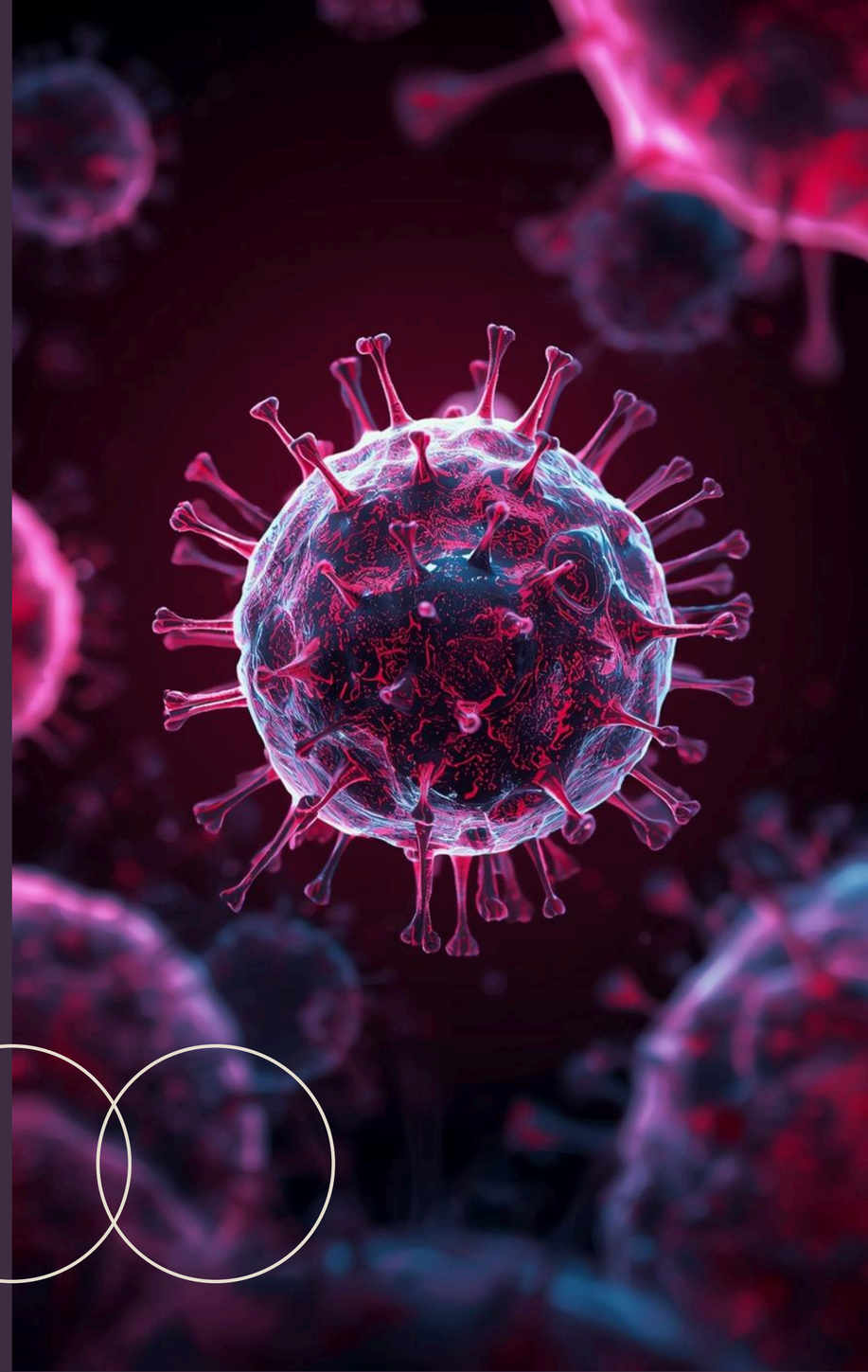
FIGHTING BREAST CANCER WITH DATA



Date: 05/11/2025

BUSINESS UNDERSTANDING

- Breast Cancer remains one of the leading causes of cancer related deaths among women worldwide.

- Early detection and accurate prediction of patient outcome are essential in improving survival rate and guiding treatment decisions.

PROBLEM STATEMENT

- Doctors often face challenges in diagnosing breast cancer early and accurately because manual image analysis is slow and prone to human error.
- There is a need for a data-driven tool that intergrate image and clinical information to improve diagnostic accuracy and support personalized treatment decisions.

PROJECT OBJECTIVE

- Develop a CNN model that classify breast cancer images as benign or malignant.
- Analyze clinical data to identify key factors affecting patient survival and treatment outcomes.
- Build a predictive machine learning models for survival and treatment effectiveness.
- Demostrate how integrated image-based & clinical models improve diagnosis and prognosis.

DATA UNDERSTANDING

This project uses 2 types of data:

- Tissue Images - Microscopic images of breast tissues.
- Patient Data - Clinical and genetic information from patient records.

