



ARTIFICIAL INTELLIGENCE

PAPER 1

(THEORY)

Maximum Marks: 70

Time Allotted: Three Hours

Reading Time: Additional Fifteen Minutes

Instructions to Candidates

1. You are allowed an **additional fifteen minutes** for **only** reading the question paper.
2. You must **NOT** start writing during reading time.
3. This question paper has **9 printed pages**.
4. There are twelve questions in the paper. All questions are compulsory.
5. This paper is divided into **three sections: A, B and C**.
6. **Section A** has **fourteen subparts** which are very short answer questions. Each question carries 1 mark.
7. While attempting **Multiple Choice Questions** in Section A, you are required to **write only ONE option as the answer**.
8. **Section B** has **seven questions** which are short answer questions. Each question carries 4 marks.
9. **Section C** has **four questions** which are long answer questions. Each question carries 7 marks.
10. **Internal choices** have been provided in **three questions in Section B** and in **two questions in Section C**.
11. The intended marks for questions are given in brackets [].

Instruction to Supervising Examiner

Kindly read **aloud** the Instructions given above to all the candidates present in the examination hall.

Note: The Specimen Question Paper in the subject provides a realistic format of the Board Examination Question Paper and should be used as a practice tool. The questions for the Board Examination can be set from any part of the syllabus, though the format of the Board Examination Question Paper will remain the same as that of the Specimen Question Paper. The weightage allocated to various topics, as given in the syllabus, will be strictly adhered to.

SECTION A – 14 MARKS

Question 1

- (i) Which of the following is an example of sentiment analysis? **(Understand)** [1]
- (a) Translating reviews from Hindi to English
 - (b) Grouping similar reviews into clusters without labels
 - (c) Extracting names of products and brands from reviews
 - (d) Detecting whether a review is positive, negative, or neutral
- (ii) Choose the statements that justify the preference for HSV / HSL colour model over RGB for colour detection. **(Understand)** [1]
- (P) Lightness and value are not considered.
 - (Q) Saturation is separated from brightness.
 - (R) Hue is separated from brightness.
- (a) Only (P) and (Q)
 - (b) Only (Q) and (R)
 - (c) Only (P) and (R)
 - (d) (P), (Q) and (R)
- (iii) The learning paradigm which involves an AI system, learning to maximise long-term rewards through a sequence of decisions is: **(Recall)** [1]
- (a) Deep learning
 - (b) Supervised learning
 - (c) Unsupervised learning
 - (d) Reinforcement learning
- (iv) Given below are two statements marked Assertion and Reason. Read the statements carefully and choose the correct option. [1]
- Assertion:** The ReLU activation function is preferred to sigmoid in the hidden layers of a deep neural network.
- Reason:** For large positive or negative inputs, the sigmoid function produces near zero derivatives, causing weight updates in early layers to become negligibly small, a problem known as the vanishing gradient. **(Analyse)**
- (a) Both Assertion and Reason are true, and Reason is the correct explanation for the Assertion.
 - (b) Both Assertion and Reason are true, but Reason is not the correct explanation for the Assertion.
 - (c) Assertion is true and Reason is false.
 - (d) Both Assertion and Reason are false.

- (v) A face-recognition system at an airport achieves 99% overall accuracy but only 70% accuracy for passengers with different skin tones. [1]

