

# SCO INTERNATIONAL OLYMPIAD

## CLASS 10 AI QUESTION PAPER

SCO International AI Olympiad | Set S | 2025-26

A carefully formatted question paper for schools, teachers, parents, and students, with answer keys, explanations, and visual learning aids for AI, machine learning, NLP, computer vision, and responsible AI.

**Designed from Class 10 AI pathways and aligned with SCO's platform flow for guided preparation, practice, reporting, and future-ready academic growth.**

- exam-ready AI concepts covering machine learning, NLP, computer vision, ethics, and Python AI tools
- case-study based reasoning for healthcare, autonomous systems, recruitment, language translation, and surveillance
- visual prompts placed inside question blocks to support understanding and revision before PDF publication

ML Metrics	Clustering	NLP	Transformers	Fairness
Python AI	OpenCV	CNN	Case Study	Ethics

# Candidate Guidelines

## SCO International AI Olympiad 2025-26 | Class 10 | Question Paper Set S

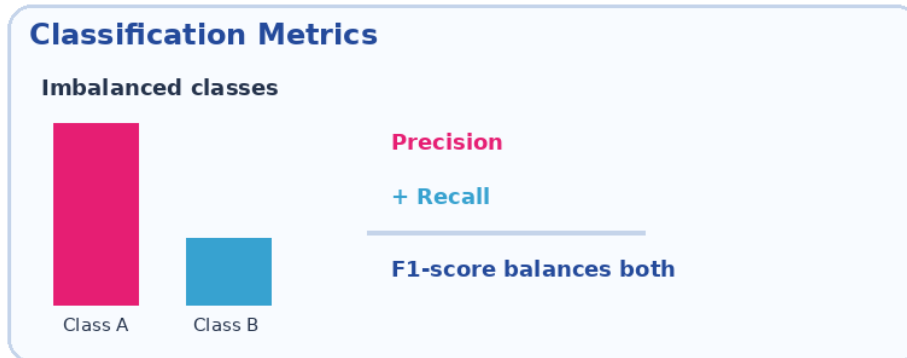
- Total Questions: 50 | Suggested Time: 1 hour
- Before the exam begins, complete the OMR sheet with name, school code, class, roll number, registration ID, and contact number where applicable.
- Each question has only one correct answer. Select one option only.
- Use an HB pencil or a blue/black ballpoint pen to mark answers on the OMR sheet, as instructed by the invigilator.
- No calculator is required. All questions must be attempted. There is no negative marking in this sample format.
- At the end of the test, hand over the OMR sheet and question booklet to the invigilator if applicable.

## Paper Structure

Section	Question Range	Focus Area
General Questions	1-40	AI concepts, ML, NLP, Python, OpenCV, CNN, ethics
Case Study Based Questions	41-50	Applied AI reasoning and real-world decision-making

## General Questions

1. Which metric is most appropriate for evaluating a classification model when dealing with imbalanced datasets?

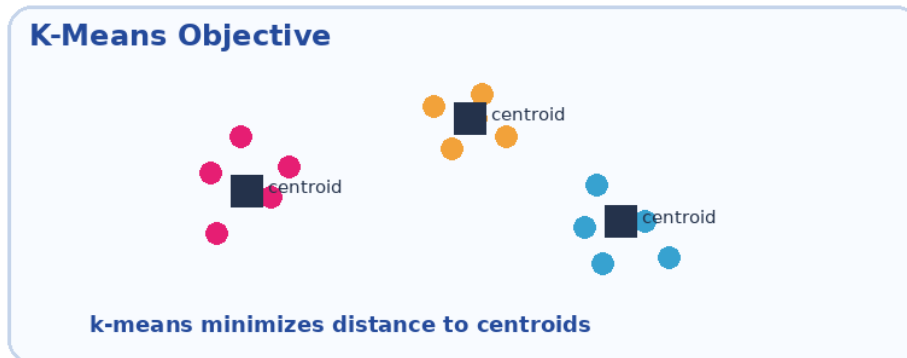


- A. Accuracy
- B. Precision
- C. F1-Score
- D. Mean Squared Error

**Answer: C**

**Explanation:** The F1-Score balances precision and recall, making it suitable for imbalanced datasets where accuracy can be misleading.

2. In k-means clustering, which of the following is true about the objective function?



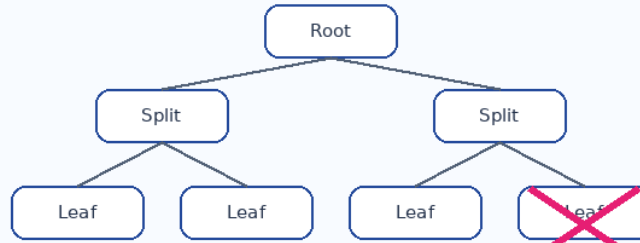
- A. It minimizes the sum of squared distances between each data point and its cluster centroid.
- B. It maximizes the variance within each cluster.
- C. It minimizes the silhouette coefficient.
- D. It maximizes the entropy of clusters.

**Answer: A**

**Explanation:** The objective function in k-means minimizes the sum of squared Euclidean distances between data points and their assigned cluster centroids.

3. Which technique is used to reduce overfitting in decision trees?

### Decision Tree Pruning



Pruning removes unnecessary branches to reduce overfitting.

