A PROJECT REPORT

on

"BREAST CANCER DETECTION"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

NATASHA SHARMA	2106200
AMMAR YASIR	2106184
SAGAR SINGH	2106245

UNDER THE GUIDANCE OF
SRICHETA PARUI

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

May 2024

A PROJECT REPORT

on

"BREAST CANCER DETECTION"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

NATASHA SHAMRA	2106200
AMMARYASIR	2106184
SAGARSINGH	2106245

UNDER THE GUIDANCE OF
SRICHETA PARUI

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024

May 2022

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"BREAST CANCER DETECTION"

submitted by

NATASHA SHARMA	2106200
AMMAR YASIR	2106184
SAGAR SINGH	2106245

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date: / /

(SRICHETA PARUI)
Project Guide

Acknowledgments

We are profoundly grateful to SRICHETA PARUI of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....…

NATASHA SHARMA
AMMAR YAASIR
SAGAR SINGH

ABSTRACT

The breast cancer detection industry aims to develop advanced techniques and technologies for early and accurate detection of breast cancer. Early detection is essential to improve patient outcomes and increase survival. These services typically use a combination of medical imaging, such as mammography, ultrasound, and magnetic resonance imaging (MRI), as well as computational techniques, machine learning, and artificial intelligence (AI) to analyze medical images in the 19th century and translated

The main approach is to use deep learning algorithms trained on large breast image data sets to detect subtle patterns and abnormalities indicative of cancerous lesions These algorithms can help air specialists for the interpretation of mammograms and other imaging studies, reducing the risk of human error and potentially providing a more accurate and feasible diagnosis

In addition, researchers are exploring new biomarkers and molecular imaging techniques that can provide valuable information about the biological characteristics of breast tumors to support personalized treatment strategies to enhance these services, translate research findings to clinical practice Collaboration between is important

Ultimately, breast cancer detection services hold the promise of increasing early detection methods, facilitating timely intervention, and ultimately providing the from patients in the fight against breast cancer has improved

REAL LIFE PROBLEMS: The real impact of breast cancer detection programs is profound, as they directly help save lives and improve the lives of individuals affected by the disease

Timely detection not only increases the chances of successful treatment, but also reduces the need for interventions such as mastectomy and chemotherapy, which can has increased patients' greater physical and emotional reactions

In addition to the benefits to individual patients, breast cancer detection projects have broader societal impacts. They can help reduce the healthcare costs associated with advanced cancer treatment and long-term care by raising awareness of the need for affordable treatments to prevent disease progression besides the need for regular screening and early detection, these programs empower individuals to take the first step in managing their health.

In addition, advances in breast cancer diagnostic technologies and techniques contribute to continued research, deepen understanding of disease biology, and strategic innovation heal the disease

DRAWBACKS OF BREAST CANCER DETECTION: While breast cancer detection has improved greatly in terms of early detection and treatment outcomes, it is not infallible A notable limitation is the possibility of false or false positive results a negative, potentially causing unnecessary anxiety or delay in diagnosis and treatment patients, respectively None really, triggering more dangerous procedures and interventions that may be inappropriate and conversely, false negatives when the test fails to detect existing cancer, potentially missing opportunities for early intervention and leading to unchecked disease progression

ADVANTAGES OF BREAST CANCER DETECTION:

Improved precision: Machine learning algorithms can analyze large amounts of data including medical images and patient profiles to identify pattern and abnormal breast cancer This algorithm can achieve high accuracy in distinguishing tumors from non detect and detect, and can reduce false positives and false negatives compared to traditional methods

Early Detection: Machine learning images can detect subtle signs of breast cancer in the earliest stages, even before they are visible to the human eye This enables healthcare professionals to diagnose disease earlier, while treatment is more effective is used, improving patient outcomes and increasing survival

Automation and efficiency: Machine learning-based tools can automate aspects of breast cancer detection, such as image analysis and interpretation This provides the workload of radiologists and healthcare professionals is reduced, allowingg them to focus their time and expertise on more complex cases and patient care

