

Study of Detection of Various types of Cancers by using Deep Learning: A Survey



Dr. Navneet Malik¹, Vinod B Bharat², Sudhanshu P Tiwari³ and Jimmy Singla^{4*}

¹ Ph.D (CSE), Lovely Professional University, Punjab, India, subhash.14335@lpu.co.in

² Ph.D Scholar, Lovely Professional University, Punjab, India, vinodbbharat@gmail.com

³ Assistant Professor, Lovely Professional University, Punjab, India, sudhanshu.15813@lpu.co.in

^{4*} Associate Professor, Lovely Professional University, Punjab, India, jimmy.21733@lpu.co.in

ABSTRACT

In this survey paper we have concentrated on the detection of the cancer by using the applications of Machine Learning / deep learning. There are more than 100 types of cancer in the human body but, for study, we have considered the most common and most dangerous cancer diseases like breast cancer, lung cancer, brain cancer and Skin cancer. This paper also gives a basic introduction of machine learning and deep learning and also gives information about various deep learning algorithms. We have considered only latest references for this survey and in the last section of this paper, we have provided the summary, comments on the applications of various algorithms of deep learning.

Key words: Cancer, Convolutional Neural Network, Deep Learning, Machine Learning.

1. INTRODUCTION

1.1 Cancer

Cancer is the kind of disease that we can find almost anywhere in the body of the human being. We all know that the human body consists of trillions of cells [1]. It is a normal process of our human body to grow new cells so that they can occupy the place of old/dead cells. When any person is affected by the Cancer disease, then this normal process breaks, due to cancer, it develops new cells (when it is not necessary) and it also helps the old/dead cells to survive. Now our body is having new cells as well as the old/dead cells (which actually should have died, but due to cancerous cell it survives) this results to form a solid tumor [2]. This tumor leads to spread Cancer disease.

Cancer is one of the most dangerous diseases and is one of the causes of death. The death rate in females is higher as compared to the males. The deadliest cancer in female is

breast cancer; stomach cancer comes second, followed by the cervical cancer [3].

According to CancerIndia [4], one woman dies of cervical cancer every 8 minutes in India, if two women are newly diagnosed with breast cancer, one dies because of it in India. Due to use of tobacco there were total 3.17 lacks deaths in 2018 by considering both men and women. There are total 25Lakh Indians are living with this chronic disease.

Table 1: New Cases and Deaths due to Cancer in 2018 [5]

Type of Cancer	Recently Diagnosis cases (in Lacks)	Number of Deaths (in Lacks)
Breast Cancer	20.90	6.30
Lung Cancer	20.90	17.60
Non-Melanoma Skin Cancer	10.40	0.65
Melanoma Skin Cancer	2.60	0.60
Brain Cancer	2.90	2.41
Cervical Cancer	5.69	3.11
Prostate Cancer	12.76	3.59
Colon Cancer	10.96	5.51
Stomach Cancer	10.33	7.83
Liver Cancer	8.41	7.82
Cervix uteri Cancer	5.70	3.11
Thyroid Cancer	5.67	0.41
Pancreas Cancer	4.59	4.32

1.2 Machine Learning

Machine learning (ML), branch of AI that programs the system in such a way [6], so that the system will learn and improve its performance, automatically, from the datasets. ML concentrates on the improvement of the computer programs, which access the data and use it learn for themselves. Figure 1 shows the types of machine learning. The main purpose of ML is to train computer systems based the past experiences (Old Data) and from new experiences also, without any assistance from human being.

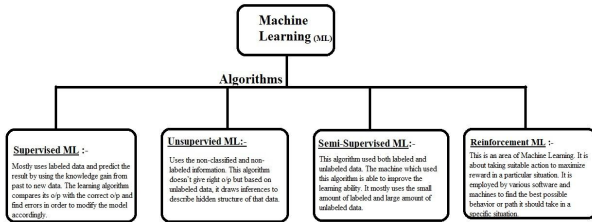


Figure 1 : Types of Machine Learning.

1.3 Deep Learning

Artificial Neural Network (ANN) is the kind of model in which the working is mostly similar to biological neural network. Deep Learning’s [7] working is somewhat similar to ANN. It is a part of ML. There are different algorithms of the Deep Learning, for example, Convolution Neural Network (CNN), Deep Belief Networks (DBN), Fully Convolution Network (FCN), which are quite useful in speech recognition, computer vision, bioinformatics, natural language processing and many other fields. Now we will overlook some of these algorithms.

