A comparative study of deep learning models for medical image classification

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042097 doi:10.1088/1757-899X/263/4/042097

A Comparative Study of Deep Learning Models for Medical Image Classification

Suvajit Dutta, Bonthala CS Manideep, Shalva Rai and Vijayarajan V
School of Computer Science and Engineering, VIT University, Vellore-632014, India

Email: vijayarajan.v@vit.ac.in

Abstract. Deep Learning (DL) techniques are conquering over the prevailing traditional approaches of neural network, when it comes to the huge amount of dataset, applications requiring complex functions demanding increase accuracy with lower time complexities. Neurosciences has already exploited DL techniques, thus portrayed itself as an inspirational source for researchers exploring the domain of Machine learning. DL enthusiasts cover the areas of vision, speech recognition, motion planning and NLP as well, moving back and forth among fields. This concerns with building models that can successfully solve variety of tasks requiring intelligence and distributed representation. The accessibility to faster CPUs, introduction of GPUs-performing complex vector and matrix computations, supported agile connectivity to network. Enhanced software infrastructures for distributed computing worked in strengthening the thought that made researchers suffice DL methodologies. The paper emphases on the following DL procedures to traditional approaches which are performed manually for classifying medical images. The medical images are used for the study Diabetic Retinopathy(DR) and computed tomography (CT) emphysema data. Both DR and CT data diagnosis is difficult task for normal image classification methods. The initial work was carried out with basic image processing along with K-means clustering for identification of image severity levels. After determining image severity levels ANN has been applied on the data to get the basic classification result, then it is compared with the result of DNNs (Deep Neural Networks), which performed efficiently because of its multiple hidden layer features basically which increases accuracy factors, but the problem of vanishing gradient in DNNs made to consider Convolution Neural Networks (CNNs) as well for better results. The CNNs are found to be providing better outcomes when compared to other learning models aimed at classification of images. CNNs are favoured as they provide better visual processing models successfully classifying the noisy data as well. The work centres on the detection on Diabetic Retinopathy-loss in vision and recognition of computed tomography (CT) emphysema data measuring the severity levels for both cases. The paper discovers how various Machine Learning algorithms can be implemented ensuing a supervised approach, so as to get accurate results with less complexity possible.

1. Introduction

Biological learning has always proved itself superior to prevailing Machine learning techniques. But with the advancement of DL, an intimidating concept, ways have been found, where researchers and scientists have got success in training various DL neural networks so as to match the outputs as those of the biological neurons. The outcome expected is hard to achieve, as biological neurons are more complex, having complicated functions in comparison to current artificial neurons. Considering the brain, first echelon of neurons receiving information are sensitive to set of blobs and edges whereas

other brain regions maybe sensitive to other complex structures, for instance faces. This is because for generating features that are composed of large information, cannot be operated on inputs directly, thus aroused the need to transform the first features (i.e. blobs and edges) again in order to obtain other complex structures that comprise of more information so as to distinguish among classes[1].

Before the existence of DL, hierarchical feature learning provided models for the concept but the models suffered from considerable problems including vanishing gradient where gradients turn out to be very small for provided learning signal(most often mentioned as α), for mere deep layers, making architectures perform below par when compared to the trivial learning algorithms like the SVMs(Support Vector Machines) [2].DL utilized strategies that worked in overcoming problems related to vanishing gradients so as to train architectures having dozens of layers including non-linear hierarchical data structures as well. The work comprised of combining GPUs and activation functions so that improved gradient flow was offered which was ample for training models, without any complications, thus increasing the interest towards the field of DL. Approaches seem to influence several domains, including speech recognition, NLP[3] and computer vision among others, providing substantial performance advances as compared to those state-of-art techniques, as far as their respective domains are concerned. Many of successful deep-learning frameworks are made of combination of distinct layers, for instance fully connected layers, convolutional layers and recurrent layers. These are trained typically with a variation of the stochastic gradient descent algorithm in consort with several regularization techniques. With the increasing popularity of DL models, various types of DL software environments came to existence, enabling the effectual development and effective implementation of worked on techniques [1].

