Nat. Volatiles & Essent. Oils, 2021; 8(4): 6416-6425

NVEO
Natural Volatiles &
Essential Oils

An Efficient Cancer Detection Using Machine Learning Algorithm

M. Arun Kumar M.E. (Ph.D)[1] , P.Gopika Ram M.E. [2] , P.Suseendhar M.E. [3] , R Sangeetha M.E[4]

[1]Assistant Professor,Department of Biomedical Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu

[2]Assistant Professor,Department of Biomedical Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu

[3]Assistant Professor,Department of Electronics and Communication Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu

[4]Research Scholar,Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu

E-mail:arunnkumarrm@gmail.com

Abstract. Machine learning algorithms based automated systems are gaining attention for creating new device applications in the field of artificial intelligence. Especially in health care systems, implementation of emerging machine learning paradigms helps in providing a potential to raise productivity, consistency, and quality of treatment. Cancer is an epidemic that has high mortality and incidence rate worldwide. Breast cancer and oral cavity cancer are most prevailing types of cancer in females and males, respectively. Biopsy is the technique to determine the ability of cancer with confidence which includes various processing steps such as grading, staging and visual inspection. The manual analysis of histopathology slides is a labour intensive task and influenced by various factors like fatigue, attention, and expertise of pathologist. However, recent developments in soft computing techniques allow to build an automated computerized diagnostic system for cancer detection. In this context, many efforts are dedicated to feature extraction step in conventional machine learning. Deep learning is the latest advancement in this direction and opens up a new horizon in the field of Machine Learning. Automatic representation of the data is the key asset of the deep learning technique but require intense training and a comprehensive well-annotated dataset for their good performance. In the view of foregoing, the present work addresses the above discussed challenges and which is freely available in public domain.

1. Key requirements

- Deep Learning
- Histopathalogical analysis
- Convolutional Neural Network

2. Introduction

In this section, a brief description on the significance of image processing techniques and machine learning algorithms is provided along with their broad usages and application, especially emphasized on cancer detection application. Additionally, the utilized feature descriptors and machine learning algorithms in conjunction with the basic working principle of deep learning and transfer learning approaches are also described. A detailed description on the architecture of employed deep learning models and performance evaluation metrics are also provided for better understanding of the work.

Images have always played a pivotal role in human life because vision is the most significant sense of human beings. Due to the fact, image processing has plenty of applications in the field of agriculture, military, security, medical, etc [1-4]. The ease in the generation of a huge amount of images owes to the advancement in the digital technologies. With such a plethora of images, image processing techniques need to deal with more complex problem and to cope with their adaptability to the problem. The diagnosis of diseases from the images is one of the examples of such complex problems.

Machine learning acts as an integral component of intelligent computer vision systems when such adaptation is required. The recent advancements in the field of 'Machine Learning' open up a new opportunity to develop intelligent computer-controlled machines and software. A large community of researchers, engineers and academicians are continuously investigating the new machine learning algorithms to employ them in designing automated systems for the applications ranging from object detection to medical diagnosis. Machine learning is a subset of 'Artificial Intelligence' which learns from the data [5]. It analyzes the existing patterns in the data to respond a situation for which they have not been explicitly programmed [6-7]. Day by day, the majority of human beings are interacting with the systems that are based on machine learning, for example, biometric systems, 'Alexa'-the voice recognition system, selfdriving cars and 'Google Suggest' or 'Auto complete'. In addition, the machine learning is aiding the designing of automated and effective systems in the field of computer vision which includes object detection, facial expression recognition, human pose estimation, crowded scene analysis and medical image analysis. In spite of noteworthy advancement in diagnosis, cancer is still a worldwide problem. Europe, Africa and America account for 23.4%, 5.8% and 21% of incidence, respectively. However, the proportion of cancer deaths in EuropeAfrica and America countries is 20.3%, 7.3%, and 14.4%, respectively (Figure 1.1). In Asia and Africa, the mortality rate is higher in comparison of the new cases due to poor prognosis, delay in diagnosis and treatment. Inability to access technical advances against the disease by poor people of developing countries is another reason behind the high mortality rate. Therefore, the development of efficient and affordable approaches for early detection, diagnosis and treatment is an urgent need to control cancer.

