ISSN: 0974-1011
A Special Issue of International Journal Of Computer Science And Applications Vol. 13, No.1, Jan-Mar 2020

A review on classification of breast cancer histopathological images using convolutional neural networks.

Gauri N. Gurav[1]
Dept. of Computer Science
Sheth L.U.J.& Sir M.V.College
Mumbai, India
guravgauri24@gmail.com

Mahendra G. Kanojia[2]
Dept. of Computer Science
[1]JJT University, Jhunjhunu
Rajasthan, India.
kgkmahendra@gmail.com

Abstract— **Breast Cancer is worldwide registered as one of the common threat to women. There is an increasing death rate in women because of breast cancer. Breast cancer can be cured or lifespan of the patient can be increased if it is detected at an early stage. It is necessary to validate patients histopathological conditions for detection of cancer. The histopathologist are the experts to examine the case. Performing the classification procedure manually is very time consuming and prone to error, based on human expertise. So to subdue this lack of accuracy and consumption of time, researchers around the world are experimenting with various soft-computing methodologies for automated diagnosis of breast cancer. Convolutional Neural Network (CNN) is impending neural network with deep learning capabilities and promising results for breast cancer classification. In this paper we give a review about CNN and hybrid-CNN based breast cancer classification models. The paper also reviews the work where deep learning environments such as GoogLeNet is used to achieve high accuracy and efficiency in detection of breast cancer.**

Keywords- Convolutional Neural Network (CNN); Deep Learning; GoogLeNet;

I. INTRODUCTION

Breast Cancer is a malignant tumour in the breast and eventually, it spreads to other organs in body. Breast cancer spreads majorly through the lymph system or through blood cells. It occurs in men and women both, but comparatively, male breast cancer is rare. Worldwide breast cancer is considered one of the most common cancer in women[20]. According to Globocan 2018 reports, among all cancers breast cancer shares 14% in women, new cases registered is 1,62,468 and 87,090 is the number of deaths due to this diseases[1]. Tarly stage detection of breast cancer is essential for the best results of the case and the survival of the patient. The process of detection is carried out by histopathologists with the help of tissue samples, it delays the diagnosis process and affected by external factors [3].

Researchers around the world are developing various techniques for early and accurate detection of breast cancer. The histopathological samples are converted into histopathological images that are classified into weather benign (non-cancerous) or malignant (cancerous). The artificial intelligence machine learning algorithms are implemented for automated breast cancer detection[13]. The convolutional neural network is recent and promising techniques for image classification. This paper gives insight of the types of CNN methodologies and its implementation in GoogLeNet [17] environment.

