

Post Graduate Diploma in Management (Hospital & Health Management)

PGDM – 2024-2026 Batch

2nd Year – 4th Semester End Examination

Subject & Code : Artificial Intelligence in Healthcare-Integrating
Biomedical Data (AIHIBD)-HIT 711

Reg. No.:

Semester & Batch : IV, 2024-2026

Date : 02-03-2026

Time & Duration : 10:30 A.M.-01:30 P.M. (3 Hrs.)

Max. Marks : 70

Instructions:

- Budget your time as per the marks given for each question and write your answer accordingly.
 - Don't write anything on the Question Paper except writing your Registration No.
 - Mobile Phones are not allowed even for computations.
-

Part A: Q.1 to Q.10 all questions are compulsory (10 X 2 Marks = 20 Marks)
One liner, MCQs, True/False

Q1. Name one Python library commonly used for numerical data analysis.

- A) matplotlib
- B) numpy
- C) tensorflow
- D) requests

Q2. What is the primary purpose of model evaluation in a clinical AI application?

- A) To make the AI model run faster.
- B) To assess the model's performance, accuracy, and reliability on unseen data
- C) To write the code for the AI algorithm.
- D) To collect more data from patients.

Q3. Which of the following is a valid Python identifier?

- A) my_var
- B) ay-var
- C) 2var
- D) my var

Q4. What is meant by "Bias" in an AI model?

- A) The model is too slow to make predictions.
- B) The model is very accurate for all types of patients.
- C) The model makes systematic errors, often favoring or disfavoring specific groups of people.
- D) The model requires a lot of memory to run.

Q5. Which of the following is an example of unstructured healthcare data?

- A) Blood Pressure reading (120/80)
- B) Patient's age in a database (45 years)
- C) A doctor's clinical note describing a patient's symptoms
- D) Body Temperature (98.6°F)

Contd...2..

- Q6. A hospital wants to use AI to find 'new clusters' of patients who respond similarly to a drug, without using existing diagnostic labels. Which approach should they use?
- A. Model Validation
 - B. Reinforcement Learning
 - C. Supervised Learning
 - D. Unsupervised Learning
- Q7. Apply the principles of 'Reinforcement Learning' to a clinical scenario. Which of the following fits best?
- A. Grouping patients into clusters based on similar genetic traits.
 - B. Predicting if a tumor is malignant based on past labeled X-rays.
 - C. Training an AI to find the optimal insulin dosage through a system of "rewards" for stable blood sugar levels.
 - D. Cleaning a database of duplicate patient records.
- Q8. Predict the outcome of using a diagnostic AI with high Sensitivity but low Specificity in a screening clinic:
- A. The AI will miss many sick patients.
 - B. The AI will catch most sick patients but will also generate many "false alarms" (false positives).
 - C. The AI will be 100% accurate for all patients.
 - D. The AI will fail to function on unstructured data.
- Q9. What is the output of
- ```
>>> item = ('Harihar',)
>>> print(type(item))
```
- Q10. The output of the python script below is same: TRUE or FALSE
- ```
print(cities[:])and print(cities)
```

Part B: Q.11 to Q.15 attempt any four questions (4 X 5 Marks = 20 Marks) - Short Notes

- Q11. A rural health clinic in India is implementing an AI tool to screen for diabetic retinopathy using smartphone images of the retina. The original model was trained on high-resolution, perfectly lit images from a multi-specialty hospital in Bengaluru. When deployed in the rural clinic, the AI's accuracy drops significantly. Demonstrate the data preparation and clinical validation steps the clinic must take to "fine-tune" this model for their specific environment.
- Q12. An ICU uses two different AI models to manage sepsis.
- Model A is a Supervised Learning algorithm that predicts the likelihood of sepsis based on historical EHR data.
 - Model B is a Reinforcement Learning algorithm that suggests real-time adjustments to intravenous fluid dosages.
- Differentiate between these two models in terms of their input-output relationship and their "Clinical Utility." Attribute which model carries a higher ethical risk regarding patient safety and explain why.
- Q13. A tech startup launches a "Direct-to-Consumer" Mental Health Chatbot. To improve the bot's empathy, the company decides to train its Large Language Model (LLM) on anonymized transcripts of private user therapy sessions without explicit renewed consent, arguing that the

data is "de-identified" and serves the "greater good" of mental health. Critique this company's decision based on the AI Ethics & Governance frameworks (GDPR or India AI Guidelines). Justify whether this practice meets the standard of "Responsible AI." Provide two specific recommendations that the company should have followed instead.