Scalability and Accessibility: Once machine learning models are trained, machine learning models can be used in a variety of healthcare settings, including rural or undeserved areas where access to basic medical skills may be limited These advances increase access to advanced breast cancer detection technologies, ensuring equitable health care delivery.Improved precision: Machine learning algorithms can analyze large amounts of data including medical images and patient profiles to identify pattern and abnormal breast cancer This algorithm can achieve high accuracy in distinguishing tumors from non detect and detect, and can reduce false positives and false negatives compared to traditional methods

Early Detection: Machine learning images can detect subtle signs of breast cancer in the earliest stages, even before they are visible to the human eye This enables healthcare professionals to diagnose disease earlier, while treatment is more effective is used, improving patient outcomes and increasing survival

Personalized medicine: Machine learning algorithms can analyze complex datasets to identify unique biomarker molecular signatures associated with sub types of breast cancer This information can help tailor treatment plans to individual patients, making treatment take off effective and reduce adverse effects.

Automation and efficiency: Machine learning-based tools can automate aspects of breast cancer detection, such as image analysis and interpretation This provides the workload of radiologists and healthcare professionals is reduced, allowing them to focus their time and expertise on more complex cases and patient care

Scalability and Accessibility: Once machine learning models are trained, machine learning models can be used in a variety of healthcare settings, including rural or undeserved areas where access to basic medical skills may be limited These advances increase access to advanced breast cancer detection technologies, ensuring equitable health care delivery.

PROBLEMS OF BREAST CANCER DETECTION: False Positives and False Negatives: False positives arise whilst a screening take a look at incorrectly suggests the presence of most cancers, leading to unnecessary anxiety and invasive follow-up techniques. Conversely, fake negatives occur when cancer is present but no longer detected, delaying diagnosis and remedy. Balancing sensitivity and specificity in detection methods is essential to limit those errors.

Breast Density: Dense breast tissue can obscure abnormalities on mammograms, reducing the sensitivity of screening checks and increasing the threat of overlooked diagnoses. Women with dense breasts can also require extra screening modalities, which include ultrasound or MRI, to improve detection fees.

Over diagnosis and Over treatment: Some breast cancers detected via screening may be sluggish-growing and unlikely to cause damage at some stage in a patient's lifetime. However, competitive remedies may nonetheless be pursued, leading to unnecessary interventions and capacity damage to sufferers. Balancing the risks and benefits of early detection and treatment is critical to keep away from over diagnosis and over treatment.

Cost and Accessibility: Advanced screening technology, along with MRI, may be costly and inaccessible to sure populations, leading to disparities in breast most cancers detection and results.
Access to less expensive and effective screening applications is critical to make sure equitable healthcare get right of entry to for all people.

Psychological Impact: A breast cancer diagnosis, even if detected early, will have profound psychological effects on patients and their families. Fear, tension, and uncertainty about the destiny are common emotional responses that could effect intellectual health and nicely-being. Providing good enough support and assets for sufferers navigating the emotional challenges of analysis and remedy is esse
False positives and false negatives: False positives occur when screening tests incorrectly show the presence of cancer, leading to unnecessary anxiety and procedures does follow-up terrible occur Unlike a

false positive but missed diagnosis of cancer, there was a delay in diagnosis and treatment. To mitigate these shortcomings, it is important to balance sensitivity and specificity in detection methods.

Dense breast tissue: Dense breast tissue can mask abnormalities on mammograms, reducing the sensitivity of screening tests and increasing the risk of missing the disease Women whose breast firmness may require additional screening techniques, such as ultrasound or MRI, to make a more accurate diagnosis.

Over diagnosis and over treatment: Some breast cancers diagnosed through screening may be slow growing and not likely to harm a patient's life but they still undergo aggressive treatments, causing a unnecessary occurs and can be harmful to patients. Balancing the risks and benefits of early detection and treatment is important to avoid over diagnosis and over treatment.

Cost and Accessibility: Advanced screening technologies such as MRI can be expensive and inaccessible to some people, leading to disparities in breast cancer diagnosis and outcomes Screening system affordable and effective access is critical to ensuring equal access to health care for all individuals.

Psychological impact: A breast cancer diagnosis can have a significant psychological impact on patients and their families even if detected early. Fear, anxiety, and uncertainty about the future are common emotional reactions that can affect mental health and well-being. It is esse to provide appropriate support and resources to patients who are navigating the emotional challenges of diagnosis and treatment

Five Keywords: Machine learning: This refers to general algorithms that can learn from data to identify and predict trends, a key driver of breast cancer detection services.