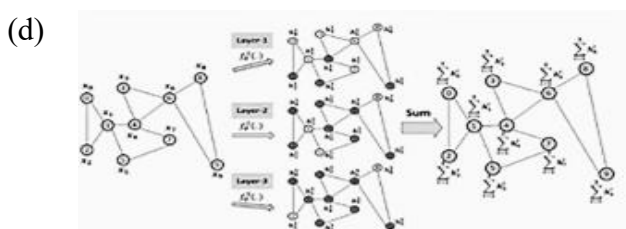
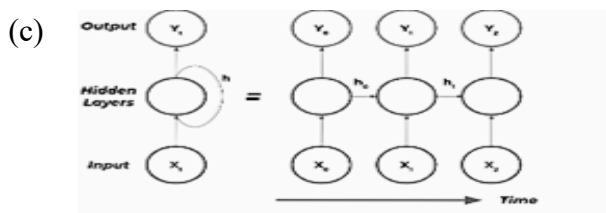
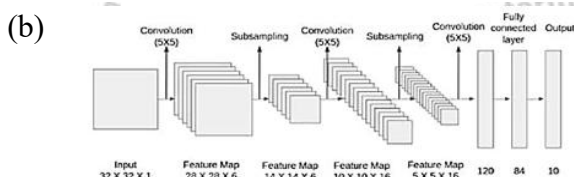
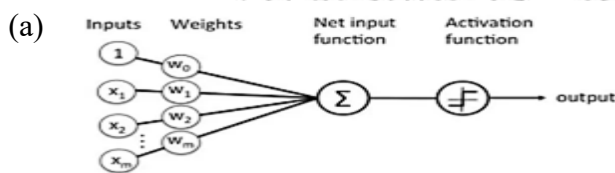
Which one of the following is **MOST** likely the cause for this bias? (Evaluate)

- (a) Insufficient computing power at the airport
- (b) Faulty algorithm that cannot model non-linear facial features
- (c) Under-represented training data, embedding demographic bias
- (d) Overfitting to the training data across all demographic groups
- (vi) Given below are two statements marked Assertion and Reason. Read the statements carefully and choose the correct option. [1]

Assertion: Medical diagnosis systems based on Machine Learning can improve the efficiency of healthcare services.

Reason: Genetic information is one of the sources of data that can be analysed by Machine Learning models. (Analyse)

- (a) Both Assertion and Reason are true, and Reason is the correct explanation for the Assertion.
- (b) Both Assertion and Reason are true, but Reason is not the correct explanation for the Assertion.
- (c) Assertion is true and Reason is false.
- (d) Both Assertion and Reason are false.
- (vii) Which of the following Neural Network (NN) architectures is suited for image classification tasks? (Understand) [1]



- (viii) A neural network used for recognising handwritten digits contains an input layer, one or more hidden layers and an output layer. The hidden layers transform the input information before the final prediction is made. [1]
What is the primary role of the hidden layers in such a neural network? (Understand)
- (a) To display the final predicted class
 - (b) To divide the dataset into training and testing sets
 - (c) To store the original input data without modification
 - (d) To learn and transform intermediate features from the input data
- (ix) A bank uses a Machine Learning system to identify unusual transaction patterns that may indicate identity theft. Which practical implication of Machine Learning is evident here? [1]
(Understand)
- (a) Fraud Detection
 - (b) Medical Diagnosis
 - (c) Workflow automation
 - (d) Predictive Maintenance
- (x) The Computer Vision technique that can be used to count the number of people in a photograph is: [1]
(Recall)
- (a) Object tracking
 - (b) Object detection
 - (c) Facial recognition
 - (d) Image classification
- (xi) Name the Deep Learning model commonly used for image classification. (Recall) [1]
- (xii) A machine learning model incorrectly classifies a genuine credit card transaction as fraudulent. Name the type of prediction made by the model. (Recall) [1]
- (xiii) What is the purpose of the **fit()** function while training an Artificial Neural Network in Keras? (Understand) [1]
- (xiv) A hospital uses Machine Learning to analyse patient records for disease prediction. Why should the model be updated periodically with new patient data? (Understand) [1]

SECTION B – 28 MARKS

Question 2

- (i) AgroAI Pvt. Ltd., a smart farming company, collects data such as soil moisture, temperature, rainfall, and Crop Health Index (CHI) to determine the time of irrigation. [2]
How can a regression model help AgroAI Pvt. Ltd. decide the time of irrigation? (Apply)
- (ii) A data scientist designs a neural network consisting of 30 input neurons, 12 hidden neurons and 1 output neuron. Calculate the total number of weights and biases required for this Neural Network. (Apply) [2]

Question 3

Read the given passage carefully and answer the questions that follow.