II. LITERATURE REVIEW

K. Kumar and A. C. S. Rao [2] extracted patches of images and used a convolutional neural network and classified the image into benign or malignant for breast cancer detection. The BreakHis database was used with 9,109 breast tumour images for classification, 90% accuracy is reported. The classification accuracy of CNN depends on the extraction of features in different layers with the variation in parameter S. Angara, et al.,[3] presented the neural network breast cancer classification for whole-slide histopathological images. The classification process includes patch generation. The BreakHis database with 7909 images were used to carry out the experiment. The research states that deep learning techniques can do an accurate classification of the histopathological images. In 2018 S. K. Jafarbiglo, et al.,[4] proposed a system for classifying histopathological images based on nuclear atypia criterion. The proposed method does data augmentation, data processing and feature extraction with CNN. Researchers could achieve an accuracy of 84.23%.Convolutional neural network model with image enhancement methodologies was proposed by A.-A. Nahid, et al.,[5] in the year 2017, where three machine learning models were tested for classifying histopathological breast images, the machine learning models were Conventional CNN model, Merge CNN Model and MaxMin Convolutional Model. The classification was performed using the BreakHis breast image dataset and the most accuracy was shown by the conventional model. F. A. Spanhol, et al.,[6] trained the CNN model using image patches and reported a better efficacy rate for histopathological image classification. Random selection and sliding window mechanism was implemented for image patches extraction and CNN model was trained using the BreakHis database. The proposed experiment inferred that improved accuracy was obtained by CNN when compared to traditional machine learning models. The imbalanced data, under-sampling, Over-sampling (WR), ADASYN and SMOTE sampling methods were used by M. S. Reza , et al.,[7] for breast cancer detection of histology images. The highest accuracy 0.8613 was achieved by Over-Sampling (WR) method which also increased the CNN performances substantially and acts as a better method in most cases. M. F. Jamaluddin, et al.,[8] proposed a automated diagnosis system were CNN was successfully implemented to classify whole slide breast carcinoma images, where cells were ER-stained. A region of 1200 cells was used and overall accuracy of 88.8% and AUC score of 97.5% was reported. The fCNN network was applied for pixel-wise region segmentation using TCGA breast cancer dataset where 92 images were cropped from 20 cases. The performance of fCNN was reported to be 20X faster than the original CNN. H. Su, et al.,[9] , M. Alkhaleefah and C.C. Wu [10] proposed RBF-SVM and hybrid CNN model for breast cancer classification. The authors had used CNN for feature extraction and classification, after feature extraction was done by radial basis function-based support vector machine. These research proved that the hybrid technique produced better results then the pre-trained models such as AlexNet and GoogLeNet. Work done by D. Lévy and A. Jain [11] proposed techniques that help in classifying mammogram images with pre-segmented breast masses. The convolutional neural network and transfer learning combined with a set of data enhancement techniques such as pre-processing and augmentation helped in the adequate classification of the cases. DDSM database was used to carry out the comparison between the Baseline, AlexNet and GoogLeNet models having 1820 images from 997 patients obtaining an accuracy of 0.604, 0.890 and 0.924 respectively. GoogLeNet outperformed other models in the experiment by a fair margin. In 2017 J. Chang, et al.,[12] implemented transfer learning method on google's inception V3 model for identification of breast cancer. The CNN deep neural network model ConvoNet was tested on 7909 microscopic biopsy images of 82 patients opted from the BreakHis database. The trained model gave an accuracy of 0.83 for benign and 0.89 for malignant class.

A unique study by S. Guan and M. Loew [13] reported that transfer learning had faster performance comparative to conventional CNN methodologies. Another comparative study by M. Żejmo, et al.,[14] used AlexNet and GoogLeNet models for dignosis of breast cancer in microscopic cytological images of 50 patients having 25 benign cases and 25 malignant cases. The analysis concluded that GoogLeNet had acquired an accuracy of 83%. A study by Y. Sun, et al., [15] present a deep neural network technique to identify DCIS. Similar to [14] GoogLeNet was reported to achieve the highest accuracy. Recently in 2019 J. Zhang, et al.,[16] studied image segmentation and classification of breast cancer photoacoustic images, in this study three machine learning techniques were compared SVM, AlexNet and GoogLeNet. The 217 breast cancer images of mammogram image database from LAPIMO was used for classification, in which AlexNet and GoogLeNet respectively achieved accuracy about 87.69% and 91.18% and SVM model achieved 84% accuracy which states that the pre-trained neural networks works better at classification. Similar to [14-16], H. Garud, et al.,[17] presented the efficiency of GoogleNet over other models for classification of high-magnification multi-views based on breast cytology cell images. The classification achieved the accuracy of 89.71%. This paper [17] stated that by using transfer learning GoogLeNet adapts visual features and ROI level classification accuracy of 80.76% is reported. Hybrid RNN and CNN model for breast cancer image classification with the spatial correlations between patches through RNN was introduced by R. Yan, et al.,[18]. The experiment was performed on ImageNet dataset which achieved 83.1% average accuracy in patch-wise and 90.5% average accuracy in image-wise method. Spatial correlations classification of breast cancer images presented by J. Ye, et al.,[19] used CNN models inception V3, inception V4 and ResNet for comparative study .Another study proposed by H. M. Ahmad, et al., [20] compared transfer learning methods for diagnosis breast cancer such as AlexNet, GoogLeNet and ResNet50 amongst which ResNet outperforms all the other methods. Image and Patch wise classification was carried out and ResNet50 generated accuracy of 83% and 85% respectively.