Deep Belief Network (D.B.N.)

D.B.N. is a Kind of Deep Neural Network, in which we can observe multiple layers between Input and the output. It follows layer by layer training procedure. It follows the Restricted Boltzmann Machine (RBM) mechanism. When we consider a single RBM, one layer is visible (Input) but other layer is hidden (output). These two layers are connected with each other in VISIBLE-HIDDEN way. But when we consider the second RBM, then the hidden layer of first RBM, is visible to the second one. No Two layers are connected in VISIBLE-VISIBLE or HIDDEN-HIDDEN way. This is the most effective algorithms of Deep Learning.

Convolution Neural Network (C.N.N.)

C.N.N. is consist of multiple layer because it also belongs to the Deep Neural Network. Its functionality is based on the artificial neural network [8].

Apart from input layer and output (softmax), it mainly it consists of 4 layers,
 a) Convolution Layer mainly accepts the input, processes it and passes the output to the next layer. There may be more than more convolution layer present. These are totally hidden layers.
 b) Max Pooling layer is essential layer in C.N.N. as its main functionality is to downsampling in space. It acts on all the neurons of the convolution layer. It uses the maximum value of the each and every cluster of the neurons.
 c) Threshold Layer is present after each convolution layer; it performs element wise operation. The output of this layer is having the same dimension as that of input.
 d) Fully connected layer use to classify the input based on the features extracted in previous layers. The output of this layer is used in softmax layer to predict the correct output.

2. STUDY OF VARIOUS DEEP LEARNING METHODS IN IMAGE BASED CANCER DETECTION:

2.1 Breast Cancer

A Cancer which is developed by breast tissues is nothing but Breast Cancer. The main sign of Breast Cancer is that the patient may feel the swelling in the breast, this swelling is called as breast lump. Other symptoms of breast cancer are nipple sore/discharge, change in size/shape of breast or slight change in the color of the breast skin. In this type of cancer pain is rare. Various stages of breast cancer is shown in figure 2.

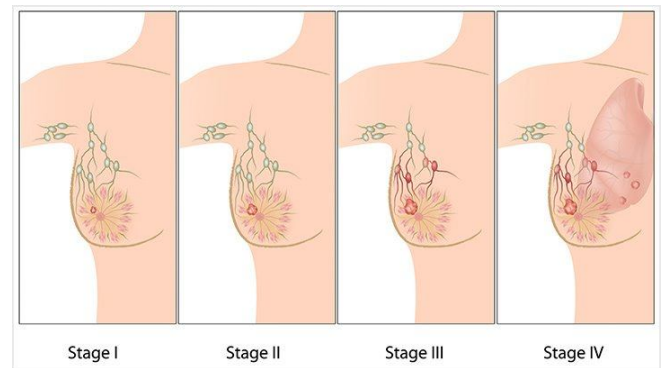


Figure 2 : Stages of Breast Cancer [33]

Chen et al. [9] concentrated their study on mitosis detection. They have used CaffeNet model and FCN model for the detection of mitosis from the histopathological images. Here authors have used MITOSATYPIA 12 & 14 dataset [10] for their study. Albayrak et.al. [11] have concentrated their study on the mitosis in breast cancer. Mitosis is the process of cell division in the human body, this process generates the two identical cells of the original parent cell. They have proposed a feature extraction algorithm. This algorithm helps to detect the mitosis by using the microscopic verification of the tissue for

the detection of the disease. The authors have used dataset from the MITOS-ATYPIA-14.

Xu et al. [12] have considered the nucleus for their study. They have used auto encoders latest version of algorithm i.e. Stack Sparc Auto Encoder algorithm which helps to differentiate the nucleus of various breast cancer cells.

Wichakam et al. [13] have combined the benefits of the two most important algorithms of deep learning i.e. CNN and SVM. They have applied this algorithm for the mass detection on mammograms. In this paper authors have used INbreast dataset for their study.

Ertesun et al. [14] have proposed the system which supports deep learning and it helps to detect the mammographic images for the mass. It actually finds that the particular section is having mass or not. Authors have used DDSM dataset [15] for their study.

