

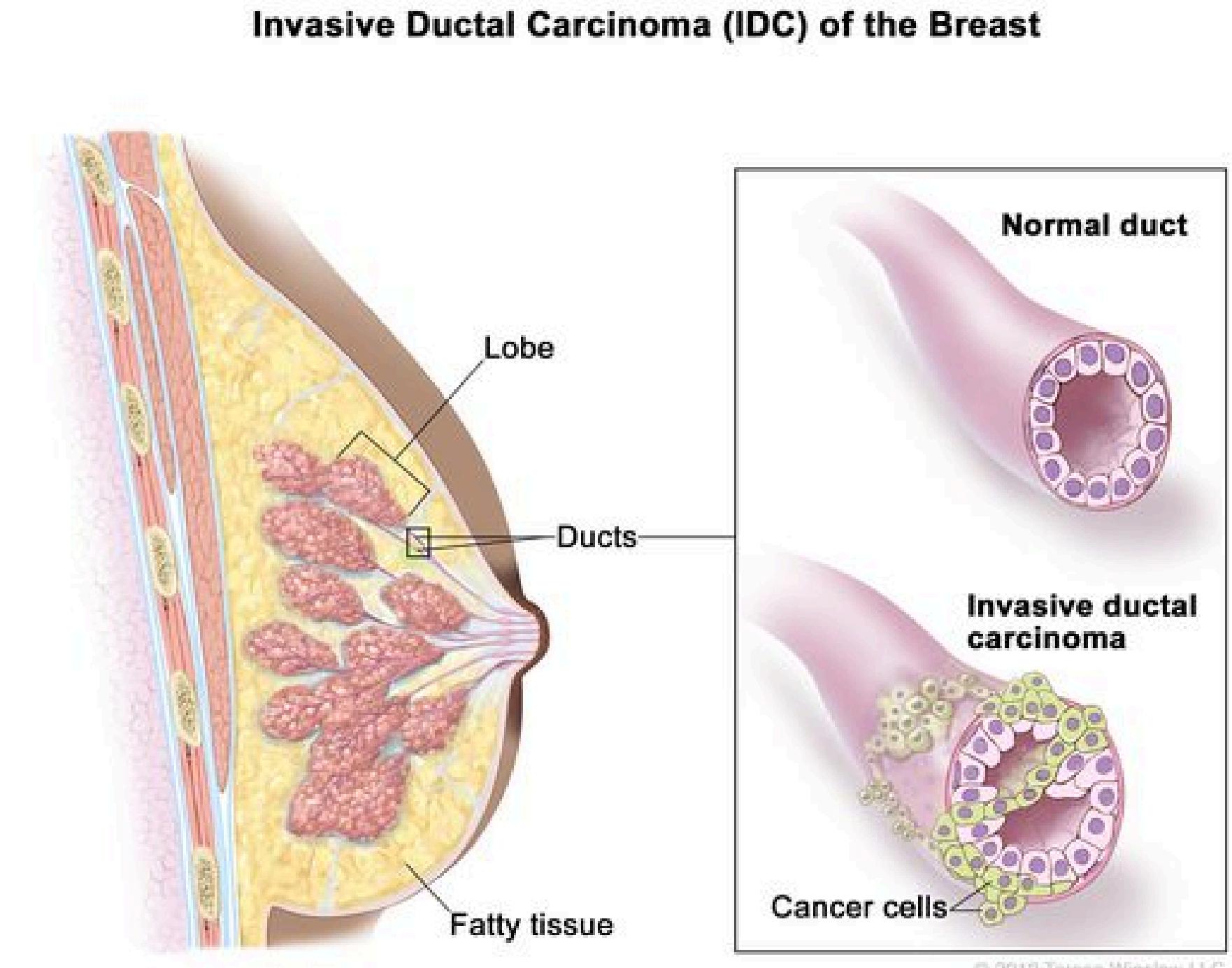
PREDICTING INVASIVE DUCTAL CARCINOMA(IDC) IN HISTOPATHOLOGY SLIDES USING DEEP LEARNING MODELS

AUTHORS: Frida Oycho, Felix Njoroge, Winny Chemusian,
Mataen Surupai, Sammy Warah & Christine Ndirangu



Background - Breast cancer Statistics

- In Kenya, the age-standardized rate of breast cancer incidence is **58.4 per 100,000 women**, making it **the most common cancer in Kenya**. Incidence **6000** (*KEMRI*)
- The reported mortality rate from breast cancer in Kenya was **7.1 per 100,000 women**. 2500 pa
- Invasive ductal carcinoma is the **most common type of breast cancer**.