M. K. Jain and A. K. Agarwal [21] has analysed CNN and machine learning models to recognise breast cancer. GoogLeNet, VGG, AlexNet CNN models and SVM, PCA, KNN machine learning models were used in the study. The BreakHis database was used at different magnifying factors for the comparisons and result stated that pre-trained networks perform better.

III. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network is a classification and object identification technique. It is based on architecture of artificial neural network [2]. Dataset of images are given as an input to train the network. The network learns the features based on the set parameters. The volume and grade of images in training data set has a greater impact on the efficiency of the network to learn. The training data set works as an outline to which the convolutional neural network refers while operating on testing set images. The testing dataset tests the convolutional neural network to check the working capability of the trained model. If the testing phase result is acceptable the training of the model stops, else the network is re-trained with new parameter setting. This process is recurrently applied the acceptable accuracy is achieved. Once the model is trained and tested is validated using unseen data. Data validation confirms the performance and accuracy of the designed convolutional neural network design and it is ready for future use.

ISSN: 0974-1011
Vol. 13, No.1, Jan-Mar 2020

IV. ARCHITECTURE OF CNN

Convolutional neural network is a multilayer recurrent artificial neural network [9].

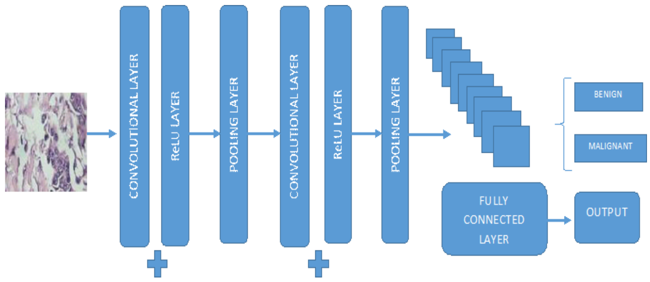

Figure 1:CNN architecture for breast cancer detection.

Layers of CNN works towards feature extraction. The network comprises of convolutional, pooling, ReLU and fully connected layers. Figure 1 shows the architecture of CNN for breast cancer detection. The histopathological image is given as an input to the model. The model is trained as binary classifier to classify the input image as malignant or benign based on trained parameters.

A. *Convolutional Layer*

Convolutional layer is the main layer of the CNN. Convolution is a mathematical operation where dot product between the two matrices are calculated [4]. In CNN, image given as an input is converted into matrix where each cell represents corresponding pixel value. Kernel is another sub-matrix use to convolute the image matrix. Kernel is a primary matrix used in generating feature map. Convolutional layer extracts desired features from the given image, it forms a matrix of the image pixels by learning the image features. The convolutional layer extracts locally interrelated pixel values by running the convolutional kernel with a defined set of weights on the image [6]. The below diagram is a basic representation if an image pixel matrix of 5*5 getting convolved in to 3*3 feature map.

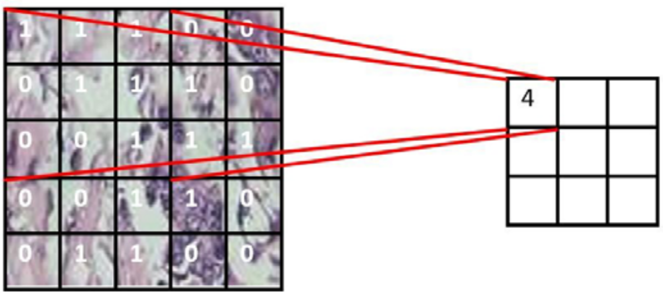

Figure 2: Working of convolutional layer

The image is convoluted with different kernels to identify features. In diagnosis of breast cancer using histology images nucleases are the object of interest. features related to nucleases are extracted for diagnosis of cancer. The convolutional layers are interlinked with each other. Output od pre-convolutional layer is given as input to post convolutional layer. This connection pattern can go deeper and learn the desired features maps. Hence, CNN is deep learning neural network.

