Scientific Paper

Multi spectral classification and recognition of breast cancer and pneumonia

Aditya KAKDE¹,², Nitin ARORA², Durgansh SHARMA³, Subhash Chander SHARMA⁴

¹ ² School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India
³ ⁴ Electronics & Computer Discipline, Indian Institute of Technology, Roorkee, India

E-mail address: adityakakde100@gmail.com

(received 28 June 2019; revised 4 September 2019; accepted 15 October 2019)

Abstract

According to the Google I/O 2018 key notes, in future artificial intelligence, which also includes machine learning and deep learning, will mostly evolve in healthcare domain. As there are lots of subdomains which come under the category of healthcare domain, the proposed paper concentrates on one such domain, that is breast cancer and pneumonia. Today, just classifying the diseases is not enough. The system should also be able to classify a particular patient’s disease. Thus, this paper shines the light on the importance of multi spectral classification which means the collection of several monochrome images of the same scene. It can be proved to be an important process in the healthcare areas to know if a patient is suffering from a specific disease or not. The convolutional layer followed by the pooling layer is used for the feature extraction process and for the classification process; fully connected layers followed by the regression layer are used.

Key words: convolutional layer; pooling layer; multi spectral classification; breast cancer; pneumonia.

Introduction

Pneumonia is a condition in which a fluid or pus gets filled in the air sacs (Figure 1). It mainly targets the people who are infants or are 65 and above years old. The symptoms related to pneumonia are chest pain during breathing, fever, sweating, fatigue, shortness of breath and many more. Various types of organisms like viruses and bacteria can cause pneumonia. The main cause of pneumonia is bacteria which can be generated if one has a cold or flu. The second is the bacteria-like organisms which is not that much severe. The third is the fungi which weaken the immune system and mainly target those who are suffering from chronic health problems. Last is the virus which targets those people who come under the category of infants. A serious case of pneumonia can also lead to life-threatening situations.

Similar to the above dangerous disease, there is also one more life-threatening disease known as breast cancer. It comes under the domain of cancer, which on later stages becomes incurable and can lead to death. It is the condition in which the cancer cells arise in the milk duct and gets confined there itself. It can occur in both men and women but is mostly seen in women. These cancer cells are called tumour cells which can even invade the other parts of the breast. The early stage of breast cancer is called Ductal Carcinoma in situ (DCIS) [2,3] where 40% cases are found where tumour cells have invaded the other parts of the breast (Figure 2). These cells divide very rapidly than the healthy cells and get accumulated in the form of a lump or mass.

Thus, it is very important to predict if a person is having cancer or pneumonia. It can be a challenging task to map the features in the diseased data, thus convolutional neural network will be perfectly suited to accomplish this task. This network was proposed by Yann LeCun et al [4]. The convolutional neural network was actually inspired by the mammal’s visual cortex. The convolutional layer sees the image by scanning one complete image through the receptive field. Through the receptive field, the convolutional layer extracts the most important features from the image. It also uses an activation function which helps to decide and forward the most sensible data. Thus ELU, also called an exponential linear unit, is used in this case. Now this activation function is subjected to the pooling layer. The pooling layer helps to downsample the images which we get from the output of convolutional layer to get most important features. It is followed by the classification layers which help to classify the images. As convolutional layer can only extract important features and can’t tell about that image, thus classification layer plays an important role in this network. These classification layers only consist of the neurons which are joined with each other via many-to-many connections.
Figure 1. How fluid gets stored in the sacs under pneumonia conditions. To avoid the attack of bacteria, white blood cells gets accumulated with these bacteria, the air sacs within the lungs are filled and thus causes a problem in breathing. Source: [1].

Figure 2. How tumour cells which are accumulated in the mild duct and causes Breast Cancer. These tumour cells migrate to the other parts of the breast and results in invasion such as in regions containing stromal cells. Thus this invaded part is known as Invasion Ductal Carcinoma (IDC) and non-invaded part as Ductal Carcinoma in situ (DCIS). Source: [2].

Literature review

Yann LeCun et al. [4] presented a paper on Convolutional Networks for Images, Speech, and Time-Series where they proposed that there is no need for handcrafted feature extraction method if Convolutional Neural Network is used for image recognition. But there is no need to normalize the image for size and orientation. It is also specified that fully invariant recognition is still beyond reach.

