Available online at : http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 3, issue 5, pp. 973- 979, May -2023
https://doi.org/10.53633/ijmasri

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

BREAST CANCER PREDICTION USING BIOPSY IMAGES

Khushboo Goel[1], Kriti Mukhija[2], Prachi Bansal[3], Ms. Deepti Gupta[4]and Ms. Kavita Saxena[5]

[1, 2,3] Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi
[4,5] Assistant professor, Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi
Email: khushboogoel.312@gmail.com, kriti.mukhija@gmail.com, prachibansal2111@gmailcom

Abstract

Breast cancer is one of the top causes of death among women. Early detection of breast cancer can lead to an efficient prognosis. The standard method of diagnosing cancer relies on detecting cancer-related anomalies in tumor biopsy images. The ability to recognize the landmarks in the visual artefacts included in the slide images is necessary for detection. Machine learning-based medical image processing algorithms are continuously being created to aid in the detection of tumors. Machine learning and artificial intelligence are essential for the analysis of cancer images. It has aided medical professionals in drawing quick and precise conclusions on the type of cancer. In order to automatically distinguish between malignant and benign tumors using biopsy images, this research proposes an effective convolutional neural network-based (CNN-based) model. A fresh dataset was used to train the Resnet50 architecture for feature extraction. The outcome showed that the suggested model performed at the cutting edge with 97% accuracy. Thus, a pre-built Resnet50 architecture-based CNN model that is trustworthy, accurate, and consistent has been created.

Keywords: Classification, histopathology images, convolutional neural network, breast cancer, BreakHis

Introduction

According to the International Agency for Research (IARC), breast cancer accounts for 11.7% of all new instances of cancer globally. In most nations, breast cancer has the highest incidence and fatality rates among women. Through early disease discovery and treatment, this high mortality rate can be decreased. Fine needle aspiration (FNA), a manual method of detection, is a rapid and easy cytology approach to find cysts or cancer by studying the fluid characteristics retrieved from the suspected location.

This FNA-related pathological difficulty motivates scientists to develop a brand-new, autonomous, efficient approach for analysing and identifying malignant cells. The convolutional neural network

(CNN), which made a substantial contribution to the automated detection of tumour images and offers additional support to healthcare professionals so they may make informed judgements, has made consistent progress over the past ten years. In order to diagnose and predict diseases, research is concentrating on computer-aided diagnostic (CAD) techniques. According to CNN, a diagnosis of breast cancer is made by looking at images obtained using various medical devices, such as mammography, breast ultrasonography, magnetic resonance imaging (MRI), or open histology. A typical CAD system has four common steps: (i) input images, (ii) pre-processing, (iii) feature extraction, and (iv) classification. After these steps, results are produced.

Researchers offered a dataset and used a two-class computer-based classification in the previously published work, obtaining accuracy between 80% and 85%. Given this context, the goal of this study is to present a highly efficient classifier that can accurately identify malignant and benign tumours in breast histopathology photos. Fig. 1 displays some illustrations of both benign and malignant tumours.

This system, which employs Resnet-50, has a 97% accuracy rate in classifying the kind of breast cancer

