International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

BREAST CANCER DIAGNOSTIC WITH DEEP LEARNING TECHNIQUES

Swetha.A[1], Student

Rajeshwari.P[2], Assistant Professor

Department of Software System,

Sri Krishna Arts and Science College

ABSTRACT: Breast cancer is one of the most prevalent and potentially fatal forms of cancer affecting women globally. Early and accurate diagnosis plays a crucial role in improving patient outcomes and survival rates. Deep learning techniques have demonstrated remarkable success in various medical applications, including breast cancer diagnosis. These abstract highlights recent advancements and applications of deep learning methods in the field of breast cancer diagnostics. This paper reviews the utilization of deep learning algorithms for breast cancer diagnosis from medical imaging modalities such as mammograms, ultrasounds, and MRIs. CNNs (Convolutional neural networks) have proven particularly effective in extracting intricate patterns and features from medical images, enabling the differentiation between malignant and benign breast lesions. Transfer learning, a technique where pre-trained models are fine-tuned for specific tasks, has significantly accelerated the development of accurate diagnostic tools.

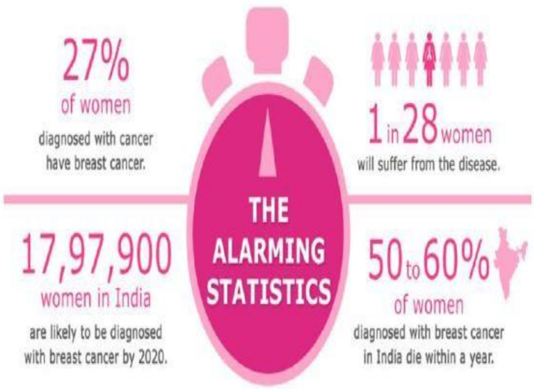
Keywords: Breast cancer ,Deep learning, Convolution neural networks, computer-aided detection,

1.INTRODUCTION:

The evolution of human civilization on planet earth was a miracle and it took millions of years in transforming an ape into to a human of modern era. A human body can be perceived as a set of interlinked as well as interdependent organs or components. Where each organ has its specific functionality. The deviation of such components from their normal behaviour may lead to a disease or it may cause to a disease and as per the archaeological evidences, mankind has started the usage of plants as medicine 60,000 years ago as a remedy. So, the growth and span of diseases is also as old as human civilization [1]. First vaccine across the world was developed in year 1796 by Edward Jenner for the treatment of smallpox. And, with the rapid growth of science and technology, lot many medicines were developed for the welfare of the human society as well as for the animals. But, the major thrust and revolution in the domain of medical sciences came after amalgamation of the technologies like Internet, Artificial Intelligence, Machine Learning, Cloud Computing, Big Data Analytics, Information and Communication Technology and Internet of Things in medical sciences. As a result, the medical services have become dependent to some extent on the equipment and devices to ensure the delivery of more accurate services to the society. The technology has changed the human perception about the diagnosis for a disease as well, i.e. the diagnosis after disease detection with the detection of disease at early stage and its diagnosis especially in the case of chronic diseases. It will not only reduce the death rate but also improves the quality of life. Hence, a new term “e-health” was coined in year 1999 with the intent to deliver numerous health care services by using the tremendous potential of ICT. [2]

2.SURVEY:

The spread of detected breast cancer amongst women is a very serious concern for the entire global society because it holds second position worldwide. Various statistics shows that, breast cancer is responsible for 25% of the total type of cancers and responsible for 15% deaths across the globe. It can be cured when detected at the early stage up to 100%, but it is more likely to be dangerous when detected at the last stage and the survival rate gets reduced to 15%. [3] In breast cancer the cell characteristics of the human body starts changing their nature and tends to behave abnormally. It is estimated that the number of cases of breast cancer globally will attain a huge growth from 1.4 million since 2008 to 2.1 million in 2030. Breast cancer is classified into two stages benign and malignant. Tumor at benign stage generally considered as non-cancerous and it will not penetrate to the other sections of the human body, whereas malignant tumor has the ability to penetrate into other parts of the body if it is not diagnosed properly or at early stage.[4] According to Lulu Wang traditional approaches are not effective for identifying breast cancer at an early stage due to the soft tissue nature of breast cancer. In Indian scenario it has been observed that, breast cancer is holding first position among the Indian females with a rate of 25.8 per 100,000 females and has the mortality rate of 12.7 per 100,000 females. The projection rate of this cancer reveals that the number of cases of breast cancer will go high as 1797900 till 2020 and the survival rate in India is also poor as compared with other nations. This may be due to various factors like improper medical facility for the treatment of breast cancer, economic constraints, social factors, lack of awareness about this disease etc.. the statistics of the breast cancer in India [5]



