European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 2020

Multi-Stage Classification Technique for Breast Cancer Detection in Histopathology Images using Deep Learning

Nagamani Gonthina[1], C.Jagadeeswari[2A], Prabhavathi V[3B], Sneha B[4A]
[1,2,4]*Assistant Professor, BVRIT Hyderabad College of Engineering for Women*
[3]*Assistant Professor, Malla Reddy Engineering College for Women*

Abstract— This research paper proposes the past decenary, substantial improvement in computational ability and betterment in algorithms for analysis of Images has gained vast fame in resolving challenges in the area of medical diagnosis. Subsequently, computerized tissue histopathology at present is becoming tractable towards the implementation of digitized analysis of images and deep learning methods. Cancer is a cluster of disorders involving irregular cell maturation with the capability to conquer or proliferate to other organs of the body. Detection of cancer in the earlier stages is a exacting task due to which many people are prone to death. Treatment of cancer benefits from the pace, perfection of Deep Learning-obliged practice of diagnosis. Deep Learning techniques are utilized to diagnose the features of progressed carcinoma with enhanced perfection compared to individual pathologist. This paper suggests a deep convolution neural network for categorizing a tissue as malicious, there after segregate the tissue then ultimately perform multi-class detection and classification of Breast Cancer disease and its stages in histopathology images.

Index Terms— histopathology, image analysis, Deep Learning, Convolution Neural Network

1. INTRODUCTION

Once a rare disease, cancer is spreading so fast in the modern world which is the unspoken fear of current generation. A constant advancement related to cancer research is being carried out. The International Agency for Research on Cancer, depending upon 2002 data predicted as the amount of carcinoma sufferers reached 25,000,000 and the American Cancer Society in 2004, declared that customarily carcinoma will restore heart disease as the major source of expiry [1]. When cancer is diagnosed in the early stages – before it has a chance to metastasize or colonize, other healthy organs and tissues of patients particularly have a better possibility of survival. Unfortunately, when observable symptoms initially occur, these cancers are well ahead the early stages of development. Early stage recognition and endurance duration is the at most significant module of a process in medical findings for patients with malevolent disease. producing an skilled treatment decision begins with the stage, or progression, of the disease. The stage of Cancer is one of the most important constituents in estimating therapy course of actions.

Breast cancer is the following dominant cancer where Lung cancer is the first leading one in the world. It originates within the cells of the breast as a cluster of cancer cells that have the ability to conquer surrounding cells or proliferate (metastasize) to other organs of the body. 62% instances of breast cancer are identified at an isolated phase, for whom the 5-year longevity gauge is 99%. Currently, an approximation of 41,760 female would expire due to breast cancer in the U.S. even few males suffer with breast cancer. The lifespan threat for U.S. male relate to 1 in 1,000. An approximate of 2,670 males would be detected with breast cancer in current era in the U.S. and nearly 500 would expire. Once a human is diagnosed of malicious lump or identified to have breast cancer, the medical management group will decide breast cancer phase to convey the extent of severity.

Breast cancer is distinctive principal factors of expiry for female globally. As reported by World Health Organization (WHO), the amount of cancer occurrences anticipated in 2025 would reach 19.3 million. In

Egypt, cancer is an escalating issue, that to mainly breast cancer. Breast cancer can be identified through various tests, together with a mammogram, ultrasound, MRI and biopsy.

Histopathology is an investigation of a biopsy or sectional mass pattern by a radiologist. The examination of images of histopathological specimen is utilized in furnishing the ultimate comprehensive identification of various problems, one among which is cancers. Present histopathology application has different limitations, as it is extremely moderate and frequently convey modest assent among radiologists. With respect to such circumstance, Computer–Assisted Diagnosis (CAD) for histopath figures would be a unique demanding area in analysis of medico-biological images. CAD for histopathology images aid to overcome aforesaid limitations, as histopathology images are specified with monotonous shapes by different ranges which will be exceptionally matched for analysis of texture and programmed identification. The blend of conventional prognosis techniques with estimation of analysis of data constitutes a chance to lower the job of pathologists while even balancing the accomplishment.