3. Related Works

Despite recent advances in molecular biology with the advent of genetic markers, the histopathological analysis always remains a standard method in the BC diagnosis [13-15]. In the process of histopathological analysis, pathologists visually inspect the histology slide under a microscope to determine a particular pattern that they learned from the previous slides of the disease. Hence, a visual inspection of histology slides is based on pattern recognition approach and subjective opinion of the pathologists. Analysis of histology slide is a time-consuming task and requires great care in the interpretation of the disease [10, 16]. New cases of cancer are projected to rise globally. Due to matter the fact, pathology services are continuously pressurized to provide quality analyses.

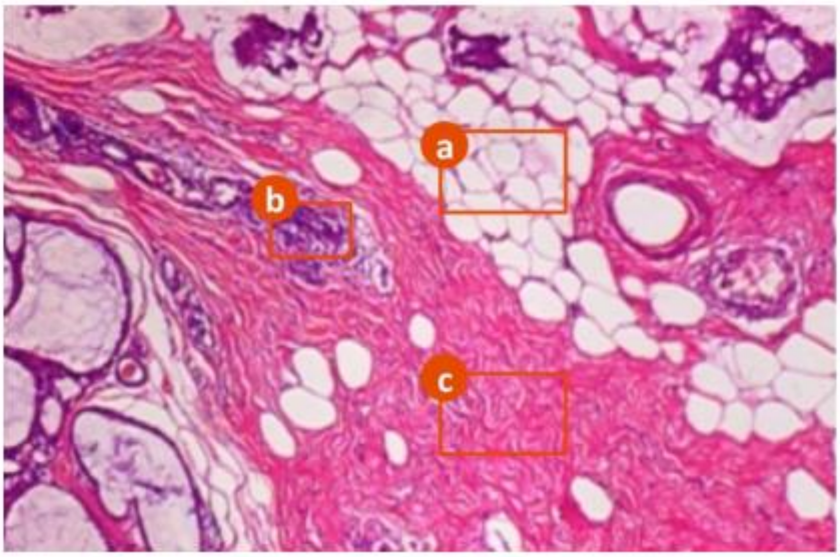

Fig. 1 Representation of different tissues: (a) fat tissue, (b) cell nuclei, and (c) collagenrichstroma

To address the challenge of a potential workforce crisis, there is a need to enrich the critical services that histopathology offers to cancer care. In recent time, digitization of tissue samples has become possible due to digital pathology and computational algorithms which includes machine learning and image processing techniques for analyzing the image [19-20]. New machine learning techniques have also been evolved and derived new innovations in healthcare by assisting in developing computer-aided detection and diagnosis systems. High performance computing system and low-cost memory devices are the fundamental factors behind the evolution of soft-computing techniques [21]. Improvement in image processing techniques and machine learning algorithms also provide potential to develop histopathological image-based computer-aided detection and diagnosis systems that can add a new value in pathology services by making it more objective, productive and coherent in diagnosis. These systems can also be used as a second opinion in pathology in order to complement thedecision of the pathologists. In near future help in maintaining the data concerning the variety of cancer types to diagnose the sufferer form healthy one.

4. Proposed Methodology

Wearehappyforauthorstosubmitsupplementarydataattachmentstoenhancetheonlineversionsofpublishedarticles.Supplementarydataenhancementstypicallyconsistofvideoclips,animationsorsupplementarydatasuchasdatafiles,tablesofextrainformationorextrafigures.

Defined training protocol is used in CNN topology. At the entire magnification levels Image distribution in dataset is not uniform, therefore before the training process balancing have been done. At each magnification level Down-sampling is applied to the classes. To determine the ability of the model in classifying the data based on the magnification testing has been performed independently

Table. 1 Description of CNN Architecture

Layer	Type	Activation	Regularization	No. of Filters	Filter Size	Pooling Size/Stride	Zero padding
Conv1	Convolutional	ReLU	-	32	3x3	-	-
Conv2	Convolutional	ReLU	-	32	3x3	-	2x2
Conv3	Convolutional	ReLU	-	64	3x3	-	2x2
Max-Pool	Pooling	-	-	-	-	2x2/2	-
FC1	Fully-Connected	ReLU	Dropout	128 nodes	1x1	-	-
FC2	Fully-Connected	ReLU	Dropout	128 nodes	1x1	-	-