Statistics of Breast Cancer in India

Deep learning models using the autoBioSeqpy tool with the Keras backend were presented, trained, and estimated by Yue L. *et al* [2020]. The classification of biological sequence simple DL tool auto BioSeqpy is used. The major benefits of this tool is its capability of easy development and estimating different DL methods. Numerous ready-torun applications for users are provided by this tool, making the DL architectures simple and easy [6]. Ma D. *et al* [2021] developed a creative method for the automatic diagnosis of breast cancer. Raman spectra were gathered from 20 patient’s breast samples for analysis of spectrum. At that time, for classification, one dimensional CNN model was established and trained. The finest classification performance with 92% of overall diagnostic accuracy, the 98% of sensitivity and the 86% of specificity, was accomplished by

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

using this model. 1D-CNN and Raman spectroscopy combination for breast tissues classification is efficient and automatic [7]. Zhang Z. *et al* [2021] presented a new voting convergent difference neural network (V-CDNN) which is dissimilar from furthermost existing NN, the neural dynamic learning algorithm was adopted by V-CDNN, in which the computation efficiency is highly improved and diagnosis accuracy rate is also increased. From the results 100% average diagnosis accuracy is achieved by V-CDNN, that is the uppermost accuracy between the methods existed in the open database [8]. Sun S. *et al* [2020] designed new ensemble deep learning process called LSTM-B, in which (LSTM) and NN are combined and strategy of bagging technique for the purpose of obtaining exact results for the rate of exchange forecasted and to progress rate of exchange trading profitability. Accuracy of forecasting and profitability of potential trading are improved significantly by LSTM-B ensemble deep learning method. Jagekar A. *et al* [2020] presented Computing of Quantum based DL methods for error analysis. In this approach, the difficulties of computation handled by conservative data-driven procedures done on traditional computers are rectified by this method's unique capabilities. Deep belief networks are combined with the fault diagnosis model from which the features are extracted at various levels. The QC-based deep learning approach relishes greater performance of fault detection and diagnosis with acquired average rates of detecting fault is 79.2% for CSTR and for TE process is 99.39% [10]. Wang H. *et al* [2020] processed an EHR-based predictive model for the estimation of probability of distant recurrent in BC patients. The features extracted and clinical features are combined by traditional ML classifiers and KCNN. An automated approach was obtained, so that prediction of BC using image processing and DL method are possible. For segmentation of BM in US, a DL method was presented by Byra M. *et al* [2020]. Disparities in size of breast mass and characteristics of the image changes the automatic segmentation problematic. SK U-Net CNN was introduced to address this problem. The Selective kernel is aimed towards the network's receptive field adjustment through the mechanism of attention, and integrates the extracted feature maps 4extracted having enlarged and conventional convolutions. For recognition of breast mass, deep learning methods developed in this method and its results acts as the significant steps [11]. To improve the diagnosis, classifying BC lesion on the basis of DL integrated with CAD was designed by Al-Antari M. *et al* [2020]. YOLO detector capabilities are represented in the detection of breast lesion's evaluated results to realize 99.17% of accuracies and F1-scores of 99.28% for the dataset DDSM and accuracy of 97.27%, F1-scores of 98.02% for the dataset INbreast. Models of classification are boosted by YOLO detector boosted for achieving hopeful performance in the diagnosis of breast lesion and used for the development of possible system of CAD motivated towards diagnosis of breast cancer [12]. Timmana H. *et al* [2020] presented a DL-based system for BC detection. Imbalance of dataset, skewness, and data insufficiencies in the dataset are resolved by this method. In addition, optimal accuracy is achieved by the hyper-parameters, which needed for training the model and also comparison of the existing machine learning models together with decision tree algorithms, Logistic regression and KNN with this model to analyse its performance [13]. Hai J. *et al* [2020] presented an algorithm for directly selecting the pathologic mammograms. Extraction of Features at low level and supervised learning algorithm. CNN is planned, extraction of semantic features at high-level. The new CNN end to end optimization was performed by the combination of multilevel features extracted for making the various portion of network learnt to give attention to features at different level. As results establish that performance of this algorithm is produces higher outcomes as compared to other CNN models having pathological images.