- A. Increasing the tree depth
- B. Pruning
- C. Using a linear activation function
- D. Applying dropout

**Answer: B**

**Explanation:** Pruning removes unnecessary branches in a decision tree to reduce overfitting.

4. Which of the following best describes the difference between supervised and unsupervised learning?

### Learning Types

#### Supervised



#### Unsupervised



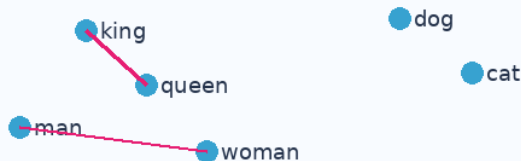
- A. Supervised learning uses labeled data, while unsupervised learning uses unlabeled data.
- B. Supervised learning is used for clustering, while unsupervised learning is used for classification.
- C. Supervised learning uses unsupervised models, while unsupervised learning uses supervised models.
- D. Both require labeled datasets.

**Answer: A**

**Explanation:** Supervised learning requires labeled data to map inputs to outputs, while unsupervised learning works on unlabeled data to find hidden patterns.

5. Which algorithm is commonly used for word embeddings in NLP?

### Word Embeddings



Word embeddings place related words near each other.

- A. k-means

- B. Word2Vec
- C. Linear Regression
- D. Random Forest

**Answer: B**

**Explanation:** Word2Vec creates word embeddings using neural networks, capturing semantic relationships between words.

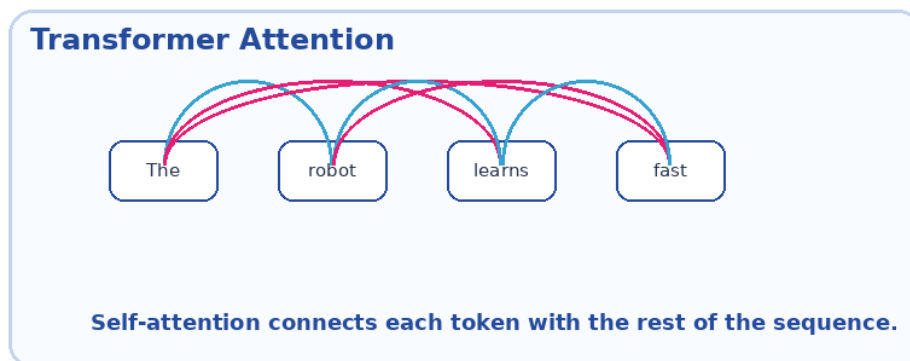
6. Which of the following NLP techniques helps identify parts of speech in a sentence?

- A. Named Entity Recognition
- B. Tokenization
- C. POS Tagging
- D. Sentiment Analysis

**Answer: C**

**Explanation:** Part-of-Speech (POS) tagging identifies and labels the grammatical roles of words in a sentence.

7. What is the primary advantage of using transformers in NLP over traditional RNNs?



- A. They process words sequentially.
- B. They use attention mechanisms to process entire sequences simultaneously.
- C. They require less memory.
- D. They are limited to short text sequences.

**Answer: B**

**Explanation:** Transformers use attention mechanisms to capture relationships across entire sequences in parallel, improving efficiency and context understanding.

8. Which of the following tools is used for stemming in NLP?

- A. BERT
- B. Porter Stemmer
- C. WordNet
- D. Spacy

**Answer: B**

**Explanation:** The Porter Stemmer algorithm reduces words to their root forms by removing suffixes.

9. Which of the following best defines "AI fairness"?

- A. Ensuring that AI models perform perfectly.
- B. Minimizing bias and ensuring equitable treatment for all groups.
- C. Creating highly accurate AI systems.
- D. Preventing adversarial attacks.

**Answer: B**

**Explanation:** AI fairness focuses on reducing bias and ensuring equitable outcomes across different demographic groups.

**10.** Which is NOT an example of algorithmic bias?

- A. Unequal outcomes for different racial groups in hiring algorithms.
- B. Errors due to noisy training data.
- C. Higher false-positive rates for a specific gender in a medical diagnosis model.
- D. Lower performance for underrepresented groups in an AI system.

**Answer: B**

**Explanation:** Algorithmic bias involves systematic inequities, while noisy data is a separate issue related to data quality.

**11.** Which Python library is primarily used for symbolic mathematics in AI development?

- A. NumPy
- B. TensorFlow
- C. SymPy
- D. PyTorch

**Answer: C**

**Explanation:** SymPy is used for symbolic computation, including algebra and calculus, essential for understanding AI concepts like gradients.

**12.** What will the following Python snippet output?

```
import numpy as np
x = np.array([1, 2, 3, 4])
y = x ** 2
print(np.dot(x, y))
```

#### Dot Product Output

$x = [1, 2, 3, 4]$

$y = x^2 = [1, 4, 9, 16]$

$\text{dot}(x,y) = 1 \times 1 + 2 \times 4 + 3 \times 9 + 4 \times 16$

**$= 1 + 8 + 27 + 64 = 100$**

- A. 40
- B. 64
- C. 90
- D. 100

**Answer: D**

**Explanation:** The array y is x squared, so y = [1, 4, 9, 16]. The dot product is  $1 \times 1 + 2 \times 4 + 3 \times 9 + 4 \times 16 = 1 + 8 + 27 + 64 = 100$ .

