

# Cancer vision: Advanced Breast Cancer Prediction with Deep Learning

---



A PROJECT REPORT SUBMITTED BY:

**TEAM ID: LTVIP2023TMID04383**

**TEAM SIZE: 4**

**TEAM LEADER: MANOHAR PREMKUMAR**

**TEAM MEMBER: MANCHURI LAKSHMIVARAPRASAD**

**TEAM MEMBER: PALAKONDU PAVITHRAREDDY**

**TEAM MEMBER: POTHULARAJU DIVYASREE**

## Table contents:

### **1. INTRODUCTION**

1.1 Project overview

1.2 Purpose

### **2. IDEATION & PROPOSED SOLUTION**

2.1 Problem Statement and Definition

2.2 Empathy Map Canvas

2.3 Proposed Solution

### **3. REQUIREMENT ANALYSIS**

3.1 Functional Requirement

3.2 Non Functional Requirements

### **4. PROJECT DESIGN**

4.1 Data Flow Diagrams

4.2 Solution & Technical Structure

4.3 User Stories

### **5. CODING AND SOLUTIONING**

5.1 Feature 1

5.2 Feature 2

### **6. RESULTS**

### **7. ADVANTAGES AND DISADVANTAGES**

### **8. CONCLUSION**

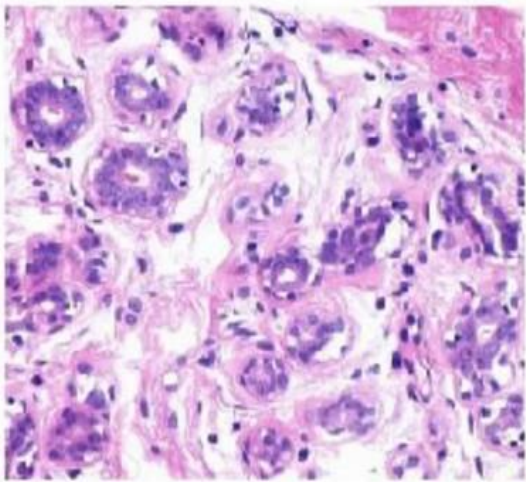
### **9. FUTURE SCOPE**

### **10. APPENDIX**

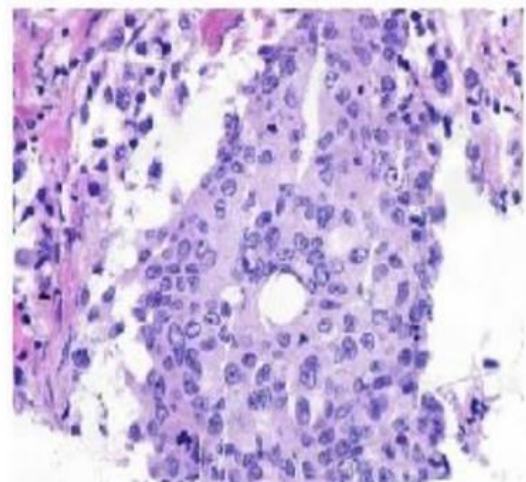
## INTRODUCTION

Breast cancer is one of the main causes of cancer death worldwide. Computer-aided diagnosis systems showed potential for improving the diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible. The goal is to classify images into two classifications of malignant and benign. As early diagnostics significantly increases the chances of correct treatment and survival. In this application we are helping the doctors and patients to classify the Type of Tumour for the specific image given with the help of Neural Networks.

### BENIGN



### MALIOGNANT



## 1.1 Project Overview

- **Problem:** Breast cancer is one of the leading causes death in women. Early detection and treatment are essential for improving survival rates
- **Solution:** This Project will develop a CNN model to help detect breast cancer earlier. The System will use deep learning to classify mammogram images as malignant (harmful) or benign (harmless).
- **Benefits:** It will have the potential to improve the accuracy of breast cancer diagnosis. This could lead to earlier treatment and improved survival rates for women with breast cancer

The project will be conducted in three phases:

1. **Data Collection:** A dataset of mammogram images will be collected from a variety of sources. The images will be labeled as malignant or benign by a team of expert radiologists.
2. **Model development:** A deep learning model will be developed to classify mammogram images. The model will be trained on the labeled dataset of mammogram images.
3. **Evaluation:** The performance of the model will be evaluated on a held-out test set of mammogram images. The model will be evaluated on its accuracy, sensitivity, and specificity.