The DL model works in three basic steps which involves: taking some data as input, training a model based on input data, and using trained model for making predictions concerned with the new data. The procedure of training the model can be stated as learning process in which the model is made exposed to unfamiliar new data at each step [6]. For every step, model predicts and receives feedback concerning the accuracy of predictions generated. The feedback is used for correcting errors made while prediction. If a parameter of model is tweaked, resulting in a correct prediction, model may end up portraying the previous prediction wrong, even if it was stated correct initially. Thus, representing learning process as a back-and-forth method in parameter space. It might take several iterations for training the model, with a blend of noble predictive performance. The process continues till these predictions from the model stops to improve[4].

The study done here focuses on working with DNN (Deep Neural Networks) in addition to more advanced CNNs (Convolution Neural Networks). DNN (Deep Neural Network)[8] is an ANN(Artificial Neural Network) having several hidden layers of components between input and the output layers. DNNs are distinct in terms of their depth from earlier used shallow networks, which comprised of single hidden layer neural networks, i.e. number of the node layers under the aegis of which, information is passed in multistep process considering pattern recognition. Here, for each layer of the node is trained on different set of attributes or features relying upon preceding layer's output. Since the nodes can aggregate as well as recombine the features from prior layers, the nodes are able to recognize multifaceted features, when further dig in neural networks is made.

The concept of feature hierarchy comes in view, which here relates to hierarchy of growing complexity in addition to abstraction, making DNNs capable of managing voluminous, high-dimensional datasets having billions of constraints, making use of non-linear functions. With so much in positive, DNNs suffered from the problem of vanishing gradient, which made researchers tend towards adoption of CNNs subsequently. Convolution is basically mathematical process describing rule for mixing two functions or fragments of information. It considers, the input data (or feature map) and convolution kernel in combination so as to form a transformed feature map. Convolution is more often interpreted as filter, where kernel filters feature map for data of definite kind, for instance one kernel may work for filtering the edges, discarding other information which is irrelevant to it. The CNN hence filter inputs resulting in useful information. The convolutional layers consist of parameters

that can be learned in order to adjust these filters automatically so as to extract the relevant information.

2. Dataset and Literature Survey

Researchers have gone far when the works on DNNs and CNNs are discussed, CNNs being comparatively newer than DNNs. Various studies on DL frameworks have been going on for evaluating performance when these are employed on single-machine (for both GPU and CPU settings). The results showed that two frameworks namely, Torch and Theano, are easily extensible, where Torch outperforms any deep architecture considering CPU, closely followed by the Theano [9]. Research describes functionalities of watershed planning system, i.e. WRESTORE, where stakeholders can jointly optimize the best management conventions on to watershed. For this this work with user modeling components which utilize neural network approach, like DL [10]. Study proposes novel algorithm centered about DL-neural networks [11], which uses suitable activation functions along with regularization layers, showing expressively enhanced accuracy as compared to existing recognition methods for Arabic numerals. Letter refer to a multilevel DL architecture, targeting crop type and land cover classification from multi-temporal and multisource satellite images based on the unsupervised NN, which is used to restore missing data because of shadows and clouds, in addition to optical imagery segmentation [12].