**13.** Which of the following is NOT a key feature of OpenCV?

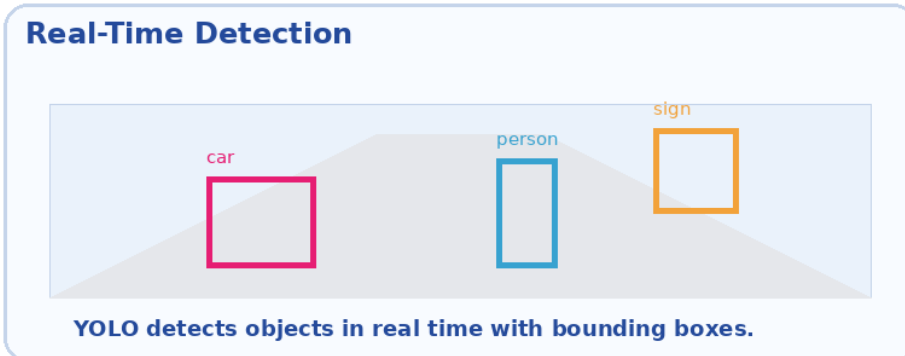
- A. Image Processing
- B. Video Capture
- C. Deep Learning Model Training

**D. Machine Learning Library Integration**

**Answer: C**

**Explanation:** OpenCV is primarily used for computer vision tasks such as image processing, video capture, feature extraction, and model inference support. It is not primarily a deep learning model training framework like Keras or PyTorch.

**14.** Which of the following CNN architectures is most suitable for real-time applications?



- A. VGGNet
- B. ResNet
- C. YOLO
- D. InceptionNet

**Answer: C**

**Explanation:** YOLO (You Only Look Once) is optimized for real-time object detection.

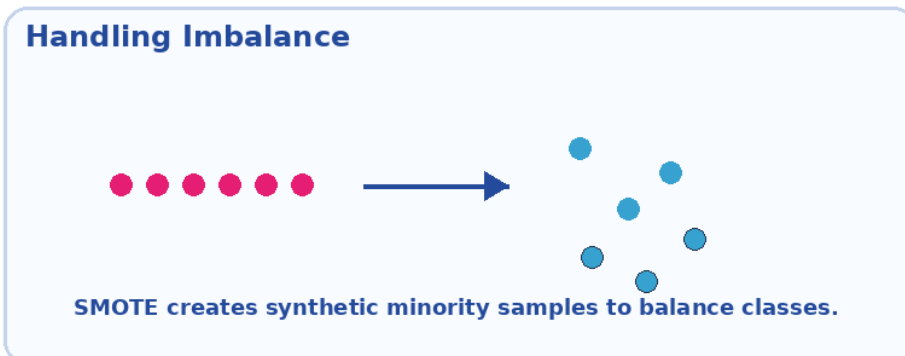
**15.** Which activation function is most commonly used in CNNs for its non-linearity and efficiency?

- A. Sigmoid
- B. Tanh
- C. ReLU
- D. Softmax

**Answer: C**

**Explanation:** ReLU (Rectified Linear Unit) is computationally efficient and mitigates vanishing gradient issues.

**16.** In classification problems, which technique is used to handle imbalanced datasets by increasing the minority class samples?



- A. SMOTE (Synthetic Minority Oversampling Technique)
- B. Bagging
- C. Gradient Boosting
- D. Dimensionality Reduction

**Answer: A**

**Explanation:** SMOTE generates synthetic samples for the minority class to balance the dataset and improve model performance, especially in imbalanced classification problems.

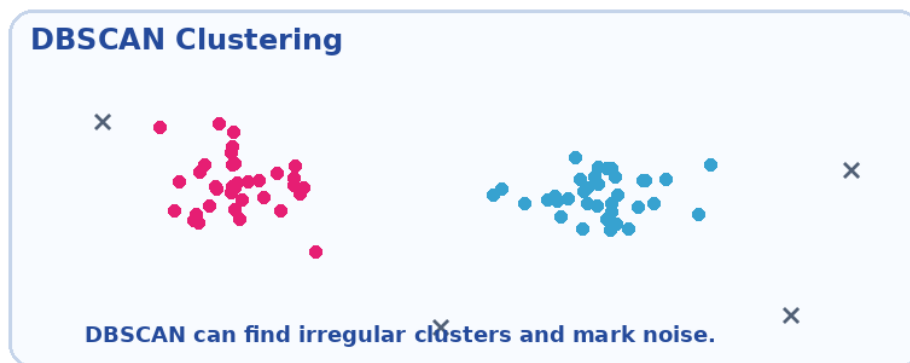
17. Which of the following statements about decision trees is FALSE?

- A. Gini Impurity is used to measure the quality of a split in a decision tree.
- B. Decision trees always produce the same results regardless of the dataset's randomness.
- C. Overfitting can be reduced using pruning techniques.
- D. Decision trees can handle both numerical and categorical data.

**Answer: B**

**Explanation:** Decision trees are sensitive to dataset randomness; small changes in the dataset can lead to entirely different splits.