Table. 2 CNN Training with Different Optimizer Performance Analysis

Optimizer	Error-function	Accuracy (%)	Iterations
Gradient Descent	Cross-Entropy	69.4	426
AdaGrad	Cross-Entropy	46.8	426
Adam	Cross-Entropy	**83.67**	**700**
RMSProp	Cross-Entropy	76.9	146

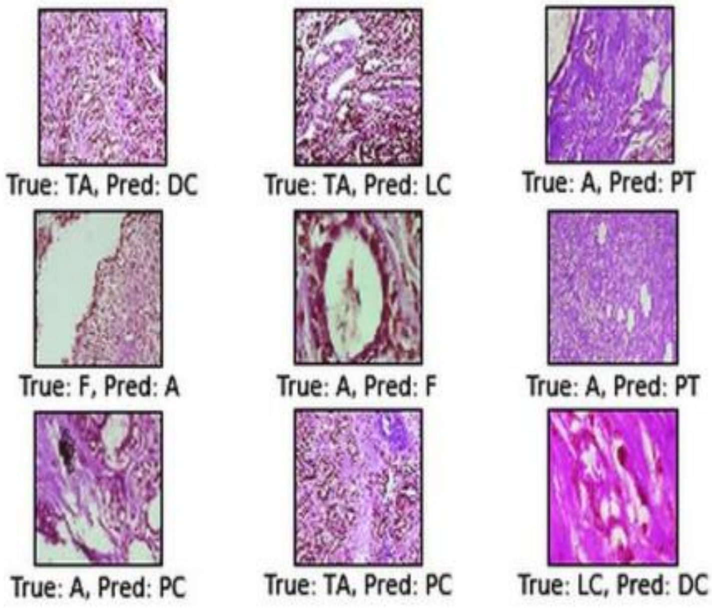

Figure 2 Illustrations of miss-classified images produced by the proposed CNN

5. Results and Discussion

4.1 Existing work Comparison

We presented the proposed model performance is used for independent magnification. The dataset size is increased in addition, data augmentation is performed, centre with the three angles: 90^o, 180^o, and 270^o by rotating the images about their [161-162].

Recognition Rate at	Layers				
	40x	**100x**	**200x**	**400x**	**Average**
Image Level	90.4±1.5	86.3±3.3	83.1±2.2	81.3±3.5	85.3

Table.3 Rate of Recognition Image at Level Image Independent Magnification

For training the increased size of the dataset is the used and it is the reason behind this enhanced performance.

Recognition Rate at	Strategy	Layers				
		40x	**100x**	**200x**	**400x**	**Average**
Image Level [85]	1	79.9±1.5	80.8±1.5	**84.0**±1.5	80.7±1.5	81.4
	2	80.6±1.5	81.0±1.5	82.7±1.5	**80.8**±1.5	81.3
	3	81.8±1.5	82.3±1.5	82.4±1.5	80.3±1.5	81.7
	4	**89.6**±1.5	**85.0**±1.5	82.8±1.5	80.2±1.5	**84.4**

Table.4 Image Level at Image Recognition Rate

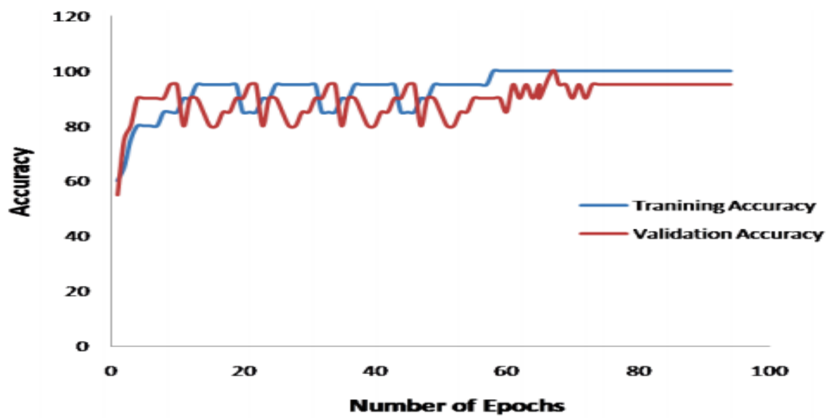

Fig. 3 Validation accuracy and Training curves at the time of training

4.2 Proposed Results

By means ROC curve the network performance for the layer wise fine tuning approach is analyzed. By using AUC and APS the performance evaluation of the network is executed.

