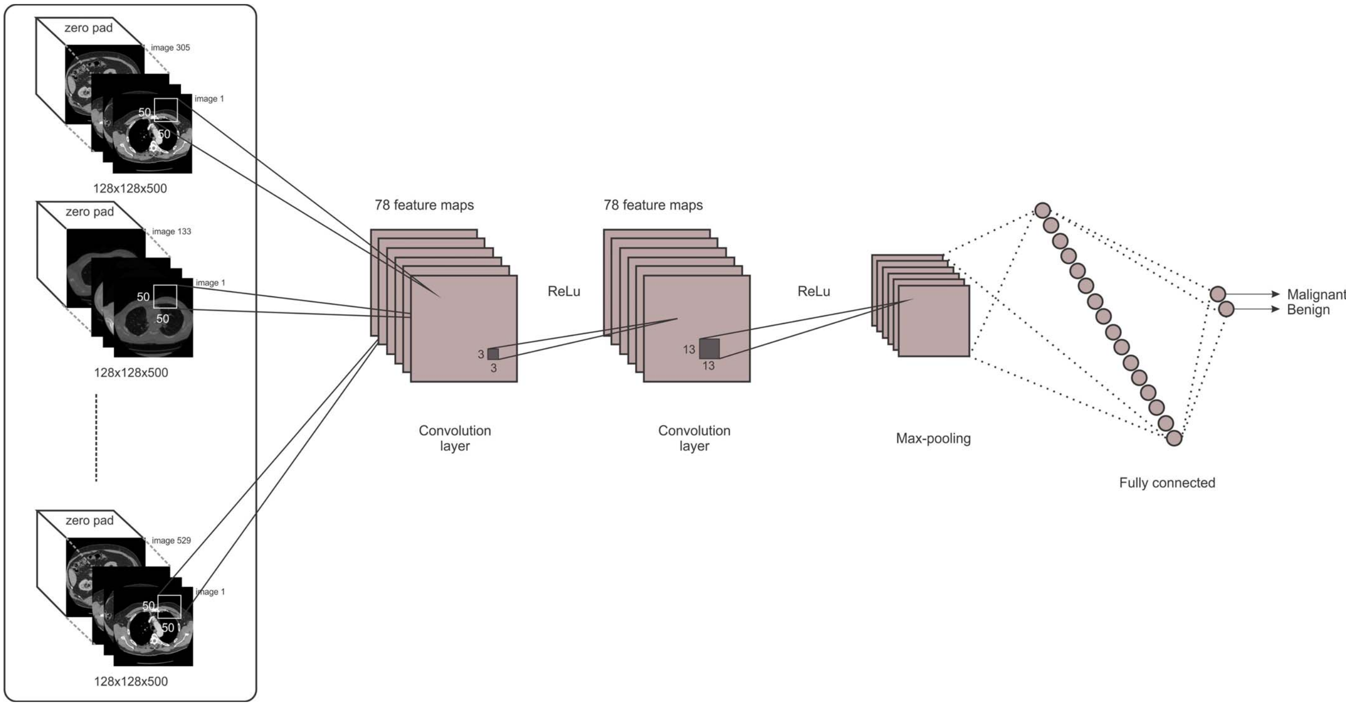

Fig. 2. Architecture of CanNet

Convolution Layer 1: The data in 3-D hdf5 format forms the input to the first convolution layer. This layer has a kernel size of 50x50 with a stride of 6. The output of this layer produces 78 features. The weight filler is set to a 0.01 Gaussian distribution change and the bias is set at constant zero. This output is then fed to the Rectified Linear (ReLu) layer to bring all the negative activations to zero. The primary application of this layer is to detect the lowest level features, e.g., whether there is calcification in some area of the image.

Convolution Layer 2: The first Convolution layer output is fed into the second having a kernel size of 3x3 and a stride of 1. This layer pads the data with one enclosure of zeros. The weight filler is the same as convolution layer 1 and the bias is set to a constant value of 1. Also, this layer is followed by a ReLu layer. This layer is intended to make use of the information predicted from the previous layer and detect the pattern of calcification - e.g., popcorn, diffuse etc. From the training phase, it will hence learn as to which among the patterns are benign, and which are malignant. In this way the CNN achieves two objectives - it learns features hierarchically, and it eliminates the need for specific feature engineering.

