International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning-Based Classification of Histopathology Images for Cancer Diagnosis

[1]Shatakshi Lall, [2]Dr. Anand Gudur, [3]Dr.Ankur Jethlia, [4]Dr Versha Prasad

Submitted:23/03/2023 **Revised:**28/05/2023 **Accepted:**11/06/2023

Abstract: In this study, we apply deep learning techniques to create a complete CAD system for accurate and efficient IDC/non-IDC categorization. This CAD system has two distinct classification methods: machine learning-based classification using a variety of classifiers and deep learning-based classification using a specially constructed convolutional-neural network (CNN). Kaggle, a publicly accessible benchmark database, is used to accomplish this study. Accuracy, sensitivity, specificity, false positive rate, classification error, and precision are only few of the performance metrics used to assess machine learning and deep learning classifiers. Accuracy and sensitivity are selected as the primary characteristics by which the best classifier is evaluated. The purpose of this new saliency detection method is to aid in the diagnosis of invasive ductal carcinoma (IDC) by using IDC histopathology pictures to train deep learning algorithms.

Keywords: *Invasive ductal carcinoma (IDC), convolutional-neural network (CNN), Kaggle database*

1. Introduction

Breast cancer, which develops in tissue cells, is now the most lethal illness affecting females worldwide. Because of its lethal effects, most nations, and notably the industrialized ones, are investing in breast cancer identification at an early stage, when it has a better chance of being treated successfully. Recognizing breast cancer early increases the likelihood of a full recovery from the disease. The proper diagnosis of breast cancer is hampered, however, by the disparity between the rising number of patients and the shortage of experienced pathologists. Research in the field of modern medical image processing has shown encouraging findings, which might improve the interpretation of breast biopsy pictures by highlighting potentially harmful areas. By flagging suspicious or ambiguous areas in whole slide images, computer aided detection (CAD) has been shown to be a useful tool for breast cancer detection and classification, allowing for the identification of cancer and reducing the mortality rate among women with breast cancer. Using computer technology, these CAD systems can automatically identify anomalies in biopsy pictures. After being given the training datasets, the classifier generates probability plots for classes, which are then tested on the test dataset to determine whether or not the photos should be classified as IDC. Cancer is a major health concern since it is so deadly. This malignancy begins in the cells that comprise the genetic makeup of tissues. Breasts are only one region of the human body where tissues may be discovered. Cells are created and split as needed by the body to maintain life and promote growth. When healthy, regularly dividing cells age and shrink, they die and are replaced by new cells, thus the total number of cells always remains constant [1]. This procedure does not always go as expected. Some new cells are made when they are unnecessary, while some old cells do not die off to make room for new ones. Cancer, the abnormal formation of cells into a mass of tissue commonly known as a lump, tumor, or growth, is further classified into various categories, one of which is breast cancer[2]. One of the leading killers of women in the last several decades has been breast cancer. Early identification of breast cancer is a priority in many nations since it may greatly increase a patient's chance of living a long, healthy life once the disease has already spread. However, accurate diagnosis of breast cancer is made difficult by a disparity between the rising patient population and the shortage of trained pathologists. Since breast escalation disease is classified pathologically based on nuclear morphometric structures and cellular architectural patterns, distinct pathologies continue to face difficulties in categorization due to the variability in tumor development patterns. Misdiagnosis

1Asst. Professor, School of Pharmacy Graphic Era Hill University, Dehradun Uttarakhand 248002,
slall@gehu.ac.in
2Krishna Institute of Medical Sciences, Krishna Vishwa Vidyapeeth "Deemed To Be University" Karad Malkapur, Karad (Dist. Satara), Maharashtra, India. PIN – 415539
anandgudur@gmail.com
3Assistant Professor Department of Maxillofacial surgery and Diagnostic sciences Diagnostic Division College of Dentistry Jazan University Jazan,Saudi Arabia
jethliaankur@yahoo.co.in
4Assistant Professor School of Health Sciences C.S. J. M. University Kanpur"