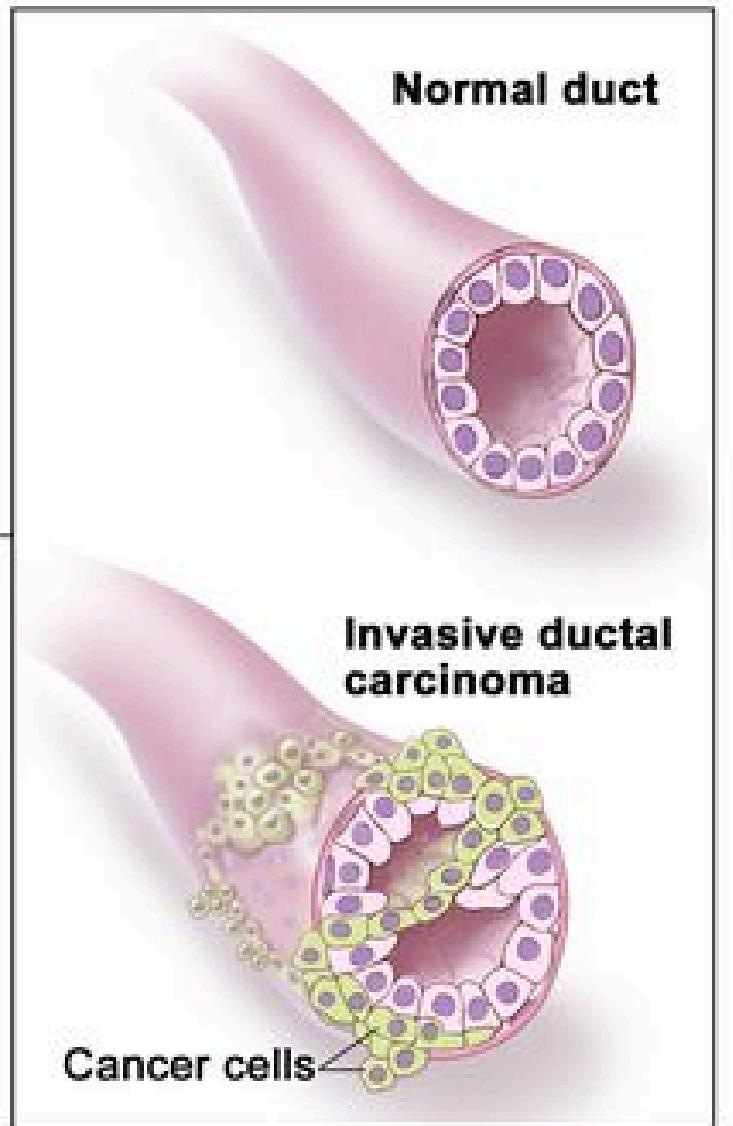
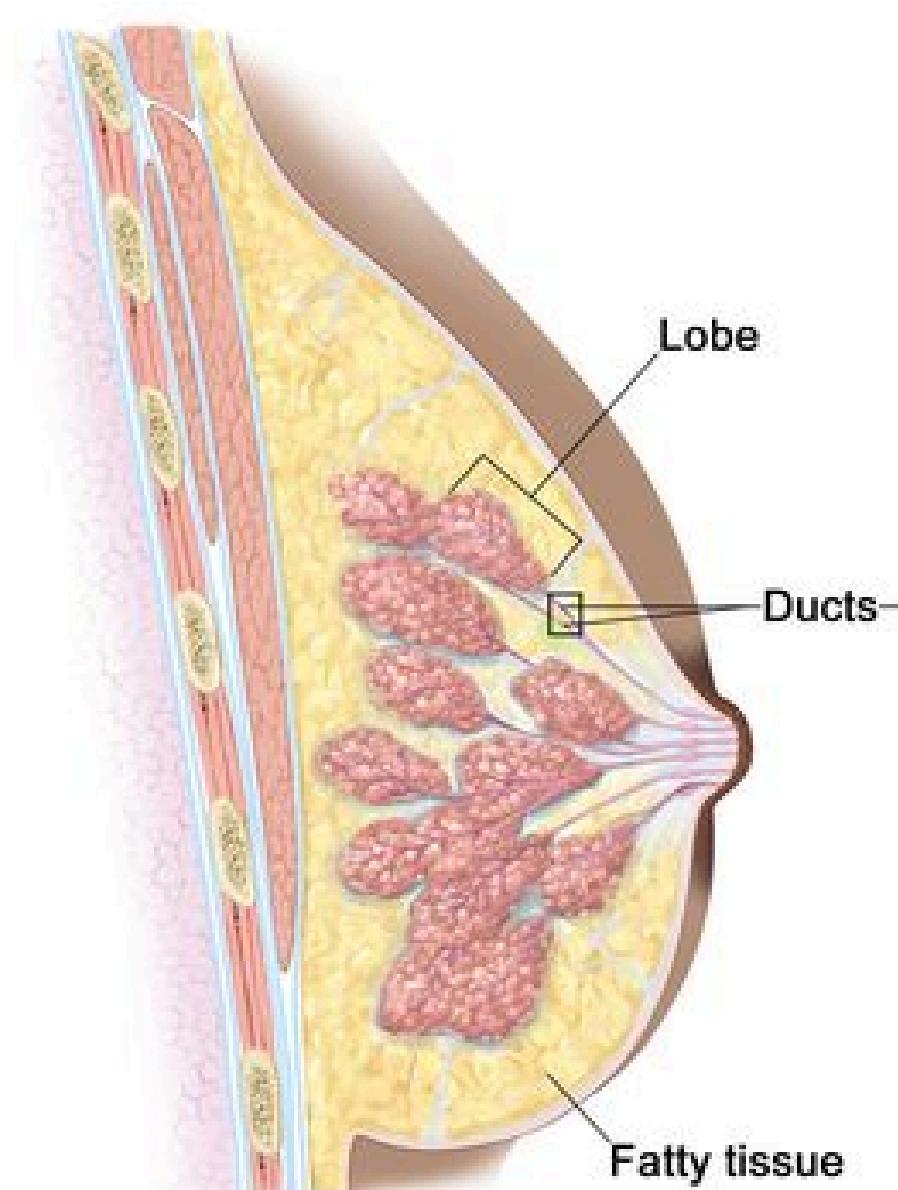