18. Which clustering algorithm is most suitable for datasets where clusters have irregular shapes and varying densities?



- A. k-means
- B. Hierarchical Clustering
- C. DBSCAN
- D. PCA

**Answer: C**

**Explanation:** DBSCAN (Density-Based Spatial Clustering of Applications with Noise) excels at identifying clusters of varying shapes and densities, unlike k-means, which assumes spherical clusters.

19. In the context of Natural Language Processing, what does "Human Language vs. Computer Language" primarily involve?

- A. Translating human language into binary code for direct execution by a computer.
- B. Converting natural language into structured formats that machines can process.
- C. Comparing human languages with programming languages to find similarities.
- D. Replacing human languages with formal computer instructions.

**Answer: B**

**Explanation:** NLP involves transforming unstructured natural language into structured formats (e.g., tokenization, parsing) to enable computational analysis and processing.

20. Which of the following is a key challenge in building machine learning models that understand human language?

- A. Lack of sufficient labeled data for training.
- B. Contextual ambiguity in natural language.

- C. Processing efficiency in high-dimensional datasets.
- D. All of the above.

**Answer: D**

**Explanation:** Human language understanding in ML is challenging due to limited labeled data, contextual ambiguity (e.g., words with multiple meanings), and computational overhead caused by high-dimensional representations (e.g., embeddings).

21. In NLP, what is the primary advantage of using subword tokenization techniques like Byte Pair Encoding (BPE)?
- A. They reduce the computational complexity of models.
  - B. They allow the model to handle out-of-vocabulary (OOV) words effectively.
  - C. They improve the syntactic parsing of sentences.
  - D. They eliminate the need for embeddings.

**Answer: B**

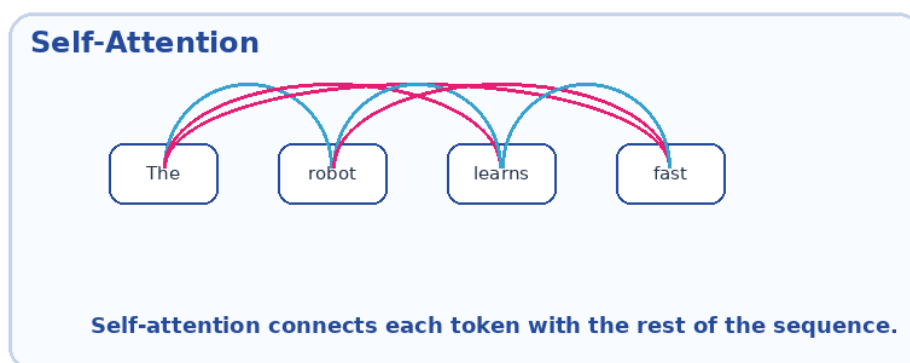
**Explanation:** Subword tokenization techniques like BPE split words into smaller units, enabling the model to understand and generate unseen or rare words by combining known subwords.

22. Which of the following is NOT an objective of Word2Vec's Skip-Gram model?
- A. Predicting the center word from its surrounding context words.
  - B. Generating word embeddings based on their semantic similarity.
  - C. Maximizing the probability of context words given a target word.
  - D. Learning a continuous vector space representation of words.

**Answer: A**

**Explanation:** Skip-Gram works by predicting context words from the center word, not vice versa, which is the objective of the CBOW (Continuous Bag of Words) model.

23. In Transformer architectures used in NLP, what role does the self-attention mechanism play?



- A. It aligns tokens in a sequence with external references.
- B. It calculates relationships between all tokens in a sequence to capture contextual dependencies.
- C. It reduces the dimensionality of the input data.
- D. It replaces embeddings with positional encodings.

**Answer: B**

**Explanation:** The self-attention mechanism computes the relationship between all tokens in a sequence, allowing the model to capture long-range dependencies and contextual meaning efficiently.

24. Consider a dataset where the text includes frequent stop words like "is," "the," and "and." What is the main drawback of removing these stop words for a sentiment analysis task?

- A. It reduces computational efficiency.
- B. It can remove important syntactic information.
- C. It eliminates features that could contribute to sentiment detection.
- D. Both B and C.

**Answer: D**

**Explanation:** While removing stop words can improve efficiency, it may also discard syntactic and sentiment-related features, as some stop words contribute to the meaning and sentiment of the text.

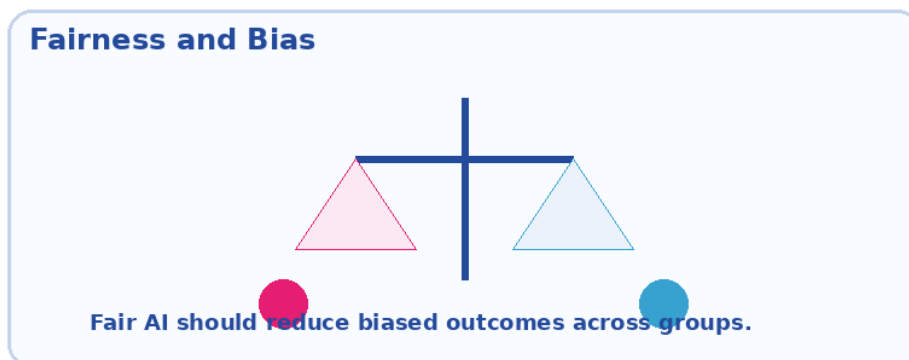
25. In Named Entity Recognition (NER), what is the primary limitation of rule-based systems compared to machine learning-based approaches?

