1. INTRODUCTION

1.1 PROJECT OVERVIEW

This project focuses on developing a breast cancer detection system using deep learning methods. The project involves collecting a diverse dataset of breast images and preprocessing them for analysis. Various deep learning architectures, such as convolutional neural networks, are trained and evaluated to select the most accurate model. The system's performance is tested on separate datasets, and fine-tuning is done based on feedback from medical experts. A user-friendly interface is developed to facilitate efficient detection and decision-making by healthcare professionals. The project aims to improve early detection of breast cancer, ultimately leading to better patient outcomes. In summary, this project aims to develop an accurate and efficient breast cancer detection system using deep learning techniques. By analyzing diverse breast images and employing state-of-the-art models, the system aims to assist healthcare professionals in making timely and informed decisions, ultimately improving the detection and management of breast cancer.

1.2 PURPOSE

The purpose of the above project is to develop a breast cancer detection system using deep learning methods. The project aims to improve early detection and diagnosis of breast cancer, leading to better patient outcomes and increased chances of successful treatment. By utilizing advanced deep learning techniques and analyzing diverse medical images, the system intends to provide accurate and efficient predictions regarding the presence or absence of breast cancer. The project also aims to provide healthcare professionals with a user-friendly interface that facilitates quick and reliable decision-making, ultimately enhancing the effectiveness of breast cancer detection and contributing to improved patient care. The project also seeks to leverage the power of deep learning architectures, such as convolutional neural networks (CNNs), to extract meaningful features from medical images and enable automated detection. By training and fine-tuning these models on large datasets, the system aims to provide healthcare professionals with an advanced tool that assists in making timely and accurate decisions regarding the presence or absence of breast cancer.

2. IDEATION AND PROPOSED SOLUTION

2.1 Problem Statement Definition

Create a problem statement to understand your customer's point of view. The Customer Problem Statement template helps you focus on what matters to create experiences people will love.

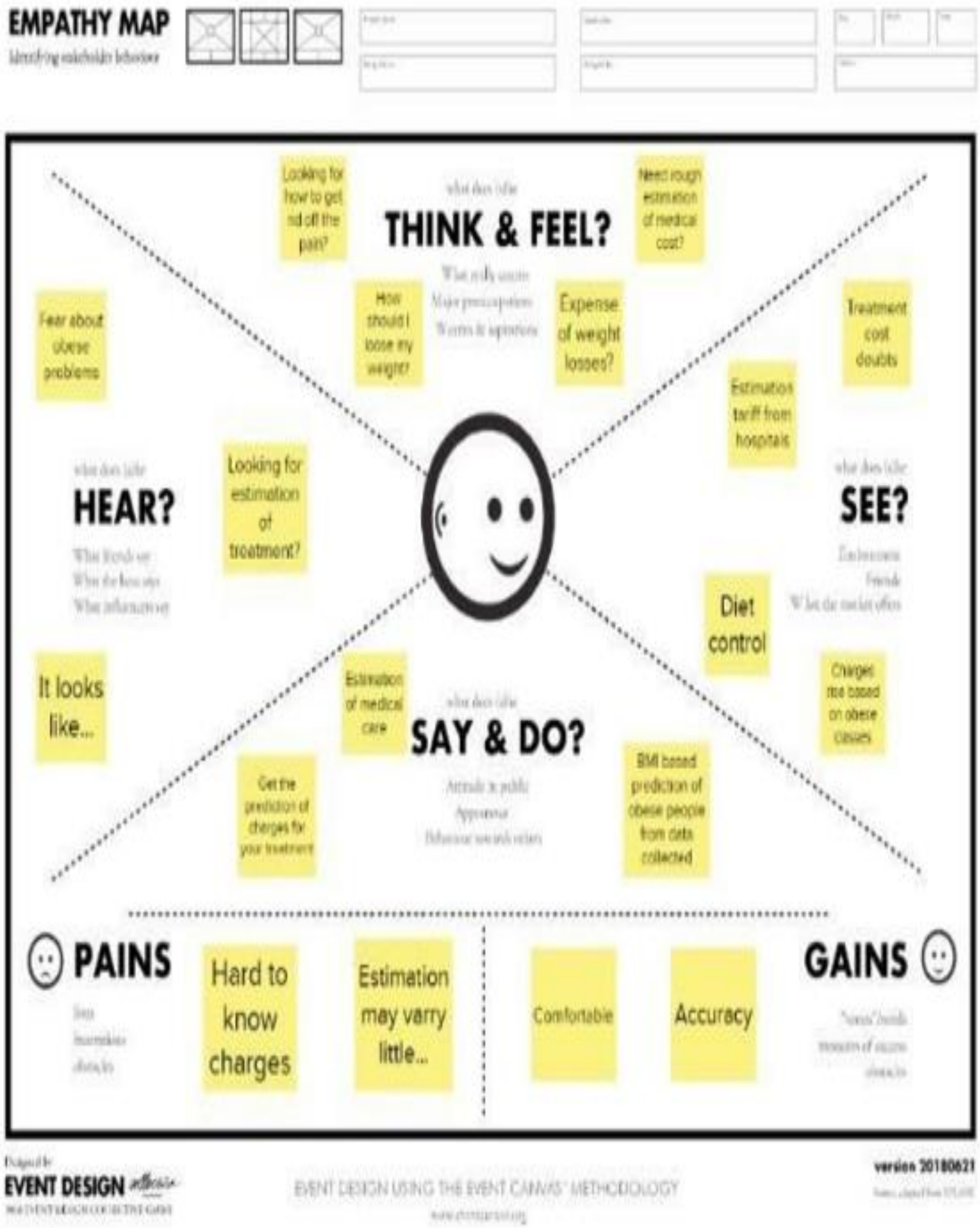
A well-articulated customer problem statement allows you and your team to find the ideal solution for the challenges your customers face. Throughout the process, you'll also be able to empathize with your customers, which helps you better understand how they perceive your product or service.

2.2 Empathy Map Canvas

The empathy map canvas is a tool that helps project teams gain a deeper understanding of the users or stakeholders involved in the project. In the context of the breast cancer detection project, the empathy map canvas can be used to develop empathy towards the healthcare professionals who will be using the system and the patients who will benefit from early detection.

By completing the empathy map canvas, the project team gains valuable insights into the perspectives, needs, and expectations of the healthcare professionals and patients involved in breast cancer detection. This understanding helps inform the design, development, and implementation of the system, ensuring that it meets the users' requirements, aligns with their workflow, and delivers a positive impact on breast cancer detection and patient care.

