



Breast Cancer Detection Using Super Pixel and Faster LSTM_CNN Algorithm

Prof. Pallavi Chaudhari

*Department of Computer Engineering
Sandip Institute of Technology and Research Centre Nashik,
India
pallavi.chaudhari@sitrc.org*

Prajwal Chaudhari

*Department of Computer Engineering
Sandip Institute of Technology and Research Centre Nashik,
India
prajwalchaudhari302@gmail.com*

Priti Borse

*Department of Computer Engineering
Sandip Institute of Technology and Research Centre Nashik,
India
prtiborse13@gmail.com*

Prathamesh Jalgaonkar

*Department of Computer Engineering
Sandip Institute of Technology and Research Centre Nashik,
India
prathmeshjalgaonkar30@gmail.com*

Vrushali Desale

*Department of Computer Engineering
Sandip Institute of Technology and Research Centre Nashik,
India
desalevrushali24@gmail.com*

Abstract

Breast cancer remains a formidable health challenge for women worldwide, underscoring the critical importance of timely and precise detection methods. This study explores the fusion of long short-term memory (LSTM) and convolutional neural network (CNN) architectures to differentiate between malignant and benign breast tumors. Drawing from a comprehensive dataset comprising patient demographics and tumor characteristics, including age, size, shape, and margin, alongside diagnostic outcomes, the model employs LSTM-CNN synergy. While LSTM captures sequential patterns in clinical data, CNN excels in extracting intricate features from mammographic images. This hybrid model synergistically combines the strengths of both architectures, promising improved sensitivity and specificity in breast cancer detection. Experimental validation on benchmark datasets underscores the model's efficacy, signaling its potential as a robust tool for early and accurate breast cancer diagnosis.

Keywords

Breast Cancer Detection, LSTM, CNN, Deep Learning, Classification.

INTRODUCTION

Breast cancer is a significant health concern affecting millions of individuals worldwide, emphasizing the critical need for early detection and diagnosis. In recent years, advanced technologies and machine learning algorithms have played a pivotal role in improving the accuracy and efficiency of breast cancer detection. This study explores the integration of logistic regression and Convolutional Neural Network (CNN) algorithms for enhanced breast cancer detection. Logistic regression, a classical statistical method, is employed as a baseline model. It leverages patient data, such as age, family history, and other relevant features, to predict the likelihood of breast cancer. This initial step allows for a comprehensive analysis of individual risk factors and establishes a foundation for comparison with more advanced techniques. In parallel, CNN, a deep learning algorithm well-suited for image analysis, is introduced to process mammographic images. The CNN algorithm excels in feature extraction and pattern recognition, making it particularly effective in identifying subtle abnormalities indicative of breast cancer. The integration of CNN enhances the model's ability to analyze the intricate details within mammograms and improve overall accuracy. To illustrate the performance of the combined logistic regression and

CNN approach, visual representations such as ROC curves, confusion matrices, and precision-recall curves are employed. These graphical elements provide a clear and intuitive overview of the model's classification performance, allowing for a nuanced evaluation of its strengths and limitations.

Literature Survey

In the medical domain, technology influence is immense, resulting in the processing of medical data electronically. Here the primary concern is security and time consumption while dealing with images. It raises the need for an automated system to retrieve and process images automatically. Medical diagnosis is a vital concept among which the diagnosis of the tedious disease has gained top attention among the research community. This paper focused on sensational breast cancer diagnosis using image processing methodologies. The most effective method of treating disease diagnosis is working with patient records in the form of mammography images.[1]

In this research, the authors are using machine-learning algorithms that will be used to compare the accuracy here. We will present a prediction model using machine-learning models including LR, RF, DT, etc. that are tested on the Breast Cancer Wisconsin dataset, and different datasets. By using the results of this research, the efficiency of various strategies can be defined, and the most precise technique may be applied to accurately forecast breast cancer.[2]

In this section, some studies in the literature are mentioned. The machine learning algorithms and results in the literature developed using the Breast Cancer Wisconsin dataset are shown in Table 1. According to Table 1, generally Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (K-NN) and Naïve Bayes (NB) machine learning algorithms were used.[3]