© 2019 Terese Winslow LLC
U.S. Govt. has certain rights

Background - IDC

- IDC begins in the lining of the milk ducts (thin tubes that carry milk from the lobules of the breast to the nipple) **and spreads outside the ducts to surrounding normal tissue**
- IDC can spread through the blood and lymphatic system

Invasive Ductal Carcinoma (IDC) of the Breast



© 2019 Terese Winslow LLC
U.S. Govt. has certain rights

Breast cancer in Kenya

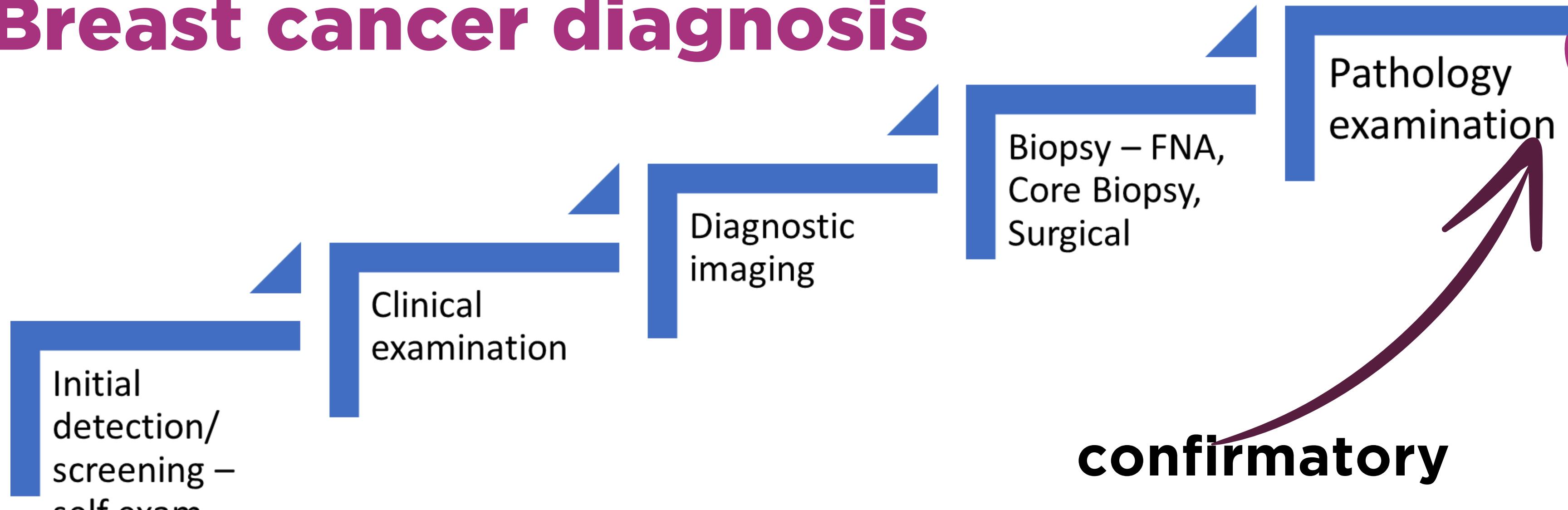
- In Kenya, like in many other countries, the incidence of breast cancer, including IDC, has been rising.
- Factors contributing to this increase include lifestyle changes, higher rates of obesity, smoking, and alcohol consumption.
- The high mortality of breast cancer is often due to factors such as the late presentation, **delayed diagnosis** and limited access to specialized treatment



NATION

We are lagging behind in
breast cancer treatment

Breast cancer diagnosis



- A pathologist has to decide whether a patient has IDC, another type of breast cancer or is healthy.
- In addition the pathological slides needs to be located to find out how advanced the disease is and which grade should be assigned.
- This has to be **done manually** and **is a time consuming process**. Furthermore the pathologists in the country are few and the **decision depends on the expertise of the pathologist and his or her equipment**.

