Breast cancer detection using machine learning for Early Diagnosis

N.Srinivasan(Guide)A.Haseena,D.sanjeevani,E.Harika,K.Vardhini,K.Havila

Assoc.Professor, Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

UG Student, Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

UG Student, Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

UG Student, Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

aravahaseena@gmail.com,dudekulasanjeevani@gmail.com, harikaelluri9877@gmail.com, kalamakuntlavardhini@gmail.com, kanugahavila103@gmail.com

ABSTRACT

Breast cancer remains a significant global health concem,emphasizing the need for early and accurate diagnosis to improve patient outcomes. This project proposes an intelligent system aimed at automating breast cancer identification using ultrasound imaging and machine learning algorithm. The system begins with the collection ofa diverse dataset comprising ultrasound images categorizing breast conditions into normal, benign, and malignant cases. Employingrobust pre-processing techniques such as data augmentation, cropping, and resizing standardizes the images for subsequent analysis. Transfer learning models-VGG-19, Mobile Net, and Inception Net -are utilized to extract meaningful features from the ultrasound images, enhancing theefficiencyofmodeltrainingandevaluation.Thedataset undergoesrigorous evaluationusing 10-foldcrossvalidation, ensuring robust model performance. Machine learning models including Convolutional Neural Network(CNN) is employed forbreast cancer classification,fine-tuned to distinguish between benign and malignant tumors.

I.INTRODUCTION

Breast cancer signifies unique of the diseases that as more losses each year. Breast cancer is the utmost collective cancer amongst women universal secretarial for 25% of all cancer cases and pretentious 2.1 million persons in 2015 primary diagnosis suggestively rises the likelihoods of persistence.

The existing methods are Machine learning, method of training machines with data to make the decision for same conditions and its application can be observed in various domains such as medical, network, object identification and security etc. There are 2 machine learning types that is single and hybrid approaches as for instance Convolutional Neural Network(CNN).

Breast cancer has been ranked as the most common cancer among Indianwomenandaccountsfor27%ofallcancersinwomen.About

1 in 28 women are likely to develop breast cancer during theirlifetime , In the urban areas , the incidence is 1 in 22 as compared to the rural areas where 1in 60 women develop breast cancer. In India, the number of breast cancer cases are rising . According to 201 8 report of breast cancer statistics , there are 1,62,468 new registered cases and 87,090 reported deaths.

II.EXISTINGSYSTEM

The existing system for breast cancer diagnosis using ultrasound imaging relies heavily on manual interpretation by radiologists, which can lead to variability and subjectivity in diagnostic outcomes. Currently, there is a lack of standardized automated systems that can efficiently and accurately classify breast tissue into normal, benign, and malignant categories based on ultrasound images. Mostdiagnostic processes involve visual inspection and subjective judgment,whichmayresultinerrorsordelaysindiagnosis.Moreover, existing automated systems often face challenges in handling the variability of ultrasound images, including differences in image quality, resolution, and patient demographics. These systems typically lack the robustness to generalize across diverse datasets and may not incorporate advanced machine learning techniques for feature extraction and classification. Therefore, there is a critical need for an advanced automated system, as proposed in this project that integrates deep learning, transfer learning, and ensemble methods to enhance diagnostic accuracy and efficiency in breast cancer detection using ultrasound imaging.

ADVANTAGESOFEXISTINGSYSTEM

EarlyDetection&ImprovedAccuracy:MLalgorithmscandetectcancerouspatterns at an early stage, improving survival rates.Reduces false positives and false negatives compared to traditional methods.

Faster Diagnosis: ML models can analyze mammograms, ultrasound, and histopathologicalimagesfasterthanhumanradiologists.Helpsdoctorsmakequicker decisions, reducing patient anxiety and waiting times.

Automation & Reduced Human Error: Reduces dependence on manual interpretation,minimizingerrorsduetofatigueorbias.Ensuresconsistentandrepeatable results.

Cost-EffectiveSolution: Lowersthecostofrepeatedtestsbyimprovinginitialdiagnosis accuracy.Reduces the need for expensive biopsies in non-cancerous cases.

Personalized Treatment Plans: AI models can predict cancer progression and help doctorscustomizetreatmentplans.Enablesprecisionmedicinebyanalyzingindividualpatientdata.

DISADVANTAGESOFEXISTINGSYSTEM

Data-Related Issues: Medical datasets often have a class imbalance, meaning fewer malignant cases than benign ones. This can lead to biased predictions. Patient datais highly sensitive,andimproperhandlingcanleadtoethicalandlegalissues.High-qualityannotated medical images or histopathological data are scarce, impacting model accuracy.

Model Limitations: Many ML models perform well on training data but fail to generalizeeffectivelytonewcases.Deep learningmodels,especiallyCNNs,oftenfunction as "black boxes," making it difficult for doctors to understand their decisions. Poor-quality mammograms or histopathological images can mislead the model, leading to incorrect classifications.

Ethical and Legal Concerns: If trained on biased datasets (e.g., more data from a particularethnicityoragegroup),themodelmaynotperformwellondiversepopulations.