Model	Features	Magnification Factor			
		40X	100X	200X	400X
BOW+SVM	DSIFT	18.77	17.28	20.16	17.49
	SURF	49.65	47.00	38.84	29.50
LLC+SVM with SPL(0)	DSIFT	48.46	49.44	43.97	32.60
	SURF	55.80	54.24	40.83	37.20
LLC+SVM with SPL(1)	DSIFT	47.44	44.32	44.46	32.86
	SURF	54.61	53.92	48.10	38.12
LLC+SVM with SPL(2)	DSIFT	44.54	51.68	44.30	35.54
	SURF	53.75	44.30	45.30	40.88
BOW+CNN	DSIFT	41.80	38.56	49.75	38.67
	SURF	53.07	60.80	70.00	51.01
LLC+ CNN	DSIFT	60.58	57.44	70.00	46.96
	SURF	80.37	63.84	74.54	54.70
Fractal Dimension + SVM	Fractal Dimension	55.60	---	---	---
Deep Learning	CNN + Original Data	86.34	84.00	79.83	79.74
Deep Learning	CNN + Augmented Data	83.79	84.48	80.83	81.03
Deep Learning	CNN + SVM	82.89	80.94	79.44	77.94
Deep Learning	CNN + Ensemble Model	88.23	84.64	83.31	83.98
Deep Learning	CNN Features + KNN	70.48	68.00	70.08	66.38
Deep Learning	CNN Features +	75.43	71.20	67.27	65.12

Table. 5 Existing methods using break his dataset Comparison

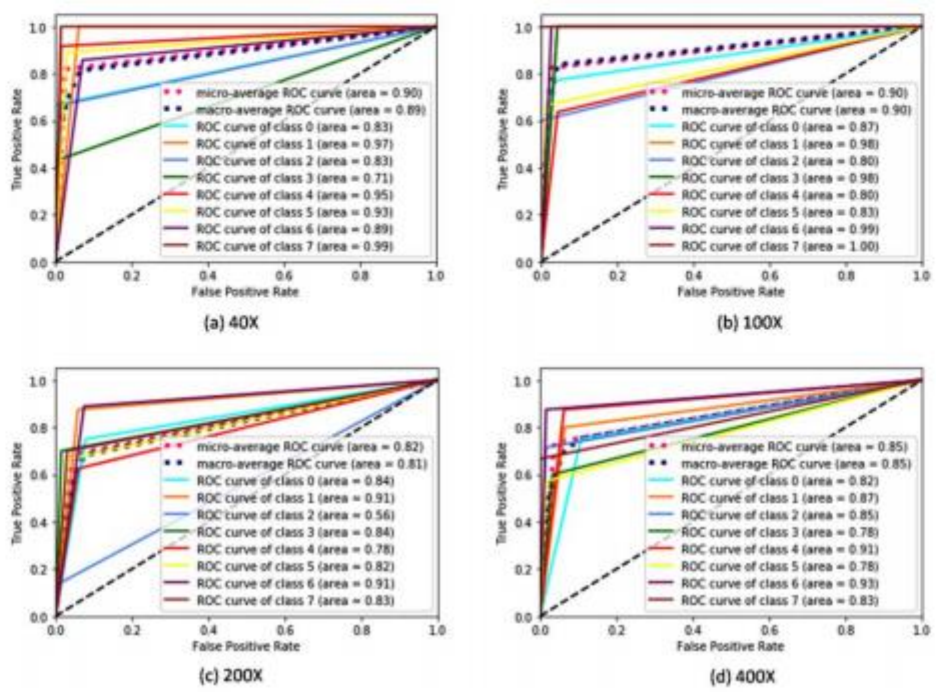

Fig 4 Multi-classification with appropriate layers For ROC curves

For feature extraction the CNN model was used. The conventional machine learning and deep learning approach utilizes hybrid model.

