ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse44922020.pdf
https://doi.org/10.30534/ijatcse/2020/44922020

Analysis of Mammography for Identifying Cancer Cells using Convolution Neural Networks

KollaBhanu Prakash [1], Lakshmi Kalyani. N[2], Pradeep Kumar Vadla[3], Naga Pawan YVR[4]
[1]Koneru Lakshmaiah Education Foundation, India
[2]VNR VignanaJyothi Institute of Engineering and Technology
[3]B V Raju Institute of Technology, Narsapur, Medak, Telangana
[4]Anurag Engineering College, Ananthagiri (V&M), Kodad, Suryapet(Dist), Telangana
drkbp@kluniversity.in

ABSTRACT

Breast cancer is the largest causes of women's death. it is the most commonly diagnosed cancer worldwide. the main modules in breast cancer prediction are data collection data pre-processing feature selection and classification.it is important to detect breast cancer as early as possible. The chances of survival of the patient is high advanced classification techniques and artificial intelligence methods has largely been used for breast classification.in the first part we will be reviewing our breast cancer histology image dataset and we will split the python script into three sets those are a training set, a validation set, a testing set. Next we will use keras libraries to define a convolutional neural network.

Key words: Breast cancer, Convolution Neural Networks, Keras, ROI selection, Deep Learning.

1. INTRODUCTION

Breast cancer is leading cancer in women which leads to loss of mortality [1]. As per the research of World Health Organization (WHO) the total number of breast cancer [5] patient's cases will increase to 19.3 million by 2025. The clinical correlation of tumor grades are shown Figure 1.

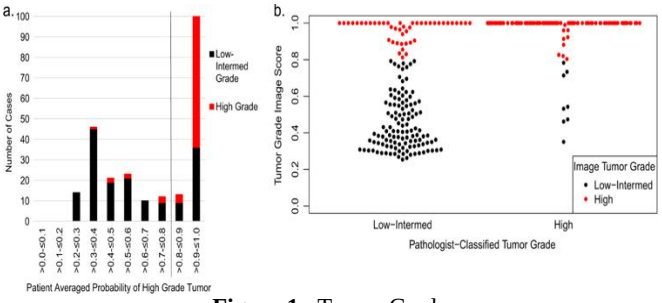

Figure 1: Tumor Grades

These are formed by nipples, ductules, and nodules. The epithelial tumors develop inside the lobules and also the ducts. There will be spread inside the body once it is started. Cancer initial stage begins from unusual cell growth and might quickly attack the other tissues. If the nuclei of the malignant tissue grow more significant than the normal tissue, it could cause serious effects in the future. Whereas benign cases are considered noncancerous which is non-threatening, they have very few chances to turn in to cancer status. These calls can be separated from the body nearly 50% to 60% of breast cancer patients are diagnosed in advanced stages. Though we have the perfect time period for curability, many of the patients don't take it as serious, since the symptoms are not indicated in early stages.

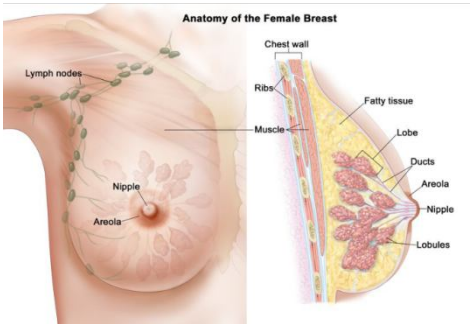

Figure 2: Breast Anatomy

The breast anatomy [2] illustrated in Figure 2.

2. RELATED WORK

The stated system involves the following phases.

a. Image acquisition
b. Image pre-processing
c. ROI selection
d. Feature extraction and selection
e. Classification

In pre-processing the mammogram image is processes to remove noise [4]. The filter with noise less than 0.2 doesn't work well. In bilateral filter image depth is extremely increased. Nonlinear digital filtering is a method which is an adaptive median filter.

ROI selection: In this phase the output form of pre-processing phase is used to discover ROI and it is

implemented by k-means each cluster is considered as the soft clustering algorithm. In Feature extraction, the features used for classification is a type of average intensity and SDM(Standard Mean Deviation). Texture analysis of mammograms helps to identify texture feature information. Feature adaptation based wrapper based classification is the feature selection is implemented by wrapper classification. The population evaluation is tested for each generation.