An automobile manufacturer installed sensors in its assembly-line machines. The sensors continuously recorded temperature, vibration levels, and operating speed. For several months, Machine Learning algorithms analysed this data and identified patterns that occurred before a machine failure. As a result, maintenance teams were alerted in advance, reducing downtime by 40%.

- (i) Which practical implication of Machine Learning is described above? (Recall) [1]
- (ii) Illustrate *two* benefits of using the above-mentioned AI driven approach for the automobile manufacturer. (Understand) [2]
- (iii) Explain the importance of historical data for the above machine learning system. (Understand) [1]

Question 4

Attempt either subparts (i) and (ii) or (iii) and (iv).

- (i) Explain the process of Backpropagation in an Artificial Neural Network. How does it help in updating weights to reduce prediction error? (Understand) [3]
 - (ii) A health-tech start-up uses a pre-trained Convolutional Neural Network to classify chest X-ray images into disease categories.
Name the appropriate AI learning paradigm used in this approach. (Recall) [1]
- OR**
- (iii) Explain the process of Forward Propagation in an Artificial Neural Network. How does information move from the input layer to the output layer? (Understand) [3]
 - (iv) What is a *Multi-Layer Perceptron* (MLP)? (Recall) [1]

Question 5

Attempt either subparts (i) and (ii) or (iii) and (iv).

- (i) Explain the importance of splitting a dataset into *training* and *testing* sets before evaluating a model. (Understand) [3]
 - (ii) Why is scaling necessary during data preparation of a neural network? (Understand) [1]
- OR**
- (iii) Explain the roles of *epochs* and *batch size* while training an Artificial Neural Network. (Understand) [2]
 - (iv) State one advantage each of using *Grid Search* and *Random Search* for hyperparameter tuning. (Understand) [2]

Question 6**[4]**

A health-tech start-up is building an AI platform to support doctors in diagnosing lung diseases from chest X-rays and clinical data. They consider three approaches:

Approach 1: Training a decision-tree classifier on tabular clinical variables (age, symptoms, lab tests, comorbidities) to predict disease labels such as “Pneumonia”, “TB”, “COVID-19”, “Normal”.

Approach 2: Using a pre-trained Convolutional Neural Network (CNN) to extract features from 224×224 pixel X-ray images followed by a small fully connected classification head on top for the same four labels.

Approach 3: Clustering historical patient records into groups based on similarity of clinical variables and imaging features, without using any labels, to explore unknown patterns in the data.

Identify the AI learning paradigm used in each of the above approaches. Justify your answer.

(Analyse)

Question 7

- (i) A bank uses an AI system to detect fraudulent credit card transactions. During a test run with 1000 transactions, the AI model flagged 120 transactions as fraud, of which 80 were truly fraud. The actual number of frauds was 100.

(a) Compute precision and recall for the above AI model.

(Apply)

[2]

(b) Explain why both recall and precision are important for Fraud Detection.

[2]

(Understand)

OR

- (ii) A Neural Network architecture was built to perform a medical diagnosis task. Following values were obtained during the classification task:

Recall = 89%, True Positive = 80, True Negative = 30 and False Positive = 20.

(a) Calculate the approximate accuracy.

(Apply)

[3]

(b) State a limitation of using accuracy as the only evaluation metric in an imbalanced data set.

(Understand)

[1]**Question 8**

- (i) An AI-assisted medical diagnosis system analyses Electrocardiograms (ECG) to detect heart attacks. The ECG of a patient who had suffered a minor heart attack showed small abnormalities. The AI system analysed the ECG and concluded that it showed ‘no signs of heart attack’.

[2]

Identify the type of error made by the AI system. State *two* possible consequences of such erroneous identification by AI system.