The conventional procedure of histopathology image construction proceeds via various steps that culminate particular shapes within the figures. Initially, sample is refined and placed on sheet of glass by adding synthetics or through frosting the samples. Thereafter, samples are stained by tints (e.g. haematoxylin–eosin, H&E), with antigens (immunohistochemistry, IHC) or through additional practices. Afterwards, the sample sheet of glass will be inspected beneath microscopic lens by the medical examiner. The examiner's perceptible examination of specimen will be the typical procedure in current generation for recognizing substantial variety of disorders in addition of categorizing various kinds of carcinomas [2].

The growth in the amount of cancer cases annually since 2005 is exhibited in Fig. 1 [3]. The inflating drift of cancer sufferers in the previous decenniums empowers many to estimate the number of cases by climax of 2020 in India. Despite of inadequacy in improvement of medical field, this disorder is enlarging as leading incurable diseases of the generation. Current styles in image processing prove how a technician will be able to the build solutions in the field of medicine. Medical image processing became increasingly sustainable process in the world. The huge evolution in information technology forms this process authentic and systematic.

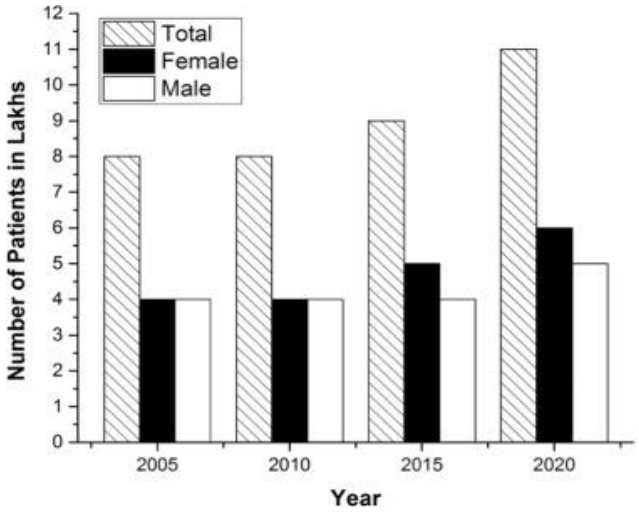

Fig.1. Complete cancer ubiquity over India year wise (anticipated instance of the period 2020).

2. ANALYSIS OF HISTOPATHOLOGY IMAGES

Histopathology cites to the minute inspection about tumor in conducive to examine the embodiment of disease. Especially, in medical field, histopathology concerns to the assessment of an effected body part cut from the patient's body or section specimen by a medical examiner, after the specimen is processed

and put on the sheet of glass. Histopathological figures are acquired via Charge Coupled Device (CCD) cam coupled microscope where an electronic digitized procedure is conducted [4]. An important perspective of various computerized methods is to acquire significant amount of information through images. These significant information include magnitude of the cell, malformations in the cell and excessive amount of cells. The important phases entangled in analysis of computerized images: Image acquisition, pre-processing, segmentation, feature extraction, feature selection, dimensionality reduction and classification. Each stage should be performed one after the other as shown in Fig 2[5]. There are various other algorithms which are computerized, accessible for histopathology image analysis of breast cancer.