Genetic algorithm [13] uses Darwin principles of survival of the fittest. These are used to determine optimization problems. It is used to find out the highest fitness value from the population helps to improve the performance of the set of feasible clarifications. The main difference between genetic algorithms and numerous conventional optimization methods. Genetic algorithm randomly tries to initialize the population. The genetic algorithm uses operations like selection, crossover and mutation.

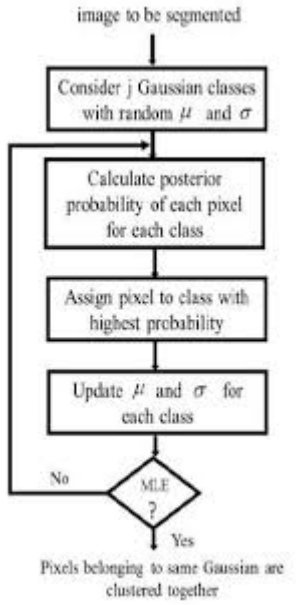

Figure 3: Image Segmentation

In genetic algorithms [13] the best fitness value is considered reproduction.Through cross over and mutation operators new individuals are created.Evaluate the new individual's fitness displacing the smallest fit population with new off-springs.

Conventional neural networks are the finest approach of the deep learning [3] which takes the specific approach in to the topology. Convolutional Neural Networks consists of multiple trained stages stacked on each other, followed by supervised classifier. Convolutional layer pooling layer and fully connected layer are the three main types of layer used to build CNN architectures. The pooling layers perform the down sampling operations to reduce the amount of computation and improve the robustness. It is amultilayer artificial neural network.

$$S_j = f\left(\sum_{i=1}^{N} I_i * K_{i,j} + b_j\right), \quad (1)$$

Identifying masses asymmetries are difficult for radiologist due to in experience, the mammograms are very sensitive analyze. A mammogram can detect breast changes that could be disease years by detecting the mammogram in the early stage, which also help them to cure the cancer.in feature extraction once the relevant features are extracted from all the digital mammograms.

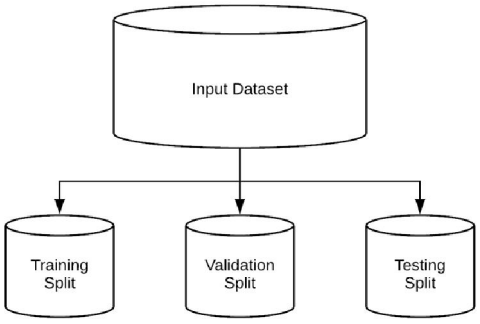

Figure 4: Image Segmentation

After extracting the relevant gestures, the final stage is to classify the obtained patterns as per the flow shown in Figure 3.

Image processing techniques are also used to remove the unwanted noise, this will make the further processing easy. Smoothing image can be done by using averaging filter.

Local binary pattern can extract the fine details from image for breast cancer detection in its early stage.

Linear Discriminant Analysis (LDA) is an efficient supervised learning method, this is also known as Fisher Discriminant Analysis (FDA)

$$\mu_k = \frac{1}{m_k} \sum_{x_k \in \prod_k} x_k \quad (2)$$

$$\Sigma_k = \frac{1}{m_k} \sum_{x_k \in \prod_k} (x - \mu_k)(x - \mu_k)^T, \quad (3)$$

$$d(x_i, x_j) = \left\| x_i - x_j \right\|_F^2. \quad (4)$$

The maximum pooling is adopted for dimensionality reduction after activation process.in order to reduce the errors and avoid overfitting to the fully connected layers drop out technique is used.

One of the unsupervised deep learning method is auto encoder that is also called as AE for every input vectorized form of the sample the representation has the sample.

$$h^{(m)} = f_\theta\left(x^{(m)}\right), \quad (5)$$

In order to compare the performance of the different methods mainly the Convolutional Neural Networks, classification methods and to extract the Distinguish features from the normal and the abnormal cases. CNN were classified by the benign and malignant cases by the deep learning methods.