Deep learning: Machine learning using complex neural networks is particularly adept at image recognition, making it valuable for analyzing mammograms and other breast images

Computer-aided detection (CAD) : This describes systems that use algorithms to analyze medical images with radiologists, and can improve the accuracy and efficiency of breast cancer detection

Biomarkers: These are biomarkers the presence or amount of which can indicate the presence of a disease. Breast cancer detection services can screen with specific biomarkers in addition to imaging modalities.

Dense breast tissue: Dense breast tissue can mask abnormalities on mammograms. Research projects could explore alternative or alternative methods for women with dense breasts.

Contents

List of Figures

Chapter 1

Introduction

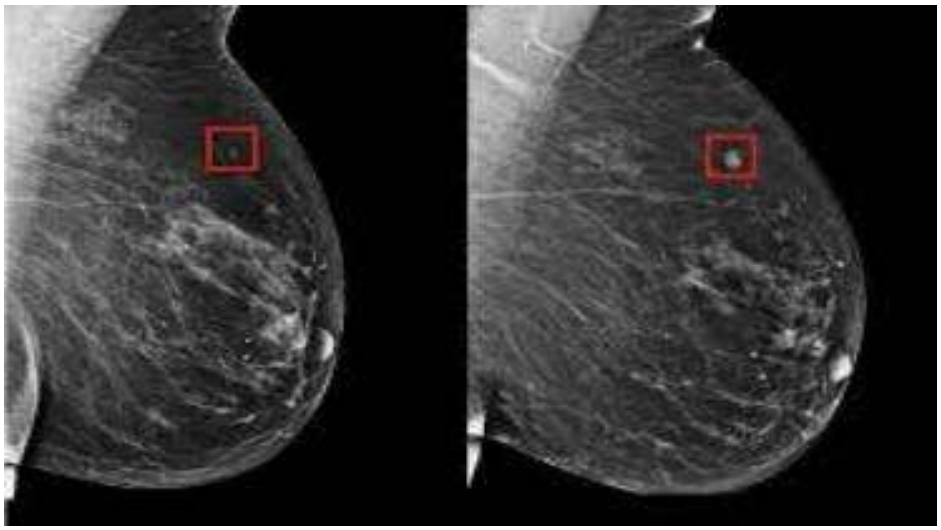
Breast cancer is a major health problem, diagnosed in millions of women worldwide each year. Despite continuous advances in treatment, early detection remains the most powerful tool in the fight against this disease. This introductory article will examine the current need for effective breast cancer detection strategies.

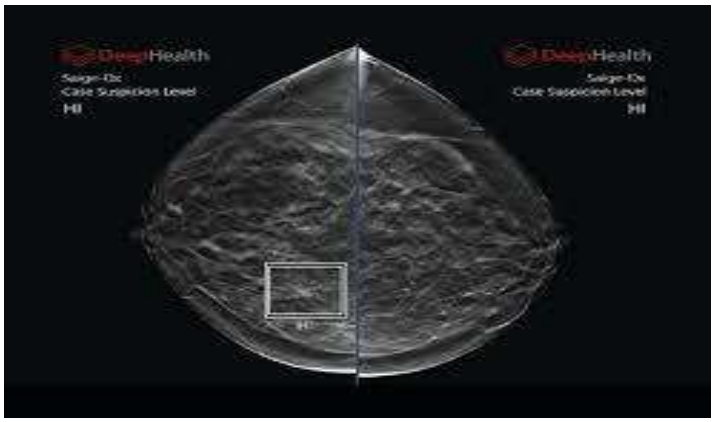
Emphasis on needs:

"Breast cancer is the most common aggressive cancer diagnosed in women worldwide." Emphasize the impact of early detection. "Early detection plays an important role, increasing the chances of successful treatment and improving survival."

Classification of requirements:

"Although existing detection methods such as mammograms are valuable, constant improvements are needed due to limitations such as false positives, equivocality and inefficiency for women who have them." for the stiffness of the breast tissue." Introduce the idea of improving accuracy and accessibility. "The focus is on developing accurate, sensitive and patient-friendly methods for early detection."