- A. They are computationally expensive.
- B. They rely on domain-specific patterns and cannot generalize well.
- C. They require large annotated datasets for training.
- D. They cannot handle multilingual text.

**Answer: B**

**Explanation:** Rule-based systems depend on handcrafted rules and patterns, making them highly specific to a domain and less capable of generalizing across different datasets or languages compared to ML-based systems.

26. Which of the following is an example of algorithmic bias in AI systems?



- A. A facial recognition system that misidentifies people with darker skin tones due to unbalanced training data.
- B. An AI-based recommendation system that prioritizes user preferences based on data without external intervention.
- C. An AI chatbot that learns inappropriate language from a specific user group.
- D. A machine learning algorithm that improves over time based on user feedback.

**Answer: A**

**Explanation:** Algorithmic bias occurs when a model produces biased outcomes due to skewed or unbalanced training data. In this case, the facial recognition system is biased because it was trained on a dataset that did not adequately represent people with darker skin tones.

27. What is the most effective approach to mitigate AI bias during the development phase of machine learning models?

- A. Increasing the size of the dataset regardless of its diversity.
- B. Ensuring the dataset includes a diverse and representative sample of all relevant demographic groups.
- C. Using complex models that cannot be easily interpreted.
- D. Continuously retraining the model on new data without validating for fairness.

**Answer: B**

**Explanation:** To mitigate AI bias, it's critical to ensure that the dataset is representative of all relevant demographic groups

to prevent skewed predictions. The diversity of the data directly impacts fairness in the AI model's outcomes.

28. In the context of AI fairness, which of the following best describes the concept of "disparate impact"?

- A. The model's ability to make accurate predictions across all classes.
- B. The unequal treatment of different demographic groups despite the same treatment in the algorithm.
- C. The unintentional creation of biased outcomes that negatively affect certain groups.
- D. The adjustment of algorithms to minimize bias and discrimination.

**Answer: C**

**Explanation:** Disparate impact refers to the unintended consequences where certain groups (e.g., based on race, gender, or socioeconomic status) are unfairly disadvantaged by the outcomes of an AI model, even when the algorithm is designed to treat everyone equally.

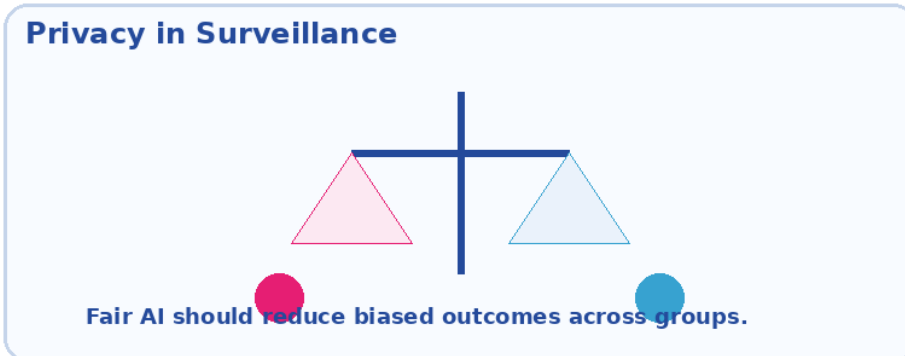
29. Which ethical dilemma arises from using AI for decision-making in sensitive areas such as hiring, criminal justice, or healthcare?

- A. Ensuring that models are explainable and transparent to the public.
- B. Balancing the trade-off between model accuracy and fairness.
- C. Guaranteeing that models can adapt to new, unseen data.
- D. Minimizing computational costs associated with model deployment.

**Answer: B**

**Explanation:** In sensitive areas, there is often a trade-off between accuracy and fairness. AI models might perform well on average but could be biased towards certain groups, leading to ethical issues in how decisions are made. Balancing these factors is a key ethical challenge.

30. Which of the following is a primary ethical concern in the use of AI in surveillance systems?



- A. AI models making real-time predictions without human intervention.
- B. The potential for privacy violations and mass surveillance without consent.
- C. The need for a centralized authority to control all AI-powered systems.
- D. The use of AI models that are incapable of adapting to new data sources.

**Answer: B**

**Explanation:** One of the key ethical concerns with AI in surveillance is the potential invasion of privacy, as AI can be used to track and monitor individuals without their knowledge or consent, raising significant privacy and civil liberties issues.

31. In Python, which of the following libraries is most commonly used for image preprocessing tasks such as resizing, normalization, and augmentation for deep learning models?

### Image Preprocessing Flow

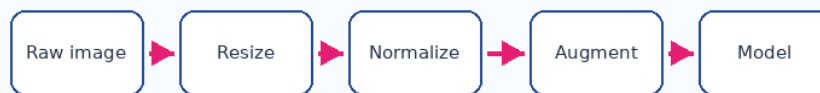


Image preprocessing prepares data before training.

- A. Numpy
- B. Pandas
- C. Matplotlib
- D. Keras ImageDataGenerator

**Answer: D**

**Explanation:** Keras image preprocessing and augmentation utilities can resize, normalize, and augment images before training. ImageDataGenerator is a commonly known Keras preprocessing tool, although modern TensorFlow workflows may also use preprocessing layers and tf.data pipelines.