6. Conclusion

The prime intention of this research work is to classify thehistopathological images using advanced machine learning techniques for an early diagnosis of cancer. The improvement in classification accuracy is the main concern and specifically targeted throughout the experimentation. A regular growth in the incidence rate of cancer and inadequacy of experienced pathologists prompt us to determine an effective machine learning approach in order to build an automated computer-aided diagnosis system for cancer detection. Moreover, histology is the most reliable but complex modality of imaging to diagnose cancer because the analysis of histology images is followed by various challenges such as baffling tissue patterns and variation in the staining process. In this context, the approach of conventional deep and machine learning are discussed, especially the convolutional neural network have been elaborated for 'binary' as well as multiple classification of BC histopathological images. Besides, another aspect of the deep learning approach 'the transfer learning' has also been conceptualized for further enhancement in the classification performance of the cancer diagnosis system. In this work, several performance metrics have been employed to determine a feasible and computationally effective machine learning approach for the application of cancer detection.

7. Future Work

The current experimental investigations have revealed that machine learning techniques can be efficiently utilized in the development of high performance automated systems for early detection, diagnosis and treatment of cancer. Though the current study presents a lot of significance of machine learning approaches in the detection of cancer; however, there is a scope for further improvement and extension of the present work. With continuous evolution of new ideas, proposals and suggestions in this area, more efficient and effective models will definitely be proposed in the near future. We are also engaged in developing other deep learning and transfer learning based frameworks. Throughout this work, a single classifier is used to make the final prediction which can be replaced with an ensemble of the classifiers. This is due to the fact that sometimes a single classifier is not sufficient for the classification of complex images and makes the system vulnerable in terms of reliability. Various ensemble learning rules such as bagging, boosting and stacking can be employed in modelling of such classifiers. The utilization of CNNs as a classifier in place of the conventional classifier for the handcrafted features training could also be a different approach to solve the task of classification.

References

[1] Liu.Y, 2018 Learning deep models for face anti spoofing Binary or auxiliary supervision.IEEE Conference on Computer Vision and Pattern Recognition. pp. (389-398).

[2] Arun Kumar M, 2020 Face Recognition System for Visually Impaired People.Journal of Critical Reviews.Vol 7. Issue 17.pp. (2760-2764), DOI: 10.31838/jcr.07.17.345.

[3] Francis, Arun&Arulselvan, M. &Elangkumaran, P. &Keerthivarman, S. & Kumar, J.. (2019). Object detection using ultrasonic sensor. International Journal of Innovative Technology and Exploring Engineering. 8. 207-209.

[4] Y. Neelaveni and G. A. Francis, "Magneto-electric dipole array with optimized antenna parameters," 2015 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 2015, pp. 1-4, doi: 10.1109/GET.2015.7453855.

[5] Francis, Arun& Wilfred, M.W. &Sekar, R.. (2019). Health monitoring with alcohol detection and ignition control system using iot. International Journal of Innovative Technology and Exploring Engineering. 8. 203-206.

[6] Arun Kumar M, 2021 An Efficient Aquaculture Monitoring Automatic System for Real Time Applications. 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE DOI:

10.1109/ICISS49785.2020.9316072.

[7] Arun Kumar M, 2020 An Efficient Patient Health Monitoring System Using IOT For Health Care Applications. Waffen-Und Kostumkunde Journal. Vol 11. Issue 8. pp. (219-222).

[8] Vignesh M, Kavin T P, A. F. G. G. C. R. K. S. (2020). Accident Detection and Alerting System Using GPS & GSM. International Journal of Advanced Science and Technology, 29(3), 3598 - 3601. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/5037

[9] Arun Francis G, Dharani S K, Manikandan P, Monica RJ and Vaishahi S K, "IOT Based Accident Identification and Alerting System", International Journal of Pure and Applied Mathematics,vol.118, No.20,pp.547-551,2018.

[10] S.Manikandan, T.Karthik, and G.Arun Francis, "Vector Monitoring Concurrent BIST Architecture using Modified SRAM Cells", International journal of scince and advance research in technology, Vol. 1 No.9 ,pp.14-17, Sep 2015.

[11] Mr.G.Arun Francis, Dr.P.Karthigaikumar, Mr.G.Arun Kumar (2020) FACE RECOGNITION SYSTEM FOR VISUALLY IMPAIRED PEOPLE. Journal of Critical Reviews, 7 (17), 2760-2764. doi:10.31838/jcr.07.17.345.

