Breast Cancer Detection using CNN

Prof. K. C. Nalavade¹, Shivani Baheti², Jayshri Hagawane³, Dnyaneshwari Harde⁴, Aditya Mehetre⁵

Department of Computer Engineering, Sandip Institute of Engineering and Management, Mahiravani, Nashik

Abstract
Breast cancer is a significant global health concern affecting millions of individuals worldwide. Early detection and prevention are crucial for improving the prognosis and reducing the mortality rate associated with this disease. This abstract introduces a comprehensive breast cancer detection and prevention system leveraging the power of deep learning and data analytics. The proposed system consists of two modules: Module 1: Image-Based Breast Cancer Detection and Prevention. In this module, users can upload mammographic or ultrasound images for the automatic detection of breast cancer. Deep learning algorithms, specifically convolutional neural networks (CNNs), are employed to analyze the uploaded images. The system classifies the images as benign or malignant, providing users with an instant diagnosis. Additionally, it offers personalized prevention and early intervention recommendations based on the identified risk factors and cancer type. The preventive suggestions may include lifestyle changes, screening schedules, and recommended medical consultations. Module 2: Data-Driven Breast Cancer Detection and Prevention. This module allows users to input personal health parameters such as blood pressure, blood sugar levels, family history, and other relevant factors into a structured form. The system utilizes these inputs to assess the user’s risk of developing breast cancer. Advanced data analysis and machine learning models are employed to correlate the provided information with breast cancer risk factors. The system then offers tailored preventive strategies and early detection recommendations to help users mitigate their risk. The system’s strength lies in its ability to combine image-based diagnosis with data-driven analysis, ensuring a holistic approach to breast cancer detection and prevention. By leveraging deep learning and AI technologies, it can provide accurate and timely assistance to users in their efforts to identify and mitigate breast cancer risks. Such a system holds great potential in improving the early detection and prevention of breast cancer, ultimately contributing to better health outcomes for individual.

Keywords: Breast Cancer, Image, Data parameter

INTRODUCTION
The project idea is to develop an integrated Breast Cancer Detection and Prevention System that harnesses the capabilities of deep learning, data analytics, and artificial intelligence. The system will consist of two modules: an Image-Based Breast Cancer Detection Module and a Data-Driven Breast Cancer Prevention Module. In the Image-Based Module, users can upload mammographic or ultrasound images, and the system will employ convolutional neural networks (CNNs) to provide instant diagnoses, classifying images as benign or malignant. Additionally, the system will offer personalized recommendations for prevention and early intervention based on identified risk factors and cancer type. The Data-Driven Module will allow users to input their personal health parameters, such as blood pressure, blood sugar levels, and family history, and employ advanced data analysis and machine learning models to assess their risk of developing breast cancer.
Tailored preventive strategies and early detection recommendations will be provided to help mitigate this risk. This project aims to bridge the gap between image based diagnosis and data-driven risk assessment, ultimately contributing to better health outcomes by improving early detection and prevention efforts for breast cancer.

1. PURPOSE
   - Identify need of Project

The motivation behind undertaking this project is rooted in the urgent need to address the global health crisis presented by breast cancer. Breast cancer remains a leading cause of mortality among women worldwide, and early detection plays a pivotal role in improving survival rates. Despite significant advancements in medical imaging and healthcare, there are still challenges in achieving timely and accurate breast cancer diagnoses, particularly in resource-constrained regions. The motivation for this project stems from the desire to make a meaningful impact by leveraging cutting-edge technology and data-driven insights. By combining deep learning, data analytics, and artificial intelligence, we aim to create a comprehensive system that empowers individuals with accessible and personalized tools for early detection and prevention. The ultimate goal is to save lives, enhance the quality of life for those affected by breast cancer, and contribute to broader efforts to reduce the burden of this disease on both individuals and healthcare systems. This project is driven by a commitment to improving public health, raising awareness, and providing innovative solutions that can positively impact the lives of individuals at risk of or affected by breast cancer.