Aditya Kakde et al. [5] presented a paper on the Novel Approach towards Optimal Classification using Multilayer Perceptron where it was proposed of using a particular number of hidden layers with best possible activation function. Thus Multilayer Perceptron having three hidden layers with ELU activation function has given the best average results that are 0.014652 and 0.990671 when compared to 0 and 1. But when the alternate values are found smaller and greater, thus further investigation was done where it was found that the loss decreased much faster for MLP with three hidden layers with ELU activation function that is 0.00429 at 600 epochs. Also, 0.5 learning rate is mentioned which tends to be best suited for the classification task on XOR operation.

Djork-Arné Clevert et al. [6] presented a paper on Fast and Accurate Deep Network Learning by Exponential Linear Unit (ELUs) where they proposed a new activation function called as ELU which is capable to take negative inputs. When tested on MNIST dataset, it yields the test error of ±0.24%. Also when compared with other networks like Alexnet, DSN, NiN, Maxout, All-CNN, Highway Network, and Fract. Max-Pooling on CIFAR-10 and CIFAR-100 dataset, the proposed network gave the least test error of 27.62% on CIFAR-100. For CIFAR-10, Fract. Max-Pooling achieved a lower error rate of 4.50%.

Alex Krizhevsky et al. [7] presented a paper on ImageNet Classification with Deep Convolutional Neural Network in which they proposed a novel architecture called Alexnet. When tested on ILSVRC-2010 test set, achieved the least Top1 error of 37.5% and Top5 error of 17.0%. When tested on ILSVRC-2012 test set, CNN with seven layers achieved the least Top1 error of 36.7% and Top5 error of 15.3%. Also, it was mentioned that removing even one CNN layer can degrade the architecture and results in a 2% loss.

Hassan Ramchoun et al. [8] presented a paper on Multilayer Perceptron: Architecture Optimization and Training where they proposed an optimal method for the classification task. Sigmoid Activation function is used in the architecture which is generated by Genetic Algorithm and when tested on Iris Dataset, train accuracy increased from 96% to 98.7% whereas test accuracy remained the same that is 97.3% compared to the previous method.

Jiuxiang Gu et al. [9] presented a paper on Recent Advances in Convolutional Neural Networks in which they have provided a broad survey on improvements in a convolutional neural network in the past few years which describes pooling layer, loss, weight initializations, regularizations, optimizations, and gradient descent. Further stress was given on describing fast processing of CNNs and application of CNNs.

Dabal Pedamonti [10] presented a paper on Comparison of non-linear activation functions for Deep Neural Networks on MNIST classification task in which MNIST dataset was tested on with different learning rates that are 0.01, 0.05, 0.1 and 0.2 and then four parameters that are train loss, test loss, train accuracy, and test accuracy are compared on the basis of activation functions which are Sigmoid, ReLU, LReLU, ELU and SELU at 100 epochs. It was seen that the accuracy increased from 2% to 4% when the learning rate is from 0.05 to 0.1. A further test was conducted by initializing different weights on different numbers of hidden layers. With all four weights, Glorot uniform distribution has given the best accuracy of 0.978. It was also compared with Gaussian distribution using SELU which achieved good performance over Glorot uniform distribution.
Andreas Kamilaries et al. [11] presented a paper on Deep Learning in Agriculture: A Survey where a survey on 40 research papers was presented in which the latest techniques of CNN are discussed. Further application of Deep Learning in Agriculture was discussed which consists of Areas of Use, Data Sources, Data Variations, Data Pre-processing, Data-Augmentation, Technical Details, Outputs, Performance Metrics, Overall Performance, Generalizations on different datasets and Performance Comparison with other approaches. Also, advantages and disadvantages of Deep Learning and Future of Deep Learning in Agriculture were discussed.

Rasool Fakoor et al. [12] presented a paper on Using deep learning to enhance cancer diagnosis and classification in which an unsupervised approach for the classification task was proposed. Thirteen different datasets of Breast Cancer are used that are AML [13], Adenocarcinoma [14], Breast Cancer [15], Leukemia [16], Leukemia [17], AML [18], Breast Cancer [19], Seminoma [20], Ovarian Cancer [21], Colon Cancer [22], Medulloblastoma [23], Prostate Cancer [24] and Leukemia [25]. Four different types of unsupervised learning methods were used which are Sparse Auto-encoders, Stacked Auto-encoders, Stacked Auto-encoders with Fine Tuning and PCA+Softmax/SVM. For datasets 3rd, 5th, 6th, 7th and 12th, Sparse Auto-encoders achieved the highest performance. For datasets 1st, 4th, 8th, 10th, 11th and 13th, Stacked Auto-encoders with Fine Tuning achieved the highest performance. For datasets 2nd and 9th, PCA+Softmax/SVM achieved the highest performance.