For prediction of BC, a new data mining technique is presented by Abdar M. *et al* [2019]. This method is motivated towards expanding the automatic expert system (ES) in order to generate exact results of BC diagnosis. Through this model, performance of conventional DL method is improved. Breast Cancer detection accuracy is 100% in this method. The overfitting problem is rectified by determining few suitable values of polynomial SVM parameter [14]. Fang Y. *et al* [2019] demonstrated an approach for the BC classification. In the usage of maximum pooling layers, number of pixels attached to lesions is calculated by the architecture of CNN. Then, increase in grayscale features and texture are reflections of high scores of quality, obtained by high density of pixel regions. At last classification of breast cancer is achieved by quality scores of multi-SVM based image kernels. Results of experimental showed that this method leaves behind single recognition depends on image classification methods like pixel grayscale or gradient [15]. Herent P. *et al* [2019] stated a method for deep learning that concurrently learns the lesions detection and illustrates them and lesion characterization model was formed on the basis of single 2D T1-weighted MR images which are fat suppressed was acquired. This system achieves good performance as compared to other DL approaches. Larger and independent cohort are used for the validation of this method [16]. Hijab A. *et al* [2019] developed a DL methodology for tackling the problem of CAD of malignant BC in US images. Numerous images of malignant and benign are comprised in training data. These are utilized for deep convolutional neural network (CNN) training. 0.97 accuracy, 0.98 AUC are result of experiments obtained by this model. In the application of biomedical images, the results of deep learning was improved by the creation of models which are pre-trained [17]. Lei B. *et al* [2020] presented a new framework that depends on a self-attention mechanism for solving encounters present in the segmentation of entire breast ultrasound images. The attention mechanism for spatial- and channel-wise signals in the similar neural block was not only introduced as well as it includes the NCB to discover the non-local signals. The module for co-attention is engaged for the exploitation of 2.5D features among two slices for the further improvement of the segmentation coherence between the consecutive slices [18].

4.METHODLOGY:

1.Data Collection: Collect a large dataset of mammograms or breast images. These images can come from a variety of sources such as medical institutions, research databases, or open-access repositories. It is important to ensure that the dataset is diverse and representative of different types of breast tissue, age, and stages of cancer

2. Data Preprocessing: Clean and pre-process the data collected to ensure consistency and quality. Preprocessing steps may include resizing images to a standard resolution, normalizing pixel values, and removing artifacts or noise

3. Data labelling: Annotate the images in your dataset to indicate the presence or absence of breast cancer. This is generally a time-consuming process, as each image needs to be reviewed by medical experts who can label them accurately.

4. Data augmentation: To increase the robustness and variability of your model, apply data augmentation techniques. These techniques involve randomly applying changes such as rotation, flip, and zoom to generate additional training samples.

5. Model Selection: Choose a suitable deep learning architecture for your task. Convolutional neural networks (CNNs) are commonly used for image-related tasks, including breast cancer diagnosis. You can choose installed architectures such as VGG, Resnet, or more specialized architectures like Dens Net or Inception.

6. Model architecture design: Design the architecture of your deep learning model. This includes configuring the number of layers, the type of layers (infectious, pooling, fully connected), activation tasks, and any regularization techniques (dropout, batch normalization) to prevent overfitting.

5. SYSTEM ANALYSIS:

5.1Problem Statement:

First, the number of available medical images is much less than the number of natural images available. This is especially an issue when screening for a condition with a significantly lower prevalence, such as breast cancer in a screening setting (<1% of screening examinations result in a cancer diagnosis

Second, access to medical imaging data is guided by a number of strict policies, given that they contain patients' medical information. Sharing medical imaging data often requires hard and time-consuming effort to identify the data as well as ensure compliance with requirements from the data sharing institution.

Finally, annotation of medical imaging data usually requires the work of radiologists, who already have high demands on their time

5.2EXISTING SYSTEM:

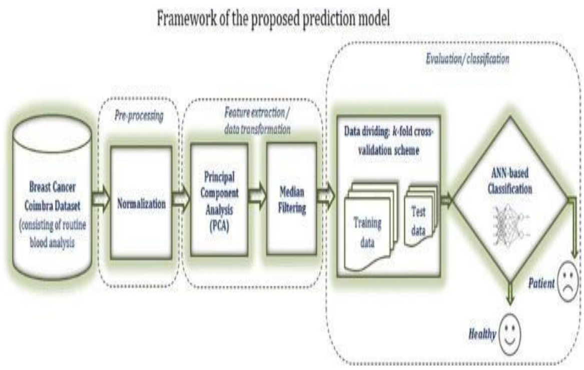
CAD, or computer-aided detection, is a computer-based technology that helps radiologists identify suspicious areas when reading digital mammograms.
The CAD system highlights these areas on the image, alerting the radiologist to carefully evaluate the area.