is exacerbated by the unequal distribution of healthcare resources.[3] Many people throughout the globe are affected by breast cancer, which may be fatal if left untreated. To avoid a mastectomy and reduce mortality, categorization is the best option [4]. Examining biological samples of tissue from a patient by a pathologist is a cautious approach to disease analysis, especially for breast cancer. Examining tissue samples, however, is a laborious and time-consuming process that might slow down the diagnostic process. Pathology slide review is a challenging process. For certain forms of breast cancer, the covenant in judgment may be reduced to 48% [5]. Pathologists have to put in a lot of time diagnosing illnesses since they have to examine several slides for each patient. However, current attempts at automating medical analysis are not geared for pathologists who lack extensive background in artificial intelligence. A pathologist may struggle to understand statistical AI terminology or other technical jargon. The possibility exists that a pathologist will be unable to make adequate use of a computer-generated report. The combination of AI and these constraints has the potential to reduce waiting times while making decisions [6][7]. Therefore, a technique for identifying breast cancer subtypes from histopathology pictures has to be created. The goal of this computer-aided diagnostic system is to help the pathologist in their search for breast cancer.

Breast Cancer and Its Types

Because of clinical and pathological development, the progression of breast cancer from its earliest stages with normal epithelial, through hyperplasma and carcinoma-in-situ, to invasive carcinoma, and finally in metastatic disease, can be visualized. Breast lesions, both benign and malignant, are shown in Figure 1. Nearly all invasive breast cancers seem to have originated from in-situ carcinomas, even if the precise processes and underlying reasons for tumor formation remain unclear. Breast lesions are classified as in situ carcinoma, invasive carcinoma, or benign proliferative based on their pathological and biological characteristics. The most effective course of therapy for breast proliferative illness requires accurate analysis or diagnosis.

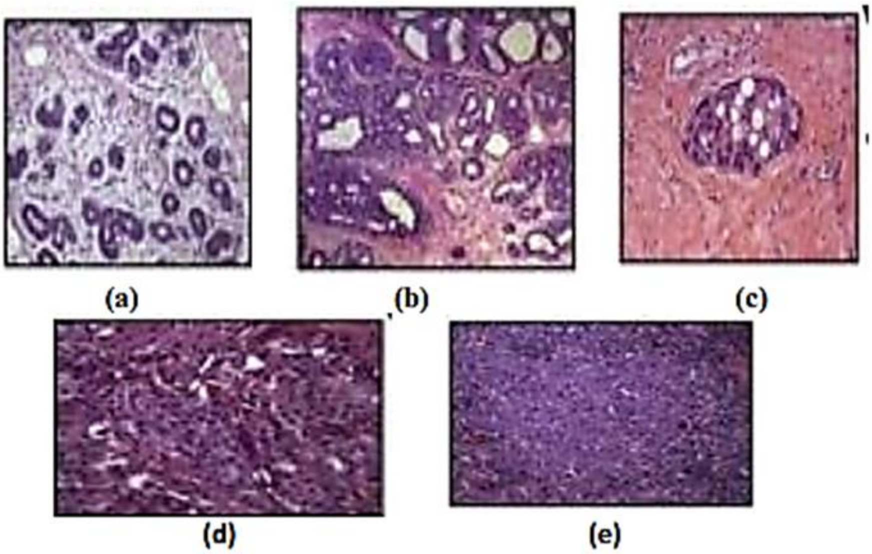


Fig. 1: Breast Cancer at Different Stages:(a) Normal-Breast, (b) Breast Hyper-plasia, (c) Ductal Carcinoma-In-Situ, (d) Invasive -Ductal Carcinoma, (e) Metastatic Disease

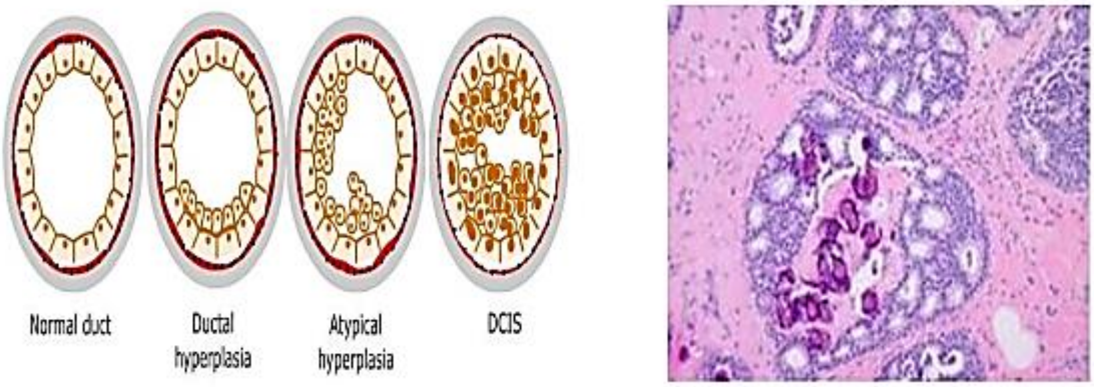
Both infiltrating/invasive and in situ/non-invasive breast carcinomas/cancers are distinguished by the unique tissue or cell configurations that characterize them. There are two distinct forms of breast cancer: lobular and ductal. Lobular carcinomas begin in the milk ducts, whereas ductal carcinomas originate in the duct glands.