## 1.2 Purpose

- **Improve diagnostic accuracy:** Enhance the accuracy of breast cancer diagnosis by leveraging computer-aided diagnosis systems and neural networks. The application aims to assist doctors in making more informed decisions based on reliable and automated tumor classification.
- **Early Detection:** The project aims to detect breast cancer at its early stages, before it progresses to an advanced form. Early detection is crucial for improving treatment outcomes and increasing the chances of successful recovery.
- **Improved Risk Assessment:** By analyzing diverse and comprehensive datasets that include genetic data, imaging results, medical records, lifestyle factors, and biomarkers, the project seeks to develop a predictive model that can provide more accurate risk assessments for individuals.
- **Supporting Healthcare Decision-Making:** The predictive model developed in the project can be integrated into clinical settings, providing healthcare professionals with valuable insights to make informed decisions about patient management and treatment plans.

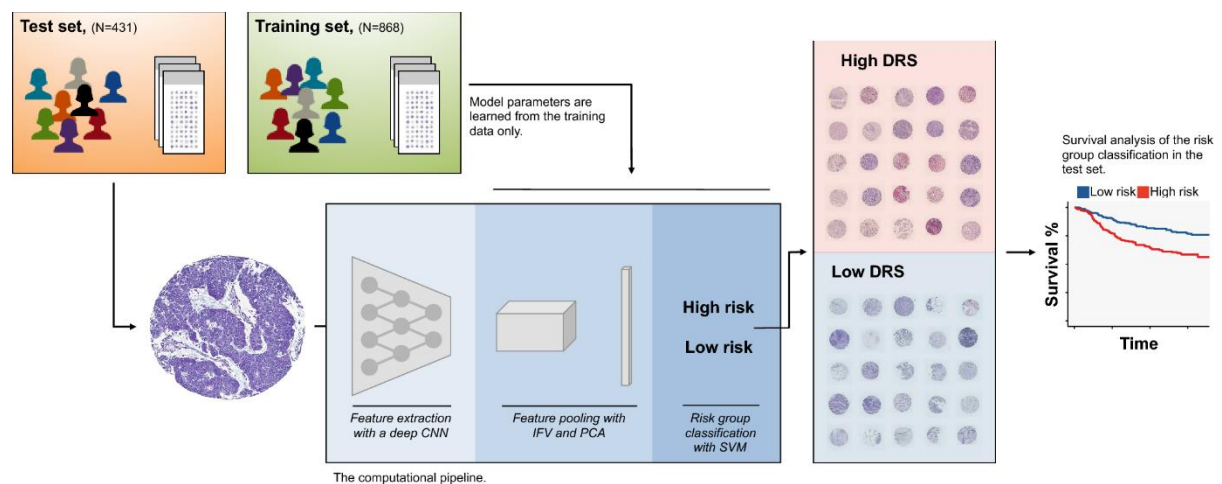
## 2. IDEATION & PROPOSED SOLUTION

### 2.1 Problem Statement and Definition

Breast cancer is one of the most prevalent and deadly cancers among women worldwide. Breast cancer is the most common cancer among Indian women, accounting for 14% to 27.7% of all new cancers detected in women in India. Early detection and accurate prediction of advanced breast cancer can significantly improve patient outcomes and increase survival rates. Traditional diagnostic methods, such as mammography and biopsy, have been effective to some extent, but they may not always provide timely and accurate results.

The objective of this project is to develop a deep learning-based predictive model that can accurately detect the presence of advanced breast cancer in patients. The model will take as input various medical data related to patients, such as clinical history, imaging results, and molecular profiling, and provide a binary classification output indicating whether the patient is likely to have advanced breast cancer or not.

### Proposed solution



The system will be developed using a dataset of histopathology images of breast tissue that have been labeled as malignant or benign by pathologists. The images will be preprocessed, handling missing values, standardizing features, and performing data augmentation (for images) to increase the dataset's size and variability. After preprocessing they will be fed into a neural network. The neural network will be trained to learn the features that distinguish

malignant tumors from benign tumors. Once the neural network is trained it will be able to classify new images of breast tissue with high accuracy.