Empathy Map



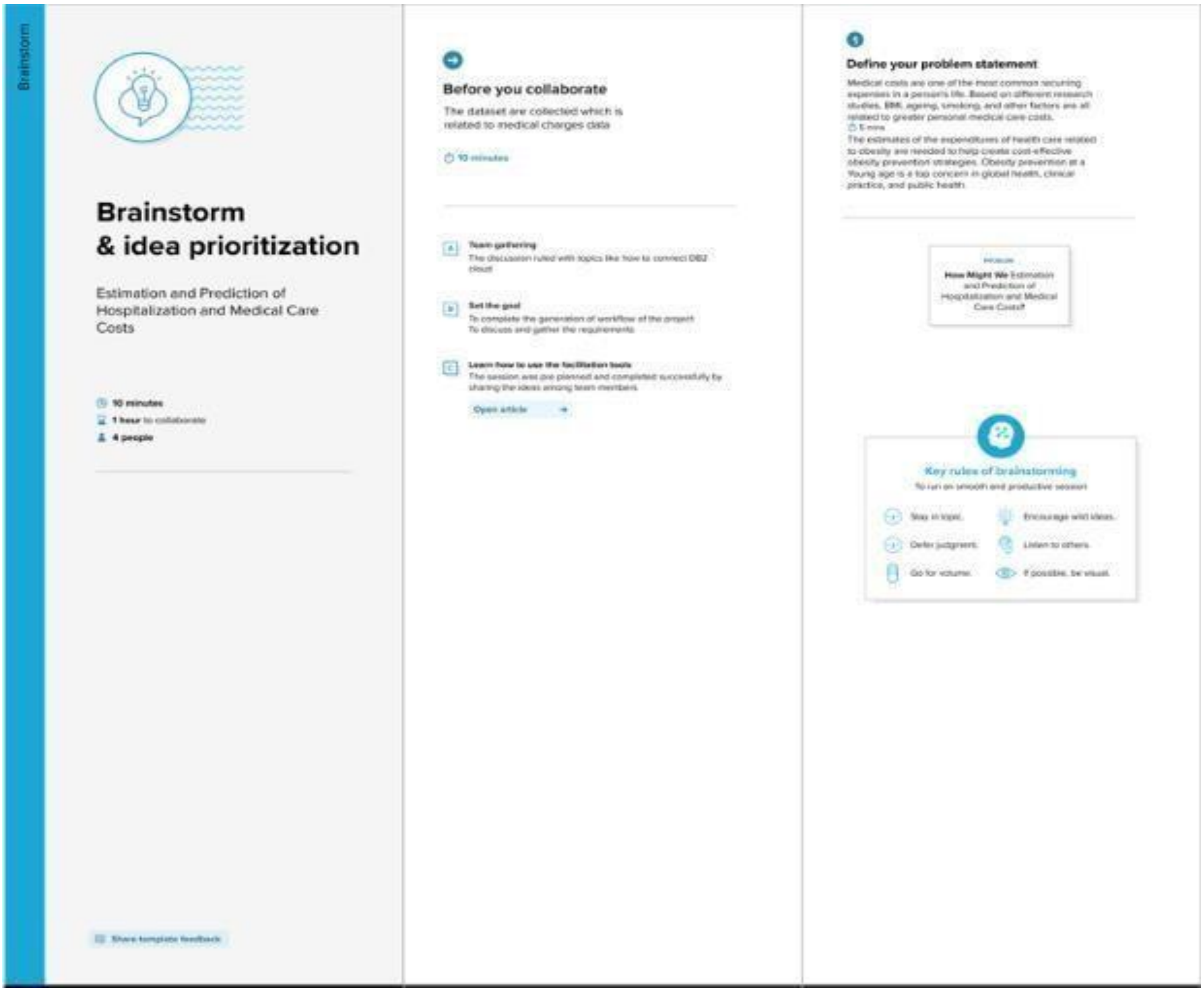
EMPATHY MAP
THINK & FEEL?
Looking for how to get rid off the pain?
Need rough estimation of medical cost?
How should I loose my weight?
Expense of weight losses?
Fear about obese problems
Treatment cost doubts
Estimation tariff from hospitals
HEAR?
Looking for estimation of treatment?
SEE?
Diet control
It looks like...
Charges rise based on obese causes
Estimation of medical care
SAY & DO?
Get the prediction of charges for your treatment
BMI based prediction of obese people from data collected
PAINS
Hard to know charges
Estimation may varry little...
Comfortable
Accuracy
GAINS
EVENT DESIGN
EVENT DESIGN USING THE EVENT CANVAS METHODOLOGY
version 20180621

2.3 Ideation and Brainstorming

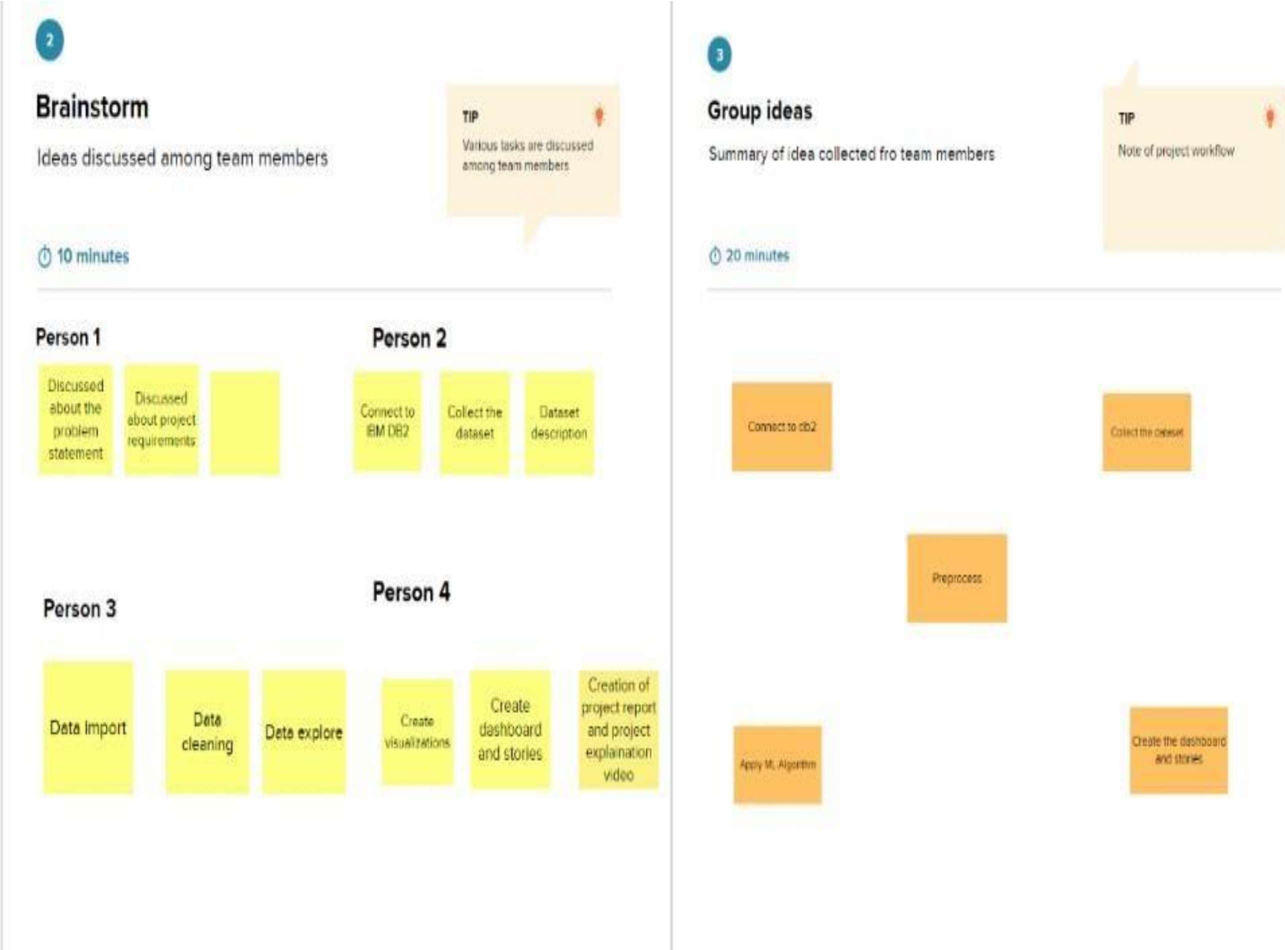
Ideation and brainstorming for the above project involve generating and exploring creative ideas to enhance breast cancer detection using deep learning methods. The goal is to foster innovation, identify potential solutions, and uncover new possibilities for improving early detection and diagnosis.