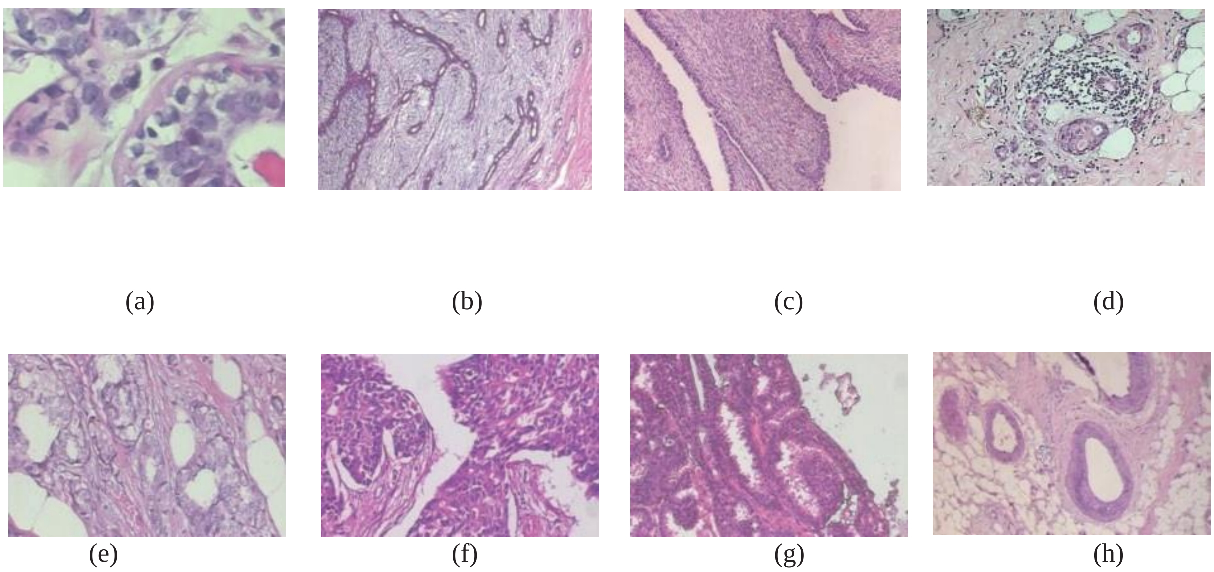

Figure 1: *Different breast cancer biopsy images (a–d) Benign tumor images. (e–h) Malignant tumor image*

Literature survey

This literature review includes the research papers that used pre-trained systems for breast cancer classification. Some substantial and relevant studies have been discussed in this section.