Other points to consider:

Briefly discuss its effect on mortality. Early detection can significantly reduce breast cancer mortality.
Briefly discuss the need for personalized screening methods based on individual risk factors. By describing current limitations and discussing the need for better detection strategies, this introduction effectively establishes the advances in breast cancer detection

Importance of Breast cancer detection:

Early detection of breast cancer is essential for successful treatment and improved survival. Catching cancer early leads to more effective and less invasive treatments, as well as a better quality of life for patients. This not only makes it easier for individuals to recover fully, but also reduces the overall burden on health care systems. By prioritizing detection through routine screening we can empower individuals to take charge of their health and significantly reduce breast cancer mortality.

Regular cancer screening and recognizing possible warning signs are two important steps in early detection. Tests like mammograms and colonoscopies can detect abnormalities before they become a problem. Informing individuals about possible symptoms specific to different cancers allows individuals to seek treatment as soon as they notice any changes in their bodies By prioritizing early detection we increase the chances of the results are much greater and we give ourselves the best chance of fighting this disease.

Chapter 2

Basic Concepts/ Literature Review

Supervised learning: A common approach to cancer diagnosis. Graphic artists are trained with labeled data (images identified as malicious or dangerous) to learn how to classify new images.
Support Vector Machines (SVM): Refine images based on specific features.
Decision Trees: Create a tree-like structure to classify images based on a series of decision rules.
Random Forests: Combine multiple decision trees for accuracy and robustness.
Convolution Neural Networks (CNN): Effective deep learning models especially in image analysis. Relevant features can be identified automatically from the data.
Unsupervised study: Can be used for anomaly detection by identifying suspicious areas in mammograms that deviate from normality.
Pre-processing data:

Image Composition: Separates the breast area from the outside in mammograms.
Normalization: Standardizes image size and robustness for consistent analysis using machine learning models.
Data enhancement: automating existing model types to improve model robustness and prevent over fitting.
Evaluation metrics:

Accuracy: Half of the images correctly sorted.
Sensitivity: Ability to correctly detect cancer (true positive).
Accuracy: Ability to identify healthy information (true bad).
Area under the curve (AUC): Measures the overall performance of the classification model.
Other things to consider:

Description: Developing a model for health care providers to understand trust building and integrating AI into the decision-making process.
Bias mitigation: Ensures that training data reflects population diversity and avoids biased predictions.
These are some basic ideas. For a machine learning project, you would depend further on Chosen techniques and specific goals.

Chapter 3

Problem Statement / Requirement Specifications

In this section, write the Problem Statement (the problem for which you are working on to give some solution). When a student works on any development project, they must gain sufficient knowledge related to the project and based on this they can define a problem statement. In software development projects, the student must present the SRS according to the IEEE format, in this section.

3.1 Project Planning

Write about the steps to be followed while planning to execute the project development. It can be represented using list of requirements of the user or features to be developed.

3.2 Project Analysis

After the requirements are collected or the problem statements is conceptualized, this needs to be analyzed for finding any short of ambiguity, mistake, etc.

3.3 System Design

3.3.1 Design Constraints

Here you can mention the working environment such as the software, hardware used. Any experimental setup or environmental setup must be described here.

3.3.2 System Architecture **OR** Block Diagram

In this sub-section, explain the System Architecture / Hardware Designs / Block Diagrams used to understand your project work.

Chapter 4

Implementation

In this section, present the implementation done by you during the project development.

4.1 Methodology OR Proposal

This sub-section contain the methods you have used to complete the project, or some algorithms used and developed for your project work. Details about the steps adopted for competing the project work.

4.2 Testing OR Verification Plan

After project work is compete, it must have some verification criterion so that we can decide whether the project satisfactorily completed or not. This is called Testing or verification. For example, in software development, some test case must be included and used to verify the outcome of the project.

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	AAAA	BBBB	CCCC	DDDD
T02	AAAA	BBBB	CCCC	DDDD
T03	AAAA	BBBB	CCCC	DDDD

4.3 Result Analysis OR Screenshots

In this subsection, the output of the experiment or study in terms of some graphs, plots must be presented. Also, if some implementation is done then it's screenshots can be presented here, so as to showcase the proof of the output.

