BREAST CANCER DETECTION USING MACHINE LEARNING AND DEEP LEARNING

PRATIKSHA NATIKAR[1], SIDDESH KT[2], KOTURSWAMY SM[3]
Student, Department Of MCA, BIET, Davangere[1]
Assistant Professor, Department Of MCA, BIET, Davangere [2]
Assistant Professor, Department Of MCA, BIET, Davangere[3]

ABSTRACT
Most recent data available, breast carcinoma of worldwide, accounting for around 900,000 deaths annually. Early detection and accurate diagnosis can increase the likelihood of positive outcomes and lower the death rate from the condition. In actuality, an early diagnosis can help stop it from spreading and spare the premature victims from getting it. Early detection and appropriate diagnosis can increase the likelihood of favourable results and lower the death rate. But as all centred around the idea of "binary grouping" (malignant and benign; no-cancer and cancer). In this study, we suggest using a computer-aided diagnostic (CAD) system that manages a database to identify and diagnose patients into three classes: non-cancerous, malignant, and no cancer.

1.INTRODUCTION

According to a credible source, the Centres for Disease Control and Prevention (CDC), among women. The chances of surviving breast cancer vary widely based on numerous factors. These are the type of tumour a woman has and the stage of the diagnosis. Breast cancer is a type of cancer that starts the cells. Most often, regions called lobules or the ducts is where the cancer first appears.

The approaches used to treat and prevent these two types of cancer. Benign cells do not have body. Because of this, if thermal imaging screening mammography is improved to a considerable extent, it may become a competitive option. A tumour result that is more conclusive, widely accepted, and suitable a screening guideline for breast cancer, research is now being done in this field. Likewise, breaking through the recently established obstacles caused by the drawn-out screening process, especially in cases where photo processing is required. This study indicate that, in order to lessen the previously mentioned shortcomings, a Convolutional Neural Network, or CNN, for thermal imaging testing and therapy.

2. RELATED WORK

Looked into imaging and genomics-based techniques for the detection of breast cancer. Furthermore, no both of these approaches.

[1]**Tiancheng J.J. Mancuso** gave various techniques utilized in histological image analysis (HIA) breast cancer diagnosis. Based on several convolutional neural network. The authors classified their study based on the dataset that was utilised. They begin with the most recent occurrence and arrange it in reverse chronological order. The results of this study indicate that the most commonly used algorithms HIA were PNNs and ANNs, with the latter being the first time around 2012.

[2]**K. Pravalika** However, most feature extraction research has relied on morphological and textural characteristics.

Its clearly very helpful in the early detection and treatment of breast cancer, which in turn resulted in more effective therapy.

Natural inspiration computing (NIC) methodologies have been proposed and implemented in the process of diagnosing a variety of human conditions. Five insect-based NIC diagnostic algorithms were presented by the authors of, who discussed their utility in detecting diabetes and cancer. According to the authors, it identified several tumours well (breast, lung, prostate and ovarian).

Combining directed ABC with neural networks helped diagnose breast cancer. [3]**C. Héry, P. Autier** authors of presented evidence that indicated the usefulness of NNs in the categorization of cancer diagnoses, particularly in the earlier stages of the disease. Their research demonstrates that the number of NNs have demonstrated some level of potential in the detection of malignant cells.
However, in order to pre-process the images, the imaging method demands a significant amount of processing resources.[4]**Sharma M** proposed methodology was designed to help physicians decide whether to proceed with a breast biopsy for a suspicious lesion in light of the mammography findings. The utilisation of Bayesian classifiers by the author a feasible substitute for several other techniques that are applicable in the medicin[5] **Alaa M. Elsayad**, it has been discovered that Bayesian networks, with their own symbolic representation, are a successful methodology.
When an RBF neural network was deployed, the method the database produced a 97 percent classification recurrence rate.

3.METHODOLOGY

This study uses a variety of methods to diagnose breast cancer. Preparing the data and building prediction models are the two primary sections of the paper. The Wisconsin Breast Cancer Dataset, which is accessible to researchers publically, is used in this work [11] V. Lahoura Using biopsy images, has 30 features and 569 samples.
The procedures that must be taken from beginning to end to construct a to forecast breast cancer shown in fig.

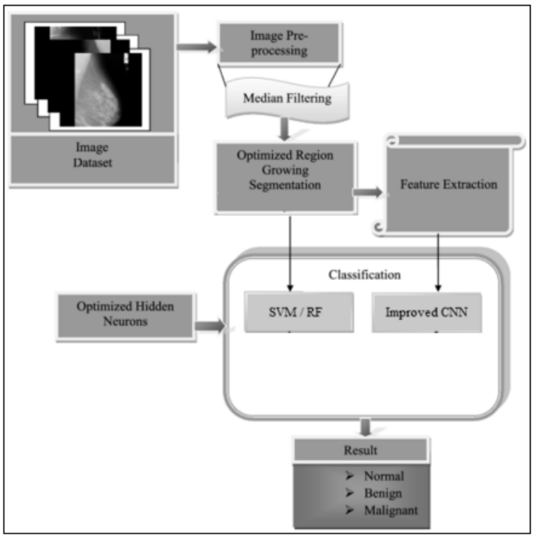

Fig.3.1 System Architecture.