(Understand)

- (ii) Explain how AI can be helpful in identifying rare genetic diseases.

(Understand)

[2]

SECTION C – 28 MARKS

Question 9

- (i) Write a Python (Keras) snippet that defines and compiles a binary-classification ANN suitable for a Medical Diagnosis task, with: [3]

- Input feature dimensions
- Two hidden layers of sizes 64 and 32 with ReLU, and
- A sigmoid output.

Use an appropriate loss function and metric.

(Create)

- (ii) A company is deploying an ANN-based model for fraud detection in credit-card transactions. Each transaction is represented by numerical features (amount, time since last transaction, merchant risk score) and categorical features (merchant category, country, device type). The final network has:

- An input layer with appropriate encodings for all features,
- Two hidden layers with ReLU activation, and
- A single sigmoid output neuron giving the probability that a transaction is fraudulent.

- (a) Explain how the categorical features *merchant category*, *country* and *device type* should be encoded. How should the numerical features be scaled before being supplied to the ANN? [2]

(Understand)

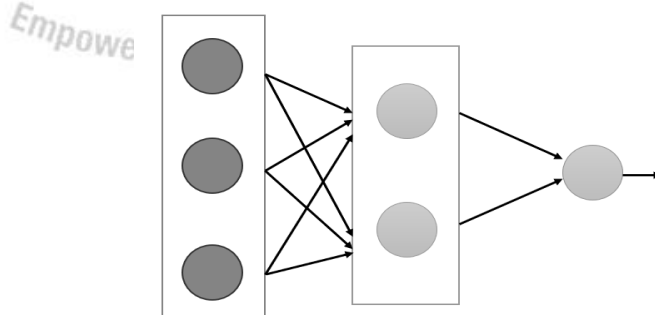
- (b) How do these preprocessing steps improve the training stability and performance of the given ANN? [2]

(Understand)

Question 10

- (i) Study the architecture of the simple Neural Network given below and write a Python program (using Keras) to build the given Neural Network architecture. [2]

(Create)



- (ii) The training accuracy of an ANN was estimated to be 98%, but the test accuracy was only 78%. [1]

Explain the concept of overfitting in the above context.

(Understand)

- (iii) A hospital develops a Machine Learning Model for disease diagnosis.

- (a) State *two* benefits of using Machine Learning in medical diagnosis. [2]
- (b) State *two* factors that may reduce the accuracy of the model in real-world use. [2]

(Understand)

Question 11

Attempt either subpart (i) and (ii) or (iii) and (iv)

- (i) Study the average response time (in seconds) of a web-based AI service for seven days given below and answer the questions based on the data.

Day	Average Response Time (in seconds)
Monday	80
Tuesday	85
Wednesday	90
Thursday	110
Friday	95
Saturday	85
Sunday	400

- (a) Calculate the *mean* and *median* response time for web-based AI service. Comment on the effect of the outlier on statistical measures. **[3]**
(Evaluate)
- (b) Compute *z-score* of response time 400 and interpret it. Assume approximate standard deviation of 110 for the data set. **[1]**
(Evaluate)
- (ii) A retail company wants to study the relationship between the amount spent on advertising and the monthly sales of its products.
- (a) State the purpose of using the Correlation Coefficient in this situation. **[1]**
(Understand)
- (b) Interpret the meanings of each of the correlation coefficients in reference to the above case: +0.92 and -0.78. **[2]**
(Analyse)

OR

- (iii) A pharmaceutical company has developed a new AI-assisted disease diagnosis system and wants to determine whether it performs significantly better than the existing system.
- (a) State the Null Hypothesis (H_0) and Alternative Hypothesis (H_1) for this study. **[2]**
(Understand)
- (b) Differentiate between Type I Error and Type II Error with *one* example each in the given context. **[2]**
(Analyse)
- (iv) The average response time of an AI driven grievance redressal is between 90 seconds to 3 minutes. If the response time exceeds 3 minutes, the system is tagged as '*slow responding system*'. The probability of the system being tagged as '*slow responding system*' for any given day is 0.1. **[3]**
- Find the probability that in 5 days, slow responses occur on exactly 2 days. Assume independence between days. **(Apply)**

Question 12

Attempt either subparts (i), (ii) and (iii) or (iv) and (v).