Machine learning techniques, which have been found to have a major impact on the identification and early treatment of breast cancer, have emerged as a hotspot for study and have been demonstrated to be a potent methodology. The authors used five machine learning algorithms in this research. There have been various empirical investigations employing machine learning and

soft computing techniques to treat breast cancer. This article provides an overview of the available technologies for advanced treatment of breast cancer, as well as an introduction to the concepts and current achievements in personalized medicine that depend on technology.[4]

Diagnosis of BC disease is a challenge for researchers. To solve this problem of breast cancer, various models and techniques such as ML, DL, and TL are used. Researchers used datasets based on mammography (X-rays), magnetic resonance imaging (MRI), ultrasound (sonography), and thermography to diagnose breast cancer disease.[5]

According to the findings, the suggested CNN outperforms estimations in recognising and requesting breast cancer for image datasets. And SVM has proven to outperform CART, NB and KNN in analysis and prediction of cancer with numerical datasets.[6]

In this paper the authors examined different machine learning techniques for breast cancer detection. They performed a comparative analysis of CNN, KNN, SVM, Logistic regression, Naïve Bayes and Random forest. It was observed that CNN outperforms the existing methods when it comes to accuracy, precision and also size of the data set.[7]

In this paper unsupervised learning techniques are utilized on the datasets of Liver, Heart, and Appendicitis. The disease detection is done using the K-Means, Agglomerative Clustering, and Fuzzy CMeans clustering techniques. The metaheuristics algorithms are used on these unsupervised algorithms for better results.[8]

Aruna et al used naïve Bayes, support vector machine, and decision trees to classify a Wisconsin breast cancer dataset and got the best result by using support vector machine (SVM) with an accuracy score of 96.99%. Chaurasia et al compared the performance of supervised learning classifiers by using a Wisconsin breast cancer dataset and naïve Bayes, SVM, neural networks, decision tree methods applied. According to the study results, SVM gave the most accurate result with a score of 96.84%.[9]

Utilizing tissue microarray (TMA) samples obtained from the primary tumor of patients (N=1299) within a

nationwide breast cancer series with long-term-follow-up, the authors train and validate a machine learning method for patient outcome prediction. The prediction is performed by classifying samples into low or high digital risk score (DRS) groups. The outcome classifier is trained using sample images of 868 patients and evaluated and compared with human expert classification in a test set of 431 patient.[10]

A specific area that has garnered significant attention is the prediction of genomics in breast cancer using MRI. Previous work on this topic utilized either imaging features manually extracted by radiologists, which is a very time consuming and subjective process, or features automatically extracted by computer algorithms. Such features include tumor texture (e.g. Haralick features), tumor shape, or enhancement dynamics [11].

An additional level of characterization interrogates the biological characterization of tumors. The emerging field of “imaging genomics” correlates radiographic imaging features with biological data, including somatic mutations, gene expression, chromosome copy number, or other molecular signatures. The maturity of genomics analyses, from a data standpoint, provides synergistic opportunities for AI-based imaging efforts.[12]

Two recent studies [45,46] developed deep learning based methods for breast cancer classification using film and digital mammograms, which were end-to-end trainable. Both studies used multi-instance learning (MIL) and modified the whole image classifier cost functions to satisfy the MIL criterion. In contrast to our approach, neither study utilized ROI annotations to train the patch classifiers first and the AUCs were lower than reported in study.[13]

In this paper, the authors analyze the performance of supervised or unsupervised learning models such as Random Forest, KNN, SVM, Logistic Regression, AdaBoost, and Perceptron. These models are used for classifying the Wisconsin Breast Cancer Dataset (WBCD) from UCI Machine learning depository. They have been tested by many researchers and were proven efficient in many aspects.[14]

In the setting of a challenge competition, some deep learning algorithms achieved better diagnostic performance than a panel of 11 pathologists participating in a simulation exercise designed to mimic routine pathology workflow; algorithm performance was comparable with an expert pathologist interpreting slides without time constraints. Whether this approach has clinical utility will require evaluation in a clinical setting.[15]

