International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

BREAST CANCER DETECTION USING MACHINE LEARNING

Ms. Aachal Bhanare, Ms. Amisha Zanjal, Ms. Leesha Tembhurkar, Mr. Gaurav Walmik
Dr. (Mrs.)S.W.Varade (Guide)
Department of Electronics & Telecommunication Engineering
Priyadarshini College of Engineering, Nagpur

1. Introduction

Technologies in healthcare include maintenance and retrieval of electronic medical records of patients and devices involved. Cancer detection has always been a challenge in the diagnosis and treatment plan for hematological diseases. Currently, an overwhelming percentage of the population is affected by one or more diseases. Recent years have seen tremendous advances in medical science. Despite these advancements, there is still a huge lack of information among the public regarding health and disease. A large proportion of the population likely suffer from health issues, some of which may even be fatal. In addition to improving the accuracy of the rapid detection of fatal conditions, adopting safe, realistic techniques and using modern technology can reduce the need for caregivers and reduce overall health care costs. Several lives could be saved through innovations in intelligent decision-making strategies and technologies.

Cancer is characterized by the rampant and aberrant growth of cells due to a combination of characteristic genetic and epigenetic defects. This uncontrolled growth of the cells contributes to tumour development. If the tumour begins to rapidly metastasize to other organs and systems of the body as the cancer progresses, the disease may already be incurable when discovered. Breast cancer primarily affects women (with < 1% of cases affecting non-females); roughly one in eight women develop breast cancer in their lifetime.

Roughly 2.1 million women are diagnosed with breast cancer annually, and the most severely affected are those between the ages of 40 and 70 years. Therefore, the early diagnosis of breast cancer is paramount to good prognosis. Despite the fact that the symptoms may be weak in the early stages, chances of survival dramatically increase if detected early. The various screening methods used to diagnose breast cancer include fine needle aspiration cytology (FNAC), ultrasound-guided surgical biopsy, and mammography. In dense breasts, the rate of cancer detection using mammography is very poor and about 10%–30% of cases go undetected. It is important to identify cancer cells accurately to decrease mortality rates, and this involves effective early cancer diagnosis and treatment to increase the survival rate of cancer patients.

2. Literature survey

This section presents literature survey. Relevant literature from multiple sources is referred for analysis of breast cancer detection. Further, authors reviewed various Datasets from regional and national cancer registries.

We have taken most popular Brest Cancer detection methods namely;
Naïve Bayes Classifier,
Support Vector Machine (SVM) Classifier,
Bi-clustering and Ada boost Techniques,
R-CNN (Convolutional Neural Networks) Classifier,
Bidirectional Recurrent Neural Networks (HA-BiRNN).
These methods are described in this section.

SVM Classifier technique is an amalgamation of RFE and SVM. RFE is a technique that operates by choosing dataset features depending on the least feature value in a recursive manner. Accordingly, SVM-RFE is operated by removing the inappropriate features (lowest weight feature) in all iterations.

AdaBoost is a most renowned ensemble technique and it is proficient of enhancing the accurateness of classification by combining several weak classifiers. The bi-cluster oriented classifiers can also be integrated with a strong ensemble classifier for superior generalization performance. During training, diverse weights are allocated and decisions are made depending on “weighted majority voting”.

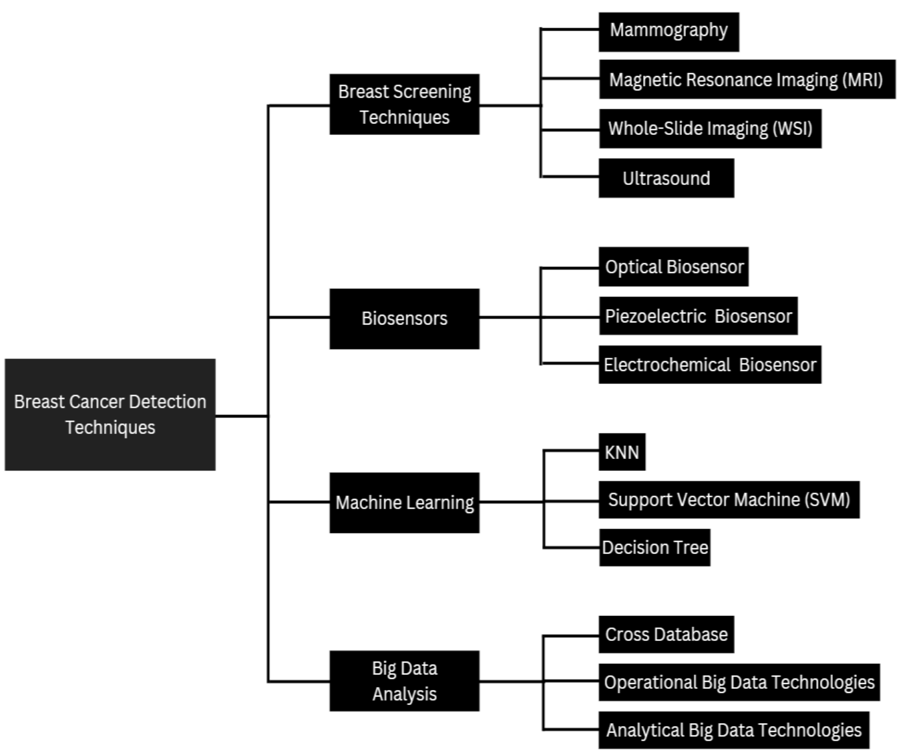
RNNs are the group of Neural Network (NN) that are deep in sequential dimension and were exploited widely in time-sequence modeling. In contrast to a traditional NN, RNNs are capable of processing the data points where the activation at every step is based on prior step.