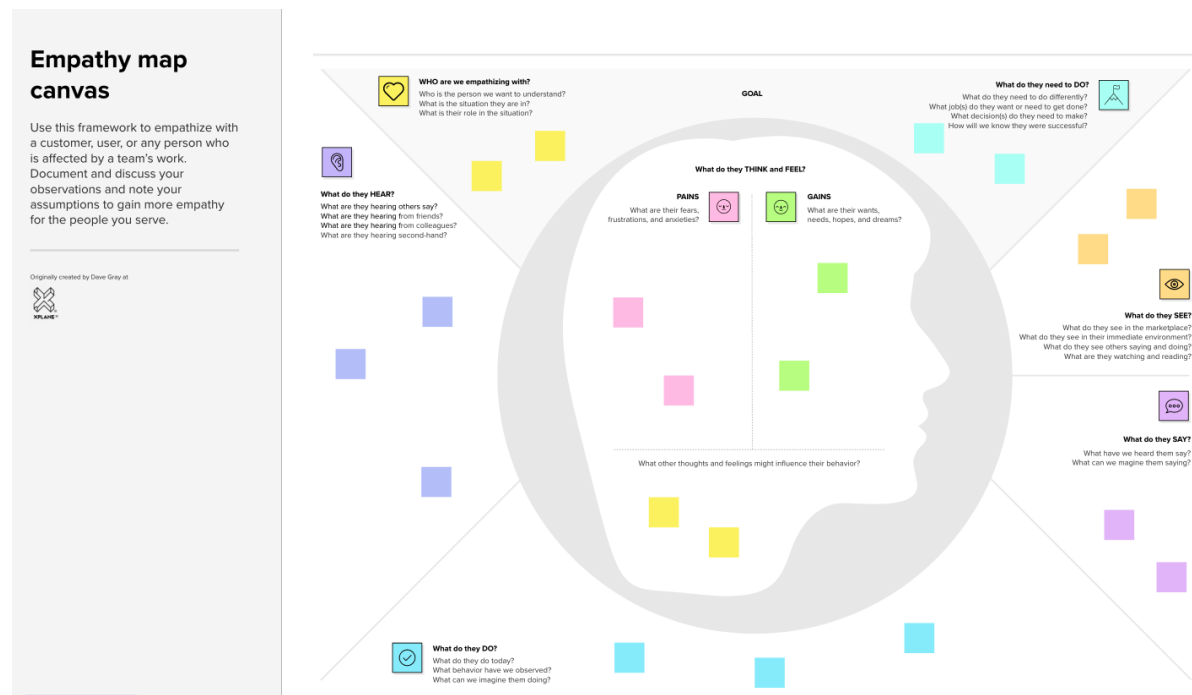
The system will be evaluated with a test set of images that have not been used to train the neural network. The accuracy of the system will be measured by the percentage of images that are correctly classified. The system will also be evaluated by its ability to detect early-stage breast cancer. We use Convolutional Neural Networks (CNN) for effective breast cancer classification.

## 2.2 Empathy Map Canvas

An empathy map is a collaborative visualization used to articulate what we know about a particular type of user. It externalizes knowledge about users in order to

- 1) create a shared understanding of user needs
- 2) aid in decision making

Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating map helps participants consider things from the user's perspective along with his or her goals or challenges.



## **2.3 Proposed Solution**

### **I. Problem Statement:**

The project's primary objective is to develop a highly accurate and efficient breast cancer prediction model using deep learning techniques. The model should be capable of analyzing mammogram images and providing early detection and prognosis of breast cancer, allowing for timely interventions and improved patient outcomes.

### **II. Data Collection and Preprocessing:**

- Gathering a diverse and extensive dataset of mammogram images from various sources, including medical databases, research institutions, and hospitals.
- Preprocessing the data to handle issues like noise, artifacts, and standardizing the image sizes for consistent model training.

### **III. Model Architecture:**

- Implementing a deep Convolutional Neural Network (CNN) model optimized for medical image analysis.
- Exploring state-of-the-art CNN architectures, such as ResNet, DenseNet, or EfficientNet, to leverage their superior feature extraction capabilities.
- Investigating the use of transfer learning from pre-trained models on large-scale image datasets like ImageNet to boost model performance.

### **IV. Data Augmentation and Handling Class Imbalance:**

- Employing data augmentation techniques (rotation, flipping, zooming) to increase the size of the dataset and prevent overfitting.
- Addressing class imbalance issues, as breast cancer cases are often a minority in the dataset, by using techniques like oversampling and generating synthetic samples.

### **V. Explainable AI and Interpretability:**

- Integrating methods for model interpretability to understand the decision-making process, such as Grad-CAM or LIME (Local Interpretable Model-agnostic Explanations).

- Enhancing the transparency of predictions to gain trust from medical practitioners and ensure ethical considerations.

## **VI. Hyperparameter Optimization:**

- Conducting hyperparameter tuning using techniques like grid search, random search, or Bayesian optimization to find the best model configuration.
- Optimizing the learning rate, batch size, and number of layers to achieve the highest prediction accuracy.

## **VII. Model Evaluation and Validation:**

- Employing k-fold cross-validation to assess the model's performance robustness and generalization capabilities.
- Evaluating the model using metrics like accuracy, sensitivity, specificity, ROC-AUC, and F1-score.

