**BREAST CANCER DETECTION**

Submitted for

INTELLIGENT MODEL DESIGN USING AI

Submitted by:

**(E22BCAU0146) SHRESHTHA RAJ**

**(E22BCAU0055) RIYA GUPTA**

**(E22BCAU0061) SIYA SIKKA**

Submitted to:

**DR. NITIN ARVIND SHELKE**

**Jan-May 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

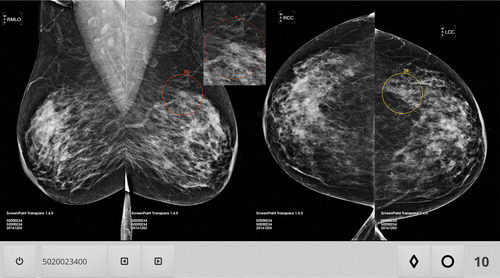
|  |  |  |
| --- | --- | --- |
| Sr. No. | Content | Page No. |
| **1.** | Introduction | 1 |
| **2.** | Related work | 2 |
| **3.** | Methodology |  |
| **4.** | Conclusions |  |
| **5.** | Future scope |  |
| **6.** | References |  |
| **7.** | Git Hub |  |

1. **Introduction**

According to the “International Agency for Research on Cancer (IARC’s) 2020 World Cancer Report”, Cancer is the primary or secondary cause of deathblow (ages 30–69) in 134 out of 183 countries .

Lung cancer is by far the leading cause of death among men and women, but in men prostate cancer is more common and in women BC is more common. As reported by IARC, cancer incidence is anticipated to inflate from 18.1 million to 29.5 million between 2018 and 2040, with mortality rising from 9.6 million to 16.4 million. Of all types of cancers, BC is one of the most colloquial and deadliest cancers among women. BC is the most prevalent cancer with high mortality and morbidity among women in the world and constitutes about 14% of all cancers1 as shown in Fig. 2. It increases the death rate among women, affecting about 2.1 million women each year. It is estimated that 6,85,000 women died due to BC in 2020. Death rates in wealthy countries are significantly higher, and they are rising in every region of the world. Cancer is the abnormal cell division due to which a mass is developed called a tumor. The tumor can be cancerous (malignant) or non-cancerous (Benign).

Breast Cancer is the malignant tumor that develops from breast cells. The exact cause is unpredictable and is hard to answer why a woman has BC, whereas some do not have it. However, it is said in the studies that BC is caused when the DNA inside the cell is damaged/mutated. “Invasive Carcinoma”, “Ductal Carcinoma”, and “Invasive Lobular Carcinoma” are the three main categories of BC. Accurate and precise diagnosis of BC is vital for the timely detection of cancer to improve survivability [1]. Many imaging modalities are continually being developed to diagnose this disease as early as possible. Medical images are the key source of information for the recognition and diagnosis of cancer diseases. Some of these modalities are used for screening purposes, some for diagnostic purposes, and a few others for adjunctive evaluation to provide field experts an additional confidence in their initial decision. The currently used imaging modalities for the diagnosis of BC are: “Mammography”, “Ultrasound”, “Histopathology”, “MRI”, “CT”, “Positron Emission Tomography” (PET) and “Thermography” (see Fig. 3). Each of these imaging modalities has its significance.

Mammography is the most commonly used method to diagnose BC in women with no signs of diseases. Mammography is the golden standard to diagnose the Breast Cancer at an early stage. Mammograms are the x-ray images of the breast. It employs low-energy x-rays for breast examination and screening [2]. In randomized controlled trials involving the general population, the Mammography screening test has been proven to reduce death rates [5]. Mammogram images have demonstrated to be technically more suitable for screening and, as a result, they can be employed for routine screening [6]. Furthermore, the fundamental Mammography technique has been amended to provide 3D scans of the breast in “Digital Tomosynthesis Mammography”. “Contrast-Enhanced Digital Mammography” (CEDM) is a recent advancement in Mammography that uses an intravenous infusion of an iodinated Breast Thermography or Thermal Imaging is an another imaging modality for BC diagnosis. Infrared Thermography is employed for different cancer detection processes. The examination of medical images for BC diagnosis by field experts and Artificial Intelligence (AI) assisted CAD systems have shown great potential applications in different fields of healthcare industry . 

**2.Related Work**

***[1]*** Magde El provided an up-to-date status of research in seven electromagnetic modalities for breast cancer detection. It is evident that EM techniques show high potential to improve the detection of breast cancer.

**[2]**Singhproposed that Naïve Bayes shows faster learning among four classifers whereas J48 found to be slower. OneR classifer is leading from other three classifers in percentage of correctly classifed instances. Te accuracy of J48 algorithm is promising in true positive and false positive rates.