[12] Francis, Arun. (2020). Home Automation Chat Bot using IOT. Waffen-und Kostumkunde. 11. 92-96.

[13] Arun Francis G., Ashok Kumar M., Cyril Silvester A. K., Kamaleshwarran A. P., Kumaravel B., "Pulsed Electric Field based Vegetable Fresher", Annals of RSCB, pp. 9918–9926, Apr. 2021.

[14] Arun Francis G., Lexmitha S., Aruna Devi N., Swathika C. K., "Embedded System based Smart Automation for Elderly and Disabled People", Annals of RSCB, pp. 9909–9917, Apr. 2021.

[15] Arun Francis G., Kanishya A., Lalithambiga M., Niveka S., Nivedha E., "Density based Traffic Light Control Using Retransmitted and Receiver", Annals of RSCB, pp. 9895–9901, Apr. 2021.

[16] Arun Francis G., Chandru L., Arun Kumar B., Nithesh Kumar R., Narendran S., "Charging Time Control Lerusingarduino", Annals of RSCB, pp. 9886–9894, Apr. 2021.

[17] Arun Francis G., Mr. S. Manikandan, Sundar V., Gowtham E., "Home Automation Using Iot", Annals of RSCB, pp. 9902–9908, Apr. 2021.

[18] Arun Kumar M, 2020 An Efficient Finger Gesture Recognition System Using Image. 4th International Conference on Electronics Communication and Aerospace Technology (ICECA). IEEE DOI: 10.1109/ICECA49313.2020.9297441.

[19] Arun Kumar M and G.Arun Francis, 2017 Survey on various advanced technique for cache optimization methods for RISC based system architecture. IEEE Xplore Digital Library. pp. (195 –

200).

[20] Sirinukunwattana.K, 2016 Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Transaction on Medical Imaging. Vol. 35. pp. (1196-1206).

[21] Arun Kumar M, 2019 An Efficient RF Energy Harvesting System for low power applications. International Journal of Innovative Technology and Exploring Engineering (IJITEE). Vol. 9. Issue-2. pp.(3510 – 3513).

[22] Francis, Arun& Kumar, Arun& Kumar, Nithesh. (2021). Simulation of Charging Time Controller using Arduino.

[23] Francis, Arun& Kanishya, &Lalithambiga,. (2021). SIMULATION OF DENSITY BASED TRAFFIC LIGHT CONTROL USING RF TRANSCEIVER.

[24] Francis, Arun&Sundar, & Gowtham,. (2021). SIMULATION OF IoT BASED HOME AUTOMATION USING ARDUINO.

[25] Francis, Arun& Kumar, Ashok &Silvester, Cyril. (2021). SIMULATION OF PULSED ELECTRIC FIELD BASED VEGETABLE FRESHER.

[26] Francis, Arun& Devi, Aruna. (2021). SIMULATION OF EMBEDDED AND IoT BASED SMART AUTOMATION SYSTEM FOR ELDERLY AND DISABLED PEOPLE.

[27] Arun Kumar M and Dr.C. ArvindChakrapani, 2017 An Survey of Low Power FFT Processor for Signal Processing Applications. Journal of Advanced Research in Dynamical & Control Systems. 15-Special Issue. pp. (633-641).

[28] Xu.J, Stacked sparse auto encoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging. Vol. 35. pp.(119-130).

[29] Musso.A, 2003 Role of frozen section histology in diagnosis of infection during revision arthroplasty. Postgraduate Medical Journal.Vol. 79. pp. (590-593).

[30] Arun Kumar M, Jeba Samuel and L.Saranya, 2015 Efficient Time Sharing of Traffic Signal Using Wireless Sensor Networks. International journal of Applied Engineering Research (IJAER) Vol. 10. No. 20. Pp. (19604-19608).

[31] Bouchelouche.K, 2017 Diagnostic applications of nuclear medicine kidney and bladder cancer. Nuclear Oncology from Pathophysiology to Clinical Applications. pp. (839- 881).

[32] Albarqouni.S, 2016 Aggnet deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Transactions on Medical Imaging. Vol. 35.no. 5. pp. (1313-1321).