## **VIII. Deployment and User Interface:**

- Developing a user-friendly web-based interface or a mobile application for easy input of mammogram images and real-time predictions.
- Ensuring the interface is secure and compliant with privacy regulations to protect patient data.

## **IX. Collaborative Partnerships:**

- Establishing collaborations with medical institutions and professionals to access expertise, validate the model's results, and ensure its clinical relevance.
- Working with healthcare practitioners to integrate the model into existing breast cancer screening programs.

## **X. Future Directions:**

- Exploring the possibility of integrating other relevant data modalities, such as patient history and genetic information, to enhance the predictive power of the model.
- Investigating the potential for the model's application in risk stratification and personalized treatment recommendations for breast cancer patients.



### 3. REQUIREMENT ANALYSIS

#### 3.1 Functional Requirement

FR No:	Functional Requirement (Epic)	Sub Requirement(Story/Sub-Task)
FR-1	User Guide	➤ Guidelines- registration, confirmation, accessing services etc, in website
FR-2	User Registration	➤ Registration through Form ➤ Registration through Gmail
FR-3	User Authentication	➤ The web application should support user authentication, requiring customers to log in using their registered email address and password. ➤ Only authorized users can access the application and their personalized features
FR-4	User Communication	➤ The web application should include features for contracting customers. ➤ This may involve sending notifications of alerts to customers regarding their uploaded images, classification results, or other relevant information
FR-5	User Acknowledgement	➤ After customers upload an image for classification, the system should provide an acknowledgement to confirm the successful submission of the image. ➤ This acknowledgement can be displayed on the web interface or communicated through email or SMS.
FR-6	User confirmation	➤ Confirmation via Email ➤ Confirmation via OTP ➤ Confirmation via SMS
FR-7	User support	➤ The web application should provide a support mechanisms for customers to contact technical support or ask questions related to the image classification process.

### 3.2 Non-Functional Requirements:

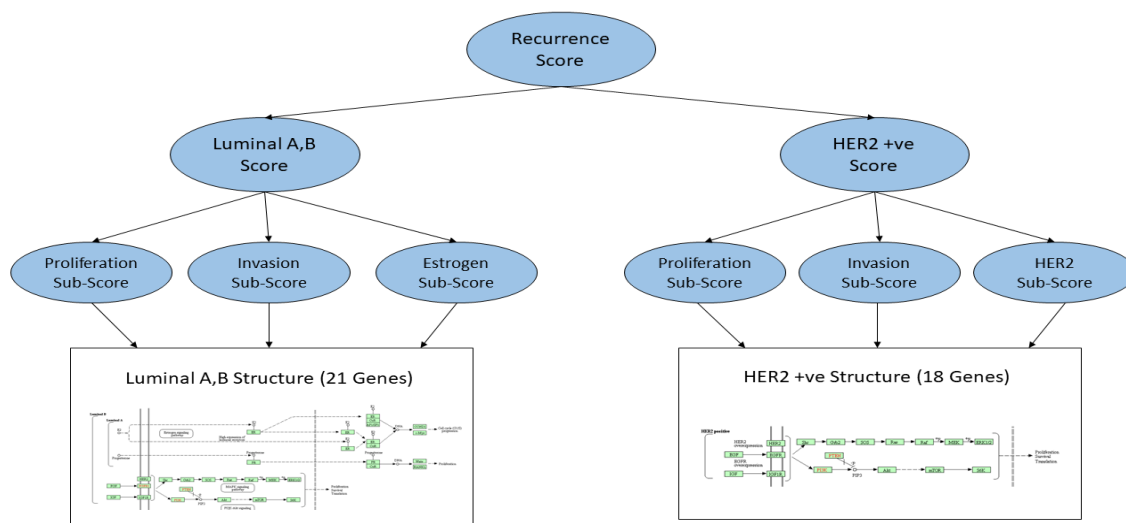
FR No:	Non-Functional Requirement	Description
NFR-1	Usability	<ul style="list-style-type: none"><li>✓ CancerVision aims to provide a user-friendly interface that is intuitive and easy to navigate</li><li>✓ The system will incorporate clear instructions and guidance, ensuring that users can effectively utilize its features and provide additional customization for individual preference and workflows</li></ul>
NFR-2	Security	<ul style="list-style-type: none"><li>✓ Cancer Vision places a high priority on datasecurity and protect patient sensitive data from anauthorized access, breaches,or tampering.</li><li>✓ This includes implementing stronger user authentication mechanisims and access controls to ensure that only authorized individuals can access the system.</li></ul>
NFR-3	Reliability	<ul style="list-style-type: none"><li>✓ Cancer Vision strives to be a highly reliable systems, minimizing errors, failures, and crashes that could adversely affect the accuracy and availability of breast cancer predictions.</li></ul>
NFR-4	Performance	<ul style="list-style-type: none"><li>✓ Cancer Vision is designed to deliver high performance, enabling quick breast cancer predictions.</li><li>✓ This includes effectively utilizing memory and processing power to ensure efficient timely prediction results.</li></ul>
NFR-5	Availability	<ul style="list-style-type: none"><li>✓ Cancer Vision prioritizes high availability aiming to minimize downtime and ensure continuous accessibility for healthcare professionals.</li><li>✓ This system will incorporate redundancy and failover mechanisms to handle hardware failures and will have robust backup and recovery processes.</li></ul>