Non-Invasive Breast Carcinoma

Breast lobules and milk ducts are safe against non-invasive cancer cells. There is no evidence that this kind of breast cancer may return to normal tissues or spread to other organs. This pre-cancer stage is often referred to as carcinoma in situ (which literally means "in the same place"). Both lobular and ductal carcinomas in situ (LCIS and DCIS) are subtypes of invasive ductal carcinoma.

Ductal Carcinoma In Situ (DCIS)

The most common kind of non-invasive cancer is DCIS, as seen in figure 2. Breast cancer develops from circulating ductal epithelial cells in a malignant mammary gland, which are lined throughout the milk ducts but lack proof of proliferation of cells beyond the duct wall.

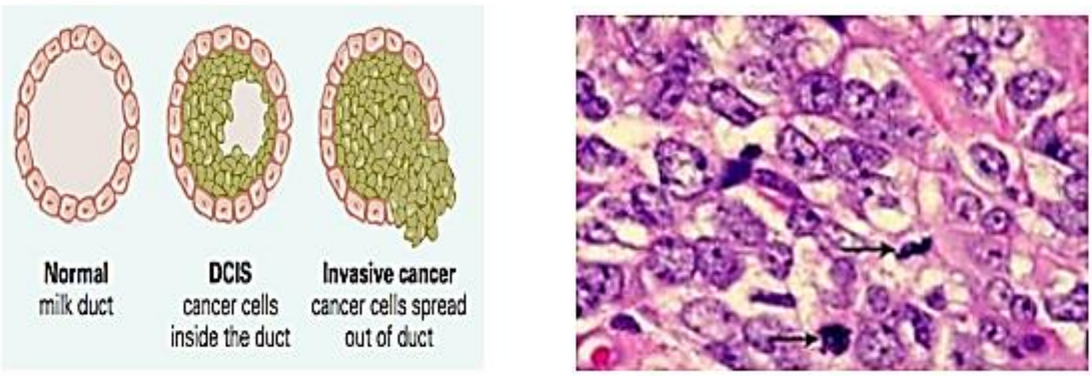


(a) (b)

Fig. 2: Ductal Carcinoma in Situ: (a) Cancer Cells in Ducts (b) Histopathology Image

Invasive Ductal Carcinoma (IDC)

This kind of breast cancer is the most common one. To be invasive, breast cancer must have invaded neighboring tissues. Figure 3 (a) depicts a typical case of infiltrating/invasive ductal carcinoma (IDC), a type of breast cancer that begins in the milk ducts and spreads laterally through the duct membrane before establishing itself in the fatty tissue directly beneath the breast. Figure 3 (b) depicts a histopathology image of this type of breast cancer. It may now spread via the lymphatic system and the circulation to other areas of the body close to the breast. Invasive ductal carcinoma has the potential to metastasize (spread to other parts of the body) over time, most often via the lymph nodes.

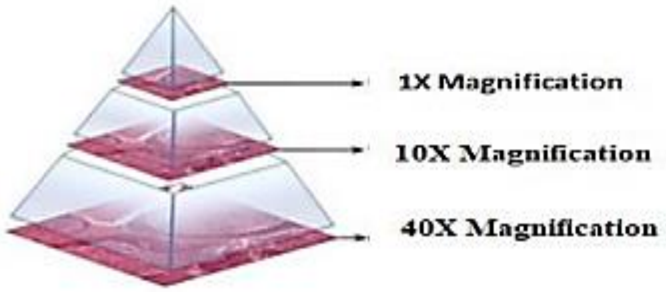


(a) (b)

Fig. 3: Invasive Ductal Carcinoma (IDC): (a) Cancer Cell in Ducts, (b) Histopathology Image