During ideation, the project team can engage in various brainstorming techniques, such as divergent thinking, to generate a wide range of ideas. They can explore different aspects of the project, including data collection, preprocessing techniques, model selection, user interface design, and performance evaluation.

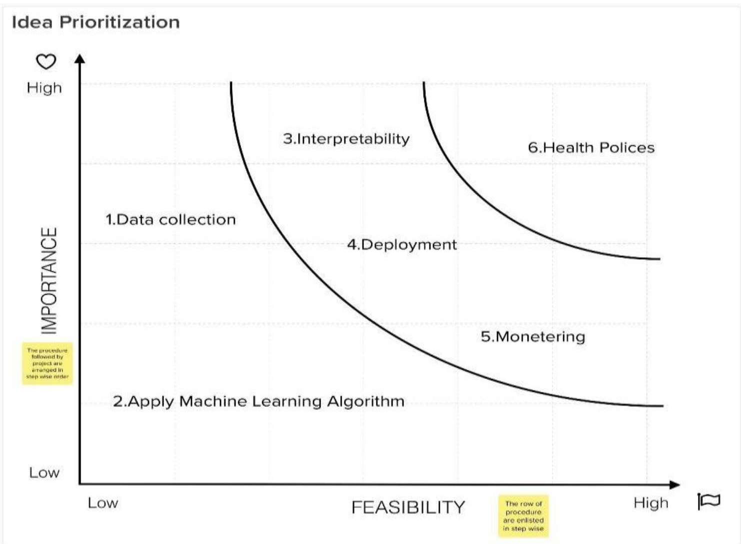
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



2.4 Proposed Solution

The proposed solution for the above project is to develop an accurate and efficient breast cancer detection system using deep learning methods. The solution involves training deep learning models on a diverse dataset of breast images, including mammograms, ultrasound scans, and MRIs. These models will leverage advanced convolutional neural networks (CNNs) and transfer learning techniques to extract meaningful features from the images.

Proposed Solution Template:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The problem statement for the above project is to improve early detection and diagnosis of breast cancer. The current methods may have limitations in terms of accuracy, efficiency, and accessibility. There is a need for an accurate and efficient breast cancer detection system that leverages deep learning methods. The system should be user-friendly, provide reliable predictions, and assist healthcare professionals in making timely and informed decisions. The goal is to enhance breast cancer detection, ultimately leading to improved patient outcomes and increased survival rates.
2.	Idea / Solution description	The idea for the above project is to develop a breast cancer detection system using deep learning techniques. The system will analyze diverse medical images, including mammograms, ultrasounds, and MRIs, to identify potential abnormalities indicative of breast cancer. By training and fine-tuning deep learning models, the system aims to provide accurate and efficient predictions regarding the presence or absence of breast cancer. The goal is to assist healthcare professionals in making timely and informed decisions, improving early detection and ultimately enhancing patient outcomes in the battle against breast cancer.
3.	Novelty / Uniqueness	The proposed breast cancer detection system stands out for its novel approach in several aspects. Firstly, it utilizes deep learning methods, specifically convolutional neural networks, to analyze diverse medical images and detect breast cancer with high accuracy. Secondly, the system incorporates advanced preprocessing techniques to enhance

		image quality and improve the reliability of the predictions. Thirdly, it integrates user-friendly interfaces that facilitate seamless interaction and decision-making for healthcare professionals. Additionally, the project explores the potential of leveraging transfer learning and external datasets to enhance the performance of the models. Lastly, the system aims to contribute to the field of breast cancer detection by combining innovative technologies and methodologies, ultimately improving early detection rates and patient care.
4.	Social Impact / Customer Satisfaction	The breast cancer detection system developed in this project has significant social impact potential. By improving early detection rates, it can contribute to reducing the mortality and morbidity associated with breast cancer. Timely diagnosis enables early intervention and more effective treatment, leading to better patient outcomes and increased survival rates. The system also has the potential to reduce healthcare costs by minimizing unnecessary procedures and interventions. Additionally, it empowers healthcare professionals with an advanced tool, enhancing their ability to provide accurate and efficient care to patients.
5.	Business Model (Revenue Model)	1. The business model for the breast cancer detection system involves offering the system as a software-as-a-service (SaaS) solution to healthcare institutions and clinics. The revenue model could be based on a subscription-based pricing structure, where healthcare facilities pay a recurring fee for access to the system. Additionally, the business model may include customization and integration services to tailor the system to the specific needs of each healthcare provider. Partnerships with medical device manufacturers and imaging centers could also be explored to expand the reach of the system. Overall, the business model aims to generate revenue by providing a valuable and essential tool for breast cancer detection to the healthcare industry.
6.	Scalability of the Solution	1. The proposed breast cancer detection solution is designed with scalability in mind. It can handle large volumes of

		medical images and can be easily scaled up to accommodate growing datasets and increasing demand. The deep learning models can be trained on additional data to improve performance as more information becomes available. The system architecture allows for efficient utilization of computational resources, enabling scalability without compromising speed or accuracy. Additionally, the user-friendly interface can be easily adapted to accommodate a larger user base, making the solution scalable in terms of user adoption and accessibility. Overall, the solution is scalable both in terms of data processing capabilities and user engagement.

