FEATURE EXTRACTION – SHAPE AND TEXTURE (95 Texture Features)

1. Introduction to Feature Extraction

In medical image processing, raw images obtained from imaging modalities such as MRI, CT scan, Ultrasound, Mammography, and Retinal imaging contain a huge amount of pixel data. However, most of this data is redundant for diagnosis.Therefore, the image must be converted into quantitative information that describes important properties of the object or tissue present in the image.

This process is called **Feature Extraction**.

Definition

Feature extraction is the process of deriving numerical descriptors from an image that represent important characteristics of objects or regions in the image.

These descriptors are called features.

Importance of Feature Extraction in Medical Imaging

Feature extraction plays a major role in Computer Aided Diagnosis (CAD) systems. It helps to:

- Identify abnormal tissues
- Detect tumors
- Analyze organ structures
- Differentiate healthy and diseased tissues
- Improve classification accuracy

Medical Image	**Extracted Features**
Mammogram	Tumor shape, texture roughness
Lung CT	Nodule size and irregularity
Retinal Image	Blood vessel structure

Medical Image	Extracted Features
Liver Ultrasound	Tissue texture patterns

2. General Pipeline of Feature Extraction

a. Medical images are obtained from imaging devices such as CT Scanner, MRI Scanner, Ultrasound Machine, Mamography System, etc.

b. Preprocessing improves the image quality. Common preprocessing techniques include Noise Removal, Contrast Enhancement, Filtering, Image Normalization.

c. Segmentation isolates the **Region of Interest (ROI)** from the background such as tumor regions and blood vessel network.

d. From the segmented region, numerical descriptors are extracted such as Shape, Texture and Intensity features.

e. The extracted features are given to classifiers such as Support Vector Machines, Neural Networks and Decision Trees. These classifiers help identify diseases.

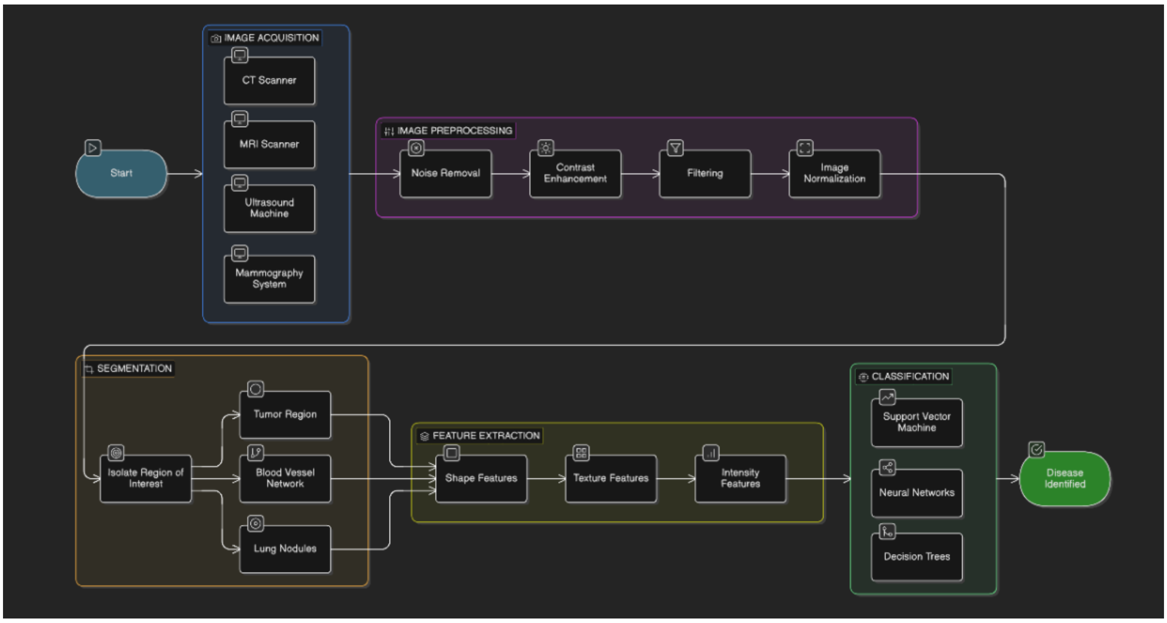


Figure 1. Block Diagram Of Feature Extraction

3. Shape Feature Extraction

Shape features describe the geometric and structural properties of objects present in medical images. These objects may represent tumors, organs, lesions, blood vessels, or nodules. Unlike texture features, shape features are derived from the boundary and spatial structure of the segmented region rather than pixel intensity patterns.