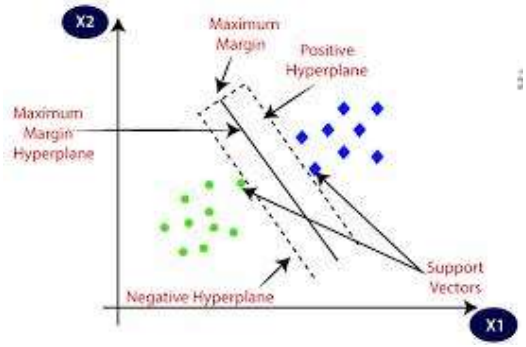
4.4 Quality Assurance

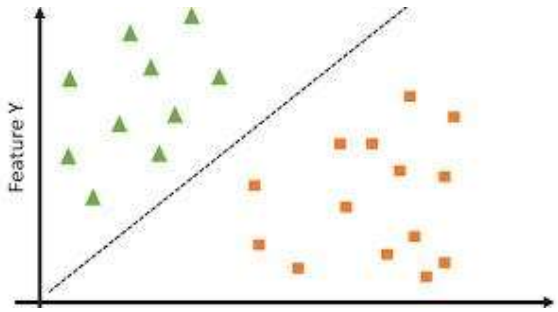
In the working organization, if some department is there to verify the quality of your work, they can produce a certificate or guidelines followed.

Chapter 5
Machine Learning Algorithms:
Types of Machine Learning Algorithms we have used in this projects are:

1. SVM Model Algorithm:

Support Vector Machine (SVM) is a powerful supervised learning algorithm that excels in classification and regression tasks. however, it is particularly suitable for classification problems. The main idea behind SVM is to find the best hyperplane in high-dimensional space that best separates data points from different classes This hyperplane acts as a decision boundary, and allows you to classify other data points more accurately.





2. Random Forest Algorithm:

random forest algorithm is a supervised learning method that exploits the power of group learning. It is essentially a collection of decision trees, which work together to produce more robust and