MOTIVATION OF PROJECT

The motivation behind the proposed breast cancer detection and prevention system stems from the pressing need to address the significant global health burden posed by breast cancer. As one of the most prevalent cancers worldwide, early detection and prevention are paramount in improving prognosis and reducing mortality rates associated with the disease. Traditional methods of detection often rely on periodic screenings and manual interpretation of medical images, which can be time-consuming and prone to human error. By leveraging the power of deep learning and data analytics, the proposed system aims to revolutionize breast cancer detection and prevention by providing users with timely and accurate diagnoses, personalized prevention recommendations, and risk assessments based on both medical imaging and personal health data.

Furthermore, the integration of image-based diagnosis with data-driven analysis offers a holistic approach to breast cancer management. By combining information from mammographic and ultrasound images with personal health parameters such as blood pressure, blood sugar levels, and family history, the system can provide a more comprehensive understanding of individual risk factors and tailor preventive strategies accordingly. This holistic approach not only enhances the accuracy of breast cancer detection but also empowers individuals to take proactive measures in mitigating their risk of developing the disease. Additionally, by offering personalized recommendations for lifestyle changes, screening schedules, and medical consultations, the system aims to empower users to make informed decisions about their health and well-being, ultimately contributing to better health outcomes and reduced healthcare costs associated with breast cancer treatment.

Risk Mitigation: Utilize advanced data analysis and machine learning models to correlate the provided health information with breast cancer risk factors, enabling the system to suggest tailored strategies for mitigating the user’s risk of developing breast cancer.
LITERATURE SURVEY:

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Title</th>
<th>Author</th>
<th>Year &amp; Publication</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breast Cancer Prediction Using Machine Learning Classifiers</td>
<td>Janaal</td>
<td>2022 5th International Conference on Advances in Science and Technology (ICAST)</td>
<td>Breast cancer is the most common type of cancer affecting women worldwide. Early diagnosis is crucial for successful treatment, recovery, and mortality reduction. Therefore, various non-invasive, minimally invasive, and invasive approaches have been developed for breast cancer detection.</td>
</tr>
<tr>
<td>2</td>
<td>Breast Cancer Prediction using Optimized Machine Learning Classifiers and Data Balancing Techniques</td>
<td>Saniya Anicesarin</td>
<td>2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)</td>
<td>The detector employs 16 rotating antennas covered with a radome contacting the surface of the breast phantom without a coupling medium. The detector can detect a breast cancer phantom target even under the imperfect contact condition between the breast skin surface and the antennas.</td>
</tr>
<tr>
<td>3</td>
<td>Breast Cancer Prognosis using Machine Learning Applications</td>
<td>Manodapu Akhil</td>
<td>2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)</td>
<td>Breast cancer is the most popular cancer for women. The sentinel lymph node (SLN) biopsy is significant for breast cancer treatment. Currently, SLN localization is normally based on visual observation and fluorescence imaging with low resolution due to the strong light scattering.</td>
</tr>
<tr>
<td>4</td>
<td>Machine Learning-Based Approaches For Breast Cancer Detection in Microwave Imaging</td>
<td>Humza Sani</td>
<td>2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)</td>
<td>In this study, we have shown the diffuse optical breast-scanning (DOB-Scan) probe, utilizing a machine learning model, can discriminate between normal and malignant breast tissues.</td>
</tr>
</tbody>
</table>
Overall, the proposed breast cancer detection and prevention system addresses a critical gap in current healthcare practices by leveraging cutting-edge technologies to provide accurate, timely, and personalized assistance to individuals at risk of developing breast cancer. By harnessing the power of deep learning and data analytics, the system has the potential to significantly improve early detection rates, reduce mortality rates, and ultimately, enhance the quality of life for millions of individuals worldwide affected by breast cancer.