Dursun Delen et al. [26] presented a paper on Predicting breast cancer survivability: a comparison of three data mining methods in which three different methods were used for the classification task that are Artificial Neural Network, Decision Tree and Logistic Regression. The results were achieved using 10 fold cross-validation and are compared on the basis of accuracy. Decision Tree has achieved the highest accuracy of 0.9362 when compared with Artificial Neural Network and Logistic Regression.

Methodology

About dataset

The datasets which are used for the training of the network are Breast Cancer [27] and Pneumonia [28]. The first dataset is of Breast Cancer consists of two classes that are Cancer and No Cancer (Figure 3). The class which consists cancer is an IDC form. IDC (Invasive Ductal Carcinoma) is the most common type of Breast Cancer. But as we are dealing with Multi Spectral Classification, both the classes consist of a particular patient’s ID. This paper took into consideration 662 training images, 10 validation images for fine-tuning and 10 test images.

The second dataset is on Pneumonia which also consists of two classes’ namely normal chest and Pneumonia chest (Figure 4). With this dataset also we are dealing with Multi Spectral Classification. So, this dataset also consists of a particular patient’s ID. This paper took into consideration 5206 training images, 10 validation images for fine-tuning and 10 test images.

Exponential Linear Unit

Exponential Linear Unit, also called an ELU activation function, was proposed by Djork-Arné Clevert et al. [3] and is given as:

$$f(y) = \begin{cases} 
\alpha(e^x - 1), & x < 0 \\
y, & x \geq 0 
\end{cases}$$

Eq. 1

The main aim behind this activation function was to take negative inputs and thus its range lies from -1 to 0 because unlike ReLU which makes the gradient 0 during the intake of negative inputs, ELU does not give this kind of results. So, ELU does not result in dead neurons. When talking of positive inputs, the mean for ELU activation function is closer to zero when compared with the other activation functions like Sigmoid, ReLU, Tanh, etc. Therefore the activation which has less positive mean value results in more learning.
Overlapping Pooling

Pooling is a layer that comes after the convolutional layer and is used for the downsampling purpose. For the output of the convolutional layer, it extracts the most important features and organises it in a matrix form. Due to pooling layers only, very large images are able to get processed for the classification or detection tasks. Pooling helps in extracting the most important features and then downsamples it which is then propagated towards the fully connected layers. Otherwise, there can be high amount of overfitting and thus can result in very poor performance. When the filter size suppose $z$ is greater than the filter size of pooling $s$, then we can call it as overlapping pooling \[4\]. It means when $z > s$ ($3 > 2$). Overlapping pooling helps in the reduction of a small amount of overfitting which is caused during the process of feature extraction.

Dropout

Dropout is a technique that reduces the occurrence of overfitting. Overfitting means when the noisy data gets accumulated with the necessary features during training time. Dropout helps to reduce it by dropping one neuron at a time, so that the other neuron is forced to learn the features. Suppose that a network has two neurons in the hidden layer and one at the output layer. When the input is feed, the first neurons learn the features and give the output with the accuracy of 80% whereas the 2nd neuron gives some random accuracy. This is due to the accumulation of noisy data with the necessary data. When dropout is applied with the network, then at each iteration one neuron is dropped so that the other is forced to learn the features. Due to this, maximum chances are there that both neurons will produce an accuracy of 80%. This makes the performance of the network much better than the previous method. Dropout is used with the fully connected layer, not with the feature extraction layers. The process of dropout is governed by two steps where first one is dropping a unit during training and second is the scaling of output which can be matched between training and testing.

With thin layer as $m$ and network without dropout as $\mathcal{M}$ can be given as:

\[
E[y_j] = \sum_{M \in \mathcal{M}} \Pr(M)y_j^M
\]

Eq. 2

Where $m$ signifies thinned network and $M$ signifies output of the thinned network.