- (i) A traffic surveillance system has to identify different vehicles in an image and determine the location of each vehicle. [2]

Which Computer Vision technique would be most suitable for this task? Explain why Image Classification alone would not be sufficient. (Apply)

- (ii) In OpenCV, the function `cv2.bitwise_and()` is applied to two images. State the condition that the two images must satisfy before this operation can be performed successfully. Also, state the data type commonly used to store HSV colour ranges such as `lower_red` and `upper_red`. (Analyse) [2]

- (iii) (a) State *one* application of Facial Recognition. (Recall) [1]

- (b) Explain *one* advantage of Object Tracking over Image Classification. (Understand) [1]

- (c) Name *one* Python library commonly used for implementing Computer Vision applications. (Recall) [1]

OR

- (iv) A school wants to improve the attendance monitoring system and effectively manage the movement of students in the school premises. The management of the school decides to introduce an AI powered camera-based system to count and identify students.

- (a) Name the computer Vision technique that can be used to locate any student. (Understand) [1]

- (b) Explain how the above technique differs from image classification. (Recall) [1]

- (c) The system has difficulty identifying students when the lighting conditions change significantly. Explain why the HSV colour model may be more suitable than RGB for colour-based detection under varying brightness conditions. (Analyse) [2]

- (v) `import cv2` # line number 1
`import numpy as np` # line number 2
`img = cv2.imread('fruits.jpg')` # line number 3
`hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)` # line number 4
`lower_red =` # line number 5
`upper_red =` # line number 6
`mask = cv2.inRange(hsv, lower_red, upper_red)` # line number 7
`result = cv2.bitwise_and(img, img, mask=mask)` # line number 8
`cv2.imshow('Red Objects', result)` # line number 9
`cv2.waitKey(0)` # line number 10
`cv2.destroyAllWindows()` #line number 11

- (a) Write the codes for lines 5 and 6. (Create) [1]

- (b) What is the purpose of the `cv2.inRange()` function in line 7 of the given code? (Understand) [1]

- (c) In line number 8, what will happen if two different images are passed to `cv2.bitwise_and()`? (Analyse) [1]

ARTIFICIAL INTELLIGENCE

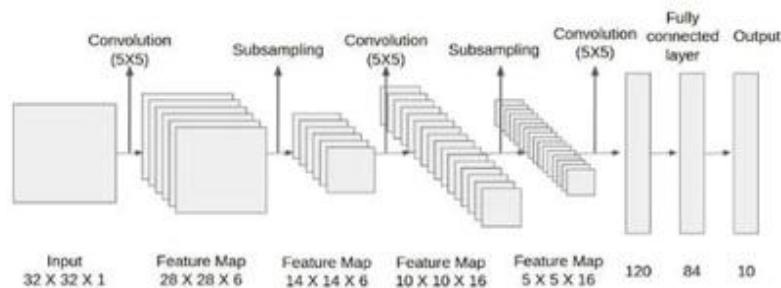
PAPER 1 (THEORY)

ANSWER KEY

SECTION A – 14 MARKS

Question 1

- (i) (d) Detecting whether a review is positive, negative, or neutral [1]
- (ii) (b) Only (Q) and (R) [1]
- (iii) (d) Reinforcement learning [1]
- (iv) (a) Both Assertion and Reason are true, and Reason is the correct explanation for the Assertion. [1]
- (v) (c) Under-represented training data, embedding demographic bias. [1]
- (vi) (b) Both Assertion and Reason are true, but Reason is not the correct explanation for the Assertion. [1]
- (vii) (b) [1]



