Name	**Abdul Naafeh**
Roll Number	**DHC-673**
Subject	**AI/Machine Learning**
Date of Submission	**05/09/2025**

Cancer Detection Using Histopathological Images

Objective

The primary objective of this task was to understand the basics of image preprocessing and the training of Convolutional Neural Networks (CNNs) for cancer detection using histopathological images.

Dataset Preparation

A subset of the Breast Cancer Histopathological Image Dataset from Kaggle was used. To simplify the experiment, 500 images were selected, equally balanced between benign and malignant cases.

- The dataset was split into training (80%) and validation (20%) sets.
- This ensured a fair distribution of classes across both splits.

Data Preprocessing

The dataset underwent preprocessing to prepare images for model training:

- Resizing: All images were resized to 128×128 pixels for the CNN model. For transfer learning using VGG16, images were resized to 224×224 pixels.
- Normalization: Pixel values were scaled to the range [0,1], improving numerical stability and faster convergence during training.
- Data Augmentation (optional): Techniques such as random flipping, rotations, and zooming were applied to increase dataset variability and reduce overfitting.

Model Training

Two models were implemented and trained:

1. Simple CNN:

- A CNN with 3 convolutional layers was designed.
- ReLU activation was used for hidden layers, while the Sigmoid activation function was used in the output layer for binary classification.
- The model was trained for 10 epochs using the Adam optimizer and binary crossentropy loss function.

2. Transfer Learning with VGG16 (Bonus):

- A pre-trained VGG16 model was used as the base.
- The top classification layers were removed and replaced with custom dense layers.
- Initially, the base layers were frozen to train only the new dense layers.
- Later, fine-tuning was applied by unfreezing selected layers, improving performance significantly due to pre-trained ImageNet features.

Evaluation

The models were evaluated using accuracy and loss on the validation dataset:

- Training and validation plots were analyzed to check for overfitting and convergence.
- The simple CNN achieved reasonable accuracy but struggled with small dataset limitations.
- The VGG16 transfer learning model showed superior performance, confirming the benefit of pre-trained models in scenarios with limited training data.

Outcome

The outcome of this task was a basic CNN classifier capable of detecting cancer from histopathological images. A one-page evaluation report was prepared to summarize the dataset, preprocessing, model training, evaluation metrics, and results.

The VGG16-based transfer learning approach provided the best results, achieving higher validation accuracy and robustness compared to the simple CNN.