## 4. PROJECT DESIGN

### 4.1 Data Flow Diagrams

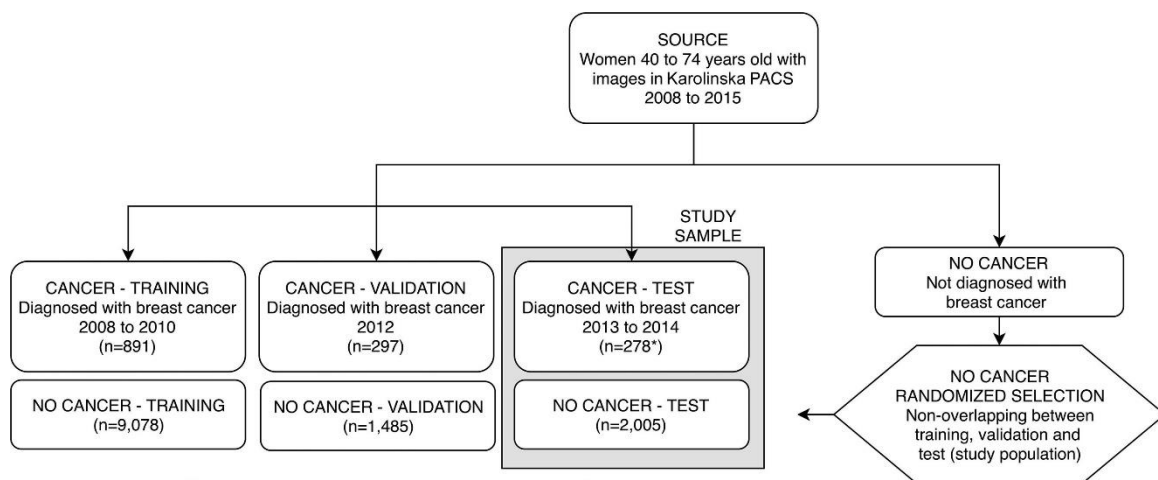
A data flow diagram is a visual representation of the information flows with in a system. It shows how data enters and leaves the system, what changes the information, and where data is stored.

### USER INTERFACE DIAGRAM



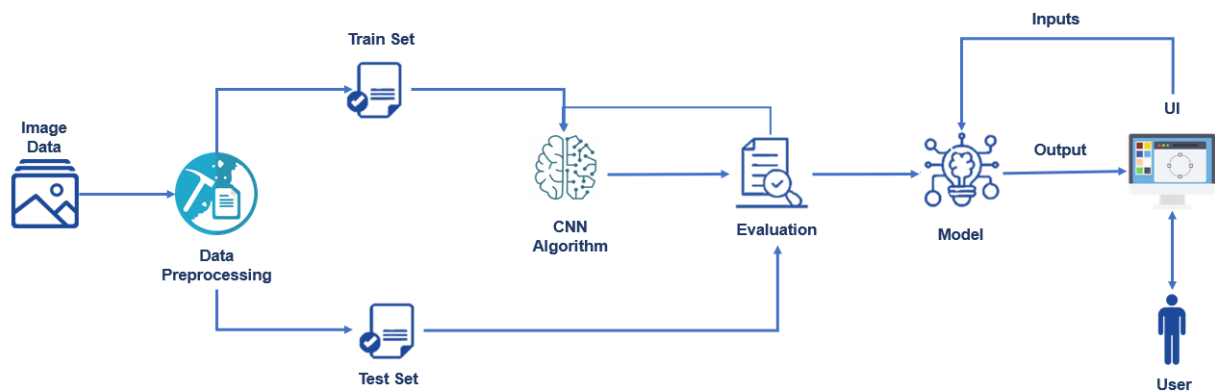
### 4.2 Solution and Technical Structure