32. When implementing a simple neural network in Python using TensorFlow, which of the following functions is used to compile the model with the optimizer and loss function?

- A. model.fit()
- B. model.add()
- C. model.compile()
- D. model.predict()

**Answer: C**

**Explanation:** model.compile() is used to configure the model with an optimizer, loss function, and metrics, preparing it for training. This is an essential step before calling model.fit() to train the model.

33. In the context of Natural Language Processing (NLP) with Python, which of the following libraries is most commonly used for tokenization and text preprocessing tasks such as removing stop words, stemming, and lemmatization?

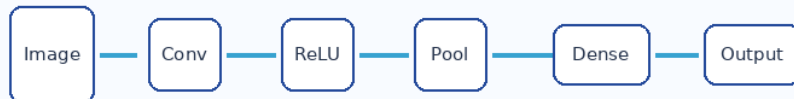
- A. NumPy
- B. Scikit-learn
- C. NLTK (Natural Language Toolkit)
- D. OpenCV

**Answer: C**

**Explanation:** The NLTK library provides extensive tools for tokenization, stemming, lemmatization, and stopword removal in NLP tasks, making it one of the most commonly used libraries in NLP.

34. Which of the following Python libraries would you primarily use to implement and train a Convolutional Neural Network (CNN) for image classification?

### CNN Image Classification



CNNs extract features layer by layer from images.

- A. Flask
- B. Keras
- C. OpenCV
- D. Matplotlib

**Answer: B**

**Explanation:** Keras is a high-level neural networks API that can be used to implement, train, and evaluate CNN models. It provides a simple interface for building CNN architectures suitable for image classification.

35. In Python, what is the purpose of the 'dropout' layer in deep learning models, and where is it commonly used?

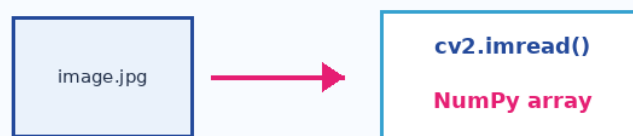
- A. It reduces the computational cost of training by skipping random layers.
- B. It prevents overfitting by randomly setting the output of some neurons to zero during training.
- C. It increases the accuracy of the model by learning more features.
- D. It accelerates the training process by using fewer neurons in each epoch.

**Answer: B**

**Explanation:** The dropout layer is used to prevent overfitting in deep learning models by randomly setting the output of some neurons to zero during training. This helps the model generalize better on unseen data.

36. In the context of using OpenCV for image processing, which of the following functions is used to load an image from a file and store it as a NumPy array?

### OpenCV Image Loading



OpenCV loads images as arrays for processing.

- A. cv2.imshow()
- B. cv2.imread()
- C. cv2.imwrite()
- D. cv2.resize()

**Answer: B**

**Explanation:** The function cv2.imread() is used to read an image from a file and store it as a NumPy array, allowing you to manipulate or analyze the image data using OpenCV.

37. When implementing a Convolutional Neural Network (CNN) for image classification, which of the following layers is used to reduce the spatial dimensions of the input image while preserving important features?

- A. Dense Layer
- B. Convolutional Layer
- C. Pooling Layer
- D. Dropout Layer

**Answer: C**

**Explanation:** The Pooling Layer (often MaxPooling) is used in CNN architectures to reduce the spatial dimensions of the input image, thus decreasing the computational load and highlighting essential features by retaining the most prominent values in each region.

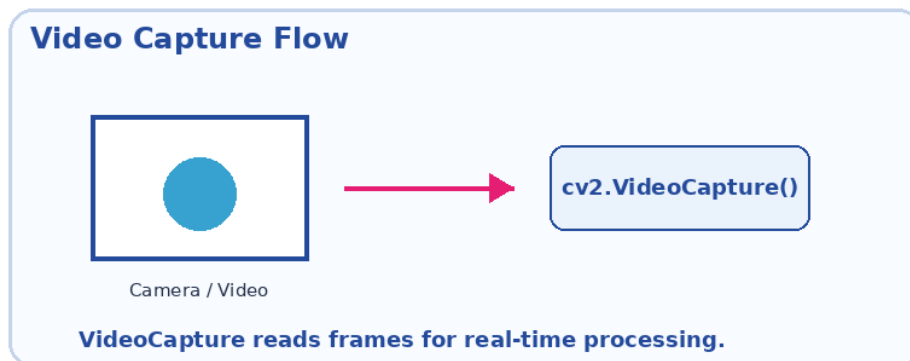
38. In a CNN architecture, which activation function is most commonly used in hidden layers to introduce non-linearity and allow the model to learn complex patterns?

- A. Sigmoid
- B. ReLU
- C. Softmax
- D. Tanh

**Answer: B**

**Explanation:** ReLU (Rectified Linear Unit) is the most commonly used activation function in CNNs, especially in hidden layers, as it helps the model learn complex patterns by introducing non-linearity and speeding up convergence during training.

39. When using OpenCV for real-time video processing, which of the following methods is used to capture video from a camera or video file?



