

Solutions

Each answer shows the steps, then the **result**.

2.1

- (i) **cleaning**
- (ii) **normalisation** (scaling)
- (iii) **feature selection**

2.2

- $y = 2.4 \times 20 + 15$
- **63**

2.3

- (i) $\frac{45 + 135}{200} = \mathbf{0.90}$
- (ii) $\frac{45}{45 + 5} = \mathbf{0.90}$
- (iii) $\frac{45}{45 + 15} = \mathbf{0.75}$

3.1 C

- no labels: it groups items

3.2 B

- weights are adjusted to reduce the error

3.3 B

- recall counts how many real cases are found

3.4

(a) Any three, e.g.:

- remove blurred, duplicate or wrongly labelled photos
- resize all photos to the same size and scale pixel values to 0–1
- encode the labels as numbers, such as healthy 0, diseased 1, and balance the classes

(b)

- this is data leakage: the date of death is not known when the prediction is needed
- the model will look excellent in testing but fail in real use

(c)

- its layers scan small patches of the image
- so it finds local features such as spots and edges, wherever they appear on the leaf

(d)

- $\text{precision} = \frac{50}{50 + 30} = 0.625$
- $\text{recall} = \frac{50}{50 + 10} \approx 0.83$
- **recall** matters more
- a missed sick plant goes untreated; treating a healthy one by mistake costs little