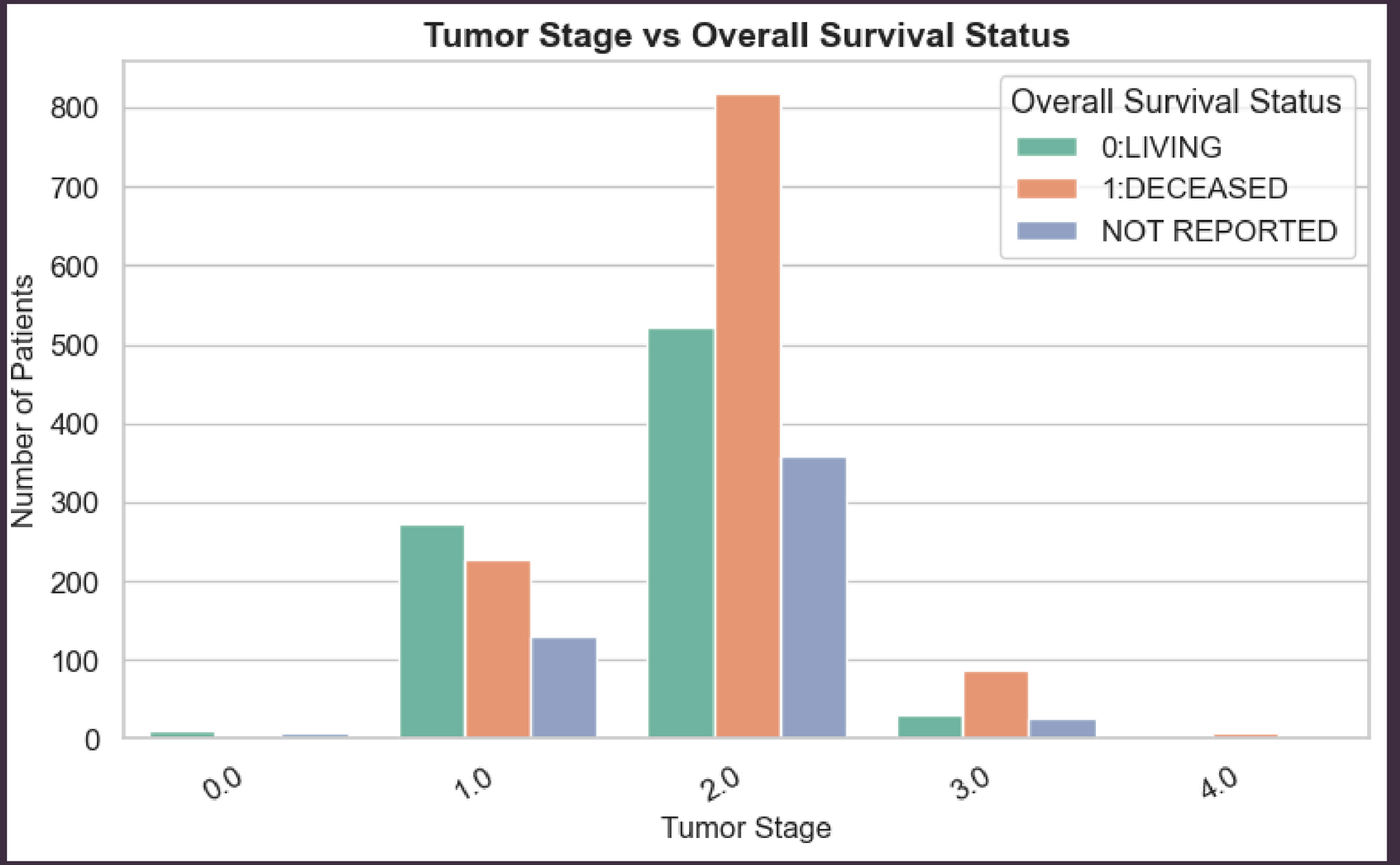
METHODOLOGY

IMAGE-BASED ANALYSIS

- Image preprocessing
- CNN model design
- Model training and validation
- Performance evaluation using accuracy, precision, recall, and confusion matrix

CLINICAL DATA ANALYSIS

- Data cleaning (handling missing values, encoding categorical features)
- Merging
- EDA and visualization
- Modeling and feature importance
- Model Evaluation


Tumor Stage vs Overall Survival Status
Overall Survival Status
0:LIVING
1:DECEASED
NOT REPORTED
Number of Patients
800
700
600
500
400
300
200
100
0
0.0
1.0
2.0
3.0
4.0
Tumor Stage

KEY INSIGHTS

1. Lower Tumor stages (0 and 1) have better survival rates compared to higher stages (2 and 3)
2. Patients with hormone receptor–positive tumors (ER+ and PR+) generally respond well to hormonal therapy and tend to have better survival.
3. Chemotherapy has a positive impact on survival
4. Older patients at diagnosis have poorer survival outcomes.
5. The CNN achieved approximately **80% test accuracy**, effectively distinguishing benign from malignant tissues.

CONCLUSION

1. Data and machine learning can work together to enhance breast cancer diagnosis and treatment.
2. The CNN model accurately detect malignant tumors, supporting earlier and more reliable diagnosis.
3. Tumor stage, age at diagnosis, type of treatment strongly influence survival outcomes
4. Combining image and clinical insights enables personalized, data-driven healthcare that can improve patients outcome

RECOMMENDATIONS

1. Adopt these data-driven tools in hospitals to help oncologists predict early detection and improve survival outcomes then recommend appropriate therapies
2. Clinicians can use the feature importance insight to identify key clinical factors & genes influencing survival.
3. Prioritize public awareness, screening programs & routine check-ups for women to promote early stage detection.
4. Combine CNN outputs with survival and treatment prediction models for multi-modal cancer analysis.

Thank you

NEURAL NEXUS

- Elsie Almadi
- Jeremy Koome
- Khalid Abdulkadir
- Tiffany Cheruto
- Meshack Mboya