Kallenberg et al. [16] have considered a mammographic image in large scale for their study. They have introduced an Auto Encoder with stacked convolution sparse. This system is incorporated with the system to improve the quality of system. Kim et al. [17] have considered the latest mammographic technique for study. The Digital breast tomosynthesis is the new technique for breast cancer detection, it helps to detect cancer at early stages by using very low x-ray system. Kim et al. have proposed deep CNN model for bilateral feature representation of DBT.

Spanhol et al. [18] have concentrated their study on the benign and malignant tumors. Benign tumor is not a cancerous tumor, but physically both of them are very similar, so most of the time, people are confused if it is a cancerous tumor or not. Benign tumor does not affect the surrounding tissues and also it does not spread to other parts of the human body. Malignant tumor is a cancer tumor, which needs to be detected at early stages, to decrease the death ratio. The authors have used CNN algorithm which is constructed with the help of AlexNet, the main aim of it is to differentiate between the cancer causing malignant tumor and non-cancer causing benign tumor. The authors have used BreakHis dataset [19] for their study.

2.2 Lung Cancer

Due to lung cancer, cells growth increases abnormally in the lung, which also leads to form a tumor in the lung. This tumor may affect the breathing ability of the human being. Identifying this cancer in early stage is not so simple. Symptoms of lung cancer are change in voice of person, frequent chest pain, weight loss, shortness in breath. (figure 3)

Hua et al. [20] have considered a Coin lesion for their study. Coin lesion is the small round/oval shape growth on the lung, its size is less than 1 inch. Authors have used 2-D Computerized Tomography images to classify these coin lesions. They have used two deep models DBN and CNN to achieve their objective. The dataset used by the authors is LIDC & IDRI.

Hussain et al. [21] have proposed the deep CNN model which is based on 3-D Computerized Tomography images and main objective is the characterization of the nodule. Based on the dimension of nodule, 2-D patches are created, then these patches are combined together to form the 3D tensor, which is then used to train the CNN model. These authors also used the LIDC & IDRI datasets.

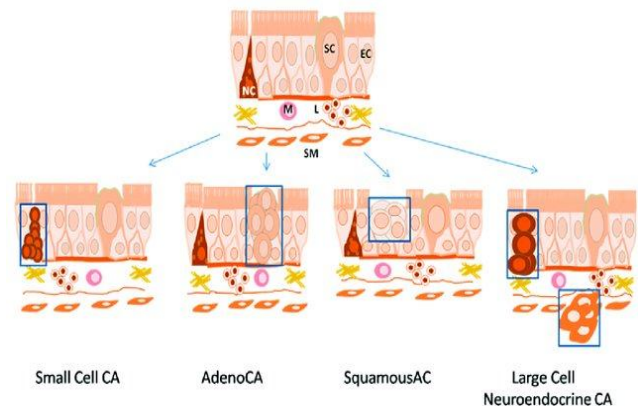


Figure 3 : Histological Types of Lung Cancer [34]

Zhu et al. [22] have used pathological images of lung cancer, through it they predict the survival time of the patient. Authors have used deep convolution neural network for it.

Setio et al. [23] by considering the objective to focus the nodule classification came up with the deep CNN model, which is trained with the help of 3-D computerized tomography images. By using 3-D image, numbers of 2-D images were created and each image is considered for the feature extraction by using single CNN model. These extracted features of 2-D images were combined together and used to feed the classifier. Along with LIDC & IDRI dataset, these authors have also used ANODE09.

From above research papers, we can say that, most of the authors have used 2-D database images to train their respective models, but Dou et al. [24] have considered the 3-D images directly, so that it can train the other proposed model. These authors have considered LIDC & IDRI datasets for their study.

Paul et al. [25] have proposed model which is going to use the pretrained CNN model, which is trained by using the non-medical datasets. These pretrained CNN were used for the feature extraction which in turn is used to train the classifier.

Zhuo Liu et al. [36] have considered the benefits of deep reinforcement learning for the early detection of the lung cancer tumors. By combining the various algorithms of the deep learning along with the reinforcement learning algorithms gives us better result. This technique is already applied successfully in various fields of computer science. The Early detection of lung cancer can help to reduce the mortality rate.

2.3 Skin Cancer

Skin Cancer is the most common kind of cancer. This cancer arises due to high growth of skin cells. In this type of cancer mostly we are not able to find the tumor also the death ratio due to this cancer is also very low.