**[3]** Jagdale proposed that Hybrid approach has been used, i.e., combination of Machine Learning and Lexicon based approach. It plays a vital role in enabling the businesses to work actively on improving the business strategy and gain an in-depth insight of the buyer’s feedback about their product.

**[4]** Kumaran proposed that he used support vector machines (SVM) and RNN with LSTM to analyze Breast cancer detection .

Table1. Summarization of Literature Review

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref** | **Techniques** | **Dataset Used** | **Performance metrics use** |
| [1] | Convolutional Neural Networks (CNN) | WSI-level annotations | 92 |
| [2] | Naive Bayes | Histopathology images | 66 |
| [3] | Naive Bayes | INbreast data | 95 |
| [4] | Feature Enhanced Attention CNN-Bi-LSTM | Breast Cancer Histology (BACH) | 95 |

**3 .Methodology:**

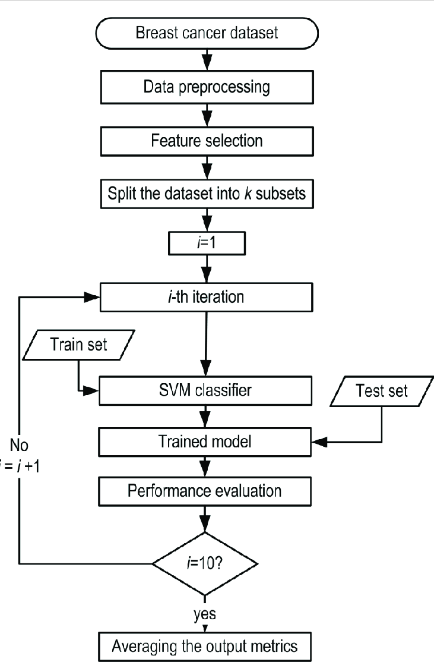


Fig.1. A framework of the proposed Breast Cancer Detection

**Outline of the experimental setup for breast cancer prediction, along with the performance metrics commonly used:**

*Data Collection: Gather a comprehensive dataset containing various features such as patient demographics, clinical attributes, and imaging data.*

*Data Preprocessing: Missing Value Imputation: Handle missing values in the dataset using techniques like mean imputation, median imputation, or predictive imputation.*

*Feature Scaling: Normalize or standardize numerical features to bring them to a similar scale.*

*Feature Encoding: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.*

*Data Split: Divide the dataset into training, validation, and test sets. The training set is used to train the model, the validation set is used for hyperparameter tuning, and the test set is used for evaluating the final model performance.*

*Model Selection:*

*Choose appropriate machine learning or deep learning algorithms for breast cancer prediction. Commonly used algorithms include logistic regression, support vector machines (SVM), random forests.*

*Experiment with different models to find the one that yields the best performance.*

*Model Training:*

*Train the selected models on the training dataset using appropriate training algorithms.*

*Tune hyperparameters using techniques like grid search, random search, or Bayesian optimization to improve model performance.*

*Model Evaluation:*

*Evaluate the trained models on the validation set using performance metrics such as:*

*Accuracy: The proportion of correctly classified instances.*

*Precision: The proportion of true positive predictions among all positive predictions.*

*Recall (Sensitivity): The proportion of true positive predictions among all actual positive instances.*

*F1 Score: The harmonic mean of precision and recall, which provides a balance between the two metrics.*