Data exploration and pre-processing, which includes techniques like label encoding and normalization, are the first steps.
An effective method for encoding the levels of the categorical features into numerical values is the Label Encoder. Every categorical feature has an encoding. d. Malignant and benign values in this paper are categorized as 0 and 1. All attribute values are rescaled in using the Normalizer Method. For this, the formula found in equation (1) is applied.
After pre-processing, data is divided into train and test sets so that models may be built. The remaining 25% was utilized for testing, and the remaining 75% was used for training.
A variety of machine learning methods, including SVM, KNN, and logistic regression, have been used to develop cancer prediction models. The result of the dataset utilized for this project can be categorized as either M (malignant) or B (benign). Because the input data is labeled, K-Nearest Neighbour is a supervised machine learning algorithm. Rather than taking the dataset's parameters into account, the test data points' classifications are based on the closest training data points.
3.1 Data Collection: Obtaining a variety of medical imaging data from clinics and hospitals, including mammograms, MRIs, and ultrasounds, is necessary for the identification of breast cancer.
3.2 Dataset: The dataset that was used had images of more than 1000 individuals, roughly 150 of whom had

breast cancer while the remaining patients did not. These photos were found on the Kaggle website.

3.3 Data Acquisition: Data acquisition for breast cancer detection involves obtaining medical imaging data from various hospitals, clinics, and research institutions. It is critical to ensure data integrity, privacy, and ethical compliance following legal requirements and securing required authorizations.

3.4 Data Pre-processing: Several crucial processes are involved in data pre-processing for breast cancer diagnosis in order to provide the best possible quality and suitability.

ALGORITHM USED

A well-liked classification and regression applications is the decision tree. It works by recursively dividing the data into subsets according to classifying the target variable (class labels for classification, for example) into discrete groups or in forecasting continuous values.

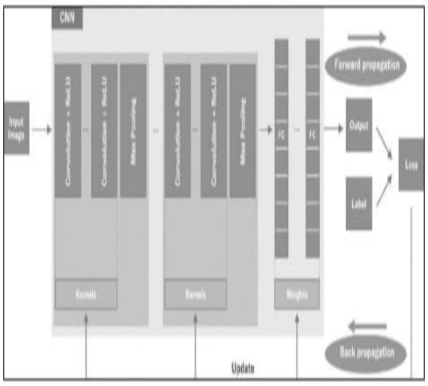

Fig 3: CNN Architecture

Splitting Criteria: This method chooses the optimal feature at each node to split the data according to a criterion like variance reduction for regression or Gini impurity for classification.

In order to optimize information gain or decrease impurity within each subgroup, it
assesses several splitting points. A decision tree creates a tree structure that predicts outcomes by recursively splitting data based on features using information gain.

Information Gain is then computed as:

IG(A)=H(D)-H(D/A)

4. RESULTAND DISCUSSION

On the Wisconsin dataset, many machine learning techniques, including K Nearest Neighbhour (KNN), Support Vector Machine (SVM), Decision tree, Naïve Bayes Logistic Regression, and Random Forest, were employed to predict the incidence of breast cancer. The Random Forest and SVM algorithms yielded a maximum accuracy of 96.5%.

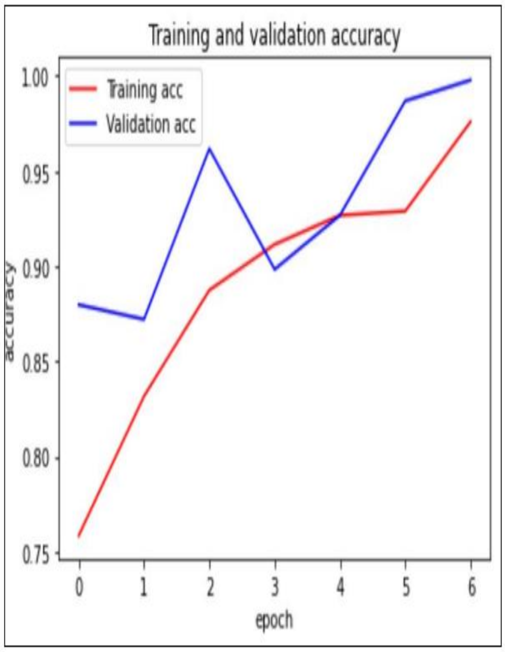

Fig.4.1.1 Comparision for Training accuracy and Validation accuracy for CNN.

