Breast Cancer Detection Using Convolution Neural Networks

1st Jimmy Nguyen
Cal Poly, Pomona
Computer Science)
Pomona, California
jimmynguyen2@cpp.edu

2nd Edgar Ortiz
Cal Poly, Pomona)
Computer Science)
Pomona, California
edgaro@cpp.edu

3rd Zihao Luo
Cal Poly, Pomona
Computer Science
Pomona, California
zihaoluo@cpp.edu

4th Britney Collier
Cal Poly, Pomona
Computer Science
Pomona, California
collier@cpp.edu

5th Ethan Ngo
Cal Poly, Pomona
Computer Science
Pomona, California
eengo@cpp.edu

***Abstract*—Early and accurate detection of breast cancer is crucial for improving patient outcomes and survival rates. In this project, our objective is to develop a convolutional neural network (CNN) model capable of identifying malignant and benign breast cancer cases from histopathological images. Using a large-scale image dataset, we propose a deep learning pipeline involving preprocessing, training, evaluation, and interpretability of the model. The dataset is taken from kaggle called BreakHis. This project will demonstrate CNNs viability in medical diagnostics and compare model performance across various metrics. Looking to solve the world**

***Index Terms*—accuracy, precision, recall, F1-score, AUC-ROC**

I. Introduction

Breast cancer remains one of the leading causes of cancer-related deaths among women worldwide. With 1 in 8 women being diagnosed with breast cancer in the United States. Early diagnosis significantly increases the chances of successful treatment and survival. Traditional diagnostic methods rely on manual analysis of histopathology slides by pathologists, a process that is time-consuming and subject to human error. Motivated by the potential of deep learning in medical imaging, we propose the development of an automated system for breast cancer classification using convolutional neural networks (CNNs). Our model will be trained and validated on a publicly available histopathological image dataset to differentiate between benign and malignant tumors. We aim to address the challenge of achieving high classification accuracy while maintaining generalizability and interpretability. By leveraging CNN's capability in feature extraction, this project contributes to the growing field of AI-powered healthcare.

II. Dataset Details

A. Breast Cancer

We utilize the Breast Cancer Histopathological Image Classification (BreakHis) dataset, which contains 7,909 images

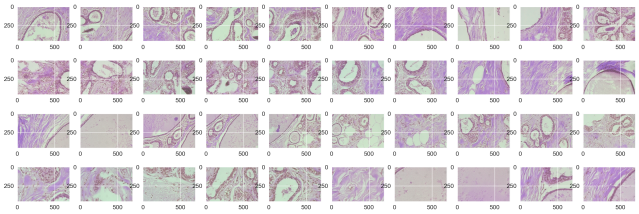

Fig. 1. Benign

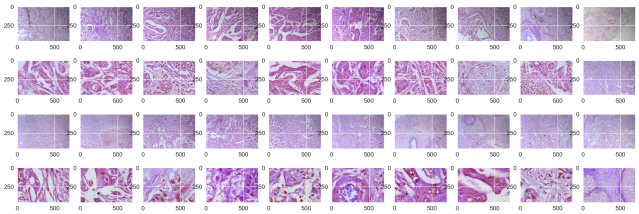

Fig. 2. Malignant

(700x460 pixels, 3-channel RGB, 8-bit PNG format). The dataset is divided into:

- Benign: 2,480 images
- Malignant: 5,429 images

Images are acquired at four magnification levels:

- 40X: 1,995 images (652 benign, 1,370 malignant)
- 100X: 2,081 images (644 benign, 1,437 malignant)
- 200X: 2,013 images (623 benign, 1,390 malignant)
- 400X: 1,820 images (588 benign, 1,232 malignant)

III. Methodologies

Our proposed methodology includes the following phases: Preprocessing:

- Begin by processing our images and label encoding them into a binary classification. 0 being benign and 1 being malignant.

Fig. 3. Unbalanced test set.

Fig. 4. Oversampled benign to balance test set.

- We also combined all our images based on Benign and Malignant instead of the type of tumor that the dataset originally categorized its dataset. This would have complicated our testing so we compressed our classes down to Benign or Malignant. The images labeled as types of benign breast tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA);and four malignant tumors (breast cancer): carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC).
- Split images into training, validation, and test set Model Architecture. With an 80 percent of Images used for training. Validation set was 10 percent of Images and Testing we allocated 10 percent of Images. We had to be careful during this phase due to problems with overfitting. When processing our data we found that we had an imbalance of malignant data showing to dominate our test set. To avoid overfitting, we over-sampled the minority class Benign.

  Training Strategy:
- Activation Function: RELU - Rectified Linear Unit The chosen activation function was used in our model to learn complex patterns. Due pixels we analyze in our images it help our model learn more complex patterns compared to linear types.
- Start with a 3 layer architecture to start off. This will help us determine our starting point in analyzing our model to determine our accuracy.
- A second architecture was chosen to fall back in-case we have a accuracies and precision less than 80 percent. Our fallback architecture we developed was a 4 layer model. Model:
- Using Tensorflow 2.10, Keras 2.10 we developed our model into 4 layers.
- To define the model we used Sequential API meaning we stacked our layers one after the other.