Classification and clustering are two widely used methods in data mining. Clustering methods aim to extract information from a data set to discover groups or clusters and describe the data set itself. Classification, also known as supervised learning in machine learning, aims to classify unknown situations based on learning existing patterns and categories from the data set and subsequently predict future situations. The training set, which is used to build the classifying structure, and the test set, which tends to assess the classifier, are commonly mentioned in classification tasks. This research mainly focuses on the classification and prediction of breast cancer. Related works regarding supervised learning techniques and their application in breast cancer diagnosis are reviewed in this section.[16]

According to the latest PubMed statistics, more than 1500 papers have been published on the subject of machine learning and cancer. However, the vast majority of these papers are concerned with using machine learning methods to identify, classify, detect, or distinguish tumors and other malignancies. In other words machine learning has been used primarily as an aid to cancer diagnosis and detection (McCarthy et al. 2004). It has only been relatively recently that cancer researchers have attempted to apply machine learning towards cancer prediction and prognosis.[17]

MATERIALS AND METHODS

Breast cancer is the most common cancer among women worldwide, accounting for 25% of all cancers. Although the overall incidence of breast cancer has been decreasing in recent years, it is still a major public health concern. The exact cause of breast cancer is unknown, but there are several risk factors that can increase a woman's chance of developing the disease. These risk factors include:

- Age: Breast cancer risk increases with age. Most breast cancers are diagnosed in women over the age of 50, but they can occur at any age.

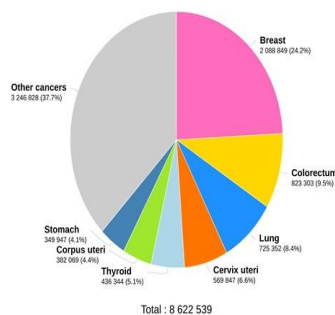
- Family history: Having a close relative (mother, sister, daughter) with breast cancer increases a woman's risk of developing the disease.
- Genetic mutations: Certain genetic mutations, such as BRCA1 and BRCA2, can significantly increase a woman's risk of breast cancer.
- Dense breasts: Women with dense breasts have more breast tissue than fatty tissue, which can make it more difficult to detect breast cancer on mammograms.
- Personal history of breast cancer: Women who have been diagnosed with breast cancer in the past are at an increased risk of developing it again.



Images CNN uses for detection

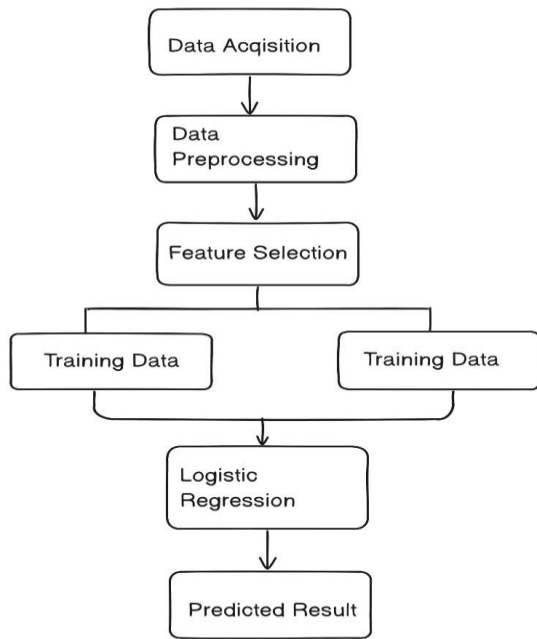
In the workflow for breast cancer detection using logistic regression and Convolutional Neural Network (CNN), patient data is initially processed for logistic regression, extracting relevant features such as age and family history. Simultaneously, mammographic images are prepared for CNN analysis through resizing and normalization. Logistic regression is employed for patient-level risk assessment, while CNN is utilized to analyze intricate details within mammograms. The predictions from both models are then integrated, providing a comprehensive approach to breast cancer detection that leverages both statistical and image-based information, enhancing the accuracy and robustness of the overall system.