B. *Pooling Layer*

The main function of the pooling layer is to size down the image by eliminating the values corresponding to redundant or unwanted features. Commonly used pooling techniques are max-pooling, average-pooling and spatial-pooling. The max-pooling is the most preferred methodology where the image is downsampled by selecting the highest values from each convolution step. The max-pooling is carried in certain steps, firstly the image input is received from the convolutional layer which has a convolved pixel matrix. Then the stride value is selected which helps the kernel to slide through the filtered image. Highest value is selected for each stride. Here the highest value is considered to be the best response for the feature map. Figure 3 is an example of max-pooling where the value of stride is 2 and a 6x6 pixel matrix is being downsized into a 3x3 matrix by selecting the highest value from each stride section. Stride 2 means the kernel is shifted 2 pixels during convolution.

1	5	6	4	2	3
3	1	7	2	4	2
7	6	5	2	3	5
6	2	1	2	6	2
1	2	5	1	4	6
7	1	1	1	2	5

Max Pooling
Stride 2

5	7	4
7	5	6
7	5	6

Figure 3: Working of Max Pooling

C. *Activation function or the ReLU layer*

The ReLU is a nonlinear activation function known as the rectified linear unit [2]. Breast cancer feature selection is based on nonlinear functions. This activation function helps in learning complex patterns in histology images. The ReLU layer activates when feature map has values above or below a certain limit. The main function of this activation function is to remove the negative values from the convolution and familiarize non-linearity[10]. For example, in figure 4 the negative values below zero are converted to 0 and the values with higher than zero threshold and kept as it is and figure 5 shows the graph plot of the matrix transformation. The ReLU layer smoothens the inputs to lessen the sensitivity of the filters. There are other activation functions as well such as tanh, maxout, sigmoid and other variants of ReLU however, the ReLU is preferred over the others because of better performance and gradient problem-solving.

10	-8	1	11
15	15	20	-4
25	45	-18	25
-35	35	20	16

Transformation function

10	0	1	11
15	15	20	0
25	45	0	25
0	35	20	16

ISSN: 0974-1011
Vol. 13, No.1, Jan-Mar 2020

Figure 4: Sample ReLU transformation

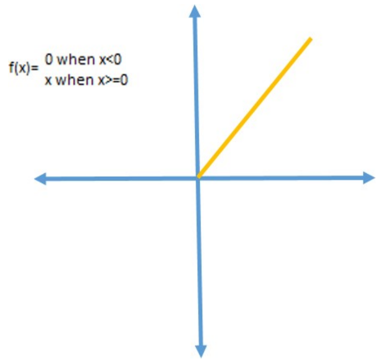

Figure 5: Response map of ReLU function

D. Fully Connected Layer

In this layer kernels are fully connected to each other. The matrix of feature map is converted into a single vector. The extracted feature vector is used to create a CNN model [10,22]. Prominent activation functions like liner, gaussian or sigmoid for classification or object recognition are included in this model. In case of breast cancer detection the output is a binary classification benign or malignant.

V. COMPARATIVE STUDY

recent years study on automation breast cancer detection using histology images and CNN has been increasing. Researchers are experimenting with various CNN and hybrid-CNN models on various platforms. The number of research in the domain also appeals for comparative understating of the work done in the past few years. Table 1 gives comparative tabulation of work done in classification and identification of breast cancer using CNN and hybrid-CNN models.