- (viii) (d) To learn and transform intermediate features from the input data [1]
- (ix) (a) Fraud Detection [1]
- (x) (b) Object Detection [1]
- (xi) Convolutional Neural Network (CNN) [1]
- (xii) False Positive (FP) [1]
- (xiii) The **fit()** function trains the neural network on the given training data by adjusting its weights to reduce prediction error. [1]
- (xiv) It should be updated to adapt the model to new patterns in recent data and maintain prediction accuracy. [1]

SECTION B – 28 MARKS

Question 2

- (i) A regression model can analyse the collected data and Crop Health Index (CHI) and fit the data to appropriate functions and parameters facilitating predictions of crop conditions. Based on the predicted value, AgroAI Pvt. Ltd. can decide the time of irrigation. [2]
- (ii) Number of weights = $(30 \times 12) + (12 \times 1) = 372$ [2]
Number of biases = $12 + 1 = 13$
Total parameters = $372 + 13 = 385$
Answer: 385 parameters (372 weights + 13 biases)

Question 3

- (i) Predictive maintenance [1]
- (ii) Benefits of using AI driven approach for the automobile manufacturer: [2]
- Real time monitoring and predictive analytics which reduces downtime.
 - The AI driven automation reduces the need for humans to do dangerous tasks thus increasing safety.
 - Real time data optimises workflows.
 - It reduces maintenance costs.
- (iii) Historical sensor data provides past patterns of machine operation and failures. The Machine Learning model learns these patterns to predict future machine failures reducing downtime. [1]

Question 4

- (i) • Backpropagation is the process by which an Artificial Neural Network learns from its errors. After forward propagation produces an output, the network calculates the prediction error (loss) by comparing the predicted output with the actual output. [3]
- This error is then propagated backward from the output layer toward the input layer, and the contribution of each weight to the error is determined using gradients.
 - The weights and biases are then adjusted, usually using an optimization method such as gradient descent, so that the loss is reduced. Repeating this process over many training cycles gradually improves the accuracy of the Artificial Neural Network.
- (ii) Supervised Learning / Supervised Deep Learning [1]

OR

- (iii)
 - Forward propagation is the process in which input data moves sequentially through the neural network to produce an output. It transforms raw inputs into predictions using weights, biases and activation functions.
 - The Input Layer receives raw data where each feature corresponds to a neuron.
 - The processed input passes through one or more hidden layers where most of the computation takes place. Each neuron performs a weighted sum of inputs and applies an activation function to capture non-linear patterns.
 - Then the data is passed forward to the output layer which generates the model's prediction [3]
- (iv) A Multi-Layer Perceptron (MLP) is a feed-forward neural network consisting of an input layer, one or more hidden layers, and an output layer. [1]

Question 5

- (i) Splitting a data set aims to ensure that the model not only focuses on learning from a specific data set but also performs well when faced with new and unseen data set. [3]
- The training set guides the model during learning process and provides the fundamental data needed to optimise a model's parameter.
 - The test set is used in the final evaluation phase of the model. The performance of the model on this data reflects how well the model is likely to perform in a real-world scenario.
- (ii) Scaling is important because it puts all the data columns into the same numerical range. This helps the model to train faster and also ensures equitable weight distribution for every input. [1]
- OR
- (iii)
 - **Epoch:** It means one complete pass through the entire training dataset during model training. Finding the right numbers of epochs is crucial for achieving good model performance without overfitting. The suitable number of epochs dictates the learning duration and improves accuracy of the model.
 - **Batch Size:** It is the number of training samples processed before the model updates its weights. It plays a crucial role in the training dynamics of a machine learning model by affecting the computational efficiency, training speed and training stability of the model. [2]
- (iv) Advantage of using Grid Search: [2]
- It systematically explores different combinations of hyperparameter values to identify a suitable configuration.
- Advantage of using Random Search:
- It can improve model performance by helping select better hyperparameters such as learning rate, batch size, or number of hidden layers.