A fast and entirely parameterized, GPU implementation of DNNs is been proposed which does not need careful designing of the pre-wired character extractors. Their work progressed by combining numerous DNNs which are trained on distinct preprocessed data in a MCDNN (Multi-Column DNN) for further boosting of recognition performance, for the [13] German traffic signs recognition. A proposed SHL-MDNN (Shared Hidden Layer Multilingual DNN), making hidden layers common across several languages while making softmax layers as language dependent. The work also showed that the transfer of learned hidden layers shared across languages may be carried out for improving recognition accuracy of the new languages, along with reductions in error. A developed a technique that adapts provided model in a conservative manner has been proposed. They considered KLD (Kullback Leibler Divergence) regularization and showed that applying mentioned regularization is equal to that of changing of target distribution, while considering conventional backpropagation procedure. Alow-rank matrix factorization [16] of concluding weight layer and relates low-rank method into DNNs for equal language and acoustic modeling. A robust technique for submerged body recognition grounded on CNNs, thus utilizing one of the DL approaches [17].

Evaluation and exploration of various CNN frameworks using the CAD (Computer Aided Detection) problems has been done, such as ILD (Interstitial Lung Disease) classification and thoraco-abdominal LN (Lymph Node) detection [18]. The work can be further extended to designing high performing CAD systems, for various medical imaging related tasks. ADL based classification approach, which combined CNNs(CNNs) and ELM (Extreme Learning Machine) for improving classification performance [19] where ELM classifier along with CNN-learned features is used as opposed to fully connected layers of the CNN for obtaining desirable results [19]. A CNN model and evaluated their system on [20] English portion of CoNLL 2012 and demonstrated that their proposed scheme achieved enhanced performance above state-of-art approaches.

Loss of lung tissue is named as Emphysema. It is categorized by as a key component of COPD, and classification of accurate emphysematous and healthy lung tissue is valuable for a further detailed analysis of the disease. Using texture analysis on CT image the objective of characterize the emphysema can processed. Supervised learning is being reliable for this texture based images; levels are given in the dataset for training & testing purpose. This domain of research grew some consideration in recent years. The dataset got is only having 168 images of binary converted CT images with textures in 16bit TIFF format having resolution of 512 x 512 pixels [21] [22].

Another dataset has been taken for concern is the Diabetic Retinopathy images. Dataset contains dissimilar resolution images over 2000×3000 pixels. The dataset contains 5 types of diabetic retinal

images. The images are being taken from FUNDUS camera. From the dataset, the model will extract the features like Hemorrhages, Hard Exudates, Cotton Wool-Spots, and Abnormal Blood Vessels by identifying this features the model will predict the level of severity level of 5 types.

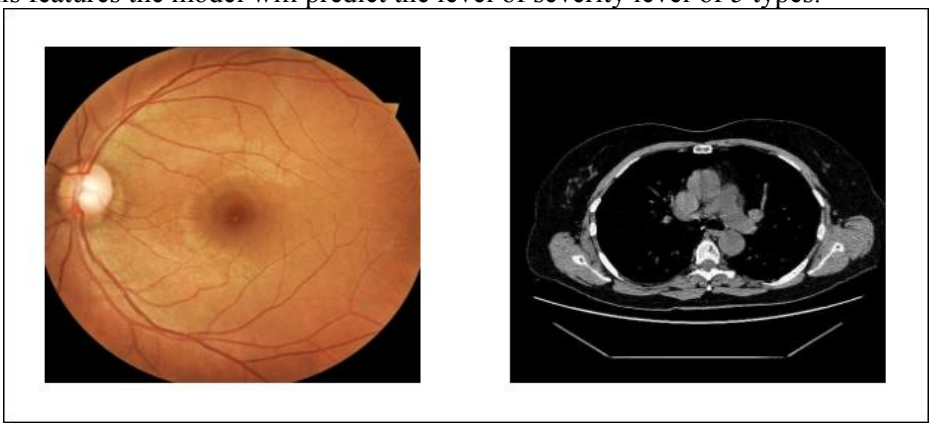

Figure 1. Dataset (Left DR & Right CT)