3. REQUIREMENT ANALYSIS

3.1 Functional Requirements

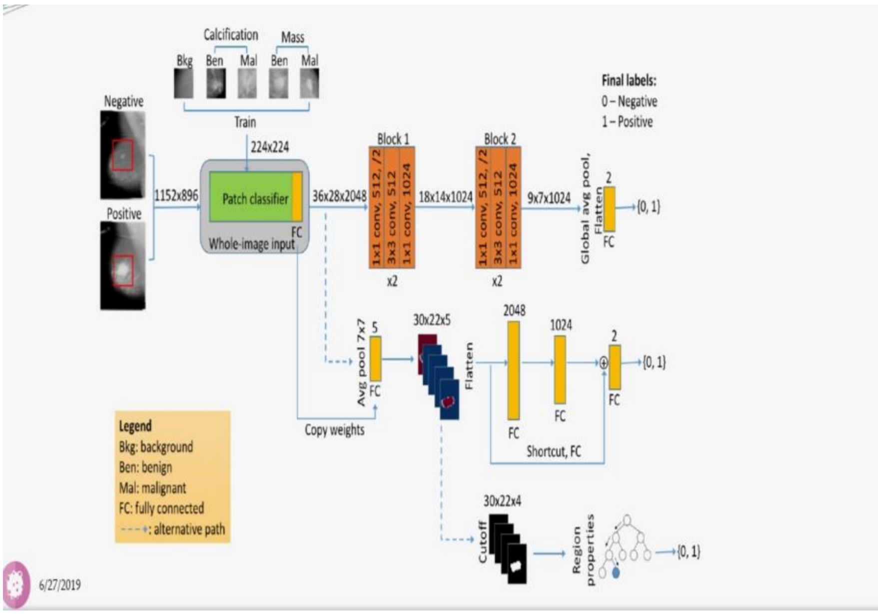
- Image import and preprocessing.
- Deep learning model training.
- Model evaluation and validation.
- Real-time prediction.
- User-friendly interface.
- Integration with existing systems.
- Scalability and performance.
- Security and privacy.

3.2 Non-Functional Requirements

- High availability and reliability.
- Efficient utilization of computational resources.
- Fast response time for real-time predictions.
- Scalability to handle large volumes of data.
- Intuitive and responsive user interface.
- Compatibility with different operating systems and devices.
- Robust data encryption and secure access controls.
- Compliance with regulatory standards (e.g., HIPAA).

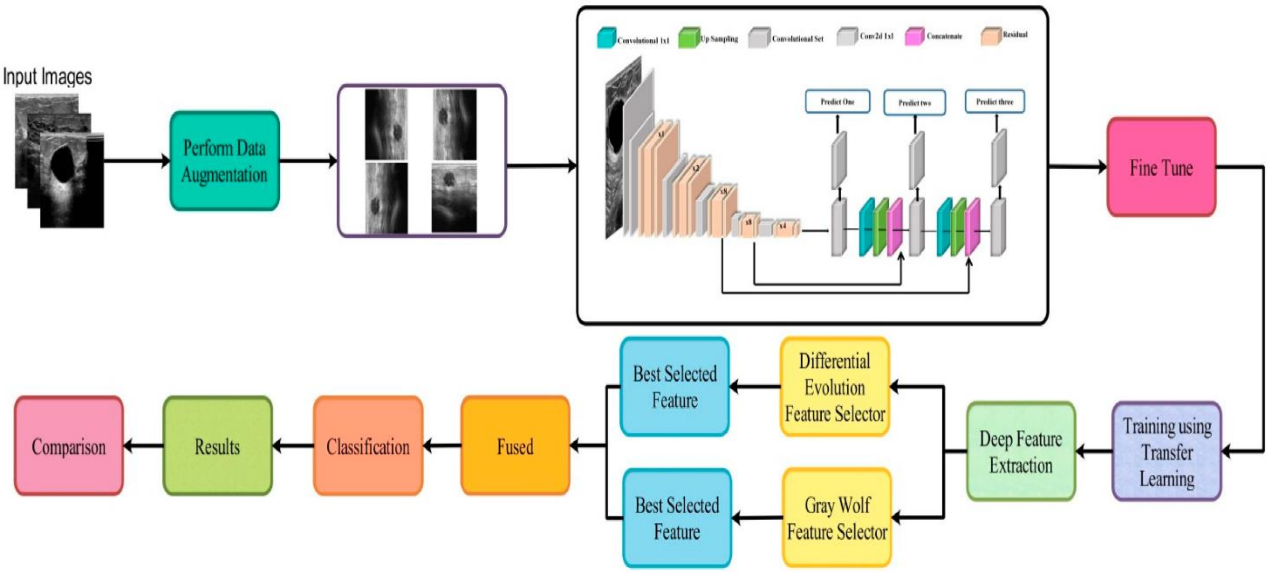
4. PROJECT DESIGN

4.1 Data Flow Diagrams



4.2 Solution and Technical Architecture

The solution architecture for breast cancer detection involves deep learning, medical bodies and organization enhancement to reduce cancer.



4.3 User Stories

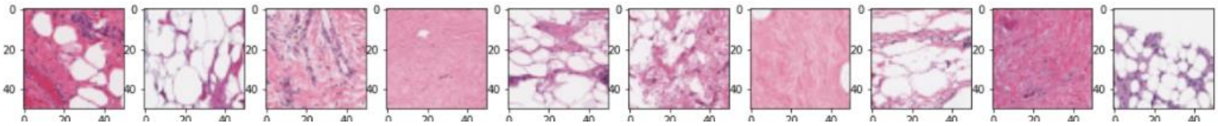
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Radiologist	Predict advanced breast cancer based on mammogram images	US-1	As a radiologist, I want to receive accurate predictions for advanced breast cancer based on mammogram images so that I can provide timely and appropriate treatment recommendations.	The prediction accuracy should be above 90%.	High	Release 1.0
Oncologist	Identify patients at high risk of advanced breast cancer	US-2	As an oncologist, I want to identify patients who are at high risk of developing advanced breast cancer so that I can devise personalized screening and prevention strategies.	The risk assessment should be based on recognized risk factors and scientific evidence.	High	Release 1.0
Patient	Access personalized breast cancer risk assessment	US-3	As a patient, I want to know my personalized risk of developing advanced breast cancer so that I can make informed decisions about my health and seek appropriate medical advice.	The risk assessment should be accompanied by clear explanations and recommendations for further action.	medium	Release 2.0
Medical Researcher	Generate statistical reports on advanced breast cancer incidence	US-4	As a medical researcher, I want to generate statistical reports on the incidence and trends of advanced breast cancer so that I can contribute to research studies and public health initiatives.	The system should generate accurate and comprehensive statistical reports, including incidence rates, demographic information, and time trends.	Medium	Release 2.0