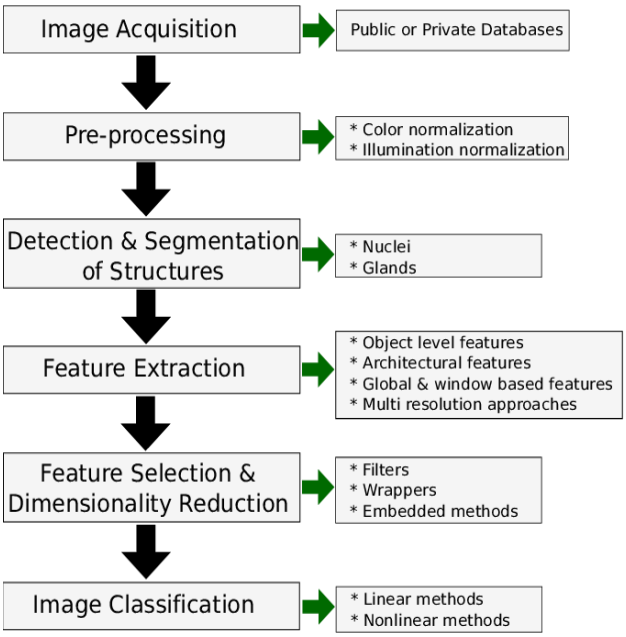

Fig. 2. Phases involved in Analysis of Digital Images

3. LITERATURE SURVEY

A. Image Acquisition and Preprocessing of histopathology image

Data can be acquired either by publicly available datasets or from private hospitals and research centers. In case of publicly available datasets, acquiring information is not a big deal as data is readily available. We need to only download the data. Various publicly available datasets are prepared by various challenges and contests. On the other hand if datasets are not available publicly, they need to be generated artificially. This is one of the major challenges in medical field. Data can be gathered from various hospitals and research centers or medical labs. But the data gathered may not be huge and may not be of common format. So all the data from various distributed sources should be gathered, converted into a uniform format and dumped in a common location so that it would be accessible easily by others.

Image Preprocessing is done to usually remove noise and unwanted data from the images and to improve the quality of image. Preprocessing could be implemented with various morphological methods such as expansion and destruction, less band filtering (median and averaging), thresholding etc. [6]. Processing in prior can remunerate for variations among images which are divergent in color, radiance and further deficiencies, like noise or outputs which are repeated due to the procedure of scanning. [5]. Various pre-processing steps are Staining normalization, Illumination normalization and Noise removal.

A. Detection and Segmentation of images

In histopathological images of Breast Cancer, appearance with count and the structural properties of few structures (such as nucleus and glands) are basic variables for estimating the existence, extremity of a pathology. [7]. The procedure of detecting Nuclei within large-scale breast cancer images is highly exciting for instance of image processing procedures because of definite varied features of cancer nuclei

like expanded and non-uniformly formed nuclei, extremely rough chromatin deprecated towards nuclei margin and evident nucleoli. [8]. Lymphocyte and epithelial cells are the necessary alternatives for nuclei. By investigating histopathological images, a complete proposition with respect to cell shape, severity of cancer maturation etc can be procured. Therefore prognosis in breast cancer enlargement is feasible. For exploring the region of interest (ROI), histopathology images should be initially segregated.

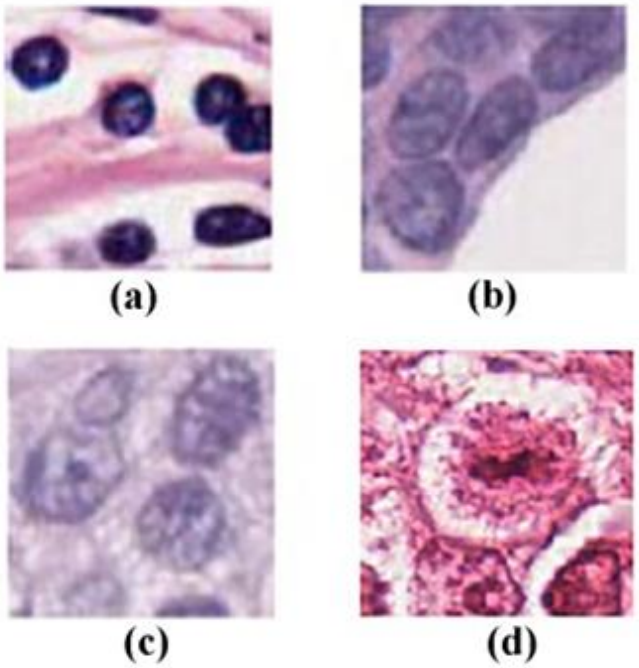

Fig.3. a) Lymphocyte (LN), b) Normal Epithelial nuclei (EN), c) Cancerous Epithelial Nuclei (CN) and d) Mitotic nuclei (MN)