3. Methodology

3.1. Processing of Data

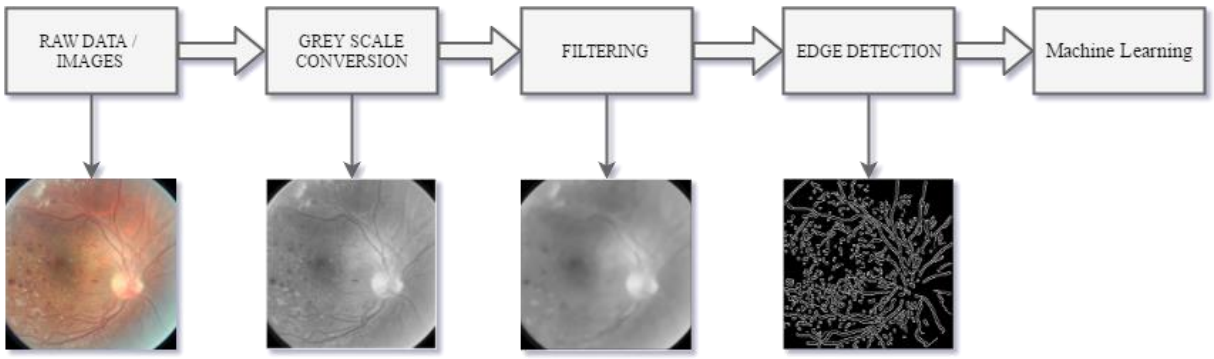

Figure 2. Image processing flow chart

The images or raw data we have extracted or collected contains noise like blur, high contrast. With this noise, it is difficult to extract desired features from the images, which may result in miss-classification of data during training, validation & testing. In order to extract proper features image processing is necessary for features detection or features extraction. The above-mentioned process is adapted for the image processing in this project. Where the raw data is first converted into grey scale i.e., converting RGB (3-bands) image into grey-scale (1-band) image. On this gravy-scaled image median filtered is applied to remove noise, so that all the pixels are normalized to the data around the pixel intensity values. From the median filtered image, we can extract features such as white lesion's and few thin veins. When we subtract the median filtered image from the grey scaled image we can get the exudates. which plays the key role in the feature extraction. Edge detection is applied on this difference data to highlight the features. Then this edge detected image is taken as input for the machine learning algorithms as the input data to proper classification.