In a study, researcher have developed a deep feature-based model (DeCAF) (Spanhol 2017) for the classification of breast histopathology images, and the pre-trained BVLC Caffenet model, was implemented as a feature extractor. Various outputs obtained from different layers were examined individually in different combinations. For the development of these deep features, logistic regression was used as a base classifier. The experimental outcome has exhibited a high accuracy rate and better performance than hand-made features for image recognition. In an experimental work (Song 2017) a transfer-learning method was employed to classify histopathological images, and a pre-trained CNN was used to encode the local features into Fisher vector (FV). Authors also introduced a new adaptation layer that enhances discriminant characteristics of features and improves classification accuracy. VGG-VD was implemented for accomplishing transfer learning. The study (Zhang 2014) presented a classification scheme to generate and assign confidence score to images of each class. The classification was based on the Kernel principal component analysis (KPCA) ensemble method, and the proposed scheme was successfully tested on two different datasets: (a) breast cancer and (b) optical coherence tomography. This study has also imposed a research challenge to find an optimiser, to reduce the fixed size parameters applied in this work. In the reported work (Krizhevsky 2017) 1.2 million high-resolution images have been categorised in 1,000 different classes through a large, deeper and focused neural network in the ImageNet (LSVRC-2010) competition. This novel model consisted of 60 million parameters, 650,000 neurons and five convolution stages. The fully connected layers applied the SoftMax function. Experiment outcome reported the lowest error rates of 37.5% and 17.0%, respectively. The research work (Barker 2016) provides an automatic two-step classifier for slide investigation. In the first step, coarse regions were analysed in the whole slide to obtain diverse spatially localised features, after which the detailed analysis of the selected tiled region was completed. This experiment used elastic net classifier (Kandemir 2014) and weighted voting classifier for cancer detection on brain image dataset. The research paper presented a probabilistic classifier that combines multiple-instance learning and relational learning. The instance-based learning was used for the classification task, whereas relationship-learning map was used for the changes in the cell formation due to cancer. This approach had been evaluated on breast and Barrett's cancer tissue microarray datasets. In another work, the researcher had developed an accurate, reliable, and multistage method for measuring the neoplastic nuclei size for pleomorphism grading (Cosatto 2008). In the preliminary stage, image quality, staining value and tissue appearance were observed. In a later stage, machine-learning methods were used to score and remove abnormality that existed in nuclear contour, for better segmentation. A study has discussed difficulties associated with multiclass classification due to broad variability, high coherency, and inhomogeneity of colour distribution. In this paper, authors had resolved these issues and proposed a breast cancer multi- classification method, which used a structured deep-learning model (Han 2017). This model includes a training stage, extracting features and optimises inter-class distances. The validation stage was used for parameter tuning, while the testing stage comprised model evaluation. The final accuracy was obtained as 93.2% for multi-class binary classification on the breast biopsy image dataset. A study has addressed an intra-embedding supervised classification method to classify histopathology images; in this innovative work (Song 2017), Fisher encodings are embedded with CNN architecture. The results of the work concluded that the proposed method can successfully resolve the issues of high dimensionality and busty visual components, which are associated with FV. This model was evaluated in lymphoma and breast cancer datasets, where highest classification accuracies on different magnification levels were reported on the breast cancer histopathology dataset. An experimental work (Gupta 2017) was focused on the use of colour-textural characteristics of the breast cancer histopathology images, for effective classification. In this work, different classifiers, namely support vector machine (SVM), decision tree, nearest neighbour and discriminant analysis, were included for creating an ensemble classifier. The proposed model was studied on various optical magnification levels, to acquire better discriminative features. A CNN-based approach was applied in (Araujo 2017) where patch-level classifications were carried out on breast cancer biopsy images. The maximum accuracies reported 77.8% for four classes and 83.3% for two classes, for instance, carcinoma/non-carcinoma through the CNN-SVM model. Another approach (Kooi 2017) had examined 74 features and their importance was inferred, based on pixel level, contrast and texture. Additional features were also studied like geometry and context for significant contribution, to achieve a better accuracy rate. In the novel work (Matlani 2019) researchers have applied CNN for correct detection of smoke in a region from video frames.

Based on earlier research work, it can be summarised that the use of machine-learning methods, especially deep learning algorithm, can be efficiently optimised for cancer cell detection, segmentation and classification. Review of literature also highlighted the importance of input image quality, feature extraction and classifier for development of an effective computer-based system. A lot of contributions have already been made in the recent years, but there is always scope for achieving better classification accuracy, which is the basis of our research work.

Functional Description

Inorder to identify and categorize the picture as malignant or benign, this model employs ResNet-50.ResNet is a deep learning model that has already been trained to classify images. ResNet – 50 features a deep architecture that is superior for image recongnition and performs very well at object detection and picture categorization. It has 50 layers and was honed using one million pictures from the ImageNet database divided into a thousand categories.

Breast cancer Dataset

In this study, we have used a BreakHis open dataset, which contains different breast cancer classes in two main categories, benign and malignant. A total of 3,880 images. Out of these, 1800 belonged to the benign and 2,080 to the malignant class. Each image has been resized to 224 pixels. The complete dataset was divided into training (3720 images), test subsets (160 images).

Proposed system

This model is trained using residual neural network, or ResNet-50 to recognise breast cancer from trained images data. Residual networks are capable of skipping one or more network levels. Typically, these skipped layers contain both batch normalization and nonlinearity. Dense networks are networks with a lot of parallel skips, whereas plain networks are networks without any parallel skips. Layers are skipped because of diminishing gradients. Prior layers' weights can be reused as the subsequent layer learns its own. While explicit weights from the upstream layer are not employed, weights from the neighboring layer's link are.