The process of dividing region of interest amid the backdrop is known as segmentation and is a complicated job. Basically segmentation might be utilized to discover nuclei, stroma and context [9]. Multiple techniques have been presented to accomplish accurate nuclei segmentation. In spite of the intelligibility, practices built on thresholding and structural operations can conduct admirably on unvaried backgrounds but they are fragile to magnitude, structure and texture modifications. Fundamental segmentation procedures such as Hidden Markov Model (HMM), Active Contour Model (ACM), watershed Model and expanded models of later are used in identification.

a. Feature Extraction

Diagnosis of disease accurately depends upon the perfection of feature extraction. There are various types of features in medical images and in histopathology images, features are mainly classified into three types .i.e. Object-level properties, Architectural properties, Global and window based properties. Object–level properties mainly rely on the reviewed pieces like nuclei or glands and on the segmentation techniques. The previous mentioned properties in several cases are withdrawn from HD images. Object– level properties are generally withdrawn for every color passage and could be clustered to magnitude and appearance properties (e.g. region, weirdness, reflection uniformity), photometric and density based properties (e.g. image range, frigidity, shade), texture properties (e.g. co– occurrence matrix, run–length and wavelet properties) and chromatin–determined properties (average and combined image solidity) [10]. Architectural properties are devised with vertices (also called nodes or points) those are joined by edges (arcs or lines). Graphs are a productive way to constitute structural data utilizing topographical properties in histopathology images. Graph constructions utilized in histopathology consists of: the Voronoi tesselation, the Delaunay triangulation, minimum spanning trees, O'Callaghan Neighborhood graphs, connected graphs, relative neighbor graphs, k–Nearest Neighbour graphs [11].

Global and window dependent properties could be associated with color, smoothness (for example co–occurrence matrices, run–extent and wavelet properties). Mean object as well as structural properties are calculated comprehensively or with windows at numerous intentions.

b. Feature Selection & Dimensionality Reduction

All the features in a model may not be required to diagnose the problem in a image. So a subset of features which are not essential may be eliminated. Feature selection is the procedure of extracting set of applicable properties to a framework design therefore decreasing training intervals, clarifying the models making evaluation simple and enhancing the rate of deduction, circumventing over fitting. We have three important feature selection procedures: filters (e.g. Knowledge attainment), wrappers (e.g. inspection monitored through precision) and implanted techniquess (where these properties will be appended else deducted from the framework based on forecasted differences). popular feature selection techniques are interpreted with consecutive advancing selection and consecutive reverse selection [12], sequential floating forward search — one amongst the repeatedly utilized process in pathology image examination[11] and sequential floating backward search [12]. Further accessible procedures are genetic algorithms, simulated annealing, boosting and grafting.

Dimensionality reduction helps in reducing the storage space and computation time. It also helps in removing multi–co linearity, hence rectifying the efficiency of machine learning techniques. Dimensionality reduction improves information visualization (decreasing to small dimensions like 2D or 3D). There are various linear data transformation techniques like Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) , many non–linear dimensionality reduction techniques like manifold learning, non–linear embedding that are constructed and shown as helpful towards analysis of histopathology image [13].