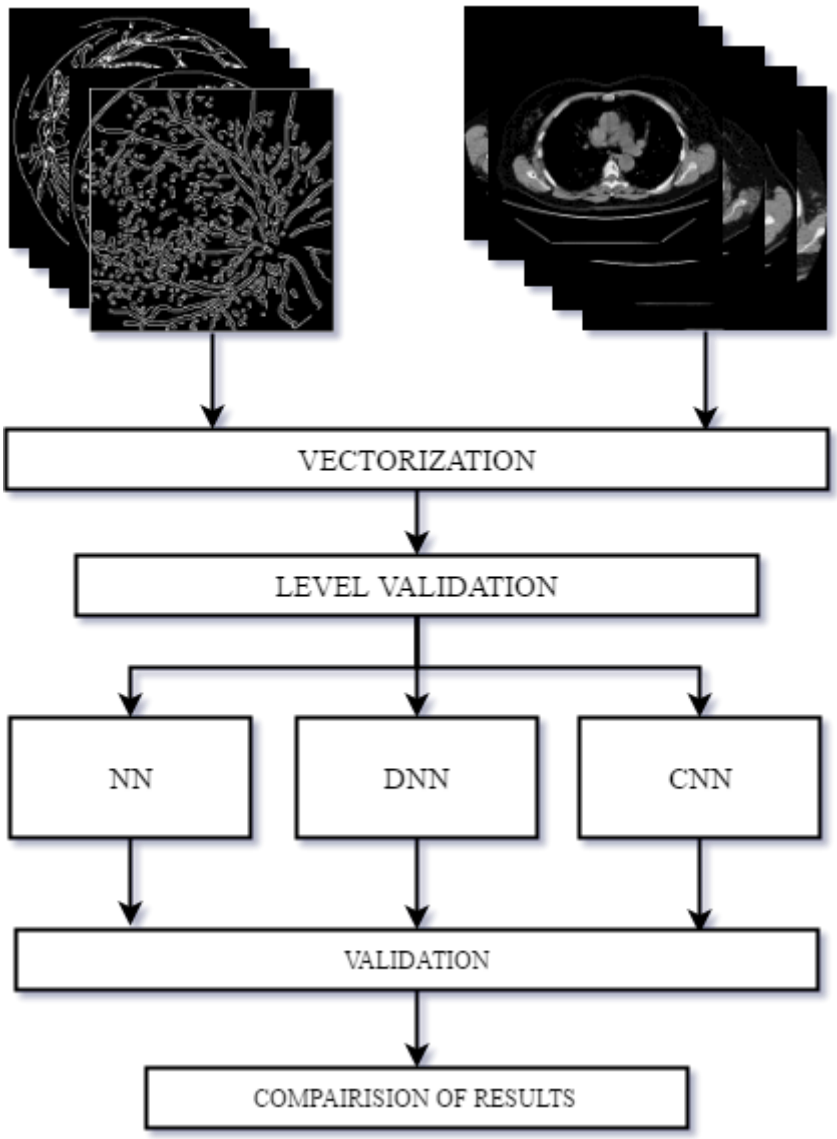

Figure 3. Flow Diagram

The preprocessed image might lose some of the features, K-means clustering method has been used on the processed images to get a validation levels through cluster analysis. The clusters then compared with the original training levels to for checking purpose of levels. The clustering has been carried out with Euclidian distance method with "MacQeens" algorithm with 100 iterations. k-means is one of the least difficult unsupervised learning calculations that take care of the known clustering issue. The idea is to apply a simple clustering algorithm and Kmeans is preferably easier to understand. The algorithm takes X = {x1, x2, x3,……., xn} with having V = {v1,v2,…….,vc} as cluster centers. Number of centers selected as 5, as 5 types of levels are having in the dataset. Before applying K-means all images are being vectorized, Vectorizing is a technique of getting all 300×300 images in a single vector. The equation of calculating new centers (v_i) as,

$$v_i = {}^{1}/_{(c_i)} \sum_{j=1}^{c_i} x_i \quad (1)$$

Where, 'ci' signifies the amount of data points in i^{th} cluster. After finding cluster centers (v_i) the distance among each data point re-calculated and new centers will be acquired.

The complete validation process of training labels will need to train the NN model. The training process as per the proposed model has been carried out on three NN models. First model is normal feed forward NN model with perceptron learning.The fundamental structure of Neural Network is a Directed Acyclic Graph (DAG) model, where,G = (N,E) such that N is set of the nodes and E is set of the edges present in a neural network. A FNN (Feed Forward NN) is the DAG, in which the neurons act as nodes and have to fulfil some

conditions: (i) If the input N has no ancestor in graph then it's considered as input neuron having single input space, given ln is set of input neurons. (ii) If the input N has p antecedents then it should have absolutely p spaces of input, one space for each of the predecessors in graph. Hence, dimension of the k^{th} space of input equals dimension of output of the kth predecessor [1]. The input of FNN is concatenation of $(x_1, x_2, \ldots, x_{|ln|})$ of input of provided neurons and likewise, the output is concatenation of outputs of output neurons which don't have successor in observed graph. Similarly, 'w', (i.e. weight vector) of whole network is concatenation (i.e. $w_1, w_2, \ldots, w_{|N|}$), of provided weight vectors of every neuron in N. Used network of FNN have 150 epochs with learning rate of 0.2, the activation function used is the Sigmoidal function, $s(t) = \frac{1}{1+e^{-t}}$