A convolution batch normalized layer and four residual blocks are then created. The output is then generated by feeding the global averaged pool output into the softmax. This architecture was specifically created using ResNet18 as inspiration. It was groundbreaking because it offered a fresh approach to a pressing issue facing deep neural networks at the time. Identity shortcut connections are included into ResNet, thus skipping the training of one or more layers and resulting in a residual block.

Data Pre-Processing Stage

The BreakHis dataset contains all of the microscopic images in.jpg format with three-channel RGB and an 8-bit depth for each channel. The photos were originally 700 × 460 pixels, but we downsized them to 224 x 224 pixels. To enable the model to train more quickly and with less memory use, we turn each image into a Numpy array. The photos were then mixed up so that the model could train on some unordered data. The BreakHis dataset is divided into two parts - training and testing. In this project, 70% of the data was used for training purpose and 30% of the data was used for testing.

Result

The ResNet-50 model identified the images into benign (IDC-) and malignant (IDC+) as shown in Fig 2. Using the "ResNet50" framework, the confusion matrix and receiver operating characteristics (ROC) curve for the BreakHis dataset are shown in Fig. 3 and Fig 4. As seen in Fig. 3, a total of 74 and 82 microscopic images for benign and malignant breast cancer, respectively, are accurately classified. Only one benign and three malignant microscopic pictures were incorrectly diagnosed using the suggested framework at the same time.

Additionally, this framework obtains an area value of 0.976 (see Fig. 4), demonstrating the model's consistency and stability.