In medical image processing, shape features are widely used because many diseases cause structural changes in organs or tissues. For example, malignant tumors often exhibit irregular and spiculated shapes, whereas benign tumors tend to have smooth and circular boundaries.

To analyze these structural characteristics, shape features are generally classified into the following categories:

- Geometric Features
- Boundary-Based Features
- Region-Based Features
- Moment-Based Features
- Structural Shape Descriptors

I. Geometric Shape Features

Geometric features describe the basic physical measurements of an object, such as size, length, and proportions. These features are calculated directly from the segmented region of interest (ROI).

Geometric features are often the first set of features extracted in medical image analysis because they provide basic information about the object.

Common geometric features include:

- Area
- Perimeter
- Major axis length
- Minor axis length
- Aspect ratio
- Diameter

- **Area**

Area represents the total number of pixels contained within the object region. It indicates the size of the structure detected in the image. Mathematically,

$$Area = \sum_{(x,y)\in ROI} 1$$

Where:

(x,y) represents the coordinates of pixels in the region of interest.

Each pixel inside the object contributes a value of 1.

If the physical pixel spacing is known, the area can also be converted into square millimeters or square centimeters.

Area is commonly used for measuring tumor size, monitoring tumor growth and quantifying organ size

- **Perimeter**

Perimeter represents the length of the boundary surrounding the object.

It is calculated by summing the distances between adjacent boundary pixels.

$$Perimeter = \sum_{i=1}^{N} d_i$$

Where:

- NNN is the number of boundary pixels.
- d_i is the distance between two consecutive boundary pixels.

Perimeter provides information about the complexity of the object boundary.

Irregular tumors often have larger perimeters relative to their area.

- **Diameter**

Diameter refers to the largest distance between any two points on the object boundary. This measurement is often used to estimate tumor size in medical imaging reports.

- **Major Axis Length**

The major axis is the longest axis of the ellipse that best fits the object. Major axis length provides information about the maximum elongation of the structure.

- **Minor Axis Length**

The minor axis is the shortest axis of the fitted ellipse.

It represents the minimum width of the object.

II. Boundary-Based Shape Features

Boundary-based features analyze the shape of the object's contour or edge. These features focus on the structure of the boundary rather than the interior region. Boundary analysis is useful for detecting irregular or spiculated tumor margins, which are often indicators of malignancy.

Common boundary-based features include:

- Circularity
- Compactness
- Roundness
- Curvature

- **Circularity**

Circularity measures how close the shape is to a perfect circle.

$$Circularity = \frac{4\pi A}{P^2}$$

Where:

- A = area of the object
- P = perimeter of the object
- π = constant (3.1416)

Interpretation:

- Circularity ≈ 1 → perfect circle
- Circularity < 1 → irregular shape

Cancerous tumors often show low circularity values due to irregular boundaries.

- **Compactness**

Compactness describes how tightly packed the object shape is.

$$Compactness = \frac{P^2}{A}$$

Where:

- P = perimeter
- A = area

Higher compactness values indicate irregular or elongated shapes.

- **Roundness**

Roundness measures how closely the shape resembles a circle.

$$Roundness = \frac{4A}{\pi D^2}$$

Where:

- A = area
- D = maximum diameter

Roundness values close to 1 indicate circular objects.

- **Curvature**

Curvature describes how sharply the boundary changes direction at a particular point. High curvature regions often correspond to spikes or protrusions on tumor boundaries.

III. Region-Based Shape Features

Region-based features analyze the entire interior region of the object, rather than just the boundary. These features help measure symmetry, solidity, and spatial distribution of the object.

Common region-based features include:

- Solidity
- Extent
- Eccentricity

- **Solidity**

Solidity measures the degree of concavity in the object shape.

$$Solidity = \frac{Area}{Convex\ Area}$$

Where:

- Area = actual object area
- Convex Area = area of the convex hull surrounding the object

A convex hull is the smallest convex shape that completely encloses the object.

Interpretation:

- Solidity ≈ 1 → smooth boundary
- Lower solidity → irregular or concave shapes

- **Extent**

Extent measures the ratio of object area to the area of its bounding box.

$$Extent = \frac{Object\ Area}{Bounding\ Box\ Area}$$

Bounding box is the smallest rectangle that completely contains the object.

- **Eccentricity**

Eccentricity describes the degree of elongation of an object.