Breast Cancer Diagnosis using Histopathology Images

Individuals who have reason to suspect they may have breast cancer due to test results or symptoms are referred for further testing, including a biopsy. Breast cancer detection may be confirmed by screening exams or biopsies. The purpose of diagnostic testing is to learn more about the characteristics of cancer cells. Breast lesions fall into one of three categories: benign proliferative, in situ carcinoma, or invasive cancer. Defining the best course of therapy for breast illness requires accurate diagnosis. Images created by the process of whole-slide imaging (WSI) are referred to as whole slide images. Whole slide photos of gigapixel size are often stored in a pyramid-like configuration of several resolutions. As can be seen in Figure 4, the digital data include several down-sampled variants of the original picture. Each pyramid picture is saved in a series of sequential files so that individual image areas may be quickly recovered. These pictures are not designed to be read with standard image libraries or experimental software because their compressed formats are too large to uncompress into RAM



.

Fig. 4: Digital Slides in Pyramid Structure

The cost of digital storage space has decreased in recent years, and technological advancements in slide scanning technology have made it possible to completely digitize the microscopic evaluation of stained, discolored tissue samples. As a result of these advancements, we have the potential for better clinical diagnosis and computerized or computer-aided diagnostics; moreover, we may benefit from remote diagnostics; immediate access to archive cases; tele-education; comfortable conversations with competent pathologists; and tele-education.

2. Literature Survey

AmirReza BabaAhmadi et.al(2022) To achieve high accuracy and minimal computing cost, this novel technique for detecting abnormal patches in breast pictures was designed. Thus, a new architecture using Xception and MobileNetV2 has been developed.This novel approach beats previously reported systems using deep learning techniques to detect IDC histopathology pictures, with a balanced accuracy of 93.4% and an F1-Score of 94.8%..

Shubham Kushwaha et.al(2021) The purpose of this work is to make breast cancer staging less complicated. Using histopathology pictures, we present a deep learning-based Convolutional Neural Network model for detecting and classifying breast cancer. Pretrained neural network DenseNet-201 is used to extract picture characteristics and provide predictions for classification in this technique. The precision of our model was 97.05 percent.

Yasin Yari et.al(2020) The model is fine-tuned with a deep classifier and data augmentation to distinguish between malignant and benign samples in binary and multi-class classification by first transferring the weights of a pre-trained DesneNet121 on the Imagenet as initial weights. The suggested model achieved up to 97% accuracy in multi-class classification using picture data. The model achieved up to 100% accuracy in image-level binary classification. Multiple performance measures for breast cancer CAD systems show that the results produced are superior than those of prior research. The suggested technique is also extensible, so it may be used to identify more illnesses in the future and combined with additional CNNs to improve its generalization capabilities.

Computer Aided Diagnosis (CAD) of Breast Cancer

The development of a diagnostic system is necessary to advance clinical diagnosis. To improve output accuracy and facilitate automated classification for entire slide pictures, a CAD system has been developed for cancer classification based on histopathology images. A computer-aided design (CAD) system is a method wherein clinicians are provided with digital help in deciphering medical pictures. One such technique is the one that uses artificial intelligence to tell the difference between healthy and cancerous breast tissue in entire slide images[30]. Histopathology image processing using a CAD system may be broken down into three primary phases. The first step in improving the accuracy of expected analysis or detection tasks is for pathologists to utilize a computer-aided design (CAD) system to segment the data. The morphology of glands is the primary criterion for grading cancer metastases, and pathologists use CAD systems to segment nuclei for diagnosis of nuclear morphology, inscriptions of segmentation of glandular structures, and cancer detection in lymph node sections. On the other hand, the primary challenge in diagnosis, especially when applied to cancer, is classification of histopathological images. Classifying lesions based on histopathological pictures is subjective and often involves assessing cancer. The categorization findings generated by a CAD system may be objective and accurate, providing the pathologist with useful information for making decisions. Breast lesion categorization, prostate cancer detection, and other uses are only a few examples. Invasive cancer, the Gleason scale, and DCIS. When all is said and done, the computerized tools used for disease prognosis and diagnosis gather relevant information that is then calculated based on subtle visual changes in the patterns of important structures in histopathology images that are otherwise indistinguishable or difficult to diagnose. This is mostly done for early-stage detection of cancer prognosis and may also aid in the subsequent clinical management of patients. Due to the many variables involved in histopathology categorization, and the fact that tissue samples are often stained unevenly in terms of color and intensity throughout complete slide pictures, the adoption of a computer-aided diagnosis (CAD) system has become more important. Additionally, the pathologist's ability to classify full slide pictures is hindered by the images' inconsistent hue and intensity. In addition, several artifacts are introduced into the digital picture of the slide and tissue sample during preparation, posing difficulties for automated assessment in certain cases. Lastly, the output accuracy of the algorithms of the CAD system for classification of histopathology pictures may be enhanced by the greater size of whole slide photos and the higher number of dissimilarity in the structure which are to be examined by whole slide images.