Where does deep learning come in?

- Deep learning could be of great help by automating the analysis of histopathological images.
- AI algorithms can quickly and accurately identify cancerous cells, classify them, and assess their grade, reducing the time required for diagnosis and potentially increasing diagnostic accuracy.
- Helps in standardizing the results, minimizing human error, and providing consistent and reliable diagnoses.
- This way one would be able to overcome the dependence on the pathologist which would be especially useful in regions where no experts are available.



Success metrics

1. **Accuracy:** A model has demonstrated with an accuracy rate of over 90% in distinguishing IDC-positive from IDC-negative samples.
2. **Sensitivity and Specificity:** model's sensitivity (true positive rate) and specificity (true negative rate) that are both above 90%, indicating reliable performance across different cases
3. **Processing Time:** The model that can analyze and provide results within seconds, significantly faster than traditional methods.

Key stakeholders

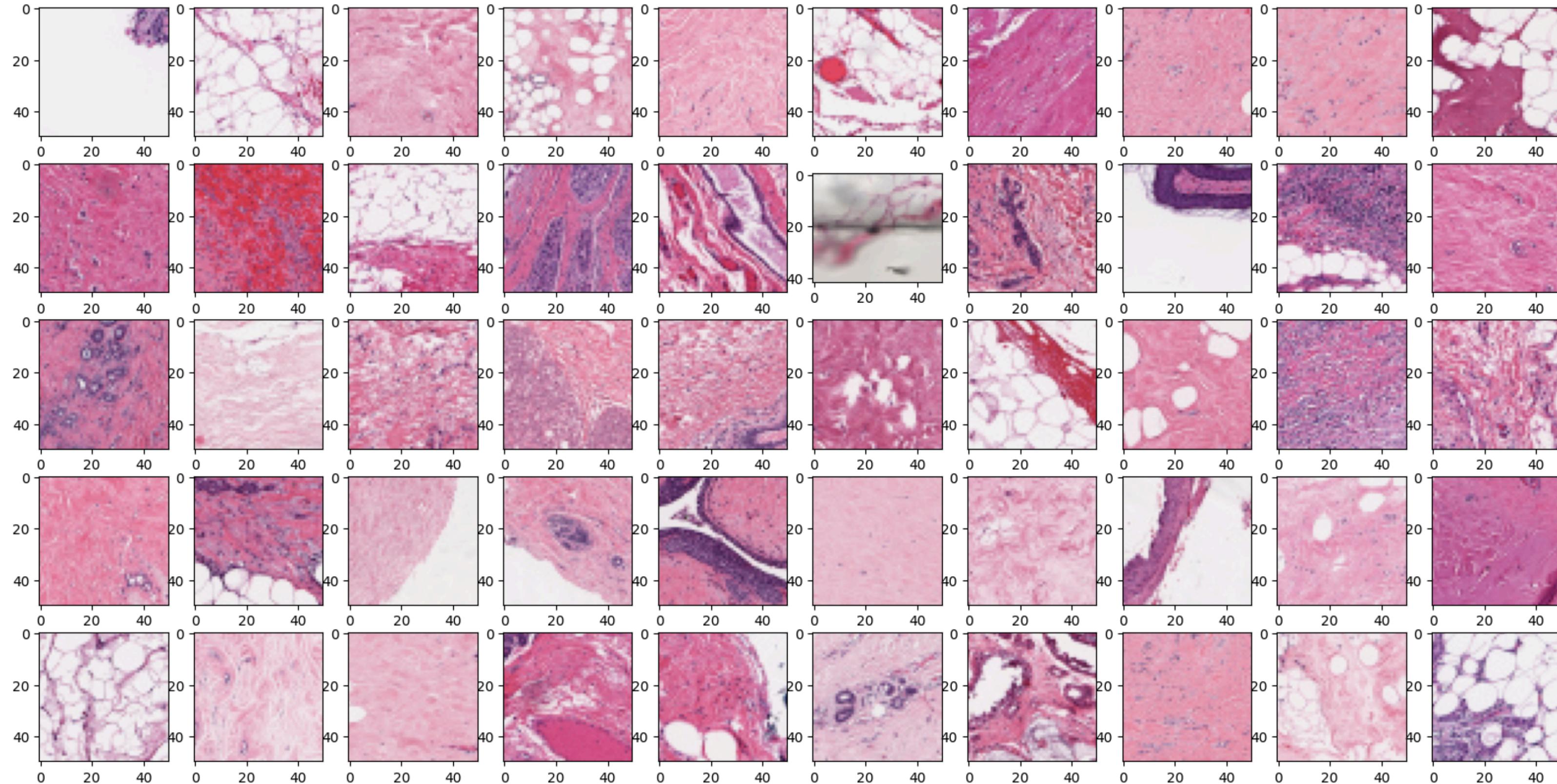
1. Pathologists
2. Oncologists
3. Healthcare institutions
4. Medical researchers