Basically there are two types of skin cancer, melanoma and non-melanoma. Non-Melanoma skin cancer (figure 4) is mostly seen affecting the human being, here the mortality rate is approximately 0.66% by considering total affected patients. Melanoma cancer is rare cancer which is mostly not seen affecting the human beings but it is more deadly cancer, approximately 23% by considering the persons affected with melanoma cancer. Melanoma cancer can affect the other parts of the body, so if melanoma cancer is detected that may lead to spread the cancer to other parts of body.

As we are not able to find the tumor in the skin cancer, so that it is more complicated to detect this type of cancer, but if proper medical treatment taken then we can detect it easily. One of best method to detect the skin cancer is by doing "Biopsy".

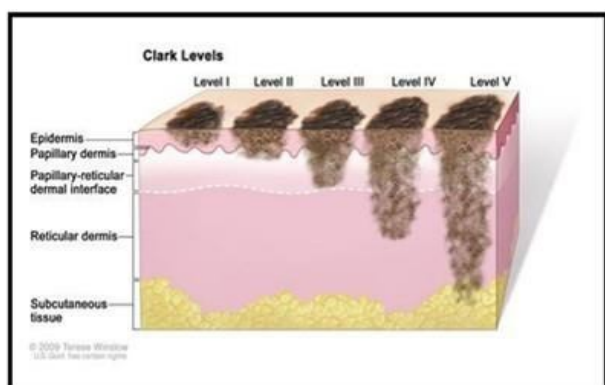


Figure 4: Non-Melanoma Skin Cancer Stage 1 [33].

Pomponiu *et al.* [26] are having objective of their to study to classify the skin cancer. They have considered the dermoscopy images for mole lesion classification and they have used DermIS and DermQuest dataset for their study. For feature extraction of skin images, authors have used pretrained CNN model along with AlexNet.

Esteva *et al.* [27] have used open access dataset in their study and for the skin cancer classification they have also used pretrained CNN model.

Massod *et al.* [28] have used labeled and unlabeled data to form the different datasets. The authors have proposed semi supervised model for the detection of the melanoma in skin cancer. In the proposed model, they have used Deep belief network (DBN) along with 2 SVMs which are trained with the help of datasets used.

Majtner *et al.* [29] have used two different SVMs for the melanoma identification. These SVMs are used to train with the ISIC datasets. The CNN model is used for the feature extraction from the images.

Yu *et al.* [30] proposed a system for the detection of melanoma from the available dermoscopy images. Fully

convolution residual network is used to find out the skin lesion from the dermoscopy images, once this process is complete then this images containing skin lesion are passed to deep residual network.

2.4 Brain Cancer

The human brain may consist of tumors, but these tumors are not always the cancer tumors which cause brain cancer. There are more than 40 kinds of tumors, which are seen in brain, all these tumors are basically divided in to two types, a) Benign and b) Malignant.

Benign type of tumors is less dangerous and not spreading to other parts of the body. But malignant type of tumors is cancerous tumor which abnormally grows the cells in the brain and also affects the other parts of the body.

Gao *et al.* [31] studies for the Alzheimer's disease early detection, this disease destroy the mental function of the brain. They have used two CNN models, 2D and 3D, and these models were trained by using the 2D and 3D Computerized Tomography images and the result is declared by combining the output of these two models.

Pereira *et al.* [32] have used magnetic resonance images to train the CNN model, in the proposed algorithm, for the segmentation of the brain tumor, they have investigated intensity normalization and augmentation.

Devkotoa B *et.al.* [35] proposes the CAD approach which helps to detect the brain tumor. Authors have used Mathematical morphological Reconstruction in their study. They have studied various images, which are preprocessed to find the regions of the brain tumor. Once the preprocessing is completed then segregated images are classified to find the type of the tumor, benign or malignant.

3. DISCUSSION AND CONCLUSION

In this paper, we have studied various types of cancers, which are responsible for the nearly 45% deaths of human being, by considering the death caused by other type of cancers. So only way to reduce this death ratio, is by early detection of cancer. We surveyed nearly 22 reference papers, in which, the main objective of the author is to provide a system (by using deep learning methods) which helps to detect the cancer early. Authors have used various online datasets such as LIDC & IDRI BreaKHis DDSM MITOS-ATYPIA-14 and some authors have used the database directly from the hospitals. This paper also gives basic information about the various types of cancers and also includes the study of good reference papers.