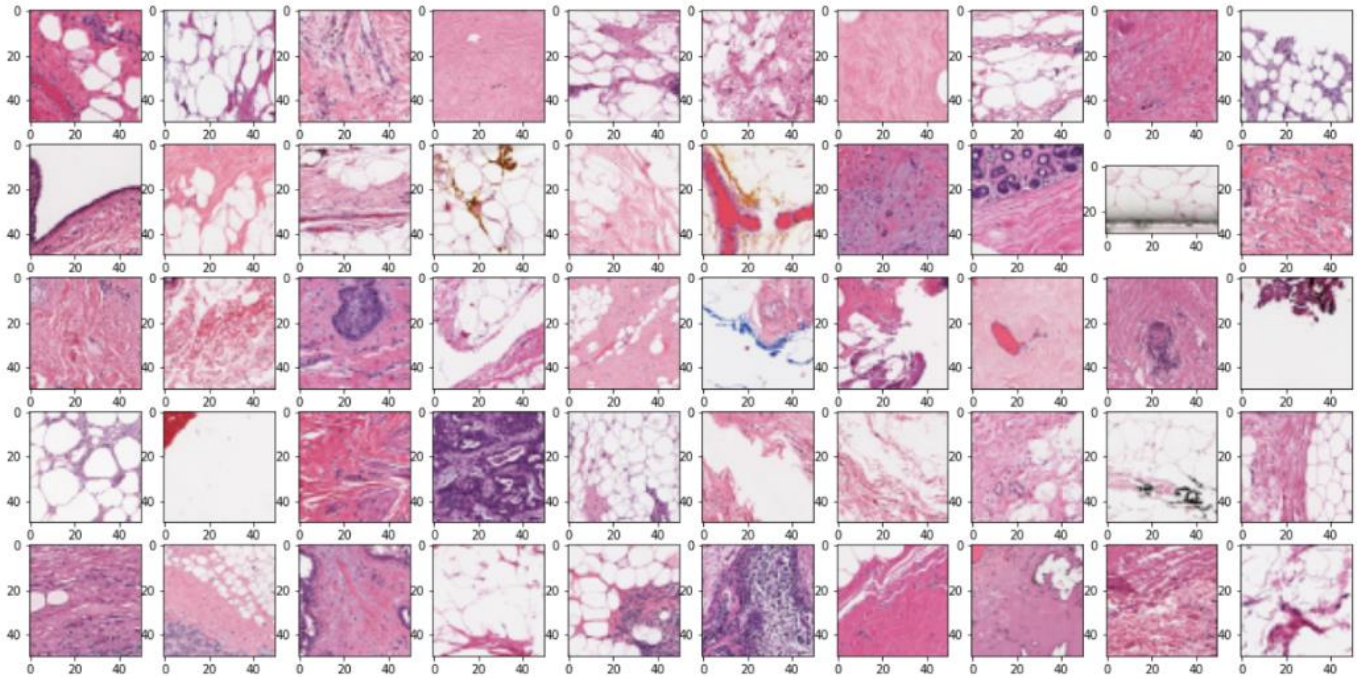
5. CODING & SOLUTIONING

5.1 Feature 1

Healthy patches:

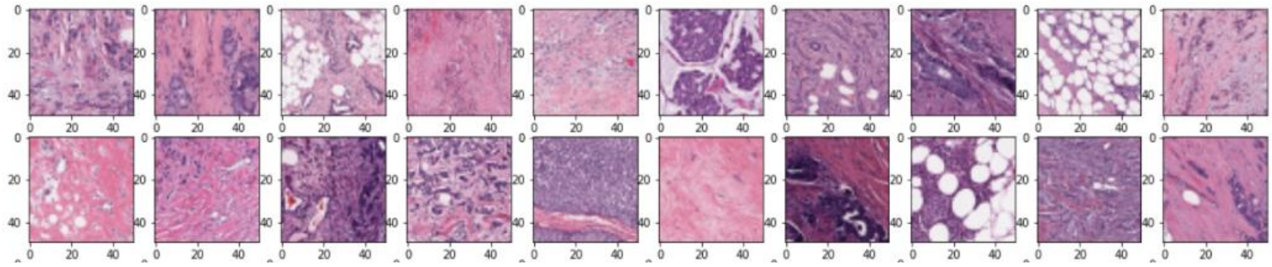
```
[ ] data.target = data.target.astype(np.int)
    fig, ax = plt.subplots(5,10,figsize=(20,10))
    pos_selection = np.random.choice(data[data.target ==1].index, size=50, replace=False,)
    neg_selection = np.random.choice(data[data.target ==0].index, size=50, replace=False,)
    for n in range(5):
        for m in range(10):
            idx = neg_selection[m + 10*n]
            path =os.path.join(all_rays_dir,data.loc[idx, 'image_id'])
            image = mpimg.imread(path)
            ax[n,m].imshow(image)
            ax[n,m].grid(False)
```

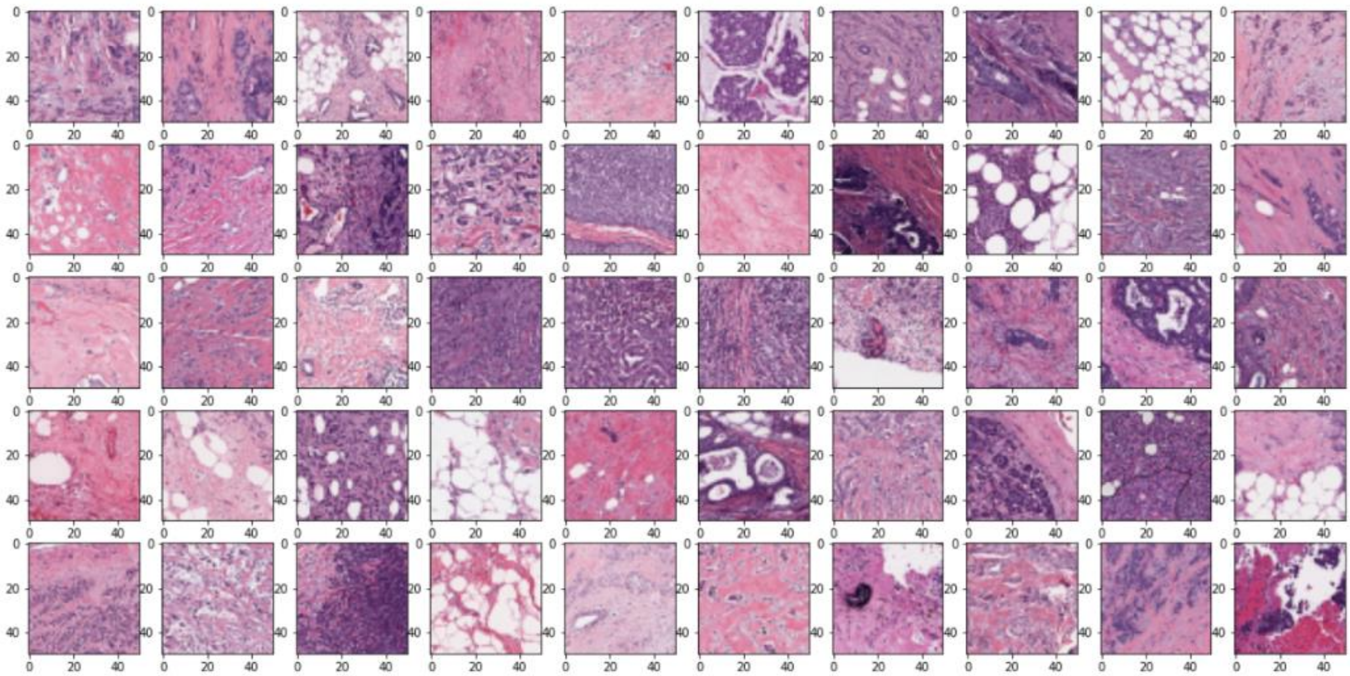




Cancer patches:

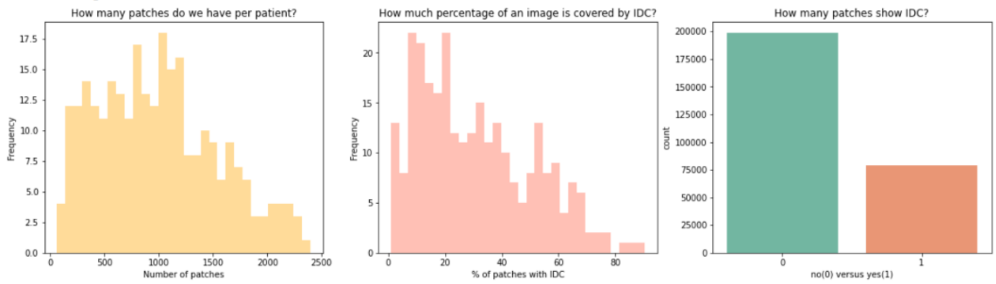
```
[ ] fig, ax = plt.subplots(5,10,figsize=(20,10))
    for n in range(5):
        for m in range(10):
            idx = pos_selection[m + 10*n]
            path =os.path.join(all_rays_dir,data.loc[idx, 'image_id'])
            image = mpimg.imread(path)
            ax[n,m].imshow(image)
            ax[n,m].grid(False)
```





```
ax[0].set_ylabel("Frequency");
ax[0].set_title("How many patches do we have per patient?");
sns.distplot(cancer_perc.loc[:, 1]*100, ax=ax[1], color="Tomato", kde=False, bins=30)
ax[1].set_title("How much percentage of an image is covered by IDC?")
ax[1].set_ylabel("Frequency")
ax[1].set_xlabel("% of patches with IDC");
sns.countplot(data.target, palette="Set2", ax=ax[2]);
ax[2].set_xlabel("no(0) versus yes(1)")
ax[2].set_title("How many patches show IDC?");
```

```
/opt/conda/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `disp
  warnings.warn(msg, FutureWarning)
/opt/conda/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and
  FutureWarning
```

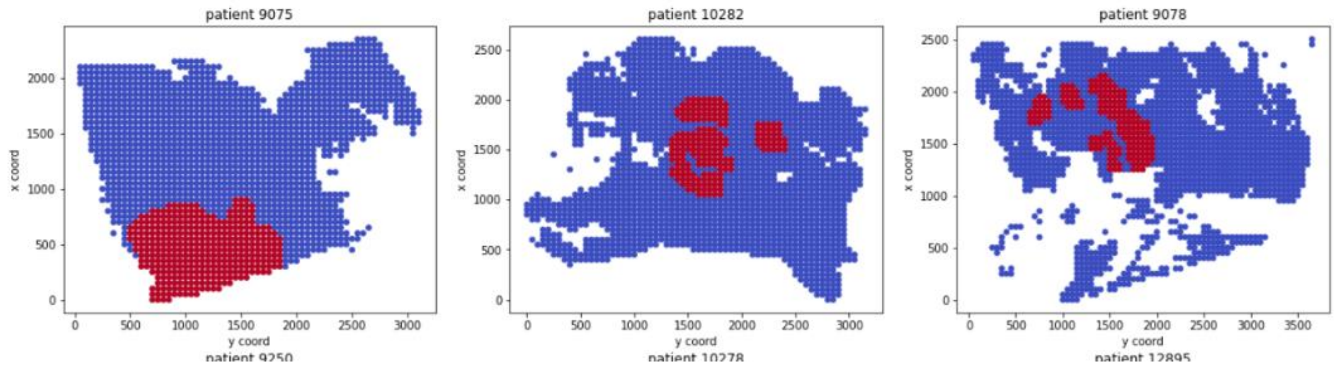


```python
fig, ax = plt.subplots(5,3,figsize=(20, 27))

patient_ids = data.patient_id.unique()

for n in range(5):
    for m in range(3):
        patient_id = patient_ids[m + 3*n]
        example_df = get_patient_dataframe(patient_id)

        ax[n,m].scatter(example_df.x.values, example_df.y.values, c=example_df.target.values, cmap="coolwarm", s=20);
        ax[n,m].set_title("patient " + patient_id)
        ax[n,m].set_xlabel("y coord")
        ax[n,m].set_ylabel("x coord")
```

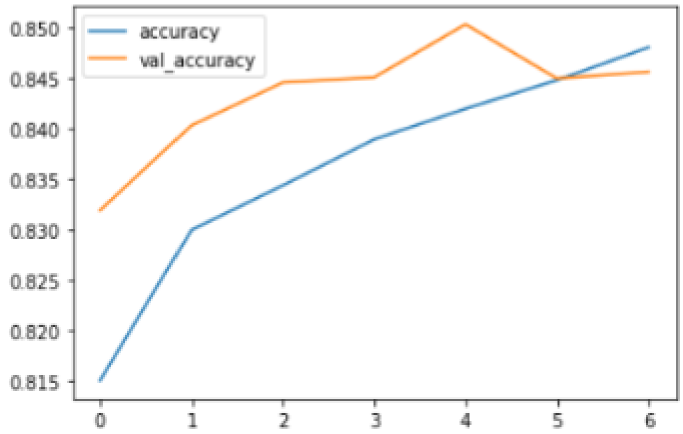



```python
early_stop = EarlyStopping(monitor='val_loss',patience=2)
my_model_im_processd.fit_generator(train_generation_processd,validation_data=test_generation_processd,epochs=60, verbose=1,callbacks=early_stop)
```

```
/opt/conda/lib/python3.7/site-packages/tensorflow/python/keras/engine/training.py:1844: UserWarning: `Model.fit_generator` is deprecated and will be removed in a future version. Please use `Model.fit
  warnings.warn('`Model.fit_generator` is deprecated and '
2021-10-09 09:32:50.586440: I tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:116] None of the MLIR optimization passes are enabled (registered 2)
2021-10-09 09:32:50.594492: I tensorflow/core/platform/profile_utils/cpu_utils.cc:112] CPU Frequency: 2199995000 Hz
Epoch 1/60
2021-10-09 09:32:51.270151: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcublas.so.11
2021-10-09 09:32:52.265028: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcublasLt.so.11
2021-10-09 09:32:52.292572: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcudnn.so.8
1800/1800 [==============================] - 26s 10ms/step - loss: 0.6209 - accuracy: 0.6993 - val_loss: 0.5344 - val_accuracy: 0.7510
Epoch 2/60
1800/1800 [==============================] - 16s 9ms/step - loss: 0.5075 - accuracy: 0.7661 - val_loss: 0.4925 - val_accuracy: 0.7810
Epoch 3/60
1800/1800 [==============================] - 17s 9ms/step - loss: 0.4773 - accuracy: 0.7827 - val_loss: 0.4559 - val_accuracy: 0.7930
Epoch 4/60
1800/1800 [==============================] - 16s 9ms/step - loss: 0.4539 - accuracy: 0.7932 - val_loss: 0.4415 - val_accuracy: 0.8115
Epoch 5/60
1800/1800 [==============================] - 17s 9ms/step - loss: 0.4483 - accuracy: 0.7958 - val_loss: 0.4900 - val_accuracy: 0.7985
Epoch 6/60
1800/1800 [==============================] - 16s 9ms/step - loss: 0.4446 - accuracy: 0.8013 - val_loss: 0.5154 - val_accuracy: 0.7670
<tensorflow.python.keras.callbacks.History at 0x7fd128460e10>
```