Question 6**[4]**

- **Approach 1 – Supervised Learning:** The model is trained using labelled clinical data and known disease outcomes to predict disease classes.
- **Approach 2 – Deep Learning:** A pre-trained CNN extracts features from X-ray images and uses them for classification. CNNs are commonly used for image-based tasks.
- **Approach 3 – Unsupervised Learning:** The model groups patient records based on similarity without predefined labels, helping identify unknown patterns.

Question 7

- (i) (a) • True Positives (TP) = 80 **[2]**

• False Positives (FP) = $120 - 80 = 40$

• False Negatives (FN) = $100 - 80 = 20$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} = \frac{80}{120} = 66.67\%$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{80}{100} = 80\%$$

- (b) **Recall** is important to identify as many fraudulent transactions as possible and minimise missed fraud. **[2]**

Precision is important to minimise genuine transactions being incorrectly flagged as fraudulent.

OR

- (ii) (a) Given: **[3]**

- Recall = 89%

- TP = 80

- TN = 30

- FP = 20

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$0.89 = \frac{80}{80 + \text{FN}}$$

$$\text{FN} \approx 9.80 \approx 10$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$= \frac{80 + 30}{80 + 30 + 20 + 10} = \frac{110}{140} \approx 78.57\%$$

Answer: approximately 78.6%.

- (b) Accuracy can be misleading in an imbalanced dataset because a model may achieve high overall accuracy by correctly predicting the majority class while performing poorly on the minority class. **[1]**

Question 8

- (i) False Negative (Type II error) [2]
- Two possible consequences are delay in necessary treatment, which may worsen the patient's condition, and failure to take timely preventive or medical action, which may increase health risks.
- (ii) Large Language Models and AI based tools can combine patient information including genetic data, clinical symptoms and doctors' notes to help diagnose rare genetic diseases more quickly and accurately. The specialised AI system can use the key symptoms to find similar cases in different medical papers and analyse the genetic variance of the patient. Thereafter it can create a short list of possible rare genetic diseases. [2]

SECTION C – 28 MARKS

Question 9

- (i) `model = Sequential([` [3]
- `Dense(64, activation='relu', input_shape=(feature_dim,)),`
- `Dense(32, activation='relu'),`
- `Dense(1, activation='sigmoid')`
- `])`
- `model.compile(optimizer='adam',`
- `loss='binary_crossentropy',`
- `metrics=['accuracy'])`
- (ii) (a) The categorical features such as merchant category, country and device type should be converted into numerical form using an encoding technique such as one-hot encoding. Numerical features such as transaction amount, time since last transaction and merchant risk score should be scaled / standardised so that they are brought to a comparable range before being supplied to the ANN. [2]
- (b) Encoding makes categorical data suitable for numerical processing by the ANN. Scaling prevents features with larger numerical values from dominating the learning process, which helps the network converge more smoothly and faster. It improves training stability, and overall model performance. [2]

Question 10

- (i) `from tensorflow.keras.models import Sequential` [2]
`from tensorflow.keras.layers import Dense`
`model = Sequential()`
`# Hidden Layer`
`model.add(Dense(2, activation='relu', input_shape=(3,)))`
`# Output Layer`
`model.add(Dense(1, activation='sigmoid'))`
`model.summary()`
- (ii) Here the model has very high training accuracy (98%) compared to a much lower and moderate test accuracy (78%). [1]
This implies that the model was trained exceptionally well on the test data but it failed to generalise to new or unseen data.
- (iii) (a) Benefits of using Machine Learning in medical diagnosis are: [2]
 - Machine Learning helps in early and efficient disease detection.
 - Data driven Machine Learning assists doctors in analysing patient data and making diagnoses.
- (b) Two factors that may reduce accuracy are: [2]
 - Poor-quality or insufficient training data.
 - Changes in real-world data / patterns compared with the training data.