The idea of using DNN model is to train the images with better network, DNN consist of multiple hidden layer that gives privilege to the network to train the model with higher accuracy considering more feature extraction. The model has three "Fully Connected" layer followed with two "Activation Layer" and 1 "Softmax Layer". Activation function here used is "ReLU" because of its performance factor then other activation function like "tanh", "sigmoidal". ReLU is a very popular function used in recent times for Computer Vision. The function can work very well with sparse as well dense networks. RELU outperforms it reduces the likelihood of vanishing gradient and sparsity, as Sparse representations are more advantageous than dense representations (As obtained from Sigmodial functions)[5]. ReLU function given below,

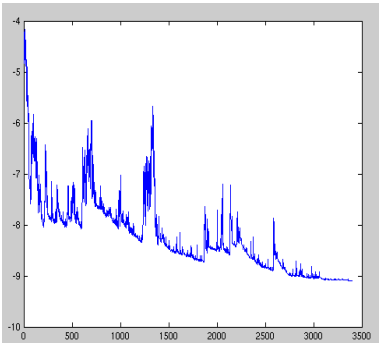

Figure 4. Stochastic Gradient Descent Fluctuation

$$f(x) = \sum_{1=1}^{\infty} \sigma(x - i + 0.5) \approx \log(1 + e^{x}) \quad (2)$$

Further the Stochastic Gradient Descent is used as it performs a parameter update for each training, converging to a potentially improved local minima. SGD performs frequent updates with a high variance that cause the objective function to fluctuate heavily. Final Softmax layer used for normalize the weights calculated by the network. The Softmax logistic function adjusts networks output weight into 0 to 1 range, function given by,

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \; for\; j = 1, \ldots . k \quad (3)$$

The influential extreme values are being normalize by the function. Also useful to identify if the is any outlaier existed in the image data or not.

Whereas for CNN, two pre-trained model LeNet [7] and VGGnet [24] network have been used. LeNet is a simple network with two convolutional layer followed with two pooling layer and three Activation layer, two fully connected layer one Flatten function and one Softmax layer. The ReLU activation function is being used here also as in DNN. On the other hand for VGGnet also ReLU is used as Activation function but the network is more dense then LeNet. The VGGnet used here consist of five Convolutional layer followed with five Pooling layer and two Fully Connected layer, in VGGnet there is no normalization (Softmax) layer is being used. For both type of medical images same methodology is being followed. Both the type of images has been resized as the CNN models suggested (300×300 for LeNet, 224*224 for VGGnet). For DR images 1000 images have been used for training and 300 images have been used for testing and for CT images out of 168 images 120 images are used for training and 48 images used for testing. This classification method is further converted into retrieval system, the work presented in [27] as generic framework for ontology-based information retrieval and image retrieval in web data will help.

4. Results and Discussion

All the model has been tested on CPU support, there is no GPU training being used. The computational time for both dataset is being compared and FFNN takes highest time for training and DNN take Low time compared to CNN models. The training accuracy for all the models given in Table 1.

Table 1. Training Accuracy of NN Models in percentages:

Image Type	FFNN	DNN	CNN(LeNet)	CNN(VGGnet)
DR	62.7	89.6	72.7	76.4
CT	42	86.3	71.7	78.3