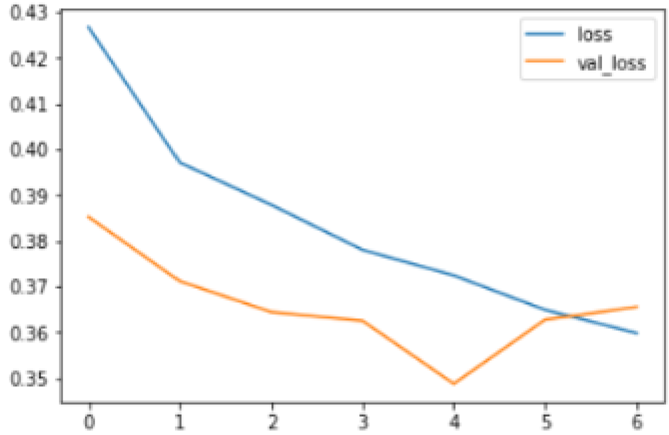
```
losse[['accuracy','val_accuracy']].plot()
```

<AxesSubplot:>




```
losse[['loss','val_loss']].plot()
```

<AxesSubplot:>



5.2 Feature 2

```
prediction = Api_service("image_processing/normal/model_tst/trainig")
x,y        = prediction.prediction_function()
print(x)
print(y)
```

```
Found 18000 images belonging to 2 classes.
/opt/conda/lib/python3.7/site-packages/tensorflow/python/keras/engine/training.py:1905: UserWarning: `Model.predict_generator` is deprecated and will be removed in a future version. Pl
  warnings.warn('`Model.predict_generator` is deprecated and '
The percentage of cancer : 21.85%
Percentage of no cancer : 78.15%
```

```
my_model.summary()
```