IDC Histopathology Image Classification Using Deep Learning

In recent years, deep learning has been widely used for medical picture classification. Deep learning entails several layers of interconnected nodes in a neural network between the input and the output. These layers provide the functions of feature detection and sequential processing. When compared to deep neural networks (DNN), which often contain numerous hidden layers, normal neural networks typically only have one or two. Histopathology pictures of breast cancer are classified and feature extraction (using image pixels) is performed using deep neural network techniques in this study. Instead of using the characteristics that were manually derived from the picture, we utilize the pixels themselves as the features. Later, several parameter analyses are performed on the multi-layered convolutional neural network used as a deep learning model for the classification of breast cancer histopathology pictures.

3. Deep Neural Network

An artificial neural network is a kind of complex computer system made up of a large number of linked nodes (or "neurons"). These neurons are placed in a hierarchical structure with each layer producing an activation of increasing real value. Figure 5 depicts a network with a hidden layer comprised of neurons activated by weighted connections from previously active neurons and an input layer consisting of a small number of neurons designed to collect a variety of data from the environment. The network's ability to learn from the input data is measured by the information provided by the output neuron. The output of a neuron is the product of its inputs in form of x, the weights that link them, and an activation function.

Rectified linear unit (ReLU) is a typically utilized activation function, or "selection of activation function," after a convolutional layer, while tanh function and sigmoid function are often employed at the network's output. By using a method called back propagation, weights are provided to neurons in a network throughout the learning process so that the network can optimally complete the job at hand.

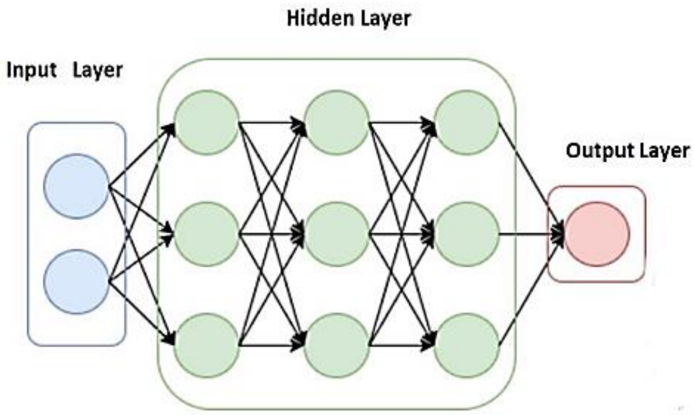


Fig. 5: Deep Neural Network

K Nearest Neighbor Classifier

Many scientists have turned to the K-NN classifier when they need to categorize data. K-NN is a kind of categorization algorithm that uses similar instances to help place new ones into categories. This is a kind of event-based learning in which familiarity with how information flows is assumed but not required.

K and the distance or separation measure used are the primary implementation controls for KNN classifiers. The correct value of the parameter "K" must be determined by the data. Choosing a large value for 'K' is the standard method for suppressing background noise in classification, however this often results in less clear separation between categories. Cross-validation is one method for selecting a good value for the parameter K, but there are others. The local estimate will suffer from confusing and incorrectly labeled points if K is relatively small. Now, we may use a bigger value of K to smooth the estimate, however this results in excessive smoothing and a drop in classifier performance owing to outliers. Finding the commonalities among the training samples is a prerequisite to identifying the k nearest neighbors. KNN classifier is no longer optimal when training samples are small, yet it takes longer to detect similarities when there are many examples. There are three approaches to fix this problem: reducing the size of the feature space, using a more condensed data collection, or using more sophisticated algorithms. The classifier is generated from the training samples alone, without any additional data. There is no difference between samples with a small number of data points and those with a huge number of data points. Thus, it fails to coordinate the true miracle in cases when the samples are not uniformly distributed.