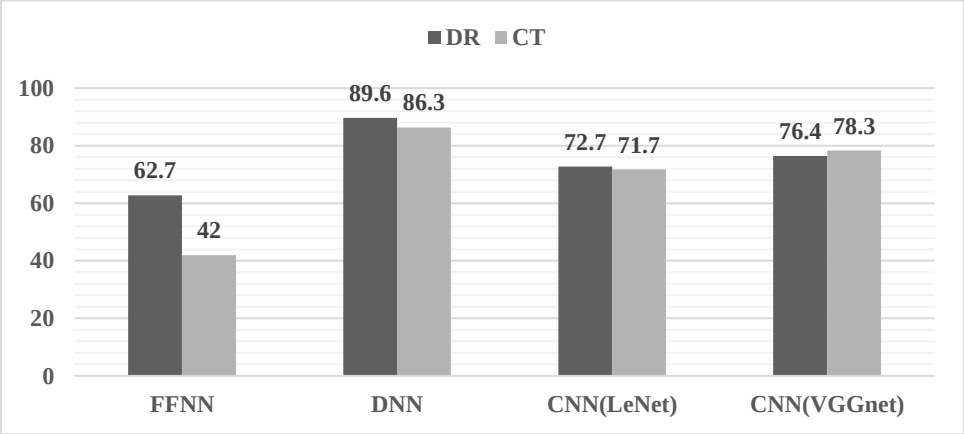

Figure 5. Training accuracy

The lack of GPU support clearly made an impact on the training accuracy of NN models. As per all the models concerns DNN out performs all other network. The CPU training degrades the performance of CNN [25] for both LeNet and VGGnet, also CPU training consumes lots of time and increases the RAM utilization because of activation functions [26].

For training models 1000 training image of DR and 120 images of CT have been taken and their training accuracy given in Table 1. For DR images training labels are compared with the K-means result of image vectors and as for image processing the accuracy of image label accuracy deviated about 20% which is 5% higher than of normal vectorized images without any processing. For CT images there were no image processing applied. Both the images have five classes of severity levels. The models have been trained based on five levels.

For testing purpose 300 DR image and 48 CT images have been selected. The accuracy due to lower training accuracy of both FFNN and CNNs accuracy was very less for them. Table 2 shows the accuracy of testing images. Through RMSE calculation validation of testing images have been calculated. From confusion matrix a pattern of identification the image levels has been noticed, for both DR and CT images the normal and extreme sever groups of images are classified with higher accuracy rate for DNN as well as for two CNN models also.

Table 2. Accuracy after RMSE calculation:

Image Type	FFNN	DNN	CNN(LeNet)	CNN(VGGnet)
DR	34.7	75.7	68	72.4
CT	28.2	77.4	64.8	69.2

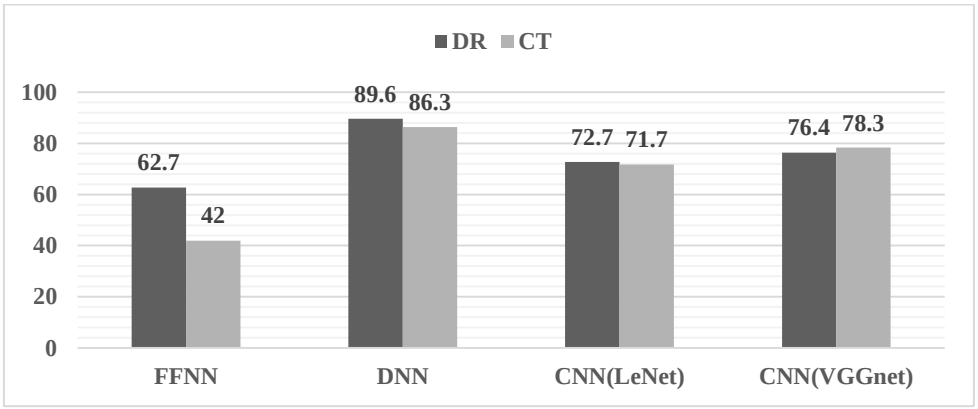

Figure 6. Testing Accuracy

5. Conclusion

The idea of this study was to compare the performance of NN models for medical image processing. For training the networks there was no GPU support which makes a major impact on the training as well as for testing and validation. The CPU training also time consuming. For validation of the training labels with Clustering improves the training accuracy. The images were used are of one band only that also makes a difference in result. In multiband images both DNN and CNN can identify the pixel intensities more accurately. As per the study suggest with the CPU training DNN performs better than other Networks. For future work the network should train with GPU support for better accuracy. The training image size should increase more for better prediction. Instead of using prebuilt models of CNN, a new CNN network design will be more appreciable.