OBJECTIVE OF SYSTEM

Tamara Saad Mohamed; Saad Mohammed Khalifah,
Breast Cancer Prediction: The Classification of Non-
- Early Detection: Develop a system capable of early Recurrence-Events and Recurrence-Events Using Functions detection of breast cancer by analyzing mammographic or Classifiers, 2022 3rd Information Technology To Enhance e- ultrasound images using deep learning algorithms, learning and Other Application (IT-ELA): Each year, the

particularly convolutional neural networks (CNNs).

Percentage of breast cancer related deaths increases dramatically. It represents most frequent kind of cancer and it is

- Accurate Diagnosis: Ensure the system provides accurate the main reason why women worldwide pass away. Any and reliable diagnoses by classifying uploaded images as improvement in the prediction and diagnosis of cancer is either benign or malignant, allowing for timely medical essential for living a healthy life. Consequently, a high level of intervention.

accuracy in cancer prognosis is essential for improving the treatment aspect and the standard of patient survival. Machine

- Personalized Recommendations: Offer personalized learning's approaches can help a lot with predicting and prevention and early intervention recommendations to diagnosing breast cancer early. They have become a popular

users based on identified risk factors, cancer type, and area for researches and have been shown to be a useful method.
Individual health parameters.

Almost researches are studies for prediction and diagnosis, the most important thing after the patient survive is to ensure the breast cancer does not recurrence in the future. Rarely studies is

- Data-Driven Analysis: Create a module for data-driven talking about this issue. This study discussing the recurrence and breast cancer risk assessment by collecting and analyzing non-recurrence breast cancer events after the patient survive, to personal health parameters such as blood pressure, blood analyze and classify this issue we used data about breast cancer sugar levels, family history, and relevant factors. And depend on the classification classifiers (functions classifier) of machine learning: (Logistic classifier, SGD classifier, SGDTex classifier, SimpleLogistic classifier, SMO classifier, VotedPreceptron classifier), to find the most effective classifier for the best result of prediction the recurrence and non-recurrence breast cancer events. The high-level classifier is SGDTex and the low-level classifier is the logistic according to the results of this paper.

Saniya Anklesaria; Unnati Maheshwari; Ria Lele; Priyanka Verma, Breast Cancer Prediction using Optimized Machine Learning Classifiers and Data Balancing Techniques, The most prevalent malignancy in females and the second greatest reason for the loss of life from cancer is Breast Cancer. Hence, a computer assisted detection (CAD) system that uses a machine learning technique to give reliable breast cancer diagnosis is required. The paper is aimed to incorporate several machine learning (ML) algorithms, including Support Vector Machine (SVM), Logistic Regression, k-Nearest Neighbour (KNN), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), and Nave Bayes (NB) with hyperparameter tuning using the Random Forest Feature Importance Method for feature selection. These models have been trained on the Wisconsin Diagnostic Breast Cancer (WDBC) Dataset. Furthermore, the dataset was balanced using both Undersampling and SMOTE from which we concluded that Undersampling gave us an overall better result. The performance evaluation parameters for the designed model are specificity, accuracy, sensitivity, F1 score, precision, recall and Area Under the Curve (AUC). The research concluded that the Support Vector Machine Algorithm proved to be the most effective model which fit our dataset with an Accuracy of 95.8% followed by KNN with an accuracy of 95.3%.

Mandalapu Akhil; P.V. Siva Kumar, Breast Cancer Prognosis using Machine Learning Applications, The maximum deaths worldwide among women is breast cancer (BC). Breast cancer that mainly develops from breast tissues. Due to current population growth in medical research, early diagnosis of cancer has become a critical issue. The likelihood of death is increasing exponentially due to breast cancer as the world's population grows. Breast cancer is the second leading severe cancers that have already been revealed. Cancer prediction therefore plays an important role in updating treatment aspects and survivability standards. As a result of machine learning methods, breast cancer prediction and early detection have been improved significantly, and these methods have been a center of research for this field. Using the Wisconsin Diagnostic Breast Cancer dataset (WBCD), 5 ML algorithms were compared: Decision tree (C4.5), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and K-Nearest Neighbour (KNN). Machine learning algorithms are being used to identify the most effective algorithms for early detection and identifying breast cancer in its earliest stages.