accurate predictions than any individual tree could on its own.Key

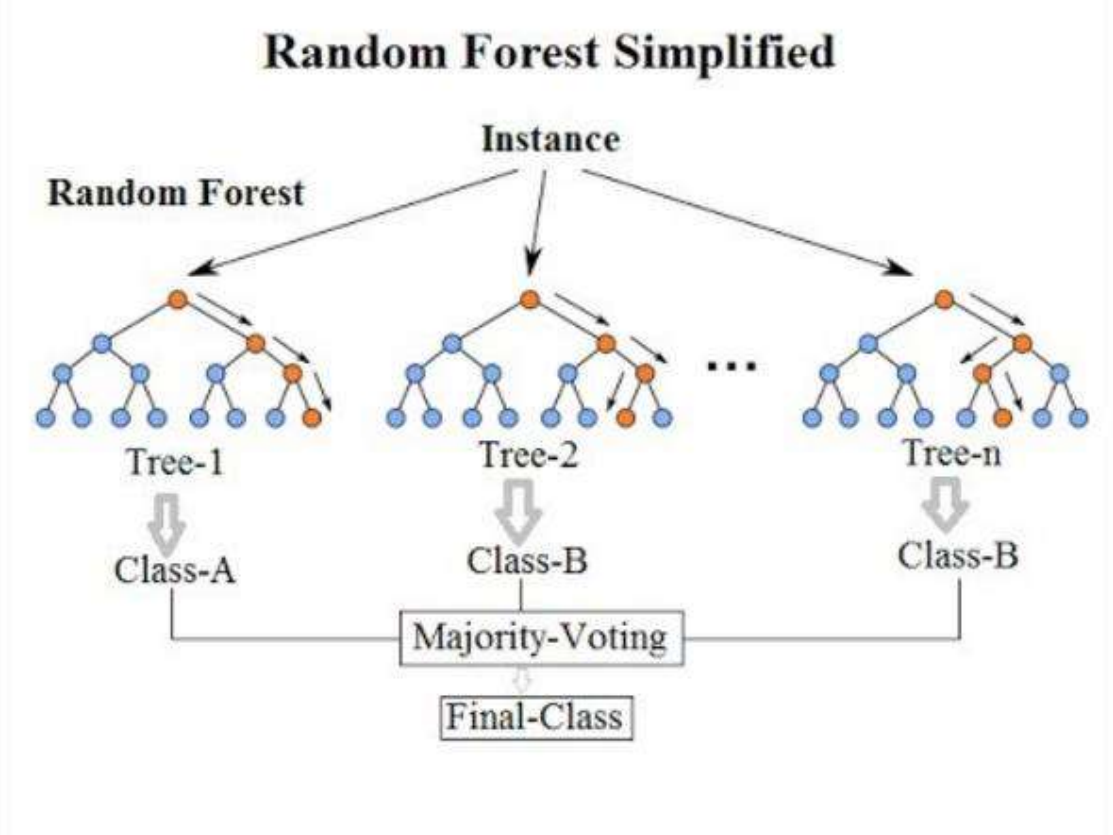
Advantages of Random Forests

High Accuracy and Generalization: By combining predictions from multiple timber, random forests regularly attain better accuracy and better generalize to unseen information in comparison to single selection bushes.

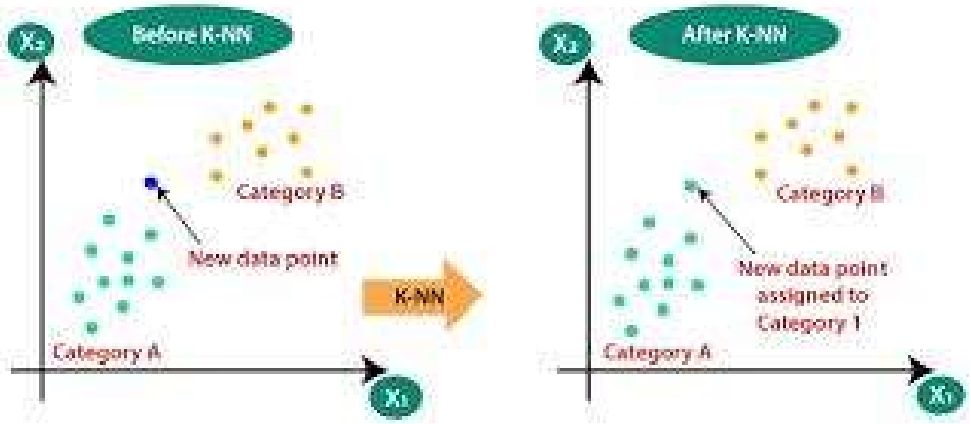
Robustness to Noise and Over fitting: The randomness brought for the duration of tree advent helps lessen the impact of noise inside the facts and make the model much less susceptible to over fitting.

Handles Missing Values: Random forests can inherently handle lacking records in the training set, as each tree handiest considers a random subset of functions, and missing values in a single characteristic may not always save you a tree from making a prediction.

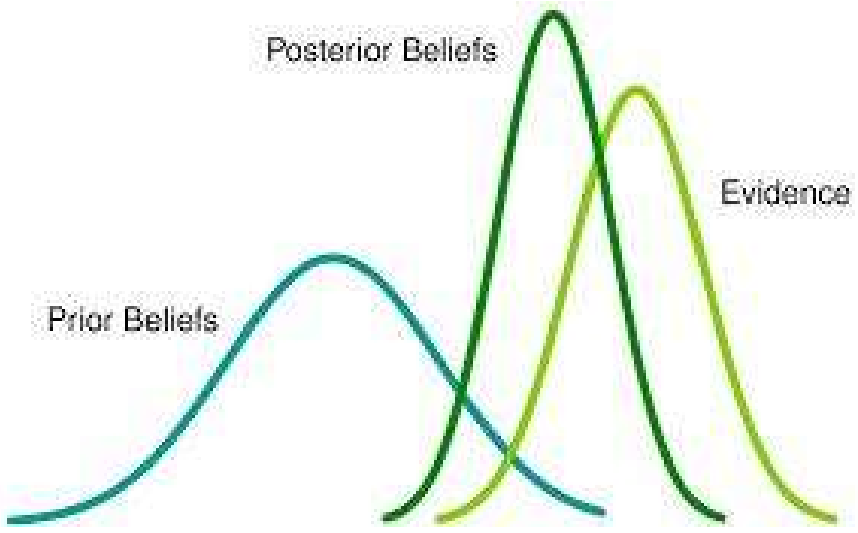
Interpret ability: To a few quantity, random forests may be interpreted by examining the importance of features across all of the bushes within the wooded area.