Assuming $\Pr(M) = 1/|m|$ and $y^M = y^\mathcal{M}$. Thus we get the output as:

\[
E[y_j] = p|m| \frac{1}{|m'|} y_j^M = py_j^\mathcal{M}
\]

Eq. 3

$|m'| = p|m|$ during the activation of $j$ where $m$ denotes the set of the network. To fulfil the outputs to be matched between training and testing is to first scale down the weights $w_{ji} = pw_{ji}$ and to scale the output training time that is $y_j' = \frac{1}{p} y_j$.

Experimental setup

This phase shows the Experimental setup of the system (Figure 5) and also gives a brief description of its working.

The system has the ability to accurately classify the input images. The system is designed in such a way that, if a patient does not have his/her account, then with a particular name an account can be generated. It is followed by checking the ID of the patient. If a patient is already having an ID, so his/her account will also be there as account name follows the order of ID_account-name. If ID does not exist, only then the account of a patient will be created. These accounts are created so that the patient’s test data can be stored. The system has the ability to fetch the patient’s data in the account means fetching of test data in real-time. The system is trained with 662 image data of Breast Cancer and 5216 image data of Pneumonia. Now getting the patient’s data in real-time, the system consists of both test data and train data. It is then able to accurately classify if a patient is suffering from that disease or not. The system is also enhanced with Google Text to Speech recognition to make it more users interactive. In the proposed network, the layer uses two stacked convolutional layers at the starting to make the receptive field strong. With the fully connected layer, Dropout is used to reduce the problem of overfitting during the classification process. At last, the regression layer is used to train left out neurons.

![Figure 5. The figure shows the experimental setup of the system. Here HL stands for Hidden Layer, Drop stands for Dropout and Reg stands for Regression Layer. Also the directory means the account of a user. From left to right signifies that the input images are processed from left convolutional layer to right means up to the regression layer and after that, the predicted or recognized image is generated.](image-url)
About architecture

The architecture is designed in such a way that the network does not become too dense. As the network becomes denser, it takes more time to get trained, especially when training in those systems which do not have GPU support. At the initial stage, two convolutional layers are used which are stacked one after another to make the receptive field more strong. Due to this, learning at the starting stage becomes better. After that alternate arrangement of convolutional layers and pooling layers is done where overall four convolutional layers and three pooling layers are used. For the classification task, three fully connected layers are used with dropout function to reduce the problem of overfitting. At last, regression layer is used to train left out neurons with Adam optimizer and mean square loss at 0.1 learning rate.

Results

This result section shows the data which was acquired during the training and testing of the data. To optimize the training of the network, Adam optimizer is used with a learning rate of 0.1. This paper took average of four parameters that are train loss, test loss, train accuracy and test accuracy up to 40 iterations.

This paper has divided these two datasets to identify if the proposed network can give better test results not only when used large number of dataset but also on a small number of datasets during training.

The proposed network was trained on two different datasets which are of Breast Cancer and Pneumonia. A huge difference in the number of datasets between these two was kept that are 662 images of breast cancer and 5206 of pneumonia for training and 10 images for testing to check the robustness of the network.

| Dataset         | Train Loss | Train Accuracy | Test Loss | Test Accuracy |
|-----------------|------------|----------------|-----------|---------------|
| Breast Cancer   | 0.0961     | 0.9038         | 0.3075    | 0.6925        |
| Pneumonia       | 0.1297     | 0.8693         | 0.1321    | 0.8450        |

The below graphs (Figures 6-9) show how the loss and training parameters vary during iterations.

Figure 6. Training loss on Breast Cancer dataset and Pneumonia dataset where x-axis signifies epochs (number of iterations) and y-axis signifies train loss. Blue line signifies breast cancer and orange line signifies pneumonia.
Figure 7. Training accuracy on Breast Cancer dataset and Pneumonia dataset where x-axis signifies epochs (number of iterations) and y-axis signifies train accuracy. Blue line signifies breast cancer and orange line signifies pneumonia.

Figure 8. Test loss on Breast Cancer dataset and Pneumonia dataset where x-axis signifies epochs (number of iterations) and y-axis signifies test loss. Blue line signifies breast cancer and orange line signifies pneumonia.
Figure 9. Test accuracy on Breast Cancer dataset and Pneumonia dataset where x-axis signifies epochs (number of iterations) and y-axis signifies test accuracy. Blue line signifies breast cancer and orange line signifies pneumonia.

Figure 11. This output is generated after the completion of the classification process for Breast Cancer dataset. It can be seen that the network has classified and recognised with Multi Spectral data. This (-) IDC shows that it is not an IDC type Breast Cancer.