IJIRMPS230521 Website: www.ijirmps.org Email: editor@ijirmps.org
Humza Sami; Mahnoor Sagheer; Kashif Riaz; Muhammad Qasim Mehmood; Muhammad Zubair, Machine Learning-Based Approaches For Breast Cancer Detection in Microwave Imaging, Detection of breast cancer at an early stage can significantly reduce the mortality rate. Microwave imaging is a promising detection tool for harmless and non-ionizing screening of breast cancer. In this work, a fast and accurate machine learning algorithm is proposed for the prediction of the breast lesion using microwave signals. Machine learning has proved itself reliable in the field of biomedical application where the diagnosis of the disease is desired. The support vector machine (SVM) algorithm with the linear and polynomial kernel is trained and tested on raw backscattered signals data. SVM with third-degree polynomial kernel obtained 99.7% accuracy that outperforms the existing conventional machine learning binary classification algorithms. Thus, the prediction of tumor presence would help the radiologist to diagnose tumor correctly at early stages.

PROPOSED SYSTEM

The system classifies the images as benign or malignant, providing users with an instant diagnosis. Additionally, it offers personalized prevention and early intervention recommendations based on the identified risk factors and cancer type. The preventive suggestions may include lifestyle changes, screening schedules, and recommended medical consultations. Module 2: Data-Driven Breast Cancer Detection and Prevention

This module allows users to input personal health parameters such as blood pressure, blood sugar levels, family history, and other relevant factors into a structured form. The system utilizes these inputs to assess the user’s risk of developing breast cancer. Advanced data analysis and machine learning models are employed to correlate the provided information with breast cancer risk factors. The system then offers tailored preventive strategies and early detection recommendations to help users mitigate their risk. The system’s strength lies in its ability to combine image-based diagnosis with data-driven analysis, ensuring a holistic approach to breast cancer detection and prevention.

By leveraging deep learning and AI technologies, it can provide accurate and timely assistance to users in their efforts to identify and mitigate breast cancer risks. Such a system holds great potential in improving the early detection and prevention of breast cancer, ultimately contributing to better health outcomes for individuals.

SYSTEM ARCHITECTURE
The preventive suggestions may include lifestyle changes, screening schedules, and recommended medical consultations. Module 2: Data-Driven Breast Cancer Detection and Prevention This module allows users to input personal health parameters such as blood pressure, blood sugar levels, family history, and other relevant factors into a structured form. The system utilizes these inputs to assess the user’s risk of developing breast cancer. Advanced data analysis and machine learning models are employed to correlate the provided information with breast cancer risk factors. The system then offers tailored preventive strategies and early detection recommendations to help users mitigate their risk. The system’s strength lies in its ability to combine image-based diagnosis with data-driven analysis, ensuring a holistic approach to breast cancer detection and prevention. By leveraging deep learning and AI technologies, it can provide accurate and timely assistance to users in their efforts to identify and mitigate breast cancer risks. Such a system holds great potential in improving the early detection and prevention of breast cancer, ultimately contributing to better health outcomes for individual

USE CASE DIAGRAM

ADVANTAGES
• Easy to used system
• Control system from anywhere
• Centralized system

SYSTEM REQUIREMENTS
• Software Used:
  1. Programming Language – Python
  2. Libraries – NumPy, Keras, OpenCV, Streamlit
  3. Database – SQLite
  5. Algorithm – CNN, SVM
• Hardware Used:
  1. Processor – i3 or above
  2. Hard Disk – 150 GB
  3. Memory – 4GB RAM
ALGORITHM
In the proposed breast cancer detection and prevention system, Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) can be utilized in complementary ways to analyze medical images and personal health data, respectively.