- A. cv2.findContours()
- B. cv2.VideoCapture()
- C. cv2.calcHist()
- D. cv2.estimateAffine2D()

**Answer: B**

**Explanation:** The function cv2.VideoCapture() is used in OpenCV to capture video frames from a camera or video file. It allows for real-time processing of video data for tasks like object tracking and face recognition.

40. In a CNN architecture, which of the following layers is responsible for applying filters to the input image in order to extract features such as edges, textures, and shapes?

## Convolution Filters

1	2	1	0
0	1	3	2
2	2	1	0
1	0	2	3

Input pixels

Filter  
(kernel)

→ features

- A. Convolutional Layer
- B. Fully Connected Layer
- C. Pooling Layer
- D. Flatten Layer

**Answer: A**

**Explanation:** The Convolutional Layer is responsible for applying various filters (kernels) to the input image to detect patterns, such as edges, textures, and shapes, in a CNN. This is a fundamental part of feature extraction in deep learning models.

## Case Study Based Questions

### 41. Case Study: Classification for Medical Diagnosis

A healthcare startup wants to use machine learning to classify medical data into two categories: "healthy" and "diseased." They use a decision tree classifier and train it with patient data that includes features like age, cholesterol levels, blood pressure, and BMI. After training the model, the startup notices that the classifier is giving biased results based on gender.

**Question: Which of the following approaches should the team use to ensure the model is more fair and unbiased?**

### Fair Medical AI

Fair AI should reduce biased outcomes across groups.

- A. Train the model without any preprocessing of the data
- B. Use a balanced dataset with equal male and female samples
- C. Ignore the gender feature during training
- D. Use an ensemble method like Random Forest

**Answer: B**

**Explanation:** Ensuring a balanced dataset with equal male and female samples will help reduce gender bias in the model. Bias often arises from unbalanced data, where the model may favor the majority group, leading to inaccurate or unfair predictions.

**42. Case Study: Clustering for Market Segmentation**

A marketing team wants to cluster customers into distinct groups based on spending behavior, age, and geographic location. They decide to use K-Means clustering. After running the algorithm, they find that the clusters formed do not align with any logical customer segments.

**Question: What might be the issue causing poor clustering results in this case?**

- A. The data was too small for clustering
- B. K-Means assumes spherical clusters, which may not fit the data's structure
- C. Clustering does not require normalization of data
- D. The algorithm is using a non-linear model

**Answer: B**

**Explanation:** K-Means clustering assumes that clusters are spherical and that they are roughly the same size. If the data contains clusters of varying shapes or densities, K-Means may not perform well. Alternative clustering algorithms like DBSCAN might yield better results.

**43. Case Study: Natural Language Processing for Customer Feedback**

A company wants to analyze customer feedback from product reviews using Natural Language Processing (NLP) to automatically categorize reviews as "positive" or "negative." They use a pre-trained Sentiment Analysis model.

**Question: What is the most likely cause of incorrect sentiment classification if the model fails to interpret sarcasm in reviews?**

- A. The model lacks a sufficient dataset
- B. The model does not recognize contextual words
- C. The model has not been trained on specific product feedback
- D. The training data used was not labeled

**Answer: B**

**Explanation:** Sarcasm often relies on the context and tone of words, which may not be captured by basic sentiment analysis models. More advanced NLP techniques like contextual word embeddings (e.g., BERT) or additional training on sarcastic text may help improve performance.

**44. Case Study: Bias in AI Recruitment System**

A company uses an AI recruitment tool to screen resumes for job applicants. The system is trained on historical data, which contains a disproportionate number of male applicants for a technical role.

**Question: What ethical consideration should the company focus on to prevent gender bias in the AI recruitment system?**

- A. Adjust the model to only consider male applicants
- B. Remove gender as a feature from the training data
- C. Only use the AI system to recommend candidates who are male
- D. Ensure the training data includes a balanced representation of all genders

**Answer: D**

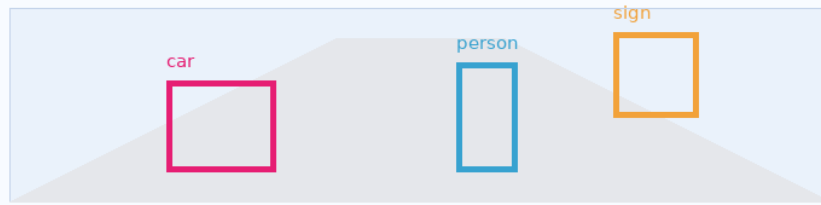
**Explanation:** Ensuring the training data is balanced and includes diverse candidates from different genders will help reduce bias. Removing gender as a feature may not solve underlying bias problems if other features correlate with gender.

**45. Case Study: Image Classification for Autonomous Vehicles**

A team is developing an image recognition system for an autonomous vehicle. The system uses a Convolutional Neural Network (CNN) to classify objects on the road (cars, pedestrians, traffic signs). After testing, they find that the model performs poorly in identifying objects in low light conditions.

**Question: Which of the following methods could improve the model's performance in low light conditions?**

### Low-Light Object Detection



YOLO detects objects in real time with bounding boxes.

- A. Use a larger dataset of images in bright conditions
- B. Apply data augmentation techniques such as image brightening and rotation
- C. Reduce the size of the training dataset
- D. Use a shallower CNN model

**Answer: B**

**Explanation:** Data augmentation techniques like image brightening, flipping, or rotation can help the model generalize better to various conditions, including low light, by artificially increasing the diversity of the training data.