Parameters	K=1	K=3	K=5	K=7
Accuracy	0.778	0.78	0.792	0.795
Classification Error	0.222	0.22	0.208	0.205
Sensitivity	0.758	0.782	0.817	0.809
Specificity	0.791	0.778	0.774	0.784
False Positive Rate	0.208	0.221	0.225	0.215
Precision	0.712	0.706	0.711	0.719

Table 1.KNN Classification values with accuracy

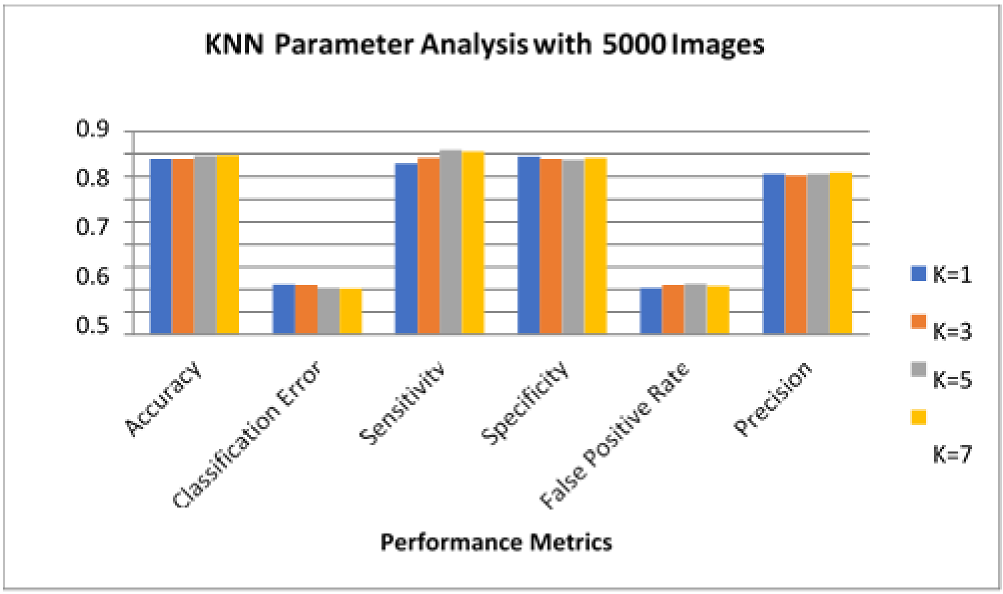


Fig. 6. KNN Performance Parameter Analysis with 5000 Images

4. Conclusion

The Kaggle database, a publicly accessible benchmarking resource, is used to assess the quality of this work. One hundred thousand histopathological photos yields an accuracy of 88.7 percent and a sensitivity of 92.6 percent. Histopathology pictures of breast cancer are effectively classified as either invasive ductal carcinoma or non-invasive ductal carcinoma with the use of a 19 K Nearest Neighbor Classifier, demonstrating improved classification accuracy and sensitivity. K-Nearest Neighbor Classifier and other machine learning methods may be used to improve classification accuracy. Therefore, the pathologist may rely on this computational approach for breast cancer categorization.

References

[1] BabaAhmadi, S. Khalafi and F. M. Esfahani, "Designing an Improved Deep Learning-Based Classifier for Breast Cancer Identification in Histopathology Images," *2022 International Conference on Machine Vision and Image Processing (MVIP)*, Ahvaz, Iran, Islamic Republic of, 2022, pp. 1-4, doi: 10.1109/MVIP53647.2022.9738774.

[2] S. Kushwaha, M. Adil, M. Abuzar, A. Nazeer and S. K. Singh, "Deep learning-based model for breast cancer histopathology image classification," *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, 2021, pp. 539-543, doi: 10.1109/ICIEM51511.2021.9445319.

[3] Y. Yari and H. Nguyen, "A State-of-the-art Deep Transfer Learning-Based Model for Accurate Breast Cancer Recognition in Histology Images," *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, OH, USA, 2020, pp. 900-905, doi: 10.1109/BIBE50027.2020.00153.

[4] C. P. Nguyen, A. Hoang Vo, and B. T. Nguyen, "Breast Cancer Histology Image Classification using Deep Learning," Proc. - 2019 19th Int. Symp. Commun. Inf. Technol. Isc. 2019, pp. 366–370, 2019, doi: 10.1109/ISCIT.2019.8905196.

[5] L. Conte, B. Tafuri, G. De Nunzio, M. Portaluri, and A. Galiano, "Breast Cancer Mass detection in DCE-MRI using Deep-Learning Features Followed by Discrimination of Infiltrative vs In Situ Carcinoma through a Machine-learning approach," no. August, 2020, doi: 10.20944/preprints202008.0179.v1.