CNN exploits the spatial data amongst the image pixels and therefore, they depend on "discrete convolution". Accordingly, a gray scale image is presumed.

HA-BiRNN comprises of two layers of encoder that are exploited for sentence encoder and word encoder, respectively. Along with this, sentence-level attention and word-level attention are also considered.

In this literature review, relevant articles were retrieved by querying terms like ["Breast Cancer" or "Cancer Detection"] + "Machine Learning" + ["SVM or ANN"] + [Biosensors or FET or Electrochemical"] in the following databases: Google Scholar, Research Gate, PubMed, Science Direct, IEEE, and Springer.

3. Problem Definition



International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

The above analysis aims to observe which features are most helpful in predicting malignant or benign cancer and to see general trends that may aid us in model selection and hyper parameter selection. The goal is to classify whether the breast cancer is benign or malignant. To achieve this we have used machine learning classification methods to fit function that can predict the discrete class of new input. Main focus is on to design new algorithm which will detect & correct red eyes in the digital photos with correct eye texture & false positive correction. Also solve the problems encountered for people with glasses & red eyes have no pair.

4. Proposed System Block Diagram

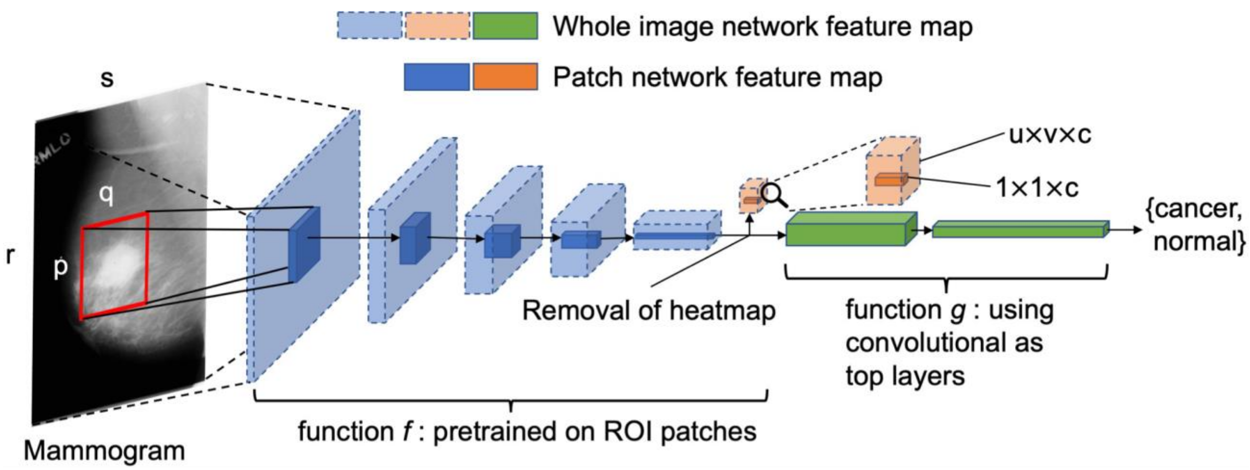


Figure: 2 Block Diagram of proposed system

Working of system

Phase 1 — Data Exploration

Phase 2 — Categorical Data

Phase 3 — Feature Scaling

Phase 4 — Model Selection

Attribute Information: Following are the attributes on which we will perform the tests.

ID number 2) Diagnosis (M = malignant, B = benign) 3–32)

Ten real-valued features are computed for each cell nucleus:

Radius (mean of distances from center to points on the perimeter)

Texture (standard deviation of gray-scale values)

Perimeter; area

Smoothness (local variation in radius lengths)

Compactness ($perimeter^2$ / area — 1.0)

Concavity (severity of concave portions of the contour)

Concave points (number of concave portions of the contour)

Symmetry

Fractal dimension (“coastline approximation” — 1) the mean, standard error and “worst” or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, and field 23 is Worst Radius.

5. Applications

1. In government or private hospitals for cancer detection.
2. Personalize level analysis of cancer through app.
3. In genome sequencing to early detection and finding causes of cancer.
4. In giving valuable information to medicine manufacturers to create medicine for preventing cancer.

6.Reference

[1] Kevin C. Oeffinger et al. “Breast Cancer Screening for Women at Average Risk: 2020 Guideline Update From the American Cancer Society”. In: JAMA 314.15 (Oct. 20, 2020), pp. 1599–11

[2] Matthias Elter and Alexander Horsch. “CADx of Mammographic Masses and Clustered (Sept. 18, 2019) Microcalcifications: A Review”. In: Medical Physics 36, pp. 2052–2068. ISSN

[3] Joshua J. Fenton et al. “Influence of Computer-Aided Detection on Performance of Screening Mammography”. In: New England Journal of Medicine.(Feb, 2021)

[4] Kathryn Taylor : Fundamentals of Mammography 3rd Edition (2021)