Module 3

Individual Task

3. Feature Extraction Thought Experiment: Select a dataset (e.g., photos, shopping lists and describe which features would be important to a machine learning model

Feature Extraction

• Feature extraction is crucial in machine learning, transforming raw data into meaningful information.

• Real-world image data is large, complex, and unstructured.

• Machine learning models cannot directly interpret raw images

• Features are extracted characteristics that help models recognize patterns and make predictions.

Selected Dataset: Photos of Flowers

• Dataset contains thousands of flower photos.

Possible model goals:

• Classify flower type (rose, tulip, sunflower, daisy, etc.)

• Detect whether the flower is wild or cultivated

• Identify flower species or variety

• Detect multiple flowers in a single image

Raw Data in Each Flower Image

- Pixel values (RGB or HSV)
- Petal and leaf shapes
- Colors and color patterns
- Petal and leaf textures
- Edges of petals and leaves
- Symmetry and spatial arrangements
- Background elements (soil, grass, pot, etc.)

1. Color Features

- Average petal and leaf colors
- Color histograms for pixel distributions
- Distinct patterns like spots or stripes
- Color contrast between flower and background

Example:

- Rose (red): The petals are deep red with a subtle gradient.
- Tulip (yellow): Bright yellow petals contrast with the green stem.

2. Shape Features

- Petal length, width, and curvature
- Flower outline: circularity and aspect ratio
- Number of petals per flower
- Leaf shape and serration patterns

Example:

- Sunflower: Large, round flower head with hundreds of small petals.
- Lily: Six long, trumpet-shaped petals curving outward gracefully.

3. Texture Features

- Smoothness or roughness of petals
- Vein patterns on petals and leaves
- Surface variation and gradients
- Petal edges sharpness and clarity

Example:

- Orchid: Smooth, waxy petals with faint vein lines.
- Marigold: Dense, ruffled petals with a slightly rough surface.

4. Spatial Features

- Relative positions of petals and leaves
- Symmetry type: radial or bilateral
- Flower-to-leaf size ratio
- Distance between multiple flowers

Example:

- Daisy: Petals radiate symmetrically around a central yellow disc.
- Hibiscus: Large single flower with widely spaced petals and prominent stamen.

5. Contextual Features

• Background elements like soil or grass

• Number of flowers in photo

• Lighting conditions in the image

• Presence of insects or objects nearby

Example:

• Lavender: Flowers growing in rows in a sunny garden.

• Daffodil: Blooming among grass with morning sunlight hitting petals.

6. High-Level Features

• Edge patterns detected by CNNs

• Petal arrangement structures

• Texture and fine details

• High-level abstract representations

Example

• Peony: CNN detects layered petals forming dense rounded clusters.

• Chrysanthemum: Edge patterns highlight tightly packed petals with radial symmetry.

7. Metadata Features

• Time of day the photo taken

• Season of the flower bloom

• Photo location or GPS coordinates

• Camera angle and zoom level

Example:

• Cherry Blossom: Photo taken in early April during full bloom.

• Sunflower: Image captured facing east at 8:00 AM.

Structured vs Unstructured Features

Image data is unstructured. But after feature extraction:

• Color histograms become structured numeric data.

• Petal and flower shape measurements become numeric values.

• Texture scores and patterns become feature vectors.

• Spatial arrangements and symmetry values are numeric features.

Thus, feature extraction converts unstructured flower images into structured input for machine learning models.

Feature Engineering

• Feature engineering means creating new features from existing ones.

Examples for flowers:

• Petal length-to-width ratio

• Color uniformity score

• Symmetry score

• Flower-to-leaf size ratio

Engineered features improve performance by giving deeper insights into the flower dataset.

Conclusion

• Feature extraction is a critical step in machine learning.

• In an image dataset of flowers, important features include color, shape, texture, petal arrangement, symmetry, and deep learning representations.

• Proper feature extraction transforms raw pixel data into meaningful numerical information.

• It helps the machine learning model classify flower species accurately.

Thus, selecting and engineering the right features directly impacts the success of the flower classification system.