Max-pooling Layer: After the convolution layer 2 comes the max-pooling layer where the most responsive node of the given kernel is extracted. The kernel size used in the proposed network is 13x13 with stride shift of 13. This is primarily intended to reduce the computational effort. Since each CT scan composes of 500 images, if we have a batch size of 50, the number of required computations can be significantly large, leading to frequent memory overloads. The max-pooling layer is used particularly to ease memory and data bottlenecks by reducing the image dimensions.

Dropout layer: The dropout layer is used in the network to prevent over-fitting. This is done by switching off random neurons in the network. Our proposed network uses a dropout layer with a drop ratio of 0.5. The intent of this layer is to improve the classification quality on test data that has not been seen by the network earlier.

Fully connected layer: A fully connected layer which provides two outputs is used. It uses Gaussian weight filler of 0.5 and a constant bias filler of 0. The two output neurons from this layer gives the classification of benign or malignancy. This layer is mainly intended to combine all the features into one top level image and will ultimately form the basis for the classification step.

Figure 2 shows the overall structure of the CanNet network that we have designed.

IV. Experimental Results and Discussion

For the creation of CanNet, we have used the Caffe deep learning framework developed by the Berkeley Vision and Learning Center [11]. The framework provides for the creation of deep networks by choosing appropriate layers and specifying the preceding and succeeding layers in the design. The inputs to the framework can be in the HDF5 format, which is particularly suitable for the representation of 3D data, such as a CT scan. The steps in preparing the data are explained in the previous section, and are the same for each CT scan. Hence, we have one HDF5 file representing all the CT scans of a patient, and each HDF5 file has the data along with the label. This label is used in both the training and testing phase.

Batch sizes are also variable, and can be set by the user. For large batch sizes, the learning process is significantly slow (requires a few days) and often terminates due to insufficient memory availability. We have used a batch size of 20 for most experiments. The training of the network is run for 1000 iterations. After every 100 iterations the network is tested for accuracy. Initial learning rate is set to 0.001 and for every 100 iteration the learning rate drops by a factor gamma=0.1

A. Results

The accuracy of classification of the CT scans is calculated for a traditional ANN with fifty neurons in the hidden layer, LeNet [4] and the proposed CanNet architecture. The learning is depicted in the graphs shown below for different iterations. ANN in Figure 3 shows an accuracy that greatly fluctuates with the iterations. LeNet, in Figure 4, however reaches a constant accuracy, but is significantly lower than our CanNet. As seen in Figure 5, shows steady increase in the accuracy as the iterations increase. Maximum testing accuracy is reported in case of CNN as tabulated in Table 1. As can be seen, we can get upto 45 % improvement in accuracy as compared to LeNet, and about 14% as compared to a traditional ANN with our CanNet architecture.