CNNs excel at image classification tasks and are well-suited for analyzing mammographic and ultrasound images in Module 1 of the system. These networks can automatically learn and extract relevant features from the images, enabling accurate classification of breast tissue as benign or malignant. For this purpose, CNN architectures can be trained using a labeled dataset of mammographic and ultrasound images, with each image labeled as either benign or malignant. During training, the CNN learns to identify patterns and features indicative of cancerous or healthy tissue, allowing it to make accurate predictions on new, unseen images uploaded by users. The trained CNN models serve as the backbone of the image-based breast cancer detection module, providing instant diagnoses and personalized prevention recommendations based on the analysis of uploaded images.

On the other hand, Support Vector Machines (SVMs) can be employed in Module 2 of the system to analyze personal health parameters and assess individual breast cancer risks. SVMs are well-suited for binary classification tasks and can effectively handle high-dimensional data. In this module, SVMs can be trained using a dataset of personal health parameters, such as blood pressure, blood sugar levels, family history of breast cancer, and lifestyle factors, along with corresponding labels indicating the presence or absence of breast cancer. The trained SVM models can then assess the risk of developing breast cancer for individual users based on their inputted health parameters. By correlating these parameters with known breast cancer risk factors, SVMs can provide tailored preventive strategies and early detection recommendations to help users mitigate their risk of developing the disease.

Overall, the integration of CNNs and SVMs into the breast cancer detection and prevention system allows for a comprehensive and multi-faceted approach to combating the disease. While CNNs analyze medical images to provide accurate diagnoses, SVMs analyze personal health data to assess individual risks and provide personalized recommendations, thereby empowering users to take proactive steps in managing their health and preventing the onset of breast cancer.

METHODOLOGY
1. Data Collection: Gather a diverse and comprehensive dataset of mammographic and ultrasound images of breast tissue. Additionally, collect relevant personal health parameters such as blood pressure, blood sugar levels, family history of breast cancer, lifestyle factors, etc.

2. Data Preprocessing:
   - Image Preprocessing: Perform standardization, normalization, and augmentation techniques to ensure consistency and enhance the quality of mammographic and ultrasound images.
   - Data Cleaning: Handle missing values, outliers, and inconsistencies in the personal health parameters dataset.

3. Module 1: Image-Based Breast Cancer Detection and Prevention:
   - Develop Convolutional Neural Network (CNN) architectures suitable for image classification tasks. This may involve architectures like VGG, ResNet, or custom-designed networks.
   - Train the CNN models using the preprocessed image dataset with appropriate labels (benign or malignant).
   - Evaluate the trained models using metrics such as accuracy, precision, recall, and F1-score.
   - Implement an interface for users to upload mammographic or ultrasound images.
Integrate the trained CNN models into the system to automatically classify uploaded images and provide instant diagnosis.

- Develop algorithms to analyze risk factors and cancer type, and provide personalized prevention and early intervention recommendations based on the diagnosis.

4. Module 2: Data-Driven Breast Cancer Detection and Prevention:

- Utilize machine learning algorithms such as logistic regression, decision trees, or ensemble methods to analyze the personal health parameters dataset and assess the user’s risk of developing breast cancer.
- Train the machine learning models to correlate the input parameters with breast cancer risk factors.
- Develop algorithms to generate tailored preventive strategies and early detection recommendations based on the assessed risk.
- Implement an interface for users to input personal health parameters and receive personalized recommendations.

5. Integration and Deployment:

- Integrate both modules into a cohesive system, ensuring seamless communication between them.
- Develop a user-friendly interface for easy interaction with the system.
- Deploy the system on a secure and scalable platform, ensuring accessibility to users while maintaining data privacy and security.

6. Evaluation:

- Conduct thorough testing and validation of the integrated system using real-world data and scenarios.
- Evaluate the performance of the system in terms of accuracy, usability, and effectiveness in early detection and prevention of breast cancer.
- Gather feedback from users and healthcare professionals to identify areas for improvement.