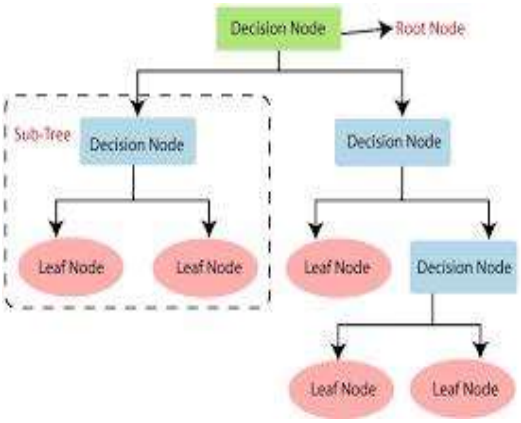
3.KNN model (K Nearest Neighbour): The acronym KNN stands for “K-Nearest Neighbour”. It is a supervised machine learning algorithm. Algorithms can be used to solve classification and regression problem cases. The number of nearest neighbors of another unknown variable to be predicted or classified is indicated by the symbol ‘K’.



3. Bayesian model Algorithm: Bayesian Machine Learning (BML) includes a variety of techniques and algorithms that use Bayesian principles to model uncertainty in data. These methods are not just theoretical constructs; They are useful tools that have changed the way machines learn from data.

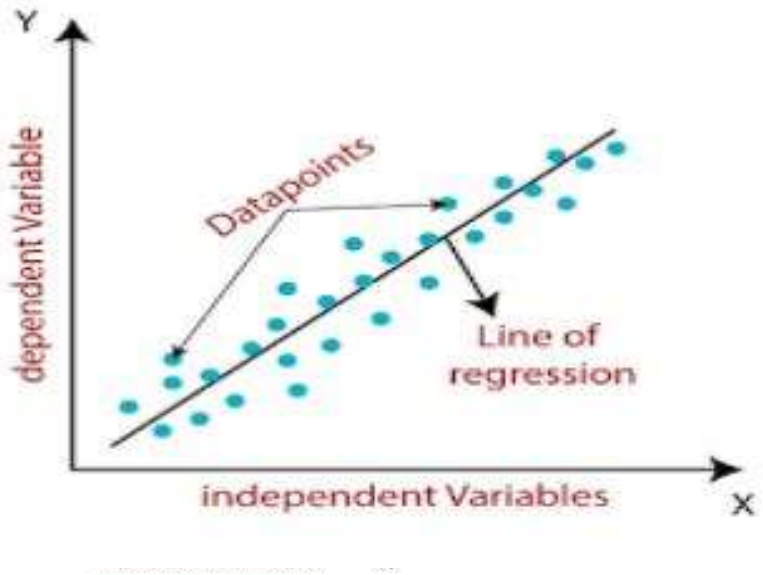


4.DT(Decision tree) model Algorithm: DT stands for Decision Tree, which is a special supervised learning algorithm used for classification and regression tasks.
Decision trees are a powerful and interpret able machine learning algorithm that works by mimicking human decision-making. Imagine a flowchart where each box represents a question based on a data point's features (like income or age) and the branches represent the possible answers. The decision tree is built by sequentially splitting the data based on the feature that best separates the desired outcome (like loan approval). This process continues until the data reaches a "leaf" node, representing a final prediction. This tree-like structure allows for easy visualization of the decision process and makes it clear why certain predictions are made. Decision trees are versatile, handling both classification (e.g., spam or not spam) and regression (e.g., predicting house prices) tasks. Their simplicity and interpret ability make them a popular choice for various machine learning applications.