REFERENCES

- 1) S G. Yabe, S Fukuda, J Nishida, F Takeda, K Nashiro, H Okochi. **Induction of functional is let-like cells from human iPS cells by suspension culture**, *Regenerative Therapy*, pp. 69-76, June2019.
<https://doi.org/10.1016/j.reth.2018.11.003>

- 2) M Gomez-Canob, T Salikaa, D Jardelc, Gary A. Aelb, G Lyratzopoulos. **Diagnostic route is associated with care satisfaction independently of tumour stage: Evidence from linked English Cancer Patient Experience Survey and cancer registration data Try My Pharma**, *Cancer Epidemiolog*, pp. 70-78, May2019. <https://doi.org/10.1016/j.canep.2019.04.011>
- 3) Indian Council of Medical Research (2019, February) Retrieved from https://www.icmr.nic.in/sites/default/files/ICMR_News_1.pdf
- 4) Cancer India (2018). Retrieved from <http://cancerindia.org.in/cancer-statistics/>
- 5) Bray F, Ferlay J, Soerjomataram I, Isabelle S, Rebecca L S, Lindsey A T, Amhedin Jamel. **Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**, *A Cancer Journal for Clinicians*, pp. 394-424, Nov2018. <https://doi.org/10.3322/caac.21492>
- 6) Dr Jimmy Singla, Syed I Manzoor. **A Comparative Analysis of Machine Learning Techniques for Spam Detection**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol 8 no 3, pp. 810-814, June2019. <https://doi.org/10.30534/ijatcse/2019/73832019>
- 7) Samir Khan, Takehisa Yairi. **A review on the application of deep learning in the system health management**, *Mechanical Systems and Signal Processing*, Nov2017. <https://doi.org/10.1016/j.ymssp.2017.11.024>
- 8) Jay Robert B. Del Rosario. **Development of Face Recognition System Using Deep Convolutional Neural Networks in a Multi View Vision Environment**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol 8 no 3, pp. 369-374, June 2019. <https://doi.org/10.30534/ijatcse/2019/06832019>
- 9) H. Chen, Q. Dou, X. Wang, J. Qin, A. Heng. **Mitosis Detection in Breast Cancer Histology Images via Deep Cascaded Network**, *Proceedings of the Thirtieth Conference on Artificial Intelligence*, February 2016, pp. 1160-1166.
- 10) Retrieved from <https://mitos-atypia-14.grand-challenge.org/dataset/>
- 11) A. Krizhevsky, I. Sutskever, G E Hinton. **Imagenet Classification with deep convolutional neural networks**, *Advances in Neural information processing systems*, pp. 1097-1105, 2012.
- 12) Xu, Xiang, R Hang, Wu. **Stacked Sparse autoencoder based framework for nuclei patch classification on breast cancer histopathology**, *Biomedical Imaging. IEEE 11th International Symposium*, 2014, pp. 999-1002. <https://doi.org/10.1109/ISBI.2014.6868041>
- 13) I Wichakam, P Vateekul. **Combining deep convolution networks and SVMs for mass detection on digital mammograms**, *8th IEEE International Conference on Knowledge and Smart Technology*, 2016, pp. 239-244. <https://doi.org/10.1109/KST.2016.7440527>
- 14) Ertosun M G, D L Rubin, **Probabilistic visual search for masses within mammography image using deep learning**, *IEEE international conference on Bioinformatics and Biomedicine*, 2015, pp. 1310-1315. <https://doi.org/10.1109/BIBM.2015.7359868>
- 15) Retrieved from <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>.
- 16) Kallenberg, Pertersen K, Nielsen M, P. Diao, C M Vachon, K Holland, RR Winkel. **Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring**, *IEEE transaction on Medical Imaging*, pp. 1322-1331, 2016. <https://doi.org/10.1109/TMI.2016.2532122>
- 17) D H Kim, Y M Ro. **Latent feature representation with 3-D multi view deep convolutional neural network for bilateral analysis in digital breast tomosynthesis**, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 927-931.
- 18) F A Spanhol, L S Oliveira, C Petitjean, Heutte. **Breast Cancer Histopathological image classification using convolutional neural network**, *IEEE International Joint Conference on Neural Network*, 2016, pp. 2560-2567. <https://doi.org/10.1109/IJCNN.2016.7727519>
- 19) Retrieved from <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>.
- 20) L. Hua, H Hsu, C Hidayati, H Cheng, J Chen. **Computer-Aided Classification of Lung nodules on Computed Tomography images via Deep Learning Technique**, *Onco Targets and Therapy*, pp. 15-22, August 2015.
- 21) S. Hussein, R Gillies, K Cao, Q song, U Bagci. **TumorNet: Lung Nodule Characterization Using Multi View Convolutional Neural Network with Gaussian Process**, *IEEE International Symposium on Biomedical Imaging*, March 2017. <https://doi.org/10.1109/ISBI.2017.7950686>
- 22) X. Zhu, J. Yao, J. Huang. **Deep convolutional neural network for survival analysis with pathological images**, *IEEE International Conference on Bioinformatics and Biomedicine*, January 2017, pp. 544-547. <https://doi.org/10.1109/BIBM.2016.7822579>
- 23) A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. van Riel. et. al. **Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks**, *IEEE transactions on medical imaging*, pp.1160-1169, 2016. <https://doi.org/10.1109/TMI.2016.2536809>
- 24) Q. Dou, H. Chen, L. Yu, J. Qin, P.-A. Heng. **Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection**, *IEEE Transactions on Biomedical Engineering*, pp.1558-1567, 2017. <https://doi.org/10.1109/TBME.2016.2613502>
- 25) R. Paul, S.H. Hawkins, L.O. Hall, D.B. Goldgof, R.J. Gillies. **Combining Deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from**