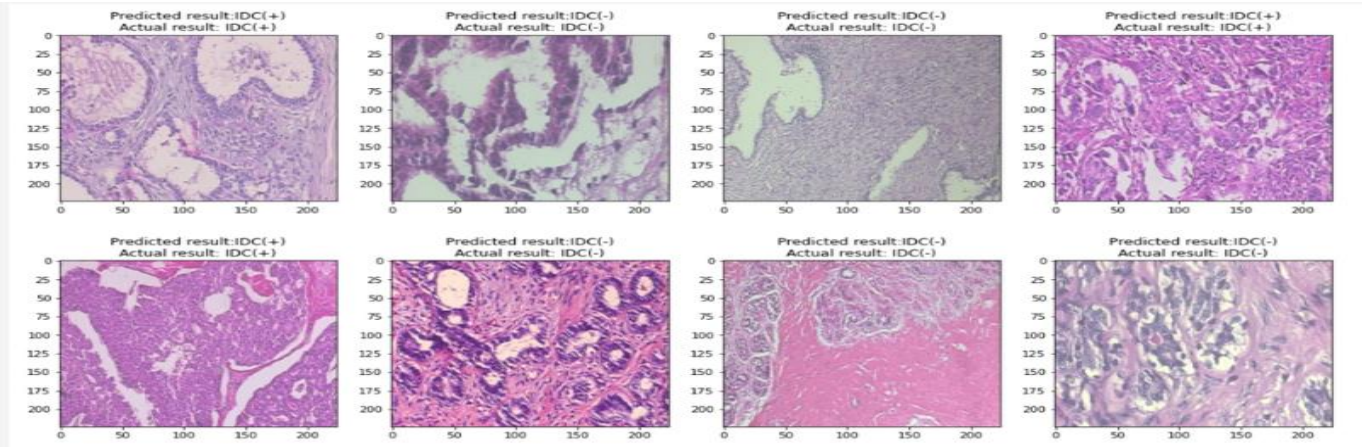

Figure 2: *ResNet-50 evaluated some microscopy images*

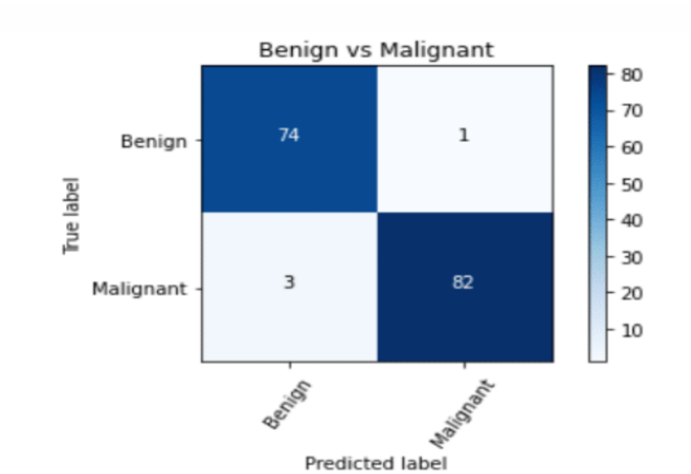

Figure 3: *Confusion Matrix*

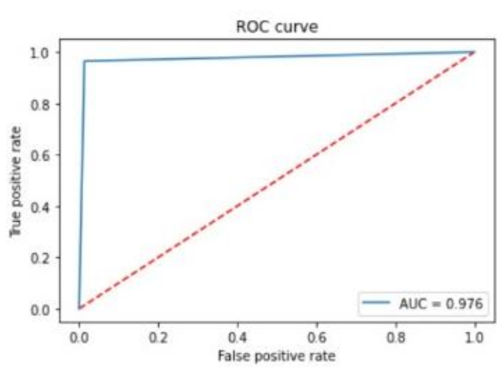

Figure 4: *Receiver operating characteristics (ROC) curve*

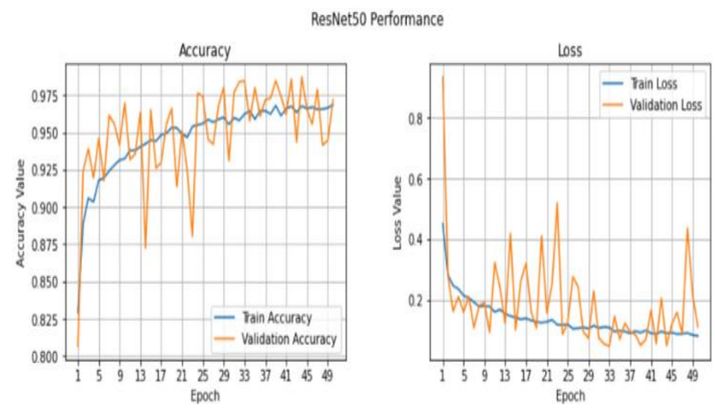

Figure 5: *Training progress for BreakHis dataset: (a) training and validation accuracy (higher is better), and (b) training and validation loss (lower is better)*

Conclusion

In this work, we have proposed ResNet-50 framework to detect breast cancer cases from microscopy images more precisely and reliably. The breast cancer detection process has multiple stages. This work focuses on using the ResNet-50 approach to analyze the dataset and categorize the pictures into benign and malignant, two predictions of breast cancer. It doesn't involve any invasive procedures at all. To determine what kind of ailment it is, it compiles previously saved images and data of the condition.

By adding additional neural layers and reducing the error percentage, ResNet enhances the effectiveness of deep neural networks. ResNet-50's accuracy and speed give it a significant advantage over other deep learning methods. Its closer ties to computer vision than those of other neural network types contribute to the fact that it performs better than other neural network types when identifying picture data.

By utilizing the ResNet-50 approach, our model ensured an accuracy of 97% with the BreakHis dataset.

Future Scope

There is significant potential for research in the field of breast cancer detection using deep learning techniques such as ResNet50. Some areas where further study may be useful include:

1. The majority of research into the identification and diagnosis of breast cancer centers on determining if a particular lesion picture is cancerous. The research has thus far concentrated on the specific issue of grouping the images into more general categories. The accuracy of models that can assist in identifying the precise kind and further educating the patient about the stage of breast cancer may be improved via future research.
2. Additionally, full-body photography may be used to identify additional body sections that may be impacted. The picture acquisition process will be automated and accelerated using this autonomous full-body photography. Unsupervised learning is a technique that looks for characteristics in dataset samples of images in order to find relationships or patterns.
3. Auto-organization strategies improve the degree of feature representation that is recovered by expert systems using convolutional neural networks. Future research may increase the accuracy of image processing systems, notably in the field of medical imaging, where even the tiniest details of characteristics are vital to a diagnosis' correctness.
4. Many areas of the world do not have access to specialized diagnostic equipment. Ways could be explored to use deep learning models to diagnose breast cancer conditions in these settings.