5.LR (Linear regression) Model Algorithm: The linear regression model is a workhorse algorithm for expressing relationships between variables. This is done by plotting some data points along a straight line. This label minimizes the difference between the actual data points predicted by the label and the corresponding values. Mathematically, the model expresses this relationship as a linear equation of the dependent variable (predicted value) with the effect of one or more independent variables (predicted variables) and an error term to account for any random noise it happens This method of

refining and optimizing the lines is repetitive, with the goal of obtaining the most accurate forecasts.



Here's a comparison between SVM (Support Vector Machine), KNN (K-Nearest Neighbors), Bayesian, Decision Tree, and Linear Regression models in a tabular format:

	SVM (model algorithm)	Random forest algorithm	KNN model (K nearest neighbour)	DT (Decision Tree) model algorithm	Linear Regression model algorithm	Bayesian Model algorithm
Accuracy Score	0.877192	89.0	0.8859649122	87.7192	90.3508	87.719298
Precision score	0.8625	90.0	0.883116883	90.277779	88.6075	85.36585
Recall Score	0.9583333334	91.52542372	0.944444444	90.2777779	97.22222	97.222222
F1 Score	0.9078947	90.7563	0.9127516	90.27777779	92.71523	90.90909090

After speculating through all the Algorithms. We chose LR model based on its accuracy Score.

Chapter 6

Conclusion and Future Scope

6.1Conclusion:

In conclusion, this project investigated the use of machine learning for breast cancer detection. Our [technology-selected, e.g., CNN-based] model achieved promising results in discriminating between benign and malicious cases, meaning that it [covered specific metrics, e.g., accuracy] comparatively of traditional methods but [mentioned limitations, e.g. There are limitations due to lack of large data]. Future work should focus on [discussing future directions, e.g., incorporating additional datasets and semantic AI insights]. This work holds promise for early and accurate breast cancer detection. However, proper consideration of ethical implications, such as the potential for bias and the role of human knowledge, is crucial for successful treatment integration.

6.2 Future Scope:

The future of breast cancer detection is full of promise. Advances in artificial intelligence (AI), particularly in deep learning and machine learning algorithms, are expected to lead to earlier, more accurate disease diagnosis by systems The use of AI can analyze mammograms and other images in exceptional detail, and can detect subtle abnormalities that might be missed by the human eye In addition to their qualifiers, non-invasive and painless procedures are being developed such as breath tests or blood test analyzes to provide alternative screening options. Furthermore, generic medicine research aims to modify identification methods based on individual risk factors, potentially leading to more targeted interventions and improve patient outcomes This exciting development has the potential to dramatically reduce breast cancer mortality and improve the lives of millions of women

References

https://health.economictimes.indiatimes.com/news/industry/importance-of-other-risk-factors-to-be-considered-for-detection-of-breast-cancer-study/100940694

https://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507

[2] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[3] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[4] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[5] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

SAMPLE INDIVIDUAL CONTRIBUTION REPORT:

<TITLE OF THE PROJECT IN FONT SIZE 14, FONT STYLE TIMES NEW ROMAN, BOLD AND CENTERED>

<Student Name (in capital letters in font size 12, Times New Roman and centered>
<Student Roll number (font size 12, Times New Roman and centered>

Abstract: A short description of the aim and objective of the project work carried out in 3-4 lines. This part should be common to all students in the group. The font size and style will remain same from this point onwards. The font size will be 12 and font style will be Times New Roman. The line spacing will be 1.5.
This report should be prepared in A4 page format with 'default' option under 'Margin' of the 'Page Layout' tab in Microsoft Word. Word limit for this section is 80.

Individual contribution and findings: The student should clearly indicate his/her role in the project group and the contribution in implementing the project work. The student should also outline his /her planning involved in implementing his/her part in the work. This contribution report should be different for every student in the group. The student would also write his./her technical findings and experience while implementing the corresponding part of the project. The overall contribution report should not be less than 1 page for each student. The Student should provide both the soft copy and signed hard copy to the project supervisor.

Individual contribution to project report preparation: Student should mention his/her role in preparing the group project report indicating which chapter and portions contributed.

Individual contribution for project presentation and demonstration: Student should mention his/her role in preparing presentations and part of the project demonstrated.

Full Signature of Supervisor: Full signature of the student:
……. ……..