Q14. A hospital trying to separate the line based on age of patient from entry gate itself. Write a python script to divide the line based on patient's age (Adult versus Geriatric) in the hospital.

Q15. In a firm the claim processing is decided on a coverage limit of 50000. Write a python script to approve or reject the claim based on the coverage limit.

Part C: Q.16 to Q.19 attempt any three questions (3 X 10 Marks = 30 Marks) - Long Notes

Q16. What are the challenges of using raw, unstructured healthcare data (like doctor's notes or medical images) directly in AI models? Describe the key steps involved in "preparing biological or healthcare data for AI" (Data Preprocessing).

Q17. A tertiary hospital is developing a clinical decision-support tool using the ID3 (Iterative Dichotomiser 3) algorithm to predict whether an ICU patient is at High Risk of Sepsis (Yes/No) within 24 hours of admission.

The model is built using the following four categorical attributes:

1. Temperature (High $\geq 38^{\circ}\text{C}$ / Normal $< 38^{\circ}\text{C}$)
2. Heart Rate (High ≥ 100 bpm / Normal < 100 bpm)
3. WBC Count (High $> 12,000$ / Normal $\leq 12,000$)
4. Systolic Blood Pressure (Low < 90 mmHg / Normal ≥ 90 mmHg)

A new ICU patient comes with Temperature = 38.7°C (High); Heart Rate = 105 bpm (High); WBC Count = 15,000 (High); Systolic BP = 88 mmHg (Low)

1. Using ID3 algorithm select which is better as root node WBC count or Temperature? (4 marks)
2. Evaluate the classification outcome for the new patient by systematically tracing the ID3 decision tree. Clearly justify each decision branch selected. (3 marks)
3. The hospital reports that during validation, the ID3 model misclassified 3 septic patients as "Low Risk." Critically appraise the potential clinical and ethical consequences of such false-negative errors in an ICU environment. (3 marks)

Q18. A tertiary oncology center is using the K-Means clustering algorithm to group breast cancer patients into risk categories to personalize treatment planning.

The dataset includes the following four attributes:

1. Tumour Size (cm)
2. Uniformity Score (1–10)
3. Lymph Node Involvement (Number of nodes)
4. Ki-67 Proliferation Index (%)

The hospital sets $K = 3$, aiming to identify:

- Cluster 1 $\rightarrow$ Low Aggressiveness
- Cluster 2 $\rightarrow$ Moderate Aggressiveness
- Cluster 3 $\rightarrow$ High Aggressiveness

After running K-Means, the centroids are:

Cluster	Tumour Size	Uniformity	Lymph Nodes	Ki-67 (%)
C1	1.5	3	0	10
C2	3.0	6	2	25
C3	5.5	8	5	60

A new patient presents with:

- Tumour Size = 5.0 cm
- Uniformity = 7
- Lymph Nodes = 4
- Ki-67 = 55%
- Evaluate which cluster the new patient most likely belongs to by comparing the clinical proximity to the cluster centroids. Justify your reasoning using the attribute values. (4 marks)
- The oncology team plans to use cluster membership to guide chemotherapy intensity. Critically appraise the potential clinical risks of relying solely on unsupervised clustering for treatment decisions. (3 marks)
- Assess whether choosing K = 3 is appropriate for stratifying cancer aggressiveness in this dataset. Provide two arguments to support your judgement. (3 marks)

Q19. A hospital data is available in csv format. Write a python scripts for the following:

- To import the data into python environment
- Describe the data variables (headings)
- If the variables in the data are:

#	Column	Non-Null Count	Dtype
0	id	500 non-null	int64
1	age	500 non-null	int64
2	height	498 non-null	float64
3	smoke	500 non-null	str
4	gender	500 non-null	str
5	caesarean	499 non-null	str
6	residence	500 non-null	str
7	disease	500 non-null	int64
8	activity	500 non-null	int64
9	birthweight	498 non-null	float64

dtypes: float64(2), int64(4), str(4)

Write a python script to give basic statistical tables: frequency tables, % frequency and their corresponding numbers.

- Write a python script to calculate measures of central tendency for continuous variables like age, height and birthweight.