### SOLUTION



\*Excludes women with (i) images from other than Hologic equipment (n=92), (ii) images later than 12 months before diagnosis (n=48), (iii) less than 4 images per examination (n=2), (iv) technical issues in processing (n=5)

## TECHNICAL ARCHITECTURE



### 4.3 User Stories

User stories help to capture specific user needs and requirements in the form of simple, actionable statements. For the project "Advanced Breast Cancer Prediction with Deep Learning," here are some user stories representing the perspectives of various stakeholders involved:

1. As a healthcare professional, I want a breast cancer prediction system that can analyze mammogram images accurately and efficiently, providing early detection and prognosis to improve patient outcomes.
2. As a radiologist, I need a deep learning model that can handle diverse mammogram images, including different imaging modalities, to ensure reliable predictions.
3. As a medical researcher, I want access to a comprehensive dataset of mammogram images for training and validation purposes to enhance the model's accuracy.
4. As a patient, I expect the breast cancer prediction system to provide transparent and understandable predictions, along with clear explanations of the features influencing the diagnosis.

By incorporating these user stories into the project development process, the team can ensure that the advanced breast cancer prediction system meets the needs and expectations of various stakeholders, making it more effective and impactful in the fight against breast cancer.

## 5. CODING AND SOLUTIONING

### 5.1 Feature1(Optimized and improved the Tensorflow API for the latest version)

TensorFlow is a powerful library for machine learning and deep learning, and optimizing its usage can significantly enhance the performance and efficiency of your project. Here are some steps you can consider:

1. **Update to the Latest Version:** Make sure you're using the latest version of TensorFlow, as it might include performance improvements and bug fixes.
2. **Model Architecture:** Choose an appropriate model architecture for your breast cancer prediction task. Depending on the nature of your data and the complexity of the problem, you might consider using convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer-based architectures.
3. **Data Preprocessing:** Proper data preprocessing can have a big impact on model performance. Ensure that you're normalizing, scaling, and augmenting your data appropriately.
4. **Efficient Data Loading:** Use TensorFlow's data loading utilities such as `tf.data.Dataset` to efficiently load and preprocess your data. This can help with optimizing data throughput during training.
5. **Mixed Precision Training:** TensorFlow supports mixed precision training, which uses a combination of float16 and float32 data types to accelerate training without sacrificing accuracy. This can be particularly useful for deep networks.
6. **Distributed Training:** If your project requires training on large datasets, consider distributing the training process across multiple GPUs or machines using TensorFlow's `tf.distribute.Strategy`. This can significantly reduce training time.
7. **TensorFlow Profiler:** Utilize TensorFlow's built-in profiler to identify performance bottlenecks in your code. This can help you pinpoint areas that need optimization.
8. **Model Quantization:** After training, you can quantize your model to reduce its memory footprint and potentially improve inference speed, especially if you're deploying the model to resource-constrained environments.
9. **TensorFlow Lite:** If your goal is to deploy your model on mobile devices, consider converting your trained model to TensorFlow Lite format. This format is optimized for mobile and embedded devices.

10. **TensorFlow Serving:** If you're deploying your model in a production environment, consider using TensorFlow Serving. It's a framework for serving machine learning models with a focus on low-latency inference.
11. **Regular Updates:** Keep an eye on updates from the TensorFlow team. They regularly release new features, optimizations, and best practices.
12. **Experimentation:** Don't be afraid to experiment with different techniques and settings. Use tools like TensorBoard to visualize your experiments and track your progress.

## 5.2 Feature 2(Inclusion of Specificity At Sensitivity Metric)

In medical diagnostics, it's important to consider both sensitivity and specificity metrics to evaluate the performance of a classification model, especially in projects like advanced breast cancer prediction. Sensitivity (also known as true positive rate or recall) measures the model's ability to correctly identify positive cases, while specificity (true negative rate) measures the model's ability to correctly identify negative cases.