These were used to improve prediction accuracy. Model accuracy and loss for the ANN model expressed graphically in relation to the number of epochs. The accuracy of the model rises as the number of epochs decreases, but the loss falls. More efficiency was achieved in the CNN model's accuracy, which was 99.3%, compared to the ANN model's 97.3%. as opposed to the machine learning algorithms mentioned below.

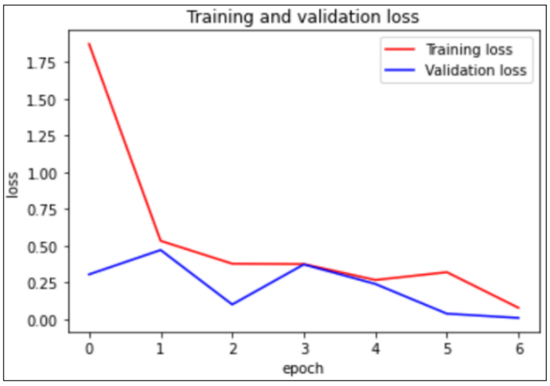

Fig.4.1.2 Comparision for Training loss and Validation loss for CNN

Deep learning was shown to be more effective when activation functions like sigmoid and ReLu were used. Activation function allowed the outcome to be obtained probability as Machine Learning techniques, which only provided the results in two labels: 1 (malignant) and 0 (benevolent).The values that were acquired utilizing the various techniques are displayed in table.

5. CONCLUSION

Our findings demonstrate the great accuracy that can be attained by end-to-end taught deep learning approaches, and their potential for seamless transfer across many mammography platforms. With the increasing quantity of publicly available training datasets and computer resources, there is a good chance that deep learning techniques will be able to significantly increase the accuracy of breast cancer detection on screening mammography.

In the course of this work, we looked into identifying breast cancer. We looked at how Random Forest, SVM, and CNN were comparable and different from one another.

It was shown that when it comes to accuracy, precision, and the quantity of data used, CNN outperforms the other methods currently in use. It was found that CNN achieved 99.67 percent accuracy, SVM achieved 89.84 percent accuracy, and RF achieved 90.55 percent accuracy. This study compares the accuracy of several that are implemented. SVM and Random Forest algorithms yielded greatest accuracy of 96.5% when it came to machine learning methods. CNN and ANN were used scenario. CNN reports an accuracy of 97.3%, while ANN reports 99.3%. The study comes to the out perform. Additionally, applying activation functions in deep learning makes the output probability-predictable, with machine learning methods. Notably, effective in identifying and classifying patterns from vast amounts of medical photos. This will help to organize and categorize the images appropriately. This will greatly improve the detection process in exchange.

These methods might be applied future to image-based datasets. A website or application may also be integrated with the system. Increasing the model's accuracy could lead to more accurate forecasts.

6.REFERENCES

1. "Deep learning management," Tiancheng J.J. Mancuso, Xiaohui Yu, J. C. Chang, S.T. C. Wong.

2. "Prediction of Techniques," Ch. Shravya, K. Pravalika, and Sk. Subhani.

3. [Global burden of breast cancer, J. Ferlay, C. Héry, P. Autier, and R. Sankaranarayanan, in Breast cancer epidemiology, ed.: Springer, 2010, pp. 119].

4. A Review of Disease Prediction, 2015. 3 Sharma M, Kaur P, and Gautam R. Prog Artif Intell 2019;8:401–24. doi:10.1007/s13748-019-00191-1.

5. Predicting breast mass severity using an ensemble of Bayesian classifiers, Alaa M. Elsayad. Computer Science Journal. 2010;6(5): 576–584.

6. Matej Mertik, Miljenko K. Bayesian network application in emergency medicine, 2008.

7. Sivapriya J, Aravind Kumar V, Siddarth Sai S, Sriram S. Machine Learning-Based Breast Cancer Prediction, 2019.

8. Digital Mammography Screening Database. The URL for the database is http://www.eng.usf.edu/cvprg/Mammography.html.
9. Curated subset of DDSM for breast imaging. CBIS-DDSM 28 can be found https://wiki.cancerimagingarchive.net/display/Public.
10. The Mini-MIAS Mammography Database. peipa.essex.ac.uk/info/mias.html.
11. Cloud computing-based framework for extreme learning machine-based breast cancer diagnosis, V. Lahoura et al., Diagnostics, vol. 11, no. 2, Feb. 2021, doi: 10.3390/diagnostics11020241.
12. Advanced Computing and Intelligent Systems, Springer Singapore, 2021, pp. 821–833, C. Kaushal and A. Singla, "Analysis.
13. "An investigation breast cancer identification through the mammography images,"G. Meenalochini and S. Ramkumar, Materials Today: Proceedings, vol. 37, pp. 2738–2743, 2021,doi:10.1016/j.matpr.2020.08.543.
14. World Health Organization. Accessed on: Feb 13, 2020. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.
15. Yi-Sheng Sun, Zhao Zhao, Han-Ping-Zhu,"Risk factors and Preventions of Breast Cancer" International Journal of Biological Sciences.