- diagnostic CT**, *IEEE International Conference on Systems, Man, and Cybernetics*, 2016, pp. 002570-002575.
<https://doi.org/10.1109/SMC.2016.7844626>
- 26) V. Pomponiu, H. Nejati, N.-M. Cheung. **Deepmole: Deep neural networks for skin mole lesion classification, (ICIP), 2016 IEEE International Conference on Image Processing**, 2016, pp. 2623-2627.
<https://doi.org/10.1109/ICIP.2016.7532834>
- 27) A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun. (2017). **Dermatologist-level classification of skin cancer with deep neural networks**, *Nature*, pp.115-118, 2017.
<https://doi.org/10.1038/nature21056>
- 28) A. Masood, A. Al-Jumaily, K. Anam. **Self-supervised learning model for skin cancer diagnosis, 7th International IEEE/EMBS Conference on Neural Engineering**, 2015, pp. 1012-1015.
<https://doi.org/10.1109/NER.2015.7146798>
- 29) T. Majtner, S. Yildirim-Yayilgan, J.Y. Hardeberg, **Combining deep learning and hand-crafted features for skin lesion classification, 6th International Conference on Image Processing Theory Tools and Applications**, 2016, pp. 1-6.
<https://doi.org/10.1109/IPTA.2016.7821017>
- 30) L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng. **Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE transactions on medical imaging**, pp. 994-1004, 2017.
<https://doi.org/10.1109/TMI.2016.2642839>
- 31) X.W. Gao, R. Hui, Z. Tian. (J2017). **Classification of CT brain images based on deep learning networks, Computer methods and programs in biomedicine**, pp. 49-56, Jan2017.
<https://doi.org/10.1016/j.cmpb.2016.10.007>
- 32) S. Pereira, A. Pinto, V. Alves, C.A. Silva. (2016). **Brain tumor segmentation using convolutional neural networks in MRI images, IEEE transactions on medical imaging**, pp. 1240-1251, 2016.
<https://doi.org/10.1109/TMI.2016.2538465>
- 33) Retrieved from <https://www.everydayhealth.com/breast-cancer/stages-what-they-mean/>
- 34) Tatiana N Z, Galina S Z, Olga S K, Ruslan A Z et. al. **Current and Prospective Protein Biomarkers of Lung Cancer. Cancers**, pp. 01-22, November 2017.
- 35) Devkootaa B. Abeer Alsadoona, Prasad P W C, Sing A K, Elchouemic A. **Image Segmentation for Early Stage Brain Tumor Detection using Mathematical Morphological Reconstructio, International Conference on Smart Computing and Communications Procedia Computer Science**, December 2017, pp. 115-123.
<https://doi.org/10.1016/j.procs.2017.12.017>
- 36) Zhou Liu, Chenhui Yao, Hang Yu, Taihua Wu. **Deep reinforcement learning with its application for lung cancer detection in medical internet of things, Future Generation Computer System**, pp. 1-9, February 2019.
<https://doi.org/10.1016/j.future.2019.02.068>