References

[1] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. “Deep learning”. MIT Press(2016)
[2] Tang, Yichuan. “Deep learning using linear support vector machines.” arXiv preprint arXiv: 1306.0239 (2013).
[3] Li, Jiwei, et al. “Visualizing and understanding neural models in nlp” arXiv preprint arXiv: 1506. 01066(2015).
[4] Bengio, Yoshua. “Learning deep architectures for AI.” Foundations and trends® in Machine Learning 2.1 (2009): 1-127.
[5] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning” Nature 521.7553 (2015): 436-444.
[6] Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. “Deep Sparse Rectifier Neural Networks.” Aistats. Vol. 15. No. 106. 2011.
[7] LeCun, Yann. "LeNet-5, convolutional neural networks." URL: http://yann. lecun. com/exdb/lenet (2015).
[8] Schmidhuber, Jürgen. “Deep learning in neural networks: An overview.” Neural networks 61 (2015): 85-117.
[9] Bahrampour, Soheil, et al. “Comparative study of deep learning software frameworks.” arXiv preprint arXiv:1511.06435 (2015).
[10] Hoblitzell, Andrew, Meghna Babbar-Sebens, and Snehasis Mukhopadhyay. “Fuzzy and deep learning approaches for user modeling in wetland design.” Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on. IEEE, 2016.
[11] Ashiquzzaman, Akm, and Abdul Kawsar Tushar. “Handwritten Arabic numeral recognition using deep learning neural networks.” Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on. IEEE, 2017.
[12] Kussul, Nataliia, et al. “Deep Learning Classification of Land Cover and Crop Types Using

Remote Sensing Data." IEEE Geoscience and Remote Sensing Letters (2017).
[13] CireşAn, Dan, et al. "Multi-column deep neural network for traffic sign classification." Neural Networks 32 (2012): 333-338.
[14] Huang, Jui-Ting, et al. "Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers" Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.
[15] Yu, Dong, et al. "KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition" Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.
[16] Sainath, Tara N., et al. "Low-rank matrix factorization for deep neural network training with high-dimensional output targets." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.
[17] Lee, Sejin. "Deep learning of submerged body images from 2D sonar sensor based on convolutional neural network" Underwater Technology (UT), 2017 IEEE. IEEE, 2017.
[18] Shin, Hoo-Chang, et al. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning" IEEE transactions on medical imaging 35.5 (2016): 1285-1298.
[19] Weng, Qian, et al. "Land-Use Classification via Extreme Learning Classifier Based on Deep Convolutional Features" IEEE Geoscience and Remote Sensing Letters (2017)
[20] Wu, Jheng-Long, and Wei-Yun Ma. "A Deep Learning Framework for Coreference Resolution Based on Convolutional Neural Network" Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.
[21] L. Sørensen, S. B. Shaker, and M. de Bruijne, "Quantitative Analysis of Pulmonary Emphysema using Local Binary Patterns", IEEE Transactions on Medical Imaging 29(2): 559-569, 2010.
[22] S. B. Shaker, K. A. von Wachenfeldt, S. Larsson, I. Mile, S. Persdotter, M. Dahlbäck, P. Broberg, B. Stoel, K. S. Bach, M. Hestad, T. E. Fehniger, and A. Dirksen, "Identification of patients with chronic obstructive pulmonary disease (COPD) by measurement of plasma biomarkers", The Clinical Respiratory Journal 2(1): 17–25, 2008.
[23] MacQueen, James. "Some methods for classification and analysis of multivariate observations" Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1. No. 14. 1967.
[24] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
[25] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[26] Chellapilla, Kumar, Sidd Puri, and Patrice Simard. "High performance convolutional neural networks for document processing." Tenth International Workshop on Frontiers in Handwriting Recognition. Suvisoft, 2006.
[27] Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. *Human-centric Computing and Information Sciences*, *6*(1), 18.