EXPLORATORY DATA ANALYSIS

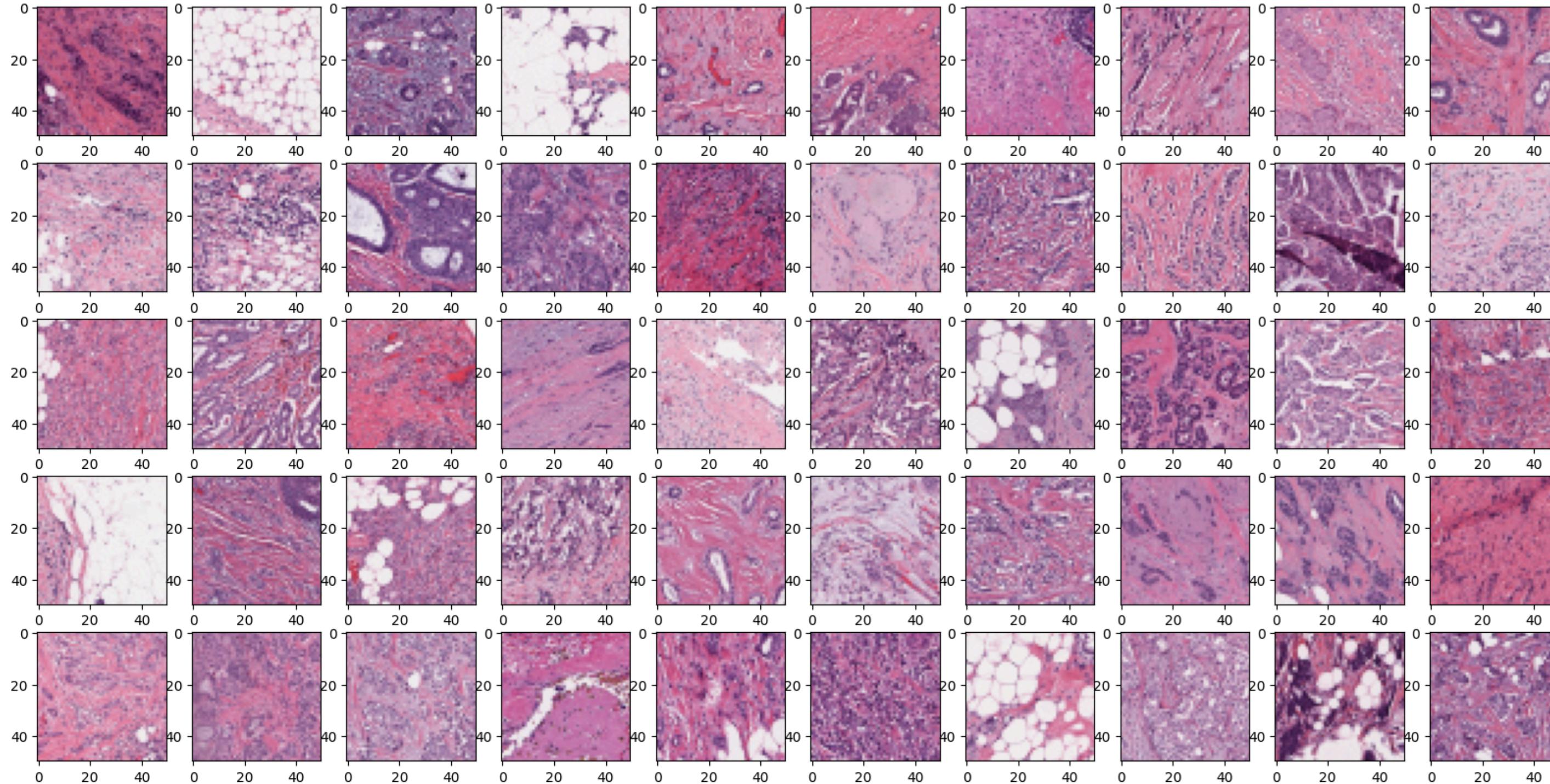
DATASET DESCRIPTION

- Data Source: **Kaggles Breast Cancer Images dataset.**
- Original data was obtained from a research article published by **Janowczyk et al. (2016) in the journal of pathology informatics.**
- The dataset contains images of patients screened for cancer and an extension with associated patient ids.
- 280 patient files with **172,203 non IDC images** and **67,434 with IDC.**