c. Classification

Classification aims at categorizing a latest sample amidst a collection of samples with support of a labeled training data. Based on the function, physiological shape, tissue composition and properties, categorization exactness defers. Random forests with a group of 18 properties (including, e.g. histogram based properties, serial relationship, total, mean, deviation, entropy, variation) procured 83% average exactness towards categorization of prostate tissue into 7 categories (Gleason grade 3, 4 and 5; benign stroma; benign hyperplasia; intraepithelial neoplasia and inflammation) [14]. There are many factors for estimating the behavior of different techniques of categorization. Classification accuracy (ϕ), Matthew's Correlation Coefficients (ρ), Specificity (τ) and sensitivity (λ) are few amidst them.

d. Deep Convolution Neural Networks

DCCN has grabbed a huge attention and fame in resolving the issues in many fields especially in medical image analysis. The medical imaging field is very quickly transforming aforesaid models to resolve and enhance a huge and variety collection of implementations. In deep learning, a convolution neural network (CNN, or ConvNet) is a category of deep neural networks, fundamentally pertained to study perceptible view. These are even called as shift invariant or space invariant artificial neural networks (SIANN), depending on their shared-weights structure and transferable uniform aspects.

4. PROPOSED WORK

The model proposed is exhibited in Fig. 4. The dataset considered is Histopathology images dataset on Breast Cancer. Once the pre-processing is completed, the input is moved via the stem module. This module consists various 3*3 convolution layers, 1*1 convolution layer, max pooling layer. We have seven 3*3 convolution layer joined at various phases , two filter-expansion layers (1*1 convolution layer). Inception-A layer includes four filter-expansion layers, three 3*3 convolution layer, and one average pooling layer. Inception-B layer consists of four filter-expansion phases, four 1*7 convolution layer, two 7*1 convolution layer and one average pooling layer. Inception-C layer includes four filter-expansion

layers, three 1*3 convolution layer, three 3*1 convolution layer and one average pooling layer. Reduction-A layer consists one filter-expansion layer, three 3*3 convolution layer, and one 3*3 max-pooling layer. The Reduction-B module consists of two filter-expansion layers, two 3*3 convolution layer, one 1*7 convolution layer, one 7*1 convolution layer and one 3*3 max pooling layer. The input and output of all these layers proceeds via filter concatenation method. The softmax layer includes four various output class: benign, stage-3, stage-2 and stage-1. The framework pulls a histopathology image as input and with draws module-wise feature characterization from the initial stem module till the final drop-out module. Depending on the feature characterization, the input histopathology image is categorized as one among the four produced stages. To estimate the deprivation of the presented architecture, we handed down cross entropy. The Softmax layer inputs the feature representation, fi and elucidates it to the output class. A probability score, pi is in addition allocated to the output class. If we describe the number of Breast Cancer stages as n, then pi is defined as

$pi = \exp(fi) / \sum_i \exp(fi)$, $i = 1, ..., n$

and $L = -\sum_i ti \log(pi)$,

where L is the loss of cross entropy for given architecture. Back propagation is utilized to compute the gradients of the architecture. If the ground truth of an histopathology image is denoted as ti , then,

$\partial L / \partial fi = pi - ti$

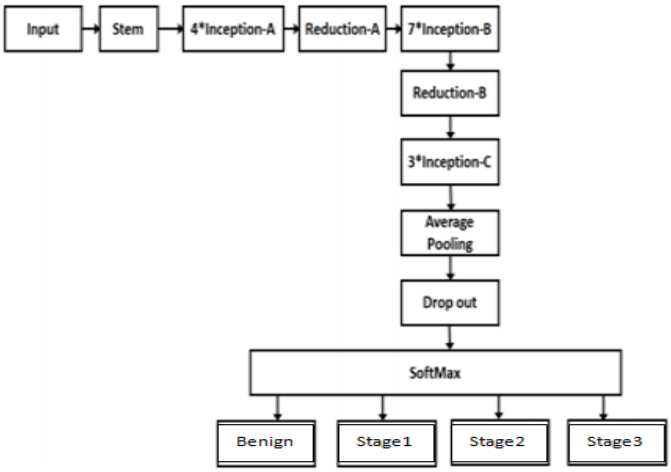

Fig.5. Structural representation of proposed Breast Cancer detection and classification framework.