It is calculated using the ellipse that best fits the object.

$$Eccentricity = \frac{Distance\ between\ foci}{Length\ of\ major\ axis}$$

The value ranges between:

$$0 \leq e < 1$$

Where:

- e=0 → perfect circle
- e=1→ elongated object

IV. Moment-Based Shape Features

Moment-based features describe the distribution of pixel intensities with respect to the object's center. These features are particularly useful because they are often invariant to translation, rotation, and scaling.

Image moments are calculated as:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y)$$

Where:

- m_{pq} = moment of order
- x,y = pixel coordinates

- $f(x,y)$ = pixel intensity value
- p,q = moment orders

From these moments, various descriptors such as Hu Moments are derived. Hu moments are widely used because they remain unchanged under rotation and scaling.

V. Structural Shape Descriptors

Structural descriptors describe the overall structure of complex shapes. These descriptors are particularly useful for analyzing blood vessels, bones, and anatomical structures.

Examples include:

- Skeleton representation
- Chain codes
- Fourier descriptors

■ Skeleton Representation

Skeletonization reduces an object to its central axis while preserving its topology. This representation is useful for analyzing vascular structures and airway networks.

■ Chain Codes

Chain codes represent the boundary of an object as a sequence of directional codes. Each direction corresponds to a movement between neighboring boundary pixels.

For example:

Direction	**Code**
Right	0
Up-right	1
Up	2
Up-left	3

Chain codes help represent the exact boundary shape efficiently.

- **Fourier Descriptors**

Fourier descriptors represent the boundary using frequency components obtained from the Fourier transform. These descriptors capture global shape characteristics and are useful for shape matching and classification.

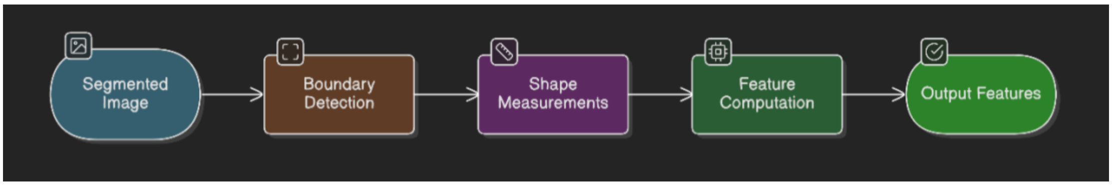


Figure 2 Flowchart of Shape Feature Extraction

4.Texture Feature Categories (Detailed Explanation)

Texture features describe the spatial arrangement and relationship of pixel intensities in an image. In medical images, tissues exhibit different intensity patterns due to variations in cell structure, density, and composition. Texture analysis helps quantify these variations.

For example:

Tissue Type	Texture Pattern
Healthy tissue	Smooth and uniform
Cancerous tissue	Irregular and heterogeneous
Fibrotic tissue	Coarse and complex

To capture these patterns, texture features are grouped into different statistical and structural categories. In many medical imaging systems (especially radiomics-based CAD systems), about **95 texture features** are extracted and grouped into the following categories:

- First-Order Statistical Features
- Gray Level Co-occurrence Matrix (GLCM) Features
- Gray Level Run Length Matrix (GLRLM) Features
- Gray Level Size Zone Matrix (GLSZM) Features
- Neighbourhood Gray Tone Difference Matrix (NGTDM) Features

I. First Order Statistical Texture Features

First-order features are the simplest type of texture features. They are derived directly from the histogram of the image. The histogram represents the distribution of gray level intensities in an image.

NOTE: First-order statistics do not consider spatial relationships between pixels. They only analyze the frequency of intensity values.

For an image with gray levels 0,1,2,...,L−1, the probability of each gray level is:

$$p(i) = \frac{n_i}{N}$$

Variables

p(i) = probability of intensity level iii

n_i = number of pixels having intensity iii

N = total number of pixels in the image

L = total number of gray levels

Important First Order Features:

- Mean (Average Intensity)

$$\mu = \sum_{i=0}^{L-1} i\, p(i)$$

Where

- μ = mean intensity
- i = gray level
- p(i) = probability of gray level

Interpretation: Mean indicates the overall brightness of the tissue region.

- Variance

Variance measures the spread of intensity values around the mean.

$$\sigma^2 = \sum_{i=0}^{L-1} (i-\mu)^2 p(i)$$

Where

σ^2 = variance

μ = mean intensity

Higher variance indicates large intensity variations in tissue.

- Entropy

Entropy measures randomness or complexity of the texture.

$$Entropy = -\sum_{i=0}^{L-1} p(i)\log p(i)$$

High entropy indicates complex and heterogeneous tissue structures, often seen in malignant tumors.