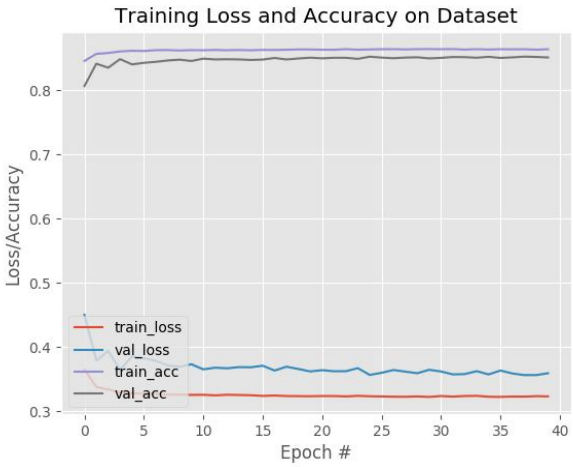

Figure 5: Training Loss and Accuracy Dataset

The computer simulation methods have given the results that, deep learning methods could train the lower dimensional and higher level features of ROIs. These provides promising classification results. CNN has the average accuracy of the 93.25 and 62.7% for abnormality and malignancy classification respectively. The details are shown in Figure 5. Many classification methods imply more features and more sophisticated method to be considered. The dataset handling is shown in Figure 4.

2.1 Stages of Cancer

The stages of cancer [14] is as follows

- Stage zero means the cancer cells have developed and they did not have time to develop to the other tissues
- In first stage the tumor had not spread to the lymph nodes.
- Stage 2a means already spread to more the 4 lymph nodes and the tumor size is less than 2cm which cannot be detected easily
- In stage 2b the tumor is greater than 2cm and less than 5cms.
- Stage 3a means the tumor is greater than in 5cms
- In stage 3b it does not spread more than 9 lymph nodes.
- Next in stage 3c the survival rate during 6 years is approximately 75%
- In fourth stage the cancer met abases are formed in bones liber brain and lungs

2.2 Signs and symptoms

The signs and symptoms of BC are as listed below:

a. Change the skin of the breast
b. Nipple retraction means if the nipple was in the other form
c. Redness swelling of the breast
d. Hard lumps of any size, shape and texture.

2.3 Treatment

There are several treatment options available

a. Target therapy
b. Chemotherapy
c. Surgery
d. Hormone therapy
e. Radio Therapy

2.4 Mammography

Finland was the first country to introduce government screening mammography program in 1987. Mammography of the Breast Cancer (BC) are detected with some changes which are difficult to distinguish from cancer. Only with non-hazardous mammography, these changes cannot be distinguishes from the cancer. The Radio Therapy of Breast / Chest options shown in Figure 6.

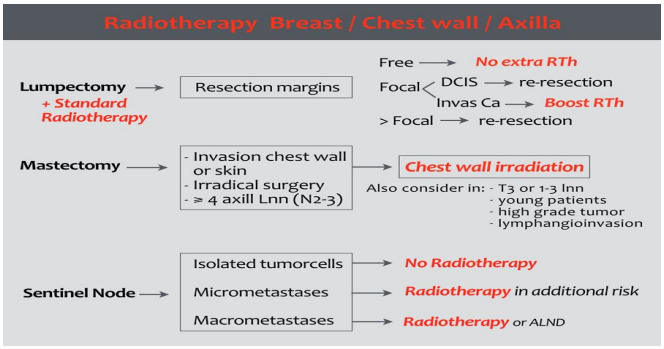

Figure 6: Radio Therapy of Breast / Chest

2.5 Deep Learning

Deep learning is inspired by the workings of the human brain.it serves to improve artificial intelligence and make many of its application possible.

2.6 Keras

Kerasis an open source neural network library. The Kerasprototype enabling fast experimentation. It has high level API and works on the top of the tensor flow. It is user friendly and extensible.

2.7 National Cancer Center Network (NCCN) Guideline

Problem description:

Stage 1: Consulting for physical examination.

Stage 2: studies with renal function tests and levels of calcium.
Stage 3: Chest X-Ray, CT scan of the chest, Bone scan for evaluation, Tumor makers may also be obtained in the makers.
Stage 4 is termed as advanced breast cancer where systematic therapy and limited surgery is needed.

Breast Cancer (BC) in pregnancies occur once in every 4000 pregnancies. They have swollen and tender breasts making it more difficult. Clinical breast exams should be a part of it. 80% of the women diagnosed will not have a family history. The treatment algorithm for BC during pregnancies is shown in Figure 7.

