AI ASSIGNMENT PART 1: HEATHCARE FOCUS

Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

Hypothetical AI Problem:  
Predicting the Risk of Breast Cancer Based on Mammogram and Clinical Data

* Objectives:
* Identify patients at high risk of developing breast cancer.
* Support early diagnosis to improve treatment outcomes.
* Reduce false positives in cancer screening.
* Stakeholders:
* Oncologists and Radiologists
* Patients and Patient Advocacy Groups

Key Performance Indicator (KPI):

Diagnostic Accuracy – Proportion of correctly predicted cancer vs. non-cancer cases (sensitivity and specificity combined).

## 2. Data Collection & Preprocessing (8 points)

* Data Sources:
* Mammogram Image Datasets – e.g., Digital Database for Screening Mammography (DDSM).
* Clinical Records – Including family history, age, hormone levels, and previous biopsy results.

Potential Bias:

Data may overrepresent certain demographic groups (e.g., Caucasian women), leading to biased model performance for underrepresented populations (e.g., Black or Asian women).

* Preprocessing Steps:
* Image normalization and resizing – To ensure consistent input dimensions for deep learning models.
* Label encoding for categorical clinical features – Such as family history (yes/no), menopausal status, etc.
* Data augmentation for images – To artificially increase dataset size and improve model generalization (e.g., rotation, flipping, zoom).

## 3. Model Development (8 points)

Chosen Model:  
Convolutional Neural Network (CNN)

Justification:  
CNNs are highly effective in image recognition tasks and can detect subtle patterns in mammograms associated with cancer.

Data Splitting Strategy:

70% Training  
15% Validation  
15% Test  
Ensure data from the same patient doesn’t appear in multiple sets to avoid data leakage.

* Two Hyperparameters to Tune:
* Learning Rate – Controls how fast the model adjusts weights; critical for convergence.
* Number of Convolutional Layers – Affects the depth and ability to extract complex features from the images.

## 4. Evaluation & Deployment (8 points)

* Evaluation Metrics:
* Sensitivity (Recall) – Ensures the model captures most true positive cancer cases (critical for early detection).
* Specificity – Minimizes false positives, reducing unnecessary stress and further testing.

Concept Drift:

Definition: Over time, medical imaging techniques or population health trends may shift, causing model predictions to degrade.

Monitoring Strategy: Regular revalidation with new mammogram images and clinical data; track changes in accuracy and recall over time.

Technical Deployment Challenge:

Integration with Hospital Systems (EHRs & PACS) – Ensuring the AI model can access and process medical images and records securely, and present predictions in real-time within existing clinician workflows.