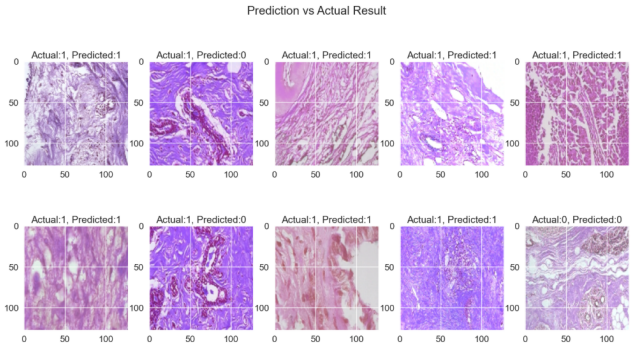

Fig. 5. Test set prediction trial.

- First Layer: 32 filters were set with a kernel size of 3x3 to scan the image to help set a base for basic patterns. Setting our activation to RELU. Setting our Max pooling size 2x2 using relu to avoid negative values and reducing the number of parameters and preventing overfitting. This helps us highlight our important features.
- Second Layer: 64 filters are set to cover more abstract features with a kernel size of 3x3 and RELU. Max Pooling layer set to prior parameters.
- Third Layer: 128 filters are set to further look into more complex details using the same kernel size of 3x3 with RELU. Max Pooling layer set to prior parameters.
- Fourth Layer: 256 filters are set as our final layer to further look into our image for missed details with a same kernel size 3x3 using RELU. Max Pooling layer set to prior parameters. Our final model design looked at features that would be missed with only 3 layers. After, running every layer we flatten our extracted features to be utilized for classification. This helps our dense layers to connect our layer and combine our extracted features which then leads to a sigmoid layer. Our sigmoid dense layer is important since it allows us to do binary classification.

IV. Results

Postprocessing phase we analyzed data and compared our different models to understand what model worked best. Results with a 4 layer Convolutional Neural Network we had the best results during our experimental phase. Relying on a fallback helped us decide on our direction of how we wanted to test our data. The final model allowed us to have an Accuracy of 89.3 percent which showed more promise compared to a 3 layer CNN. The 4 layer CNN showed less false negatives which helped us determine our final model. This is important due to False negatives having an impact in the Medical industry. Usually leading to a death sentence since a false negative putting a person at risk of dying due to undiagnosed cancer. the 4 layer model had 56 false negatives compared to 3 layer mode which had 250 false negatives.

When we originally tested with a 3 layer Convolutional Neural Network our training set lacked to show results in our

	precision	recall	f1-score	support
benign	0.31	0.22	0.26	496
malignant	0.68	0.77	0.72	1085
accuracy			0.60	1581
macro avg	0.49	0.50	0.49	1581
weighted avg	0.57	0.60	0.58	1581

Fig. 6. Confusion Matrix for 3 layer CNN.

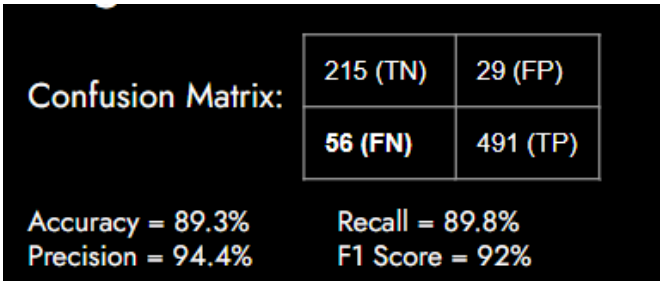

Fig. 7. Confusion Matrix for 4 layer CNN.

findings to classify images. 3 layers proved to be problematic due to its accuracy and precision being poor. Malignant precision and accuracy lead the team to agree on adding a fourth layer. Although, 3 layers would finish faster in terms of learning, but it had horrible results. This type of performance is important in the medical world due to peoples lives being at risk.

V. Experiments and Evaluation-Conclusion

In this work, we developed and evaluated a four-layer convolutional neural network for binary classification of histopathological breast cancer images from the BreakHis dataset. Our final model achieved an overall accuracy of 89.3 percent and a precision of 94.4 percent, substantially outperforming a simpler three-layer architecture, particularly by reducing false negatives from 250 to 56. These results demonstrate the model's strong predictive capability in distinguishing malignant from benign tissue. Moreover, consistent performance across training, validation, and test splits—despite varying magnification levels—underscores the network's generalizability to diverse histological samples. We further preserved interpretability by visualizing both correctly classified and misclassified patches, providing clinicians with intuitive insights into the network's decision process and fostering trust in potential AI-assisted diagnostics. We will evaluate model performance using the following metrics: Our proposed CNN model effectively differentiates between benign and malignant tumors.

- GitHub: https://github.com/EdgarO27/ProjectML/tree/main

References

[1] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: https://doi.org/10.1109/tbme.2015.2496264.

[2] buli Sai Naren, "Multi Cancer Dataset [Data set]" doi: https://doi.org/10.34740/KAGGLE/DSV/3415848

[3] li Saleem,Elahi Khan Khan Yaqoob Farooq Khattak Al-Rasheed 2023, "Breast cancer classification through meta-learning ensemble technique using Convolution Neural Networks" , volume=13, doi: 10.3390/diagnostics13132242,