Risk factors of breast cancer are dense breasts, environmental, obesity, lifestyle, lack of exercise, dietary factors.

2.8 Risk Factors

Some risk factors for breast cancer [14] are listed

i. Age
ii. Personal history of breast cancer
iii. Genetic factors
iv. Family history
v. Child bearing
vi. Menstrual history

2.9 Classification Algorithms

We have different types of classification algorithms [6], [7], [8], [9], [10], [11], [12]

i. Decision tree algorithm
ii. KernalSVM
iii. Naive Bayes
iv. Logistic regression
v. Support Vector Machine

2.10 Algorithm

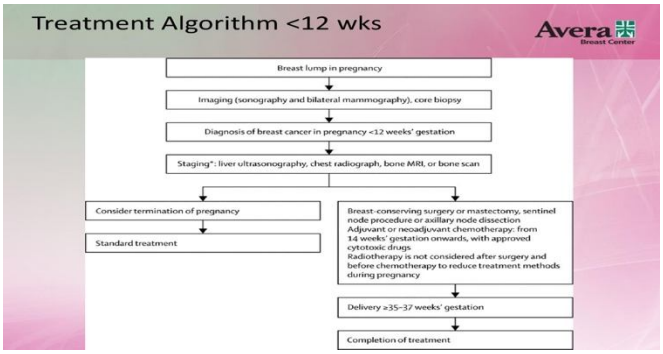

Figure 7: The algorithm for BC during Pregnancy

a. Download the zip file, unzip at the preferred location
b. Download the dataset
c. Unzip the dataset in original directory
d. Config.py: which is needed for building the data set
e. Build data set.py: in which 80% will go for training and other 20% for testing with the image data center from Keraslibraries.
f. Run the script build_dataset.py:
g. Concernet.py: This will build on convolutional neural network by following these operations Use 4*4 conventional filters

3. RESULTS AND DISCUSSION

The investigation finds the most of the conventional classifiers depend upon the feature extraction. If we achieve more than 95% accuracy by the use of machine learning models, the fact will confirm that it is an easy and a fast manner. This may lead to time of the detecting and analyzing the cancer cells to prevent death. With each new entry adding to this model the process becomes more intelligent and ensures accurate results. For feature work the other networks will suggest the very deep convolutional network.The Figure 8 [15][16], shows that Ki-67 prognostic parameter is detection of BC. Of the estimated 220 reported breast cancer cases, 63.8 percent were younger than 50 years of age. In 47 percent of patients Ki-67 was negative and 53 percent optimistic. Ki-67 rates were 1-5%, 6-14% and 15% respectively in 18%, 11.1% and 24% of patients with healthy Ki-67, respectively. In terms of tumor scale, 30.7% of patient tumors were under 2 cm, 51.4% were 2-5 cm, and the remaining (17.9%) are over 5 cm. The prevalence of grade 1, 2 and 3 tumors was 23.9%, 63.3% and 12.8% respectively. There was no major association between Ki-67 and either the age of the individual, or tumor size and grade. There was nevertheless a slightly significant association between the position of the lymph node and the expression of Ki-67. While there was a substantial association between estrogen receptor (ER) and progesterone receptor (PR) with Ki-67 status, the selection of the best therapeutic approach also needs accurate assessment of prognostic factors in BC patients.

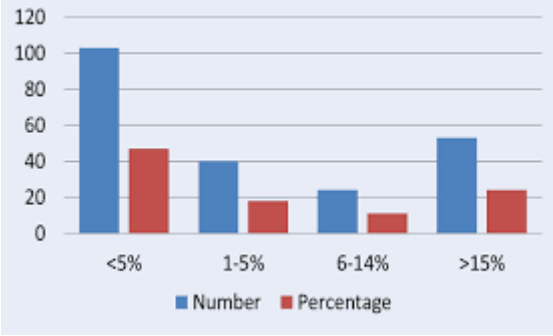

Figure 8: Ki-67 of BC Patients [15]

Stack these filters on top of each otherPerform max poolingUse depth wise separable convolution machine learning models, the fact will confirm that it is an easy and a fast manner. This may lead to time of the detecting and analyzing the cancer cells to prevent death. With each new entry adding to this model the process becomes more

intelligent and ensures accurate results. For feature work the other networks.