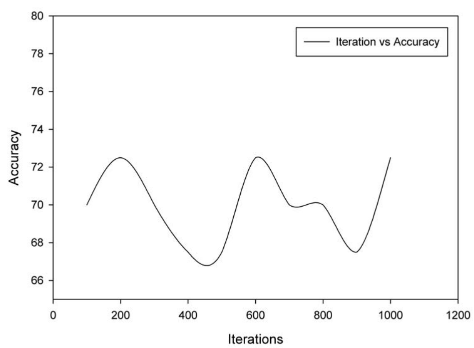

Fig. 3. Accuracy vs. Iteration for ANN

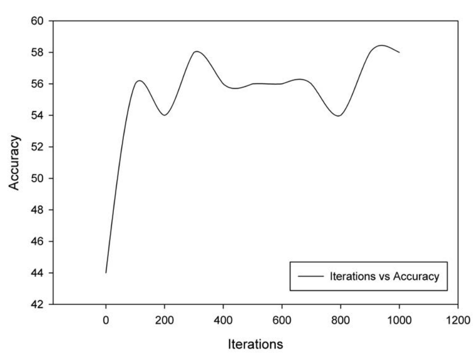

Fig. 4. Accuracy vs. Iteration for LeNet

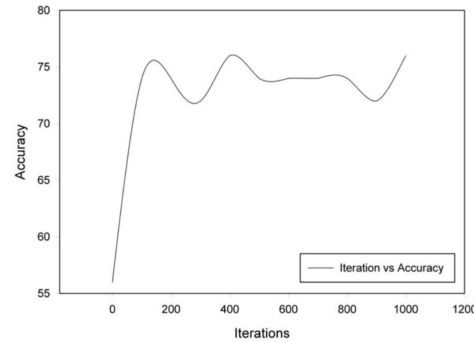

Fig. 5. Accuracy vs. Iteration for proposed CNN

TABLE I
COMPARISON OF ARCHITECTURES

Architecture	Max. Test Accuracy	Improvement
ANN	72.50%	14.50%
LeNet	56.00%	N/A
CanNet	76.00%	45.40%

B. Discussion

For LeNet, the network might be too large for the type of data we are giving it, leading to overfitting and a reduction in the accuracy when testing with non-training data. We notice similar results when increasing the number of convolution layers to more than 2 (e.g., 4) in our CanNet. Hence, through experimentation, we have determined that two consecuting convolutional layers lead to the best classification results for the proposed CanNet architecture. Further, due to the nature of the classification problem, we strongly feel that CNNs such as the proposed CanNet are much better candidates as compared to the traditional ANN architectures.

V. CONCLUSION

The translation invariance of Convolutional Neural Networks is exploited to classify the lung cancer screening thoracic CT scans efficiently. By using CNNs, one can forgo the tedious process of manually extracting features for classification which requires specific domain knowledge. The documented classification accuracy indicates that CanNet outperforms both ANN and the LeNet architecture for the given classification task. Further experimentation on various hyper parameters of the CNN could be done in order to increase the accuracy. This work presents a first step towards automating the classification procedure, and eventually becoming better than trained technicians at this particularly crucial and critical task.

REFERENCES

[1] Bram van Ginneken, Arnaud A. A. Setio, Colin Jacobs, and Francesco Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans", *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015.

[2] J. Kuruvilla and K. Gunavathi, ”Lung cancer classification using neural networks for CT images”, *Computer Methods and Programs in Biomedicine, vol. 113, no. 1, pp. 202-209*, 2014.
[3] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, and G. Wang, ”Recent advances in convolutional neural networks,” *arXiv preprint arXiv:1512.07108*, 2015.
[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE, vol. 86, no. 11, pp. 22782324*, 1998.
[5] Cruz-Roa A, Basavanhally A, Gonzlez F, Gilmore H, Feldman M, Ganesan S, et al., ”Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks,” *Medical imaging: digital pathology. San Diego, CA, USA: SPIE*, 2014.
[6] X. Yang, S. Yeo, J. Hong, S. Wong, W. Tang, Z. Wu, G. Lee, S. Chen, V. Ding, B. Pang, A. Choo and Y. Su, ”A Deep Learning Approach for Tumor Tissue Image Classification”, *IASTED Biomedical Engineering*, 2016.
[7] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, ”Mitosis detection in breast cancer histology images with deep neural networks,” *In MICCAI*, 2013.
[8] Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, et al., ”The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans,” *Medical Physics, 38: 915–931*, 2011.
[9] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F, ”The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository,” *Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057*, 2013.
[10] Armato III, Samuel G., McLennan, Geoffrey, Bidaut, Luc, McNitt-Gray, Michael F., Meyer, Charles R., Reeves, Anthony P., Clarke, Laurence P, ”Data From LIDC-IDRI,” *The Cancer Imaging Archive*
[11] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, ”Caffe: Convolutional Architecture for Fast Feature Embedding,” *arXiv preprint arXiv:1408.5093*, 2014.
[12] Fukushima, K, ”Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”, *Biological Cybernetics, Vol36, No. 4*, 1980.
[13] Kauczor, H-U, et al., ”Automatic Detection and Quantification of Ground-Glass Opacities on High-Resolution CT Using Multiple Neural Networks Comparison with a Density Mask”, *American Journal of Roentgenology, Volume 175, Issue 5*, 2000.
[14] Suzuki, K, et al., ”Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography”, *Medical Physics, Vol. 30, No. 7*, 2003.
[15] Hasler, Jennifer and Marr, Bo, ” Finding a roadmap to achieve large neuromorphic hardware systems”, *Frontiers in Neuroscience*,2013.