#### 46. Case Study: AI for Healthcare Diagnosis

A hospital is using AI to detect cancerous tumors in medical images. The algorithm is trained on a large dataset of images, but the results show that the model has high accuracy but fails to detect smaller tumors.

**Question: What could be the primary reason for the model's inability to detect smaller tumors?**

- A. The model was not trained on enough negative samples (non-tumor images)
- B. The model was overfitting to the training data
- C. The dataset does not have sufficient examples of smaller tumors
- D. The model uses too many layers in the neural network

**Answer: C**

**Explanation:** The model's failure to detect smaller tumors is likely due to an insufficient number of smaller tumor examples in the training dataset. To improve, the dataset should be expanded with more diverse examples, particularly focusing on smaller tumors.

#### 47. Case Study: Python for AI in Chatbots

A team is building a chatbot that uses Natural Language Processing to answer customer queries. They are using Python and the nltk library. However, the chatbot fails to recognize synonyms and alternative phrasing of common queries.

**Question: What feature could be implemented in the chatbot to improve its recognition of synonyms and alternative phrasing?**

### Chatbot Language Variation



Normalization and synonym dictionaries help chatbots understand va

- A. Use a machine learning model with a larger dataset
- B. Implement a synonym dictionary and text normalization
- C. Train the model using only exact queries
- D. Increase the chatbot's response time

**Answer:** B

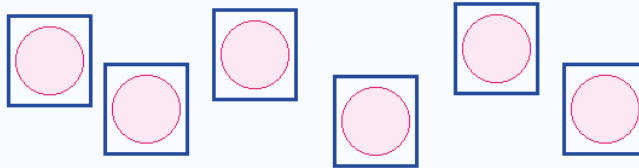
**Explanation:** Implementing a synonym dictionary and normalizing text (e.g., stemming or lemmatization) can help the chatbot recognize and handle variations in phrasing, improving its ability to understand different ways of asking the same question.

**48. Case Study: OpenCV for Facial Recognition**

A company uses OpenCV to implement facial recognition for security. However, the system struggles to accurately identify individuals in large crowds.

**Question:** What could be a potential solution to improve the accuracy of facial recognition in crowded environments?

**Face Detection Before Recognition**



**Detect faces first, then recognize each isolated face.**

- A. Increase the resolution of the camera
- B. Use a smaller dataset for training
- C. Implement face detection algorithms before recognition
- D. Reduce the number of faces in the training set

**Answer:** C

**Explanation:** A face detection step should first locate and isolate faces in a crowded scene. Recognition can then focus on each detected face region instead of processing the entire crowded image at once.

**49. Case Study: AI for Language Translation**

A company is using AI-based language translation to help global users communicate. However, the system provides inaccurate translations for languages with complex grammar structures (e.g., Hindi, Japanese).

**Question:** What could be done to improve the system's translation accuracy for these languages?

**Translation Data Diversity**



**Diverse language data improves translation quality.**

- A. Use a larger dataset with more complex languages
- B. Train the model specifically on languages with simpler grammar

- C. Use rule-based translation methods instead of AI
- D. Add more hardware resources for computation

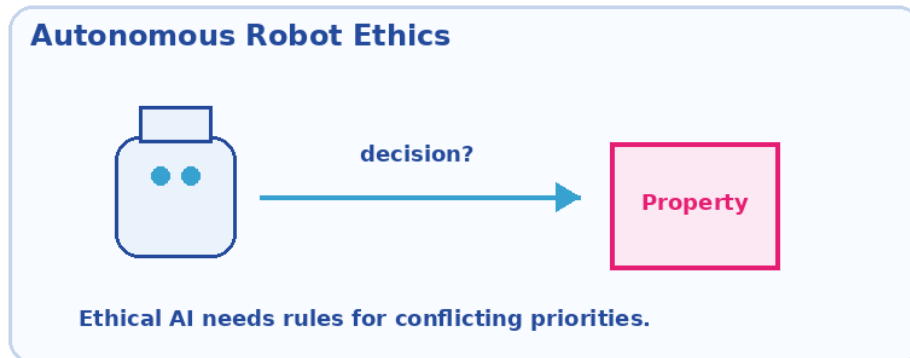
**Answer: A**

**Explanation:** Training the AI model with a larger and more diverse dataset, including languages with complex grammar structures, will help improve the model's ability to understand and translate these languages more accurately.

**50. Case Study: Ethics in Autonomous Decision-Making**

An autonomous delivery robot is programmed to avoid collisions at all costs. However, it is faced with a situation where avoiding an obstacle will result in the robot damaging property.

**Question: What ethical concern does this scenario represent?**



- A. The robot lacks ethical reasoning to handle conflicting priorities
- B. The robot is programmed to prioritize property damage over human safety
- C. The robot is not following the legal guidelines for autonomous robots
- D. The robot has not been programmed to avoid obstacles entirely

**Answer: A**

**Explanation:** The ethical concern arises because the robot lacks the ability to weigh different priorities, such as avoiding harm to people and minimizing property damage. Ethical decision-making frameworks for AI should be developed to handle such conflicting priorities responsibly.