Table 1: CNN models and their result comparison

Author	**Objective**	**Proposed Methodology**	**Result**
Kundan Kumar [2], 2018	Classification of histology images into benign or malignant.	CNN	Accurately to 90 %.
Sandeep Angara [3], 2018	Identification of breast cancer using WSI	CNN	Highest accuracy for patch -wise 5 folds.
Sanaz Karimi Jafarbiglo [4], 2018	Classification of histology images based on nuclear atypia criterion	CNN	Accuracy :84. 23%
Abdullah-Al Nahid [5], 2017	Classification of histology images into benign and malignant classes.	Retinex algorithm, Contrast and Illumination Correction,CNN	Conventional CNN model performs better than the compared models
Fabio Alexandre Spanhol [6], 2016	To obtaining better efficiency rate for histopathological image classification.	CNN	Proposed CNN model performs better then the compared models.
Md Shamim Reza [7], 2018	To classify BreakHis histopathological images	CNN	Over-Sampling (WR) Accuracy: 0.8613
Mohammad F. Jamaluddin [8], 2018	To classify cells in ER-stained histopathological whole slide breast carcinoma images.	CNN Model ,Cell Detection and PN Classification	Accuracy: 88.8% AUC score: 97.5%.
Hai Su [9], 2015	Pixel-wise segment histopathological images	fCNN	fCNN is 20X faster compared to CNN
Mohammad Alkhaleefah [10], 2018	Malignancy detection using hybrid model in mammogram images	Support Vector Machine (SVM) ,Radial Basis Function (RBF) ,CNN	Proposed Hybrid CNN performance is comparatively better than pre trained CNN networks
DanielLévy [11], 2016	Classifying mammogram images with pre-segmented breast masses as benign or malignant.	Transfer learning, CNN, GoogLeNet , AlexNet.	Accuracy: Baseline-0.604, AlexNet-0.890, GoogLeNet -0.924.
Jongwon Chang [12], 2017	Implementation of transfer learning for	Transfer learning, CNN, Google's	Accuracy benign: 0.83 Accuracy

	malignancy detection.	Inception v3	malignant :0.89
Shuyue Guan [13], 2017	To detect cancer in breast mammogram images	Transfer learning Pre-trained VGG-16 model,CNN	Abnormal vs. normal accuracy: 0.905
Micha l ˙Zejmo1 [14], 2017	Classification of breast cancer cytological images.	GoogLeNet and AlexNet, CNN	GoogLeNet accuracy : 83%
Yibao Sun [15], 2018	To use deep learning based system to automatically identify (DCIS) using deep learning methodologies.	CNN, GoogLeNet	Accuracy of GoogLeNet-V1 is comparatively better then AlexNet, VGG 11 and ResNet 18 for used dataset
Jiayao Zhang1 [16], 2018	Classification of photoacoustic breast cancer images.	SVM,AlexNet and GoogLeNet	AlexNet Accuracy: 87.69% , GoogLeNet Accuracy: 91.18% SVM Accuracy: 84%
Hrushikesh Garud [17], 2017	Recognition of breast cancer using high magnification multi-views cell images.	CNN, GoogLeNet	Accuracy: 89.71%.
Rui Yan [18], 2018	Classification of breast cancer pathological image.	Hybrid CNN and CNN	Patch-wise accuracy: 83.1% . Image-wise accuracy: 90.5%
Jiandong Ye [19], 2019	Classification of breast cancer using WSI	Inception v3, Inception v4, ResNet, VGG 16, CNN	Proposed methods shows 10% more accuracy than classical CNN
Hafiz Mughees Ahmad [20], 2019	Classification of histopathological image into benign or malignant.	AlexNet, GoogLeNet, and ResNet.	Accuracy ResNet50: Patch-wise 83% and image-wise 85% .
Mradul Kumar Jain [21], 2019	To compare CNN and machine learning models for identification of breast cancer using histology images.	CNN frameworks, VGGm, GoogLeNet, AlexNet, SVM PCA, KNN.	Comparison States: Pre-trained methods perform better.

VI. DISCUSSION

The paper presents in depth review of the work done towards diagnosis and classification of breast cancer using CNN and Hybrid-CNN models [10,18]. Results in Table 1 shows that accuracy reported by hybrid-CNN higher than conventional CNN. Transfer learning [12,13,20] is a methodology where the pretrained CNN models are used to design the customized CNN model for case based research. Promising results are reported using GoogLeNet amongst the competitive pretrained[11,14-17,20,21]. Inclination of researchers towards GoogLeNet and pretrained models are increasing. Image processing [13][20] and machine learning models [10,16,18,21] used for classification of breast cancer are limited to the trained magnification and quality of images. CNN can be trained for different magnification and quality of images [21]. Instances shows that CNN requires no image pre-processing implementations for analysis of breast cancer images [6,10]. Improvisations like the fCNN which according to the research performs 20x faster than the regular CNN [9]. The CNN is the recent times technique which has copious benefits and space to explore with.