7. Continuous Improvement:

- Monitor the system's performance in real-world settings and gather additional data to enhance its accuracy and effectiveness.
- Incorporate feedback and suggestions from users and experts to iteratively improve the system's capabilities and features.

RESULTS
DATASET INFORMATION

RSNA Breast Cancer Dataset:
The RSNA Breast Cancer Dataset available on Kaggle consists of a collection of mammographic images in PNG format. These images are standardized to a resolution of 512x512 pixels, making them suitable for analysis using deep learning techniques. Mammographic images are a crucial component of breast cancer diagnosis and screening, as they allow healthcare professionals to visualize abnormalities and potential signs of cancerous growth within breast tissue. This dataset provides a valuable resource for training convolutional neural networks (CNNs) to accurately classify mammographic images as either benign or malignant, thereby aiding in the automated detection of breast cancer. With a large number of high-resolution images, the RSNA Breast Cancer Dataset offers ample opportunities for researchers and developers to build robust deep learning models for breast cancer detection and prevention.

Breast Cancer Dataset:
The Breast Cancer Dataset available on Kaggle is a comprehensive collection of personal health data and clinical attributes related to breast cancer patients. This dataset includes features such as patient age, tumor size, tumor type, lymph node status, hormone receptor status, and other clinical variables relevant to breast cancer diagnosis and prognosis. Additionally, the dataset contains information about the recurrence of cancer and patient survival outcomes. With a diverse range of clinical attributes and outcomes, the Breast Cancer Dataset provides a rich resource for training machine learning models to assess individual breast cancer risks and predict patient outcomes. Researchers and healthcare professionals can utilize this dataset to develop predictive models and decision support systems for personalized breast cancer screening, treatment planning, and follow-up care, ultimately improving patient outcomes and quality of life.

CONCLUSION

This breast cancer detection and prevention system represents a pioneering fusion of advanced technology and medical care. Seamlessly integrating Convolutional Neural Networks for image analysis and Support Vector Machines for data-driven risk assessment, this innovative approach provides a comprehensive solution to combatting breast cancer. By swiftly analyzing medical images, the system offers rapid diagnoses, enabling timely interventions and personalized prevention strategies based on identified risk factors and cancer types. Simultaneously, through the assessment of individual health parameters, it delivers tailored recommendations to mitigate the risk of developing breast cancer. This holistic approach, powered by deep learning and machine learning algorithms, not only aids in early detection but also empowers individuals to take proactive steps towards their health, potentially transforming outcomes and reducing the global burden of breast cancer.

FUTURE SCOPE

The future scope of the proposed breast cancer detection and prevention system is promising, with several
avenues for further development and enhancement. Firstly, ongoing advancements in deep learning and artificial intelligence (AI) technologies present opportunities to refine and optimize the system's algorithms for image-based diagnosis and data-driven analysis. By leveraging larger and more diverse datasets, as well as state-of-the-art neural network architectures, the system can improve its accuracy and efficiency in detecting breast cancer at early stages, leading to more timely interventions and improved patient outcomes.

Secondly, the integration of emerging technologies such as wearable devices and Internet of Things (IoT) sensors holds great potential for enhancing the system's capabilities in monitoring and tracking individual health parameters. By incorporating real-time data from wearable devices that track physical activity, heart rate variability, and other relevant metrics, the system can provide more dynamic and personalized risk assessments, enabling users to proactively manage their health and prevent the onset of breast cancer through lifestyle modifications and early interventions.

Moreover, the expansion of the system's scope to include predictive modeling and risk stratification for other types of cancer and chronic diseases represents an exciting avenue for future research and development. By applying similar methodologies to analyze medical imaging and personal health data in the context of other diseases, such as lung cancer or cardiovascular disease, the system can potentially contribute to a more comprehensive approach to preventive healthcare, allowing for earlier detection and intervention across a broader range of health conditions. Overall, the future scope of the breast cancer detection and prevention system is characterized by continued innovation and integration of advanced technologies, with the ultimate goal of improving health outcomes and reducing the global burden of disease.

REFERENCES