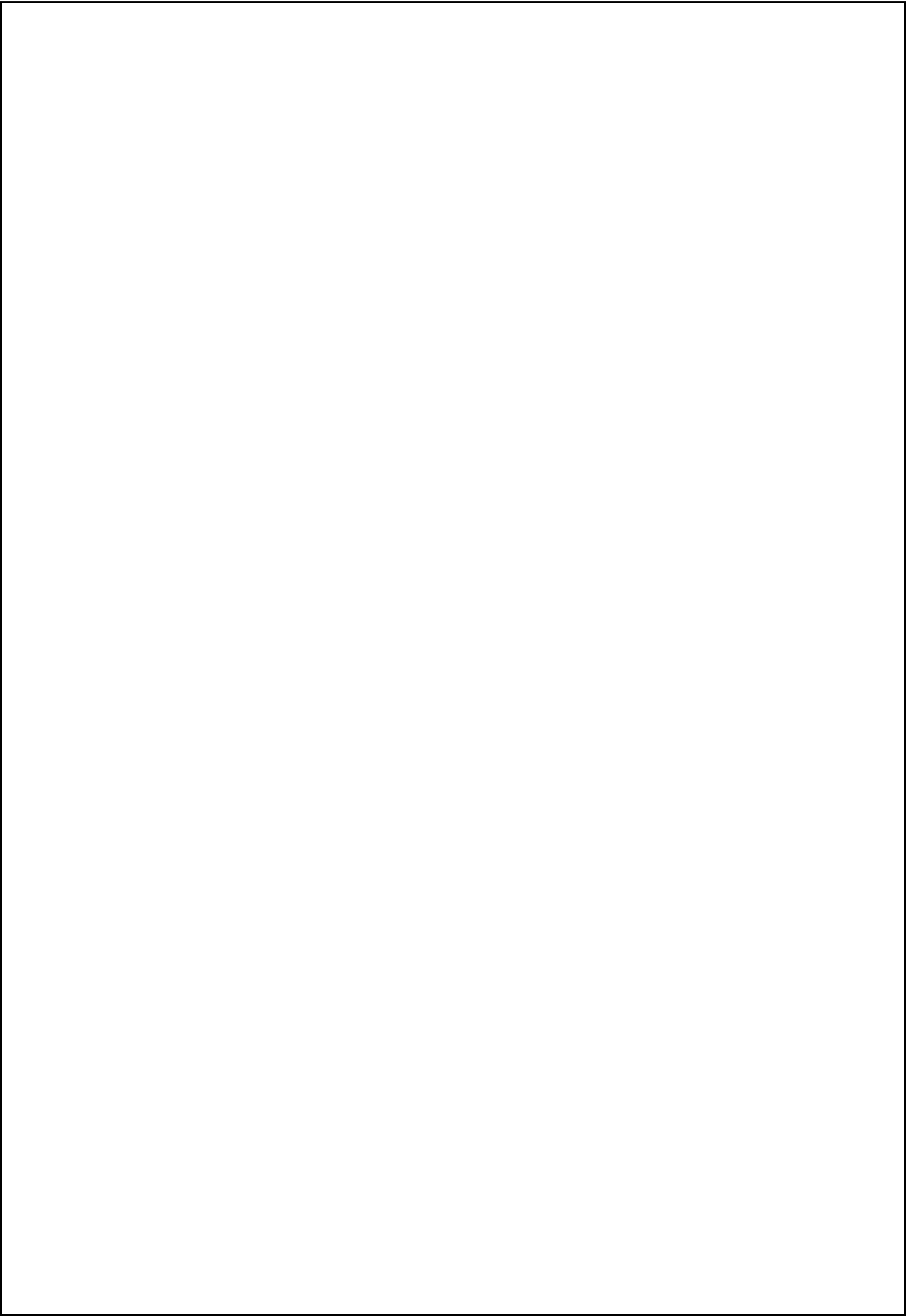
To include both sensitivity and specificity metrics in your advanced breast cancer prediction project with deep learning, follow these steps:

1. **Confusion Matrix:** Calculate a confusion matrix after making predictions on your validation or test dataset. The confusion matrix will have four components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).
2. Sensitivity and Specificity Calculation:
3.  $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$
4.  $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$
5. **Evaluate with Specific Thresholds:** Deep learning models often produce probability scores as output. By choosing different probability thresholds, you can control the trade-off between sensitivity and specificity. ROC curves and Precision-Recall curves can help you visualize this trade-off and choose an appropriate threshold based on your project's requirements.
6. **Receiver Operating Characteristic (ROC) Curve:** The ROC curve is a graphical representation of the trade-off between sensitivity and specificity. It helps you evaluate the model's performance across various threshold values. The area under the

ROC curve (AUC-ROC) is a common metric that summarizes the overall model performance.