REFERENCES

[1] http://cancerindia.org.in/cancer-statistics/

[2] K. Kumar and A. C. S. Rao, "Breast cancer classification of image using convolutional neural network," 2018 4th International Conference on Recent Advances in Information Technology (RAIT), 2018.

[3] S. Angara, M. Robinson, and P. Guillen-Rondon, "Convolutional Neural Networks for Breast Cancer Histopathological Image Classification," 2018 4th International Conference on Big Data and Information Analytics (BigDIA), 2018.

[4] S. K. Jafarbiglo, H. Danyali, and M. S. Helfroush, "Nuclear Atypia Grading in Histopathological Images of Breast Cancer Using Convolutional Neural Networks," 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), 2018.

[5] A.-A. Nahid, F. B. Ali, and Y. Kong, "Histopathological breast-image classification with image enhancement by convolutional neural network," 2017 20th International Conference of Computer

[6] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," 2016 International Joint Conference on Neural Networks (IJCNN), 2016.

[7] M. S. Reza and J. Ma, "Imbalanced Histopathological Breast Cancer Image Classification with Convolutional Neural Network," 2018 14th IEEE International Conference on Signal Processing (ICSP), 2018.

[8] M. F. Jamaluddin, M. F. A. Fauzi, F. S. Abas, J. T. H. Lee, S. Y. Khor, K. H. Teoh, and L. M. Looi, "Cell Classification in ER-Stained Whole Slide Breast Cancer Images Using Convolutional Neural Network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018.

[9] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015.

[10] M. Alkhaleefah and C.-C. Wu, "A Hybrid CNN and RBF-Based SVM Approach for Breast Cancer Classification in Mammograms," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018.

[11] D. Lévy and A. Jain, "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks," 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[12] J. Chang, J. Yu, T. Han, H.-J. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 2017.

[13] S. Guan and M. Loew, "Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks," 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2017.

[14] M. Żejmo, M. Kowal, J. Korbicz, and R. Monczak, "Classification of breast cancer cytological specimen using convolutional neural network," Journal of Physics: Conference Series, vol. 783, p. 012060, 2017.

[15] Y. Sun, Z. Xu, C. Strell, C. F. Moro, F. Warnberg, L. Dong, and Q. Zhang, "Detection of Breast Tumour Tissue Regions in Histopathological Images using Convolutional Neural Networks," 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), 2018.

[16] J. Zhang, B. Chen, M. Zhou, H. Lan, and F. Gao, "Photoacoustic Image Classification and Segmentation of Breast Cancer: A Feasibility Study," IEEE Access, vol. 7, pp. 5457–5466, 2019.

[17] H. Garud, S. P. K. Karri, D. Sheet, J. Chatterjee, M. Mahadevappa, A. K. Ray, A. Ghosh, and A. K. Maity, "High-Magnification Multi-views Based Classification of Breast Fine Needle Aspiration Cytology Cell Samples Using Fusion of Decisions from Deep Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017.

[18] R. Yan, F. Ren, Z. Wang, L. Wang, Y. Ren, Y. Liu, X. Rao, C. Zheng, and F. Zhang, "A Hybrid Convolutional and Recurrent Deep Neural Network for Breast Cancer Pathological Image Classification," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018.

[19] J. Ye, Y. Luo, C. Zhu, F. Liu, and Y. Zhang, "Breast Cancer Image Classification on WSI with Spatial Correlations," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

[20] H. M. Ahmad, S. Ghuffar, and K. Khurshid, "Classification of Breast Cancer Histology Images Using Transfer Learning," 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 2019.

[21] M. K. Jain and A. K. Agarwal, "GPU Based Digital Histopathology and Diagnostic Support System for Breast Cancer Detection: A Comparison of CNN Models and Machine Learning Models," International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April, 2019.

[22] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[23] K. Elissa, "Title of paper if known," unpublished.

[24] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[25] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[26] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.