Visualization of non IDC patches

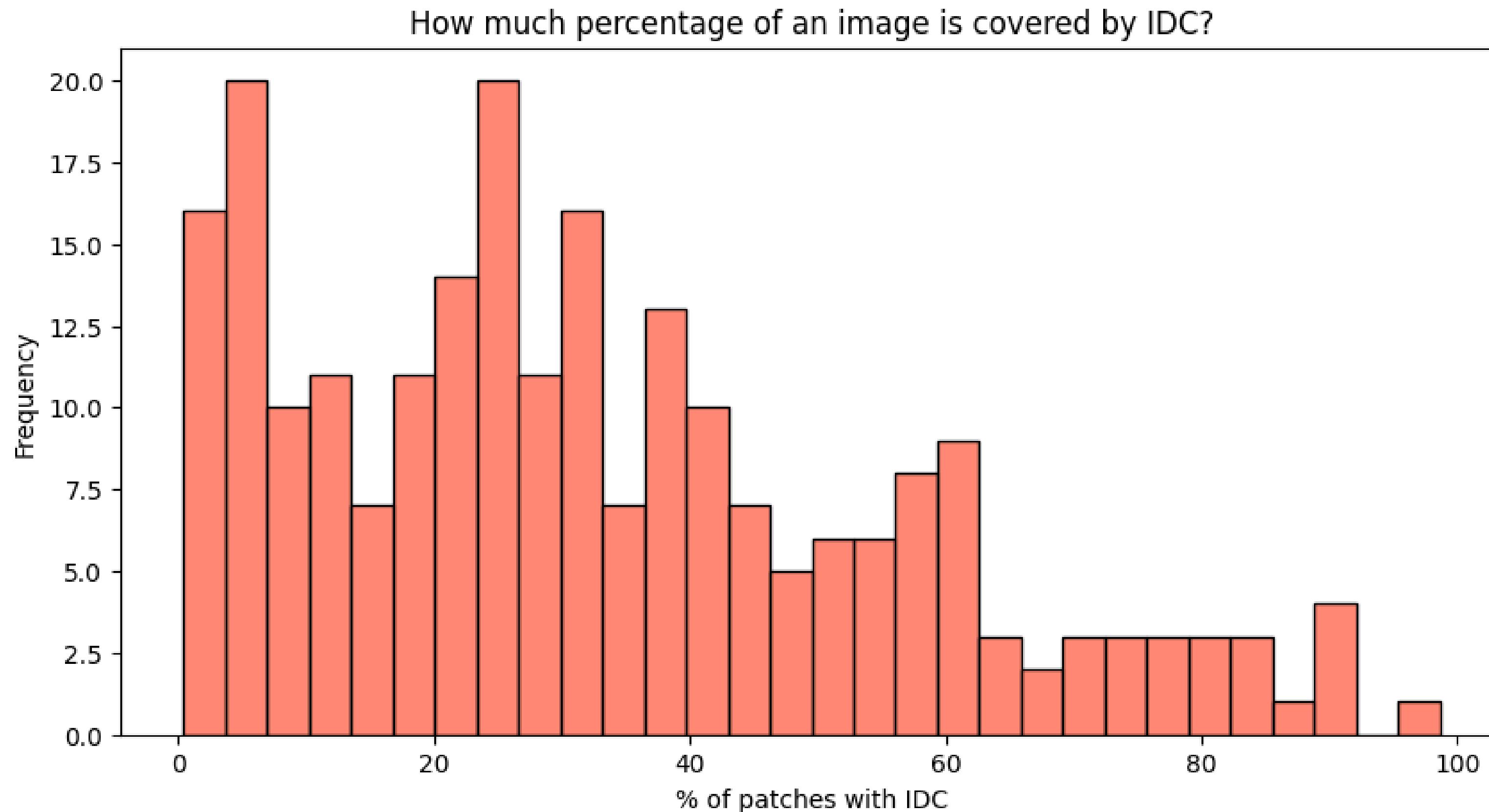


Visualization of IDC patches

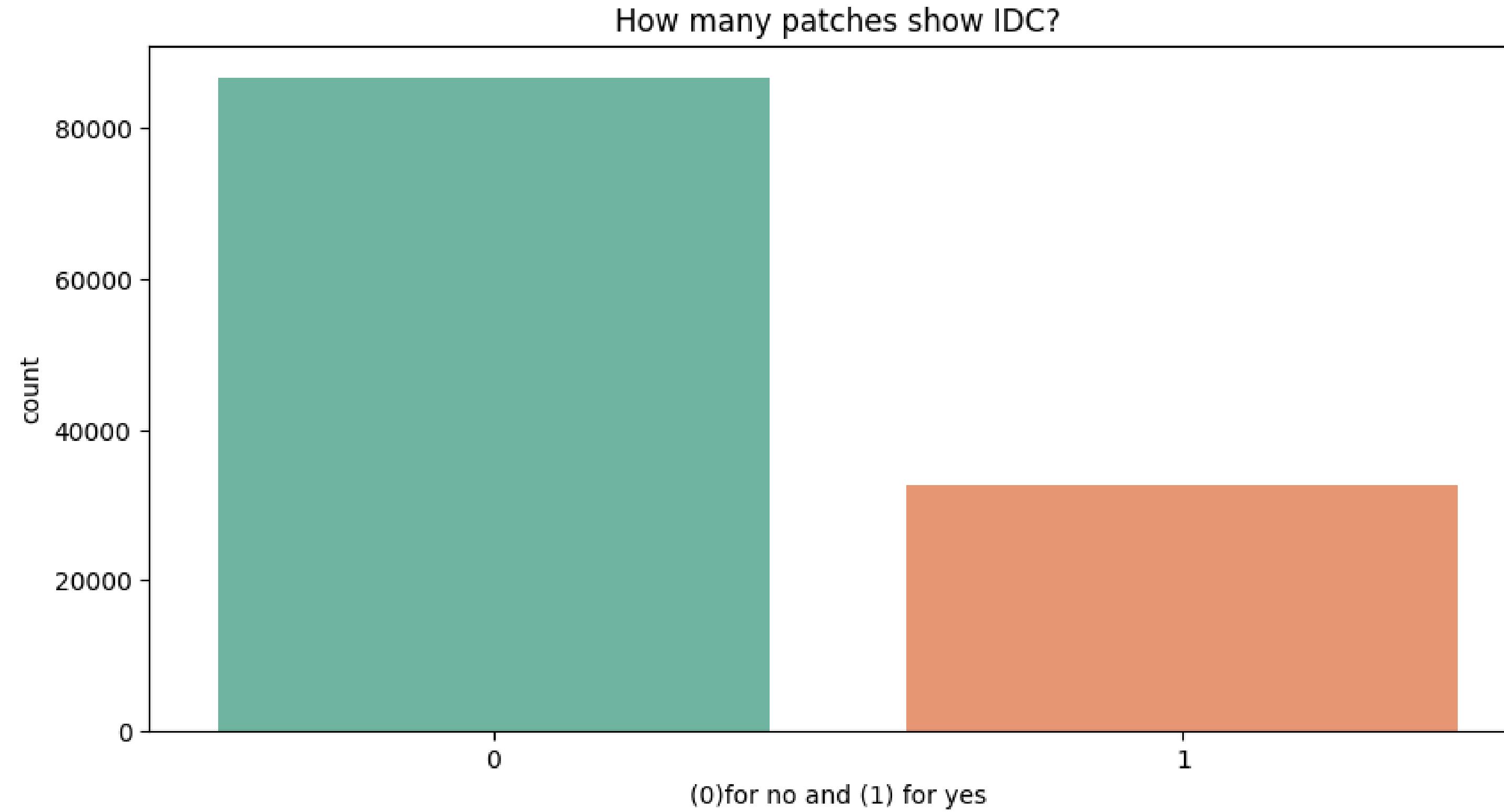


- IDC appear more purple & darker under Hematoxylin and Eosin (H&E) stain.

Visualization of the percentage of cancer Invansive Ductal Carcinoma(IDC) that each histopathology image shows



Visualization of the distribution of patches



Modelling

Preprocessing

- **Resizing:** Images are resized to a uniform size to ensure consistency in the input dimensions for the deep learning model.
- **Normalization:** Pixel values are normalized to a range of 0 to 1 to standardize the input data and facilitate model training.
- **Data Augmentation:** Techniques such as rotation, flipping, and zooming are applied to artificially expand the dataset and improve the model's robustness.

Modelling

MODEL	NUMBER OF EPOCHS	ACCURACY	LOSS
CNN	7 (with early stopping)	90.32%	0.18
RESNET50 MODEL	10	90%	0.33
EFFICIENTNET MODEL	8	80%	0.52
MOBILENET*	2	88%	0.30