[6] J. Dana, V. Agnus, F. Ouhmich, and B. Gallix, "Multimodality Imaging and Artificial Intelligence for Tumor Characterization: Current Status and Future Perspective," Semin. Nucl. Med., 2020, doi: 10.1053/j.semnuclmed.2020.07.003.

[7] Q. Liu, Z. Liu, S. Yong, K. Jia, and N. Razmjooy, "Computer-aided breast cancer diagnosis based on image segmentation and interval analysis," Automatika, vol. 61, no. 3, pp. 496–506, 2020, doi: 10.1080/00051144.2020.1785784.

[8] Yadavendra and S. Chand, "A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method," Mach. Vis. Appl., vol. 31, no. 6, pp. 1–10, 2020, doi: 10.1007/s00138-020- 01094-1.

[9] R. Krithiga and P. Geetha, "Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review," Arch. Comput. Methods Eng., no. 0123456789, 2020, doi: 10.1007/s11831-020-09470-w.

[10] A. J. Navarrete-Welton and D. A. Hashimoto, "Current applications of artificial intelligence for intraoperative decision support in surgery," Front. Med., 2020, doi: 10.1007/s11684-020-0784-7.

[11] Y. Tolkach, T. Dohmgörgen, M. Toma, and G. Kristiansen, "High-accuracy prostate cancer pathology using deep learning," Nat. Mach. Intell., vol. 2, no. 7, pp. 411–418, 2020, doi: 10.1038/s42256-020-0200-7.

[12] H. Moghadas-Dastjerdi, H. R. Sha-E-Tallat, L. Sannachi, A. Sadeghi-Naini, and G. J. Czarnota, "A priori prediction of tumour response to neoadjuvant chemotherapy in breast cancer patients using quantitative CT and machine learning," Sci. Rep., vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020- 67823-

[13] W. Sanderink and R. Mann, "Minimally invasive breast cancer excision using the Breast Lesion Excision System under ultrasound guidance," Submitted, no. 0123456789, 2020, doi: 10.1007/s10549-020-05814-z

[14] C. E. Vera-Tizatl et al., "Cell morphology impact on the set-up of electroporation protocols for in-suspension and adhered breast cancer cells," Electromagn. Biol. Med., vol. 00, no. 00, pp. 1–17, 2020, doi: 10.1080/15368378.2020.1799387.

[15] H. Chapala, "ResNet : Detection of Invasive Ductal Carcinoma in Breast Histopathology Images Using Deep Learning," no. Icesc, pp. 60–67, 2020

[16] Naga Swetha, G. ., & M. Sandi, A. . (2023). A Brand-New, Area - Efficient Architecture for the FFT Algorithm Designed for Implementation of FPGAs. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 114–122. https://doi.org/10.17762/ijritcc.v11i2.6135

[17] S. Ali, A. Tanveer, A. Hussain, and S. U. Rehman, "Identification of Cancer Disease Using Image Processing Approahes," vol. 9, no. 2, pp. 6–15, 2020, doi: 10.11648/j.ijiis.20200902.11.

[18] S. Robertson, B. Acs, M. Lippert, and J. Hartman, "Prognostic potential of automated Ki67 evaluation in breast cancer: different hot spot definitions versus true global score," Breast Cancer Res. Treat., vol. 183, no. 1, pp. 161–175, 2020, doi: 10.1007/s10549-020-05752-w

[19] K. Taunk et al., The development and clinical applications of proteomics: an Indian perspective, vol. 0, no. 0. Taylor & Francis, 2020

[20] AGYEI , I. T. . (2021). Simulating HRM Technology Operations in Contemporary Retailing . International Journal of New Practices in Management and Engineering, 10(02), 10–14. https://doi.org/10.17762/ijnpme.v10i02.132

[21] N. Maroof, A. Khan, S. A. Qureshi, A. ul Rehman, R. K. Khalil, and S. O. Shim, "Mitosis detection in breast cancer histopathology images using hybrid feature space," Photodiagnosis Photodyn. Ther., vol. 31, p. 101885, 2020, doi: 10.1016/j.pdpdt.2020.101885

[22] A. Aloyayri, "Breast Cancer Classification from Histopathological Images Using Transfer Learning and Deep Neural Networks School of Graduate Studies," no. April, 2020