Estimated number of new cases in 2018, worldwide, females, all ages

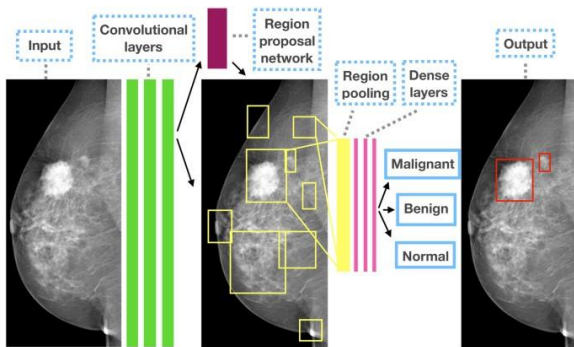


women suffering breast cancer

Convolutional Neural Networks (CNN) have emerged as a groundbreaking technology in the detection of breast cancer through medical imaging. By leveraging the intricate layers of neural networks, CNNs excel in extracting and analyzing complex features within mammographic images, allowing for the identification of subtle abnormalities indicative of breast cancer. These networks are adept at recognizing patterns and structures, enabling accurate classification and early detection of potential malignancies. The application of CNNs in breast cancer imaging not only enhances the efficiency of diagnosis but also contributes to the ongoing efforts to improve patient outcomes by facilitating early intervention and treatment.



Model

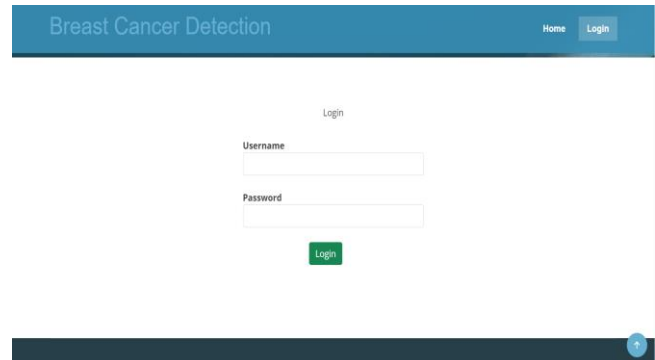


Affected Region detected in images

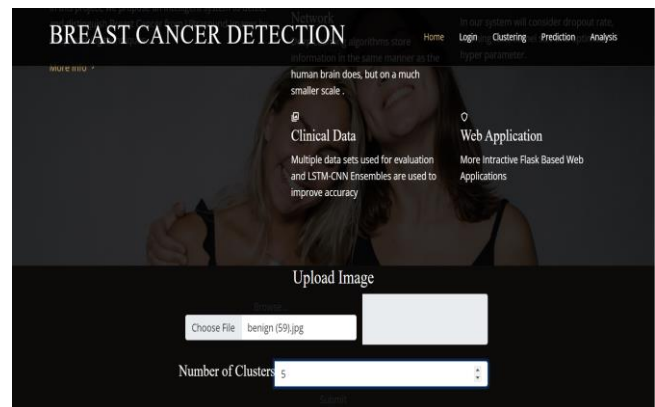
Convolutional Neural Network (CNN) offer complementary strengths in breast cancer detection. Logistic regression efficiently processes patient data, providing a statistical assessment of risk factors such as age and family history. This patient-level analysis contributes valuable insights into the likelihood of breast cancer. On the other hand, CNN excels in image analysis, capturing intricate patterns and abnormalities within mammograms that may be indicative of malignancies. The combination of these approaches creates a synergistic model that harnesses both patient information

and image features, enhancing overall accuracy and enabling a more nuanced and comprehensive detection of breast cancer. This integrated approach maximizes the strengths of each method, offering a powerful tool for early diagnosis.