We have variety of attainable combinations of the hyper-parameters for a framework. Plenty of time and labor gets wasted to deduce a steady set of values of hyperparameters for a framework. In order to decrease the aforesaid duration, hyperparameters of the InceptionV4 model is utilized rather than irregular assignment of values. The loads and constants of the inception-v4 directed in prior with association to Image Net database furnishes the network a systematic hyperparameter set.[15]. Consequently, inception-v4 has the perception of stronger characteristics detection and may utilize the same awareness for grasping properties from the minute medical image dataset.

5. CONCLUSION

Detection of Breast cancer in the early phases can increase the life expectancy of a human being. This paper proposes an architecture to examine a tissue as benign or malignant tissue and if malignant, then categorize the tissue into one of the multiple stages as stage-1, 2 and 3 where stage-3 being the case of high risk and stage-1 as low risk. Based on the stages, the type of treatments can vary and there is a possibility to increase the life expectancy.

6. REFERENCES

[1] Twombly R., “Cancer surpasses heart disease as leading cause of death for all but the very elderly”, J Natl Cancer Inst, 97(5), pp. 330–1, 2005

[2] Rubin, R., Strayer, D. S., Rubin, E., et al. Rubin's pathology: clinicopathologic foundations of medicine. Lippincott Williams & Wilkins, 2008.
[3] I. Ali, W.A. Wani, K. Saleem "Cancer scenario in India with future perspectives", J Cancer Ther, 8 (1) (2011), pp. 56-70.
[4] M. Yasmin, M. Sharif, S. Mohsin., "Survey paper on diagnosis of breast cancer using image processing techniques ", Res J Recent Sci, 2 (10) (2013), pp. 88-98.
[5] Jimenez-del-Toro, Oscar & Otálora Montenegro, Juan & Andersson, Mats & Euren, Kristian & Hedlund, Martin & Rousson, Mikael & Müller, Henning & Atzori, Manfredo. (2017). "Analysis of Histopathology Images". 10.1016/B978-0-12-812133-7.00010-7.
[6] A. Jain, S. Atey, S. Vinayak, V. Srivastava "Cancerous cell detection using histopathological image analysis", Int J Innov Res Comput Commun Engng, 2 (12) (2014), pp. 7419-7426
[7] Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., and Gurcan, M. N. "Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation." Pattern Recognition 42, 6 (2009), 1080–1092.
[8] Paramanandam M, O'Byrne M, Ghosh B, et al. "Automated Segmentation of Nuclei in Breast Cancer Histopathology Images". *PLoS One*. 2016;11(9):e0162053. Published 2016 Sep 20. doi:10.1371/journal.pone.0162053.
[9] R.C. Gonzalez, R.E. Woods, "Digital Image Processing "(3rd ed.), Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2008)
[10] Boucheron, L. E., " Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer". University of California at Santa Barbara, 2008.
[11] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. "Histopathological image analysis: A review. IEEE reviews in biomedical engineering" 2 (2009), 147–171.
[12] Pudil, P., Novovicov ˇ a´, J., and Kittler, J. "Floating search methods in feature selection." Pattern recognition letters 15, 11 (1994), 1119–1125.
[13] Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M., and Tomaszeweski, J. Automated grading of prostate cancer using architectural and textural image features. In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2007), IEEE, pp. 1284–1287.
[14] DiFranco, M. D., O'Hurley, G., Kay, E. W., Watson, R. W. G., and Cunningham, P. Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. Computerized medical imaging and graphics 35, 7 (2011), 629–645
[15] Islam, Jyoti & Zhang, Yanqing. (2017). A Novel Deep Learning Based Multi-class Classification Method for Alzheimer's Disease Detection Using Brain MRI Data. 213-222. 10.1007/978-3-319-70772-3_20.