4. CONCLUSION

These algorithms are tested on the breast cancer diagnoses on the real life problem. Through the proposed model it was found that the design decreases the significantly computing time and the feature was improved. The project ensures that the early level of detection can reduce the risk of the life time of the patient a greater extent and the treatment can be given at the right stage. The improvement of current situation with breast cancer is a big concern. Identification and regular check can save many lives such as cancer identification with computer aided techniques.

REFERENCES

[1] S. Shamy, J. Dheeba ,"A Research on Detection and Classification of Breast Cancer using k- means GMM & CNN Algorithms", , **International Journal of Engineering and Advanced Technology (IJEAT)**, ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019.

[2] Nahid, Abdullah & Kong, Yinan. (2017). "Involvement of Machine Learning for Breast Cancer Image Classification: A Survey", **Computational and Mathematical Methods in Medicine**. 2017. Pp: 1-29. 10.1155/2017/3781951.

[3] https://www.pyimagesearch.com/2019/02/18/breast-cancer - classification -with-keras-and-deep-learning/

[4] https://barisgecer.github.io/files/gecer_thesis-compressed. pdf

[5] http://www.mdpi.com/2411-9660/2/2/13/pdf

[6] Kolla, B.P., Raman, A.R., "Data Engineered Content Extraction Studies for Indian Web Pages", **Advances in Intelligent Systems and Computing**, 2019, Vol 711, pp 505-512, DoI: 10.1007/978-981-10-8055-5_45.

[7] Prakash, K.B., "Information extraction in current Indian web documents", **International Journal of Engineering and Technology(UAE)**, 2018, Vol 7(2), pp:68-71, DoI: 10.14419/ijet.v7i2.8.10332

[8] Prakash, K.B., "Content extraction studies using total distance algorithm", Proceedings of the 2016 2nd **International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016**, pp: 673-679, DoI: 10.1109/ICATCCT.2016.7912085

[9] Prakash, K.B., DoraiRangaswamy, M.A., "Content extraction studies using neural network and attribute generation", **Indian Journal of Science and Technology**, 2016, 9(22), pp: 1-10, DoI: 10.17485/ijst/2016/v9i22/95165

[10] Kolla, B.P., Dorairangaswamy, M.A., Rajaraman, A., "A neuron model for documents containing multilingual Indian texts", **International Conference on Computer and Communication Technology, ICCCT-2010**, pp: 451-454, DoI: 10.1109/ICCCT.2010.5640489

[11] P. Srikanth and D. Deverapalli, "A Critical Study of Classification Algorithms Using Diabetes Diagnosis," **6thIEEE International Conference on Advanced Computing (IACC)**, Bhimavaram, 2016, pp. 245-249.

[12] H.I. Okagbue, N.J. Peter, A.O. Akinola, C.O. Iroham, A.P. Opoko, "Smart Review of the Application of Genetic Algorithm in Construction and Housing", **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 9(1) (2020), pp: 266-273, DoI: https://doi.org/10.30534/ijatcse/2020/ 40912020.

[13] Bandara, Ravimal. "Nature Inspired Dimensional Reduction Technique for Fast and Invariant Visual Feature Extraction." **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 8(3) (2019), pp: 696–706. DoI: 10.30534/ijatcse/2019/57832019

[14] Bharat, Vinod& Malik, Dr. Study of Detection of Various types of Cancers by using Deep Learning: A Survey. **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 8, ,2019, pp: 1228-1233, DoI: 10.30534/ijatcse/2019/31842019.

[15] Kermani, T., Kermani, I., Faham, Z., &Dolatkhah, R. (2019). Ki-67 status in patients with primary breast cancer and its relationship with other prognostic factors, **Biomedical Research and Therapy**, 6(2), 2986-2991. https://doi.org/10.15419/ bmrat.v6i2.520

[16] Inwald, E. C., Klinkhammer-Schalke, M., Hofstädter, F., Zeman, F., Koller, M., Gerstenhauer, M., &Ortmann, O. (2013). "Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry", **Breast Cancer Research and Treatment**, 139(2), 539–552. https://doi.org/10.1007/s10549-013-2560-8