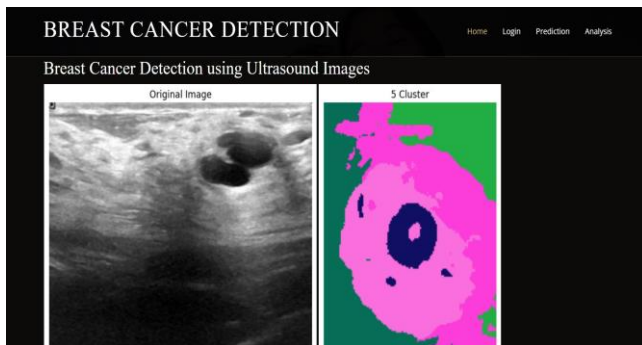
RESULT



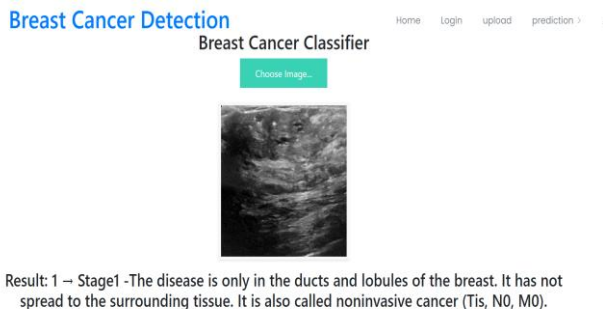
"Welcome to our platform! To get started, please log in with your username and password or register for a new account below."



"Select an image and specify the number of clusters to analyze its visual data. Our clustering algorithm will group similar elements, providing valuable insights into the underlying patterns and structures within the image."



"The image has been processed, and the clusters are visualized above, revealing distinct groupings of similar elements within the image data. Each cluster represents a unique set of visual features, providing a deeper understanding of the content and composition of the image."



"After analyzing the image, the model will provide the classification of the type of breast cancer."

VI. Conclusion

In conclusion, the utilization of logistic regression for breast cancer detection has proven to be a valuable and effective tool in distinguishing between malignant and benign tumors. This system has demonstrated its significance in the field of medical diagnostics by providing healthcare professionals with a reliable method for early detection and accurate classification of breast cancer cases. The CNN model's ability to analyze multiple features, such as tumor size, shape, and texture, has enabled the system to make informed predictions regarding the nature of the tumor. By analyzing historical data and training the model with a diverse dataset, this system has showcased its capacity to identify patterns and establish decision boundaries, making it a robust tool for breast cancer classification.

VII. FUTURE SCOPE

The future scope of a "Breast Cancer Detection Using Logistic Regression" system for distinguishing between malignant and benign breast tumors holds significant promise in the field of healthcare and medical technology. As technology continues to advance, the potential impact of such a system becomes increasingly apparent. First and foremost, the integration of machine learning and logistic regression techniques for breast cancer detection offers the possibility of enhancing early diagnosis. By analyzing a multitude of patient data, including mammograms, patient history, and genetic information, the system can provide more accurate and timely assessments. This, in turn, can lead to earlier intervention and improved treatment outcomes for breast cancer patients, ultimately saving lives. Additionally, the system's future scope extends to personalized medicine. As it continues to gather and analyze data from a wide range of patients, it can better tailor treatments to individuals based on their unique characteristics and the likelihood of malignancy. This means that treatments can be more effective and have

fewer side effects, resulting in an overall improvement in the quality of care for breast cancer patients.

ACKNOWLEDGMENT

- First and foremost, we wish to record our sincere gratitude to the Management of this college and our Respected Principal **Prof. (Dr) M. M. Patil**.
- Our sincere thanks to **Prof. (Dr) Ankita V. Karale**, Head, Department of Computer, Sandip Institute of Technology and Research Centre, Nashik.
- We express our sincere gratitude to our Guide, **Prof. Pallavi Chaudhari** for guiding us in the investigations of this project and in carrying out experimental work.