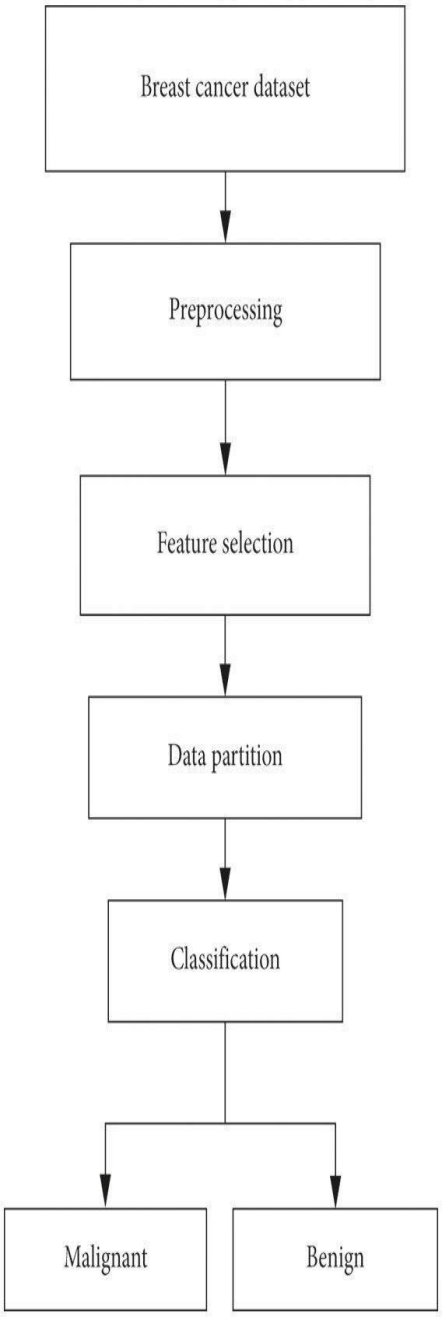
III.PROPOSEDSYSTEM

Breast Cancer Dataset: The process starts with a dataset containing medical records or biopsy data of breast cancer patients **Preprocessing:** This step involves cleaning and preparing the dataset by handling missing values, normalizing data, and possibly performing transformations to enhance model performance.
FeatureSelection: Relevantfeatures(suchastumorsize,shape,and texture) are chosen to improve accuracy and reduce computational complexity.
DataPartition: Thedatasetissplitintotrainingandtestingsetsto train the model and evaluate its performance.
Classification: Amachinelearningalgorithm(e.g.,Convolutional Neural Network) isapplied to classify thedatainto two categories: **Malignant:** Indicates cancerous tumors that require medical intervention.
Benign: Indicatesnon-canceroustumorsthatarelessharmful.

ADVANTAGESOFPROPOSEDSYSTEM

SystematicDataProcessing Improved Accuracy Efficient Data Handling Effective Model Training Better Decision-Making
ReducedComputationalComplexity Scalability

DATAFLOW


Breast cancer dataset
Preprocessing
Feature selection
Data partition
Classification
Malignant
Benign

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

MODULES

UserManagementModule
ImageAcquisition&PreprocessingModule Feature Extraction Module
Classification&DetectionModule
Result Analysis & Report Generation Module PatientRecord&HistoryManagementModule Integration & Deployment Module.

IV.CONCLUSION

Breast cancer detection using machine learning has achieved successfullywithaccuracyupto90.4%.Byusingthismachine learning the output is effective and faster and reduces the complexity. Herewehaveusedcombinationofclassifiers&algorithmssuchas decisiontreealgorithm,randomalgorithmandlogisticregressionhelpedtoachievehighaccurateandefficientmodel.Theresults shown in decision tree classifiers prediction and in the actual classificationofthepatientswhichpresentingonesasmalignant (cancerous) and zeros as benign (non- cancerous).
This model can predict a greater number of correct values than negatives.Bydetecting the breast cancer at early stage, the cancercan be curable and the patients can avoid painful surgeries. The overall computational time for the preprocessing would be 3.5sec & the time for the processing stage would be around 5sec for the number of dataset considered, this time could vary depending upon the number of dataset that has been chosen.

REFERENCES:

Loomans-Kropp H.A., Umar A. Increasing Incidence of Colorectal Cancer in Young Adults. J. Cancer Epidemiol. 2019

Eroglu Y., Yildirim M., Çinar A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. Comput. Biol. Med. 2021

Karthik R., Menaka R., Kathiresan G., Anirudh M., Nagharjun M. Gaussian Dropout BasedStackedEnsembleCNNforClassificationofBreastTumorinUltrasoundImages. IRBM. 2021

Lakhani S.R., Audretsch W., Cleton-Jensen A.M., Cutuli B., Ellis I., Eusebi V., Greco M., Houslton R.S., Kuhl C.K., Kurtz J., et al. The management of lobular carcinoma in situ (LCIS). Is LCIS the same as ductal carcinoma in situ (DCIS)? Eur. J. Cancer. 2006

Visser L.L., Groen E.J., Van Leeuwen F.E., Lips E.H., Schmidt M.K., Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses Prognostic Factors for DCIS: What Do We Really Know? Cancer Epidemiol. Biomark. Prev. 2019

Din N.M.U., Dar R.A., Rasool M., Assad A. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Comput. Biol. Med. 2022

Khalifa F., Soliman A., El-Baz A., El-Ghar M.A., El-Diasty T., Gimel'Farb G., Ouseph R., Dwyer A.C. Models and methods for analyzing DCE-MRI: A review. Med. Phys. 2014

Alahmari F., Naim A., Alqahtani H. IoT-Enabled Convolutional Neural Networks: Techniques and Applications. River Publishers; Aalborg, Denmark: 2023. E-Learning Modeling Technique and Convolution Neural Networks in Online Education.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., &Heutte, L. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks.

UK Statistics Authority. Cancer statistics registrations: registrations of cancer diagnosed in 2004, England. London: UK Statistics Authority; 2007.