5.3PROPOSED SYSTEM:

Final accuracy of the built Deep CNN model using the online extracted features reaches 91.49%. Table 1 shows the confusion matrix for the performance of the model with the test set. In the confusion matrix, predicted no, predicted yes, actual no and actual yes refer to the number of samples predicted as benign, the number of samples predicted as malignant, the actual number of benign and the actual number of malignant tumors in the test set, respectively. The confusion matrix in Table 1 shows a balanced performance where only 7 benign tumors were misclassified as malignant by the model, whereas 6 malignant tumors were misclassified as benign. A total of 62 out of 75 samples were correctly classified into their true classes. The recall plots for the Precision- model are shown in Figure 4. Precision is defined as TP/ TP+FP; where TP represents true positives and FP represents false positives. Recall, or known as sensitivity, is calculated as TP / TP+Fn; where FN denotes false negative. Thus Figure 4 shows the ratio of the number of correctly predicted malignant tumors to the total number predicted the ratio of the number of correctly predicted malignant tumors to the total number of malignant tumors versus actual malignant tumors. The accuracy metric indicates how many of the model's predictions were correct. However, if the dataset is imbalanced, a model's high accuracy rate does not guarantee its ability to equally separate classes. In particular, in the classification of medical images, it is necessary

ISSN 2581-7795

to develop a model with the ability to be applied to all classes. In these cases, sensitivity and specificity should be used to provide information about the performance of the model. Sensitivity measures the percentage of a patient's disease that is correctly predicted by the proposed model. Specificity measures the percentage of patients that do not have the disease and are correctly predicted by the proposed model. These two-evaluation metrics measure the model's ability to reduce FN and FP predictions.



4.3 fig Proposed System

6.RESULTS AND DISCUSSION:

The most intuitive and simplest performance analysing metric in classification problem to determine the accuracy and preciseness of the model is the confusion matrix. It is applicable to both binary and multi-classification. Confusion matrix by itself is not a performance metric but allows computing some valuable performance measures based on the value of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) [98]. Consider a problem of binary classification, where the person with benign cancer is represented as 0 and person with malignant cancer as 1. shows the confusion matrix, where the actual labels are presented by rows and predicted labels by the columns. In the context of considered problem, the terms TP, FP, FN, and TN have explained in a point-wise manner. i. TP: A case would be under true positive if a person is actually having benign cancer (0) and the classifying model also predicted the case as benign type (0). ii. FP: It has happened when a person with the benign type (0) cancer in actual is predicted as a case of malignant type (1). iii. FN: It is just opposite to FP, where the person with the malignant type (1) cancer in actual is predicted as the case of benign type (0) cancer. iv. TN: If a person with malignant type (1) cancer and also predicted as malignant type (1), then the case would come under the TN. Some valuable performance metrics that can be computed from the confusion matrix are accuracy, precision, recall, and F1 score which can be defined as:

$$\text{Accuracy} = \frac{\text{Correctly classified Images}}{\text{Total number of images}} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Precision}_n = \frac{\text{Images correctly classified as n}}{\text{Images classified as n}} = \frac{TP}{TP+FP}$$

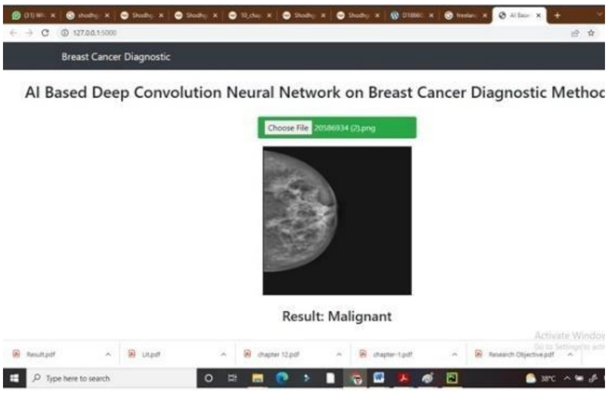
$$\text{Recall}_n(\text{Sensitivity}) = \frac{\text{Images correctly classified as n}}{\text{Images of class n}} = \frac{TP}{TP+FN}$$

$$\text{F1 score or } F_{avg} = \frac{2}{N}\sum_{n=1}^{N}\frac{\text{recall}_n * \text{precision}_n}{\text{recall}_n + \text{precision}_n}$$