Figure 12. This output is generated after the completion of the classification process for Pneumonia dataset. It can be seen that the network has classified and recognised with Multi Spectral data.
Conclusion
For the first dataset, our multi spectral classification follows the order of patient’s ID, if a patient has cancer or not and its type in form of (+) IDC and (-) IDC. For second dataset, our multi spectral classification follows the order of patient’s ID and if a patient is having pneumonia or not. After reviewing the results, it can be seen that for breast cancer data, the variations during training were not much but overall results in poor test results whereas for pneumonia dataset, the test results were good despite being fact that there was a little variation during training. For IDC type Cancer recognition, 0.6925 test accuracy was achieved which consists of 662 training data and for pneumonia recognition, 0.8450 test accuracy was achieved which consists of 5206 training data (Table I). This concludes that the proposed network can give better test results when consists of a large number of dataset during training but failed to achieve better performance during testing when used less number of dataset.

References
[1] Mayo Clinic Patient Care and Health Information. Pneumonia. https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204
[2] Micalizzi DS, Maheswaran S. On the trail of invasive cells in breast cancer. Nature. 2018;554:308-309.
[3] Mayo Clinic Patient Care and Health Information. Breast Cancer. https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470
[4] LeCun Y, Bengio Y. Convolutional Networks for Images, Speech and Time-Series. in: The handbook of brain theory and neural networks. MIT Press Cambridge; 1998.
[5] Kakde A, Arora A, Sharma D. Novel Approach towards Optimal Classification using Multilayer Perceptron. Int J Res Eng IC Soc Sci. 2018;8(10):29-38.
[6] Clevert DA, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning Exponential Linear Units (ELUs). International Conference on Learning Representations; 2016.
[7] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. NIPS’12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. 2012. pp. 1097-1105.
[8] Ramchoun H, Idrissi MAJ, Ghanou Y, Ettaouil M. Multilayer Perceptron: Architecture Optimization and Training. IJIMAII. 2016;4(1):26-30.
[9] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018;77:354-377.
[10] Pedamonti D. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. ArXiv. 2018;abs/1804.02763.
[11] Kamilaris A, Prenafeta-Boldú FX. Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture. 2018;147:70-90.
[12] Fakoor F, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. ICML Workshop on the Role of Machine Learning in Transforming Healthcare; 2013.
[13] Mills KI, Kohlmann, A, Williams PM, et al. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of am1 transformation of myelodysplastic syndrome. Blood. 2009;114(5):1063-1072.
[14] Fujiwara T, Hiramatsu M, Isagawa T, et al. ASCL1-coexpression profiling but no single gene expression profiling defines lung adenocarcinomas of neuroendocrine nature with poor prognosis. Lung Cancer. 2012;75(1):119-125.
[15] Woodward WA, Krishnamurthy S, Yamauchi H, et al. Genomic and expression analysis of microdissected inflammatory breast cancer. Breast Cancer Res Treat. 2013;138(3):761-772.
[16] Klein HU, Ruckert C, Kohlmann A, et al. Quantitative comparison of microarray experiments with published leukemia related gene expression signatures. BMC Bioinformatics. 2009;10:422.
[17] Cheok MH, Yang W, Pu CH, et al. Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. Nat Genet. 2003;34(1):85-90.
[18] Yagi T, Morimoto A, Eguchi M, et al. Identification of a gene expression signature associated with pediatric AML prognosis. Blood. 2003;102(5):1849-1856.
[19] Wang Y, Klijn JGM, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005;365(9466):617-679.
[20] Gashaw I, Grümmer R, Klein-Hitpass L, et al. Gene signatures of testicular seminoma with emphasis on expression of ets variant gene 4. Cell Mol Life Sci. 2005;62(19-20):2359-2368.
[21] Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet. 2002;359(9306):572-577.

[22] Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad USA. 1999;92(12):6745-6750.

[23] Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. 2002;415(6870):436-442.

[24] Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. 2002;1(2):203-209.

[25] Verhaak RG, Wouters BJ, Erpelinck CA, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. 2009;94(1):131-134.

[26] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med. 2004;34(2):113-127.

[27] Cruz-Roa A, Basavanhally A, González F, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. Proc SPIE 9041, Medical Imaging 2014: Digital Pathology: 904103.

[28] Kermany DS, Khang K, Goldbaum MH. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. doi: 10.17632/rscbjbr9sj.2