7. **Precision-Recall (PR) Curve:** The PR curve is another graphical representation that shows the trade-off between precision and recall (which is equivalent to sensitivity). The area under the PR curve (AUC-PR) is another metric to assess the model's performance.
8. **F1 Score and Balanced F1 Score:** The F1 score is the harmonic mean of precision and recall. However, in cases where sensitivity and specificity are both crucial, you might consider using the balanced F1 score, which takes both false positives and false negatives into account.
9. **Custom Loss Functions:** Depending on the deep learning framework you're using, you can design custom loss functions that incorporate both sensitivity and specificity. This could encourage the model to find a balance between the two metrics during training.
10. **Imbalanced Data Handling:** In medical datasets, class imbalance is common, where one class (e.g., cancer-positive cases) might be significantly fewer than the other. Techniques like oversampling, undersampling, or using weighted loss functions can help address this imbalance and improve the sensitivity and specificity trade-off.
11. **Cross-Validation:** To ensure the robustness of your model evaluation, consider using techniques like k-fold cross-validation. This involves splitting your dataset into k subsets and training/evaluating the model k times, rotating the subsets used for training and testing.
12. Remember that the choice of sensitivity and specificity trade-off depends on the specific goals of your project. In medical applications, sensitivity might be more critical to ensure that actual cases of breast cancer are not missed (minimizing false negatives). However, striking a balance between sensitivity and specificity is crucial for avoiding both false positives and false negatives.

## 6. RESULTS





## 7. ADVANTAGES AND DISADVANTAGES

### Advantages:

**Early Detection:** By integrating advanced imaging techniques with the deep learning model, the project enhances early detection capabilities. Early detection is crucial for timely intervention, leading to better treatment outcomes and potentially saving lives.

**Accurate Risk Prediction:** The project utilizes advanced deep learning techniques to analyze large and diverse datasets, allowing for more accurate risk predictions. The model's ability to identify complex patterns and relationships between risk factors leads to improved risk assessment.

**Personalized Medicine:** The deep learning model generates personalized risk profiles for individuals, enabling healthcare professionals to tailor prevention and treatment strategies based on a patient's unique risk factors and medical history. This approach supports precision medicine, optimizing treatment plans for better patient outcomes.

**Continuous Improvement:** As the project evolves and gathers more data, the predictive model can be continuously refined and improved. This adaptability ensures that the model stays up-to-date with the latest advancements in breast cancer research and data availability.

**Accessibility and Scalability:** Once developed and validated, the predictive model can be deployed across various healthcare institutions, making advanced breast cancer prediction accessible to a wider population. This scalability can have a significant impact on breast cancer care globally.

### Disadvantages:

**Data Bias:** The predictive model heavily relies on the quality and representativeness of the training data. If the training dataset is biased or lacks diversity, the model's predictions may be skewed or less accurate, leading to potential disparities in risk assessments for certain subgroups.

**Data Privacy and Security:** Handling large amounts of sensitive medical data raises concerns about privacy and security. Ensuring data anonymity and protection against potential breaches or unauthorized access is essential but challenging in large-scale healthcare projects.

**Complexity and Resources:** Deep learning models require significant computational power and resources for training and inference. Implementing and maintaining such models can be costly and resource-intensive, particularly for smaller healthcare facilities with limited budgets.

**Risk of Over-Reliance:** There is a risk that healthcare professionals might become overly reliant on the predictive model, potentially overlooking other crucial clinical indicators and factors when making decisions.

## 8. CONCLUSION

In conclusion, "CancerVision: Advanced Breast Cancer Prediction with Deep Learning" is a transformative and promising project that harnesses the power of advanced deep learning techniques to address critical challenges in breast cancer care. The project's primary objective is to develop a sophisticated predictive model capable of accurately assessing the risk of advanced or metastatic breast cancer in individuals. Through the integration of diverse datasets, including genetic information, imaging results, medical records, lifestyle factors, and biomarkers, the deep learning model aims to provide accurate risk assessments and empower healthcare professionals with valuable insights for personalized medicine. The potential benefits of "CancerVision" are numerous, ranging from improved early detection and better risk stratification to enhanced decision-making and reduced overtreatment. As "CancerVision" continues to evolve, ongoing research and refinement will be crucial to keeping the model up-to-date with the latest advances in breast cancer research and technology. Continuous improvement and real-world validation will further solidify its role as a valuable tool for healthcare professionals, supporting them in making informed decisions and positively impacting breast cancer care on a global scale.

## 9. FUTURE SCOPE

The future scope for CNN models in breast cancer detection includes potential advancements such as improved accuracy and research focusing on interpretability techniques, such as attention maps and feature visualization, which can help us to understand the decision-making process of CNN models for breast cancer detection. Developing techniques to handle imbalance, such as data augmentation strategies, specialized loss functions, and oversampling/undersampling techniques, can improve the performance of

CNN models on imbalanced datasets, ongoing advancements in imaging technologies and data availability will ultimately lead to improved early detection of cancer.

## **10. APPENDIX**

- Data set link:
- Github repository link:
- Our model:
- Project demonstration link:

### **Source code**