Reference

1. Spanhol, F. A., Oliveira, L. S. Cavalin, P. R., Petitjean, C. and Heutte, L. (2017). Deep features for breast cancer histopathological image classification. *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1868–1873. IEEE, Banff, AB. DOI 10.1109/SMC.2017.8122889.
2. Song, Y., Zou, J. J. Chang, H. and Cai, W. (2017). Adapting fisher vectors for histopathology image classifica- tion. *IEEE 14th International Symposium on Biomedical Imaging*, pp. 600–603. IEEE, Melbourne, Australia. DOI 10.1109/ISBI.2017.7950592.
3. Zhang, Y., Zhang, B. Coenen, F. Xiao, J. and Lu, W. (2014). One-class kernel subspace ensemble for medical image classification. *EURASIP Journal on Advances in Signal Processing, 2014(1),* 1–13. DOI 10.1186/1687-6180-2014-17.
4. Krizhevsky, A., Sutskever, I and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM, 60(6),* 84–90. DOI 10.1145/3065386.
5. Barker, J., Hoogi, A. Depeursinge, A and Rubin, D. L. (2016). Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. *Medical Image Analysis, 30,* 60–71. DOI 10.1016/j.media.2015.12.002.
6. Kandemir, M., Zhang, C and Hamprecht, F. A. (2014). Empowering multiple instance histopathology cancer diagnosis by cell graphs. In: Golland, P., Hata, N., Barillot, C., Hornegger, J. and Howe, R. (Eds.), *Medical image computing and computer-assisted intervention–MICCAI 2014,* pp. 228–235. Springer International Publishing: Cham. DOI 10.1007/978-3-319-10470-6_29.
7. Cosatto, E., Miller, M. Graf, H. P and Meyer, J. S. (2008). Grading nuclear pleomorphism on

histological micrographs. *19th International*
8. *Conference on Pattern Recognition*, pp. 1–4. Tampa, FL, USA. DOI 10.1109/ICPR.2008.4761112.
9. Han, Z., Wei, B. Zheng, Y. Yin, Y. and Li, K. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports, 7(1),* 1–10. DOI 10.1038/s41598-017-04075-z.
10. Song, Y., Chang, H. Huang, H and Cai, W. (2017). Supervised intra-embedding of fisher vectors for histopathology image classification. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L. and Duchesne, S. (Eds.), *Medical image computing and computer assisted intervention MICCAI 2017*, pp. 99–106. Springer International Publishing: Cham. DOI 10.1007/978-3-319-66179-7_12.
11. Gupta, V and Bhavsar, A. (2017). Breast cancer histopathological image classification: Is magnification impor- tant? *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 17–24. IEEE, Honolulu, HI, USA. DOI 10.1109/CVPRW.2017.107.
12. Araujo, T., Aresta, G. Castro, E. Rouco, J and Aguiar, P. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLoS One, 12(6),* e0177544. DOI 10.1371/journal.pone.0177544.
13. Kooi, T., Litjens, G. Van Ginneken, B. Gubern-Mérida, A and Sánchez, C. I. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis, 35,* 303–312. DOI 10.1016/j.media.2016.07.007.
14. Matlani, P and Shrivastava, M. (2019). Hybrid deep VGG-NET convolutional classifier for video smoke detection. *Computer Modeling in Engineering & Sciences, 119(3),* 427–458. DOI 10.32604/cmes.2019.04985

*****