Question 11

- (i) (a) $\text{Mean} = \frac{(80 + 85 + 90 + 110 + 95 + 85 + 400)}{7} = \frac{945}{7} = 135 \text{ seconds}$ [3]
Average response time (in seconds) arranged in ascending order:
80, 85, 85, 90, 95, 110, 400
Median = 90 seconds (middle value).
Comment: Outlier 400 strongly pulls the mean above typical values. Thus, in this case median is a better statistical measure for the normal performance of the web based AI service.
- (b) $z = \frac{400-135}{110} = \frac{265}{110} \approx 2.41$ [1]
Interpretation: About 2.4 standard deviations are above mean. This indicates an unusually slow response, potential performance issue or outlier day.
- (ii) (a) Correlation Coefficient measures the strength and direction of the relationship between advertising expenditure and monthly sales. [1]
- (b) **+0.92:** Strong **positive correlation** - higher advertising expenditure is strongly associated with higher sales. [2]
-0.78: Strong **negative correlation** - higher advertising expenditure is strongly associated with lower sales.

OR

- (iii) (a) **Null Hypothesis (H_0):** The new AI-assisted diagnosis system does not perform significantly better than the existing system. [2]

Alternative Hypothesis (H_1): The new AI-assisted diagnosis system performs significantly better than the existing system.

- (b) **Type I Error:** It concludes that the new system performs significantly better when it actually does not. [2]

Type II Error: It concludes that the new system does not perform significantly better when it actually does.

(Candidates have to state the difference with the help of relevant examples.)

- (iv) Let X denotes that the system is tagged as 'slow responding system'. [3]

Here X follows the following binomial distribution

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \text{ where } 0 \leq k \leq 5$$

With the parameters:

$$n = 5, p = 0.1, q = (1 - 0.1) = 0.9$$

Now for exactly two slow response days, $k=2$

$$P(X = 2) = \binom{5}{2} (0.1)^2 (0.9)^3$$

$$= 10 \times 0.01 \times 0.729$$

$$= 0.0729$$

Question 12

- (i) **Object Detection** will be the most suitable Computer Vision technique for this task. [2]

Object Detection can identify as well as localise individual objects in an image, generally by placing bounding boxes around them and assigning class labels such as *car*, *bus* or *truck*.

- (ii) The two images should have the same dimensions (height and width) and compatible channels for the operation. [2]

lower_red and *upper_red* are commonly stored as **NumPy arrays** (**numpy.ndarray**).

- (iii) (a) **Facial Recognition:** It identifies or verifies a person's identity using facial features. It can be used for access control / security before allowing entry to a restricted area. [1]

Other acceptable applications: airport passenger identification, smartphone unlocking, attendance system etc.

- (b) **Object Tracking** provides information about the location and movement of an object over time, whereas Image Classification only assigns a label to the entire image. [1]

- (c) OpenCV [1]

OR

- (iv) (a) Facial recognition [1]
- (b) Facial recognition identifies a specific person by matching biometrics against a known data base whereas image classification sorts any image into a broad, predefined category without identifying any individual. [1]
- (c) HSV separates Hue and Saturation from Value (brightness), making colour-based detection more robust when brightness changes, whereas RGB components are more directly affected by changes in illumination. [2]
- (v) (a) `lower_red = np.array([0, 120, 70])` # line 5 [1]
`upper_red = np.array([10, 255, 255])` # line 6
- (b) `cv2.inRange()` creates a binary mask by selecting pixels whose HSV values lie between the specified *lower_red* and *upper_red* limits. [1]
- (c) **Case 1: Images have the same size and number of channels:** [1]
- The AND operation will be performed on corresponding pixels of both images.
 - Only pixels that are present in both images (according to bitwise AND rules) will appear in the result.
 - The mask will further restrict the operation to the selected red regions.
- Case 2: Images have different sizes or different channels:**
error: Sizes of input arguments do not match

Empowering Minds & Transforming Lives since 1958