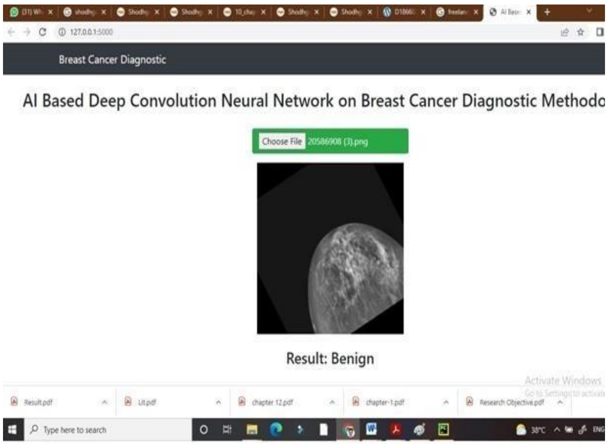
Benign	640x480	350	75	75	500
Malignant	640x480	350	75	75	500
Dataset Categories	Dimension	Training	Validation	Testing	Total

Screen Shot for Malignant:

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795



Screen Shot for Benign:



7.CONCLUSION

The coverage of this research and the literature reported in this research clearly states that breast cancer is at alarming stage across the globe and in order to prevent it there is a need for some smart systems to handle the complexities associated with this disease. The findings of this study will rebound to the society considering that machine learning plays important role in disease prediction as various techniques of machine learning are helpful in; early detection, avoidance, prediction, reduction in cost, facilitating medical practitioners to make decisions on a real-time basis and diagnose the disease at an early stage. It will be helpful for the patients as well because they may be diagnosed at initial stages of chronic disease. Also, early detection and proper diagnosis can reduce the number of deaths due to chronic disease like breast cancer.

8.REFERENCES

[1] Priya Sampathkumar. "Mayo Clinic Proceedings on Vaccines Thematic Reviews." In Mayo Clinic Proceedings, Vol. 94, No. 10, p. 1931-1933. Elsevier, 2020.

[2] H.S. Ventzer A. Bygholm. "narratives of empowerment and compliance: A study of communication in online patient support groups". In the International Journal of Medical Informatics, Vol. 82(12), pp. E386–E394, 2019.

[3] Ryan Crowley, Hilary Daniels, Thomas G. Cooney, and Lee S. Engel. "Envisioning a Better U.S. Health Care System for All: Cost of Care and Coverage." Annals of Internal Medicine Vol. 172, No.2_Supplement (2020), pp. S7-S32.

[4] Rajeshwari Sinha, Sanghamitra Patil. "Addressing the growing burden of chronic diseases in India: the need to strengthen primary care", Journal of Family Medicine and Primary Care, Vol. 6, Issue 4, October-December 2017, pp. 701-708.

[5] Henriquez-Camacho C, Llosa J, Miranda JJ, Cheney NE. "Addressing Healthy Older Populations in Developing Countries: Unlocking the Opportunity for eHealth and mHealth". emerge. Theme Pandemic. 2014;11(1):136. Published 2014 Dec 31. doi: 10.1186/s12982-014-0021-4.

[6] Anna Sanina, Alexey Balashov, Maria Rubatkova, and Daniel M. Satinsky. "The Effectiveness of Communication Channels in Governmental and Business Communication", Information Politics, Vol. 22 (2017) pp. 251–266. doi 10.3233/ip-170415. iOS Press.

[7] E. Cummings, Leonie Ellis, and Paul Turner. "Past, Present, and Future: Examining the Role of the "Social" in Transforming Personal Health Care Management of Chronic Illness." In Health literacy: breakthroughs in research and practice, pp. 287–304. IGI Global, 2017.

[8] J. Michael McGinnis, Leigh Stuckhardt, Robert Saunders, and Mark Smith, eds. best care at the lowest cost: A continuous health care learning path in America. National Academy Press, 2013.

[9] Bardhan, Indranil, Hsinchun Chen, and Elena Karahanna. "Connecting Systems, Data, And People: A Multidisciplinary Research Roadmap for Chronic Disease Management." In Management Information Systems Quarterly, Vol. 44, No. 1, pp. 185-200.

[10] Yabroff, K. Robin, Ted Gansler, Richard C. Wender, Kevin J. Cullen, and Otis W. Brawley. "Minimizing the burden of cancer in the United States: Goals for a high-performing health care system." CA: a cancer journal for clinicians Vol.69, No. 3 (2019), pp. 166-183.