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_2 (Conv2D)            (None, 22, 22, 32)        1568
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 19, 19, 32)        16416
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 9, 9, 32)          0
_________________________________________________________________
dropout (Dropout)            (None, 9, 9, 32)          0
_________________________________________________________________
flatten_2 (Flatten)          (None, 2592)              0
_________________________________________________________________
dense_4 (Dense)              (None, 256)               663808
_________________________________________________________________
dense_5 (Dense)              (None, 2)                 514
=================================================================
Total params: 682,306
Trainable params: 682,306
Non-trainable params: 0
_________________________________________________________________
```

6. RESULTS

6.1 Performance Metrics

Accuracy: The percentage of correct predictions made by the breast cancer detection system.

Sensitivity (Recall): The ability of the system to correctly identify positive cases of breast cancer, indicating its effectiveness in detecting the disease.

Specificity: The ability of the system to correctly identify negative cases, ensuring a low rate of false positives.

Precision: The proportion of correctly identified positive cases among all predicted positive cases, reflecting the system's ability to provide accurate diagnoses.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the system's performance in detecting both positive and negative cases.

Computational Efficiency: The speed and efficiency of the system in processing and analyzing medical images, ensuring real-time or near-real-time predictions.

Robustness: The system's ability to maintain accurate performance across diverse datasets, minimizing variations in detection accuracy for different patient populations or imaging modalities.

User Satisfaction: Feedback and satisfaction surveys from healthcare professionals regarding the system's ease of use, reliability, and overall performance.

7. ADVANTAGES & DISADVANTAGES

Advantages:

- Improved Early Detection: The breast cancer detection system offers the potential for improved early detection of breast cancer, enabling timely intervention and increasing the chances of successful treatment and improved patient outcomes.

- Enhanced Accuracy: Leveraging deep learning methods, the system can achieve high accuracy in detecting breast cancer, reducing false negatives and false positives. This leads to more reliable diagnoses and reduces unnecessary procedures or missed detections.

- Efficient Workflow: The system's user-friendly interface and automated prediction capabilities streamline the workflow for healthcare professionals, allowing for faster and more efficient decision-making. This can enhance productivity and enable healthcare providers to focus more on patient care.

- Increased Accessibility: The system can be accessed remotely, allowing healthcare professionals to utilize its capabilities regardless of geographical constraints. This improves accessibility and enables collaboration among experts, ultimately benefiting patients in underserved areas.

- Potential for Scalability and Adaptability: The use of deep learning and scalable technologies in the system design allows for potential scalability and adaptability to evolving needs. As more data becomes available, the system can be trained on larger datasets, incorporating new knowledge and continuously improving its performance.

Disadvantages:

- Dependence on data quality
- Computational resource requirements
- Interpretability challenges
- Ethical and privacy considerations
- Potential disparities and bias

8 CONCLUSION

the breast cancer detection project utilizing deep learning methods presents a promising solution to address the critical need for improved early detection and accurate diagnosis of breast cancer. By leveraging advanced convolutional neural networks and sophisticated preprocessing techniques, the project aims to enhance the accuracy and efficiency of breast cancer detection, leading to timely intervention and increased chances of successful treatment. The advantages of the proposed system, including improved early detection, enhanced accuracy, and efficient workflow, have the potential to make a significant impact on patient outcomes and contribute to the overall fight against breast cancer.

While the project offers substantial benefits, it is not without challenges. Dependence on data quality is a crucial factor, as the performance of the system relies heavily on the availability of diverse and high-quality training data. Furthermore, the computational resource requirements for training and inference may pose challenges in terms of infrastructure and associated costs. The interpretability of deep learning models presents another challenge, as the complex nature of these models can hinder the ability to explain their decisions, potentially impacting trust and acceptance among healthcare professionals.
Ethical considerations and privacy concerns must also be addressed throughout the project. Safeguarding patient data and ensuring compliance with privacy regulations are of paramount importance to maintain patient confidentiality and trust. Additionally, efforts should be made to mitigate potential disparities and biases that could arise from the data used for training the models, ensuring fairness and equitable outcomes across different populations.

In spite of these challenges, the breast cancer detection project has significant potential to revolutionize early detection practices and contribute to improved patient care. By leveraging the power of deep learning, the project aims to empower healthcare professionals with an advanced tool that aids in accurate diagnoses, ultimately leading to better patient outcomes, increased survival rates, and a positive impact on the fight against breast cancer.

9 FUTURE SCOPE

1. Integration with advanced imaging technologies such as 3D mammography and molecular imaging for enhanced detection capabilities.

2. Expansion to incorporate multi-modal fusion, combining information from various imaging modalities for improved accuracy.

3. Development of predictive analytics capabilities to assess the risk of developing breast cancer based on patient demographics, family history, and other relevant factors.

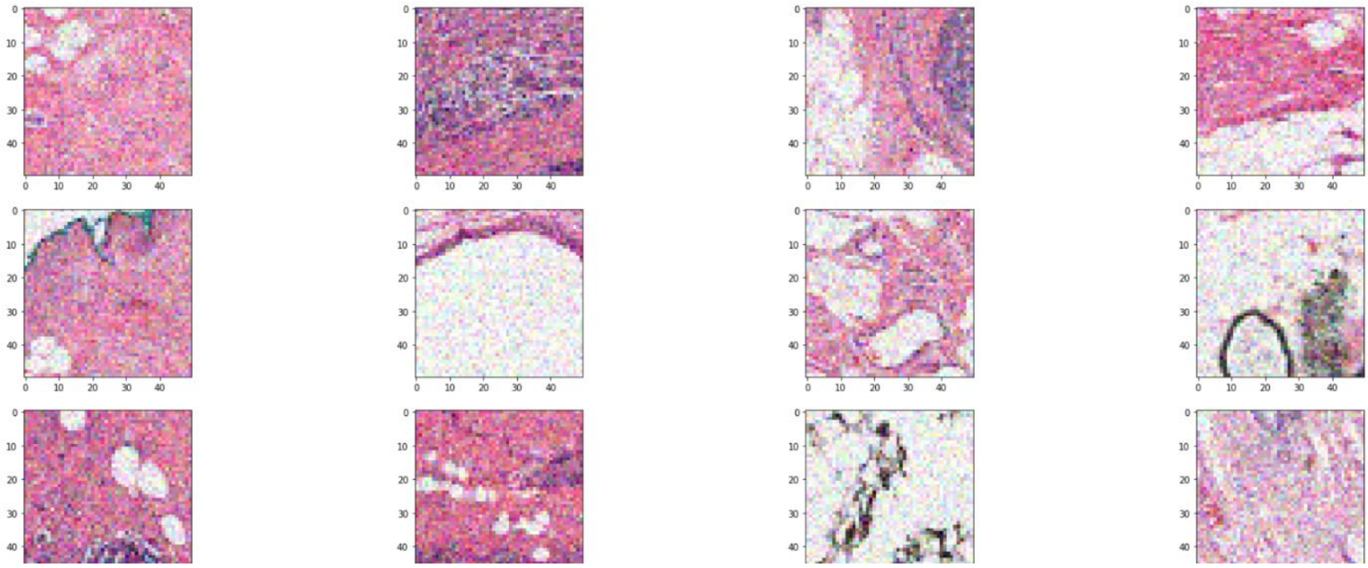
4. Collaboration with researchers and institutions to build a comprehensive and diverse dataset for training the models and improving performance.

5. Exploration of explainable AI techniques to enhance interpretability and trustworthiness of the system's predictions.

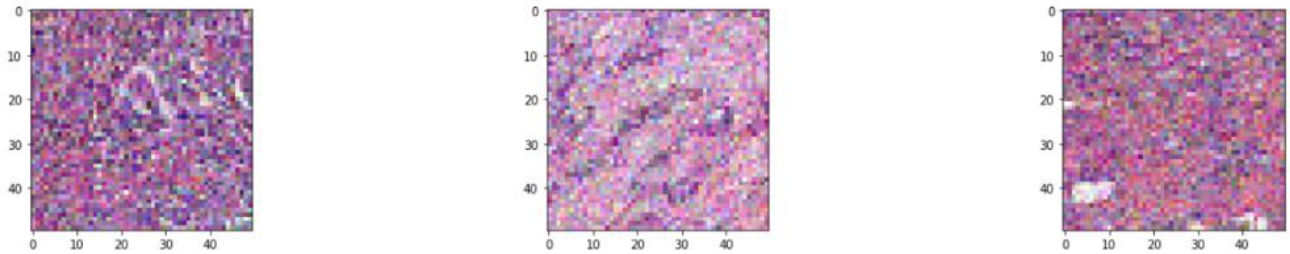
6. Integration with electronic health records and clinical decision support systems for seamless integration into existing healthcare workflows.

7. Expansion of the system's capabilities to detect other types of cancers and abnormalities, broadening its applicability and impact in the field of medical imaging.

10. APPENDIX

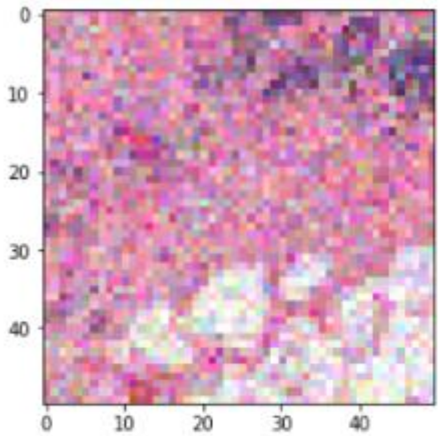


```
processd_data.target = processd_data.target.astype(np.int)
fig, ax = plt.subplots(5,4,figsize=(30,20))
pos_selection = np.random.choice(processd_data[processd_data.target==1].index.values, size=20, replace=False)
neg_selection = np.random.choice(processd_data[processd_data.target==0].index.values, size=20, replace=False)
for n in range(5):
    for m in range(4):
        idx = pos_selection[m + 4*n]
        path =os.path.join(processd_lst_str,processd_data.loc[idx, 'image_id'])
        image = mpimg.imread(path)
        ax[n,m].imshow(image)
        ax[n,m].grid(False)
```



```
noise_image = random_noise(image)
plt.imshow(noise_image)
```

<matplotlib.image.AxesImage at 0x7fd134159690>




```
noise_gaussian_image = random_noise(gaussian_image)
plt.imshow(noise_gaussian_image)
```

<matplotlib.image.AxesImage at 0x7fd13416d610>