*Model Validation:*

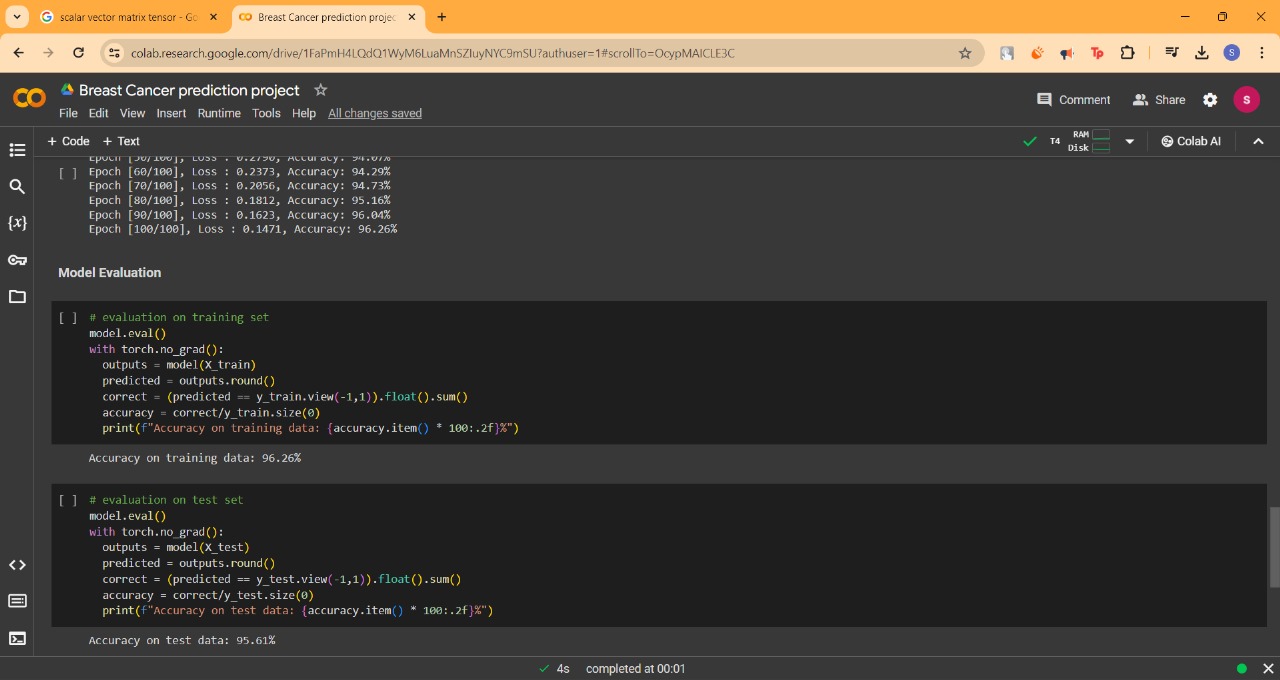
*Validate the selected model on the test set to assess its generalization performance.*

*Compare the performance metrics obtained on the test set with those obtained on the validation set to ensure consistency and reliability of the model.*

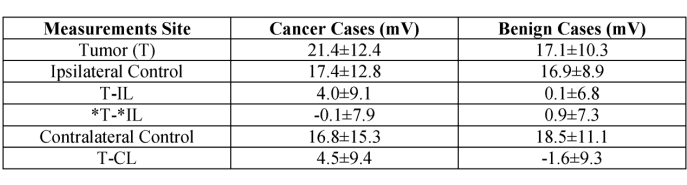
*Performance Analysis:*

*Analyze the model's performance on different subsets of the data (e.g., different age groups, tumor grades) to gain insights into its strengths and weaknesses.*

1. **Experimental Result and Discussion**



* In a numerical study, which incorporated realistic heat generation by growing tumors, the visibility of breast tumors was estimated under the assumption of ideal antennas using the Pennes equation . The result concluded that a 10-mm tumor can be detected if it is no more than 3 cm deep in the breast
* MATHEMATICAL REQUIREMENTS FOR PREDICTION

4.**Conclusions**

*To get more accuracy, we trained all supervised classification algorithms but you can try out a few of them which are always popular. After training all algorithms, we found that Logistic Regression classifiers are given high accuracy .*

*As ML Engineer, we always retrain the deployed model after some period of time to sustain the accuracy of the model. We hope our efforts will save the life of breast cancer patients.*

**5.Future Scope**

*Here are some areas where future research and development could focus:*

* *Integration of Multi-omics Data like bioinformatives*
* *Advanced Imaging Techniques*
* *Futuristic Deep Learning and Artificial Intelligence*

**7.References**

[1] Review of Electromagnetic Techniques for Breast Cancer Detection Ahmed M. Hassan, Student Member, IEEE, and Magda El-Shenawee, Senior Member, IEEE

[2] L. Tabár and P. Dean, “A new era in the diagnosis and treatment of breast cancer,” J. Breast, vol. 16, pp. S2–S4, 2010.

[3] L. Tabar, C. Fagerberg, A. Gad, and L. Baldetorp et al., “Reduction in mortality from breast cancer after mass screening with mammography: Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare,” Lancet, vol. 325, no. 8433, pp. 829–832, Apr. 1985.

[4] B. Hellquist, S. Duffy, S. Abdsaleh, L. Björneld, P. Bordás, L. Tabár, B. Viták, S. Zackrisson, L. Nyström, and H. Jonsson, “Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: Evaluation of the Swedish mammography Screening in Young Women (SCRY) cohort,” Cancer, vol. 117, no. 4, pp. 714–722, 2011.

**8.GitHub Link**

<Create a Github account and add your code, dataset and readme file>

<Past the link here>