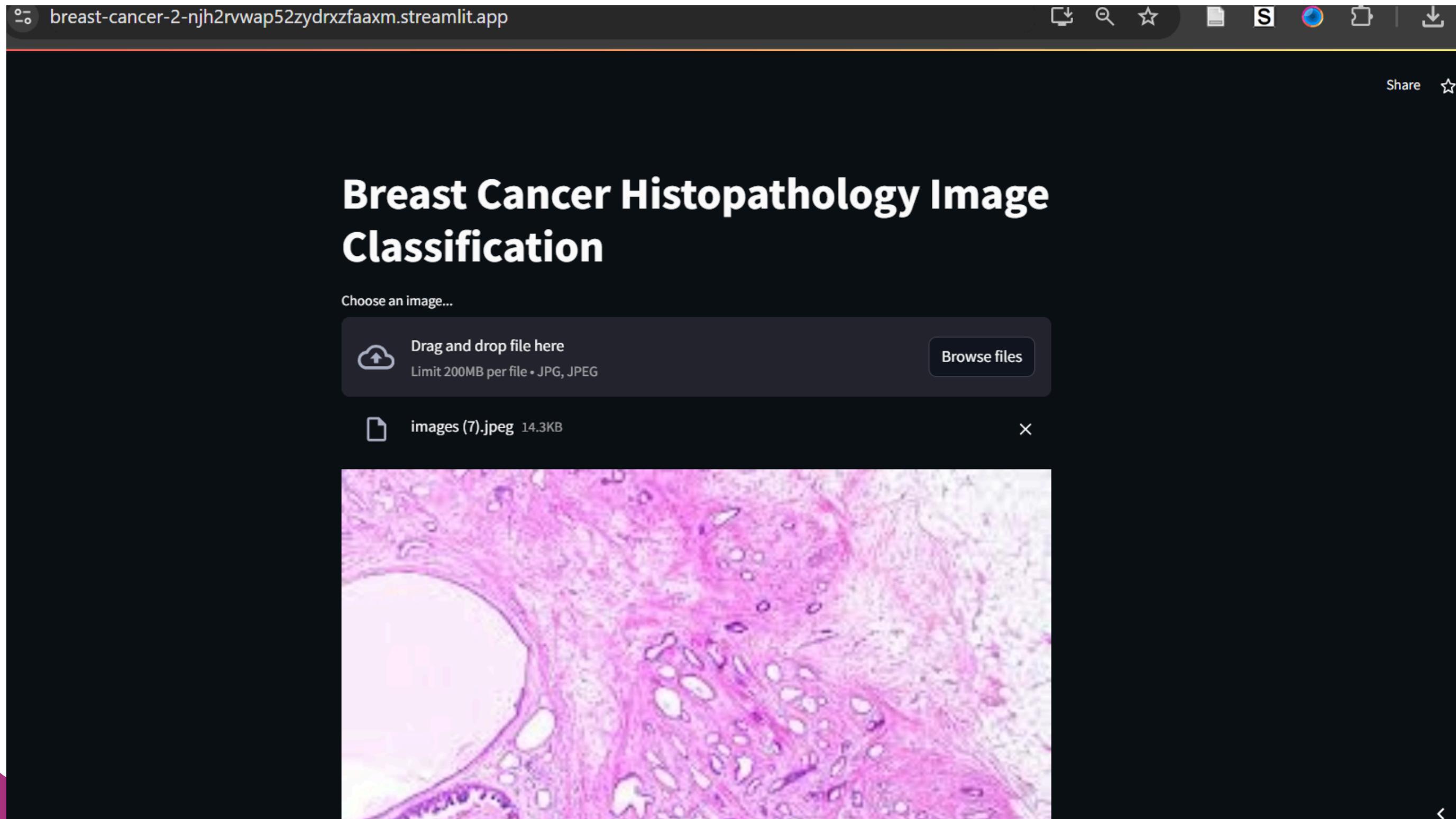
- Simple CNN was the best performing model
- *models were too computationally expensive to be completed

Real world Model Evaluation

- **Accuracy (90.32%)**: This is a strong metric, suggesting that the model correctly identifies the presence or absence of breast cancer in over 90% of cases. High accuracy is crucial for building confidence in the model's predictions.
- **Sensitivity (80.5%)**: i.e. the true positive rate/ recall, this metric indicates that the model correctly identifies 80.5% of actual positives (i.e., histopathology slides that show breast cancer). While this is relatively high, there's room for improvement, especially since higher sensitivity is critical in medical diagnostics to ensure cases are not missed.
- **Specificity (94.19%)**: This indicates that the model correctly identifies 94.19% of negatives (i.e., slides that do not show breast cancer). High specificity is essential to minimize false positives, which can reduce unnecessary additional testing and anxiety for patients.

Deployment

- This model was successfully deployed as an app hosted on **Streamlit Sharing** that classifies images as IDC or Non IDC



Limitations

- This model has a slightly low sensitivity of about 80% which means it may miss some images that are actually positive for IDC
- The model was trained on **high magnification image patches** and therefore **only works effectively in high magnification histology images.**

Recommendations & Future Work

- **Utility:** This model can be further improved; particularly when it comes to sensitivity. This can involve running more robust models on more powerful computers. Further powerful models that can accurately classify multiple histological subtypes should also be developed.
- **Validation & Regulatory approval:** Before real world deployment, a validation phase in a clinical setting through a clinical trial is necessary where the model's predictions are compared against traditional diagnostic outcomes. Adherence to data protection laws such as the Data Protection Act is also mandatory. Continuous monitoring and transparent reporting of the model's performance are required post-approval to ensure ongoing compliance and safety.
- **Practical Integration into Routine Clinical Workflow:** More groundwork need to be done in order to figure out the integration of the model into the standard diagnostic workflow of pathologists. This could involve using the model as a preliminary screening tool to prioritize cases or to double-check diagnoses, potentially streamlining workflows and reducing fatigue-related errors.

Conclusion

- In its current state, the model can be an a useful tool for assisting pathologists in making more accurate and efficient diagnoses, **provided it is used in conjunction with expert human assessment.**
- Further work should be done to improve upon this model as per previous recommendations given.

THE TEAM



Felix
Njoroge



Christine
Ndirangu



Frida
Oyucho



Winny
Chemusian



Sammy
Warah



Mataen
Surupai

THANK YOU