REFERENCES

1. Sivakumar M; M. Prathyusha; C. Sasikala; A. Anandaraj; Indirani. M: "Machine Learning Based Breast Cancer Detection Using Logistic Regression", Date of Conference: 16-17 December 2022.
2. Breast Cancer Detection Technique using Machine Learning Classifiers,28-05-2023 /10-07-2023,GUHAROY, RABEL; Hada, Dipali; Priyam, Aditya; Mehdi, Abbas; Choudhary, Divyansh (2023). Breast Cancer Detection Technique using Machine Learning Classifiers. TechRxiv. Preprint.
3. Breast Cancer Detection Using Machine Learning Algorithms Ahmet Enes KILIÇ 1 , Murat KARAKOYUN 2 1 Computer Engineering, Necmettin Erbakan University, KONYA 2 Computer Engineering, Necmettin Erbakan University, KONYA,2023
4. Identification and Diagnosis of Breast Cancer using a Composite Machine Learning Techniques Prabu kannan G1 , Abinash M.J2 , S, Suganya E3 , Sountharajan S4 ,Bhuvaneshwari R5 , K.Geetha 6 1 School of Computing Science and Engineering, VIT Bhopal University, India 2Department of Information Technology, Arulmigu Kalasalingam College of Arts and Science, Virudhunagar, India. 3Department of Information Technology, SSN College of Engineering, SH 49A, Kalavakkam, India 4,Department of CSE, School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India 6CSE, Excel engineering College,2022
5. Breast Cancer Detection and Classification Empowered With Transfer Learning Sahar Arooj 1 , Atta-ur-Rahman2 , Muhammad Zubair 3 , Muhammad Farhan Khan4 , Khalid Alissa5 , Muhammad Adnan Khan6 * and Amir Mosavi 7,8,9, 2022
6. Breast Cancer Prediction Using Machine Learning Techniques Apoorva V1,* , Yogish H K2 , Chayadevi M L3, 2021
7. Breast Cancer Detection using Machine Learning Techniques Sweta Bhise Computer Engineering SRIEIT, Goa University Shiroda, India Simran Bepari Computer EngineeringSRIEIT, Goa University Shiroda, India Shrutika Gadekar Computer Engineering SRIEIT, Goa University Shiroda, India, 2021
8. Prediction of human diseases using optimized clustering techniques,Finney Daniel Shadrach Anna University, Chennai ,Pacha Shobha R.M.D Engineering College, 2021
9. Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications Muhammet Fatih Ak Industrial Engineering Department, Antalya Bilim University, 07190 Antalya, Turkey; fatih.ak@antalya.edu.tr; Tel.: +90-242-245-0000 Received: 1 March 2020; Accepted: 14 April 2020; Published: 26 April 2020
10. R. Turkki, et.al., "Breast cancer outcome prediction with tumor tissue images and machine learning", Breast Cancer Research and Treatment, vol.177, no.1, pp. 41–52, 2019
11. Z. Zhu, et.al., "Deep learning for identifying radiogenomic associations in breast cancer", Computers in Biology and Medicine, vol. 109, pp. 85–90, 2019
12. W. L. Bi, et al., "Artificial intelligence in cancer imaging: clinical challenges and applications", CA: A Cancer Journal for Clinicians, vol. 69, pp. 127–157, 2019.



13. L. Shen, et.al., “Deep learning to improve breast cancer detection on screening mammography”, Scientific Reports, vol. 9, no. 1, pp. 1–12, 2019.
14. Q. H. Nguyen, et al., “Breast cancer prediction using feature selection and ensemble voting”, International Conference on System Science and Engineering (ICSSE), pp. 250–254, Vietnam, 2019.
15. B. Ehteshami Bejnordi, et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer”, JAMA, vol. 318, no. 22, pp. 2199–2210, 2017.
16. H. Wang and S. W. Yoon, “Breast cancer prediction using data mining method”, IIE Annual Conference Expo 2015, pp. 818–828, USA, 2015
17. Applications of Machine Learning in Cancer Prediction and Prognosis Joseph A. Cruz, David S. Wishart
Departments of Biological Science and Computing Science, University of Alberta Edmonton, AB, Canada T6G2E8, 2006