- Skewness

Skewness measures asymmetry of the histogram.

$$Skewness = \frac{1}{\sigma^3}\sum_{i=0}^{L-1}(i-\mu)^3 p(i)$$

Positive skew → histogram tail towards higher intensities.

Negative skew → histogram tail towards lower intensities.

- Kurtosis

Kurtosis measures sharpness or peakedness of the histogram.

$$Kurtosis = \frac{1}{\sigma^4}\sum_{i=0}^{L-1}(i-\mu)^4 p(i)$$

High kurtosis indicates sharp peaks in intensity distribution.

II. Gray Level Co-occurrence Matrix (GLCM) Features

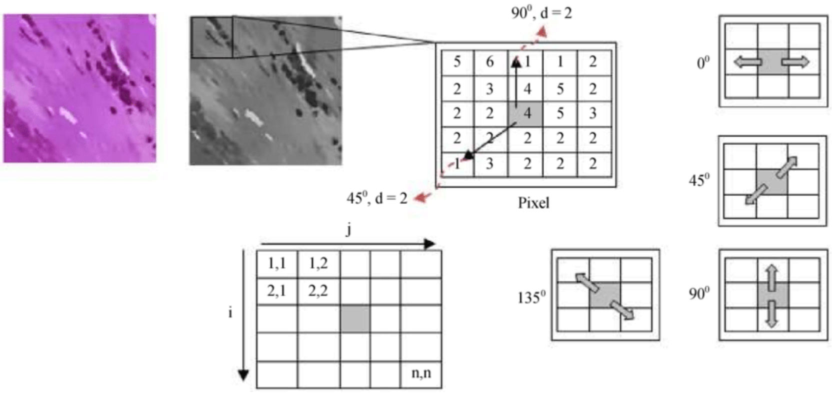
GLCM is a second-order statistical method that considers the spatial relationship between pairs of pixels. Instead of analyzing individual pixel values, GLCM measures how often two pixels with specific gray levels occur at a certain distance and direction.

For example, consider two neighboring pixels:

- Pixel 1 intensity = i

- Pixel 2 intensity = j

GLCM counts how often this pair occurs.



Parameters Used in GLCM

- Distance (d): Distance between two pixels.
- Direction**:**

- 0∘0^\circ0∘ (horizontal)
- 45∘45^\circ45∘
- 90∘90^\circ90∘ (vertical)
- 135∘135^\circ135∘

For each direction, a different GLCM matrix is computed.

Common GLCM Features

- Contrast: Measures local intensity variation.

$$Contrast = \sum_{i,j} (i-j)^2 P(i,j)$$

Where

- P(i,j)= normalized GLCM value
- High contrast → rough texture.

- Energy: Energy measures uniformity of texture.

$$Energy = \sum_{i,j} P(i,j)^2$$

High energy indicates uniform image texture.

- Homogeneity: Homogeneity measures similarity between neighboring pixels.

$$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$$

Higher values indicate smooth textures.

- Correlation: Correlation measures linear relationship between pixel pairs.

$$Correlation = \frac{\sum(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j}$$

Where

- μ_i, μ_j= mean intensities
- σ_i, σ_j = standard deviations

III.Gray Level Run Length Matrix (GLRLM)

GLRLM measures runs of consecutive pixels having the same gray level in a specific direction. A **run** is defined as a sequence of pixels with identical intensity values. Example:

Pixel sequence:
5 5 5 5 5

Run length = 5.

GLRLM records the number of runs for each gray level and run length combination.

Important GLRLM Features

- Short Run Emphasis (SRE)

Measures the distribution of short runs and presence of higher values indicate fine textures.

- Long Run Emphasis (LRE)

Measures the distribution of long runs and higher values indicate coarse textures.

- Gray Level Non-Uniformity (GLN)

Measures similarity of gray levels and higher values indicate non-uniform intensity distribution.

IV. Gray Level Size Zone Matrix (GLSZM)

GLSZM measures connected regions of pixels with identical gray levels. A **zone** is a group of connected pixels having the same intensity. Unlike GLRLM, GLSZM does not depend on direction.

Example: A connected cluster of pixels with the same intensity forms a zone.

Important GLSZM Features

- Small Zone Emphasis

Measures the distribution of small homogeneous zones.

- Large Zone Emphasis

Measures the distribution of large homogeneous regions.

- Zone Size Non-Uniformity

Measures variation in zone sizes and higher values indicate heterogeneous textures.

V.Neighbourhood Gray Tone Difference Matrix (NGTDM)

NGTDM measures the difference between the gray level of a pixel and the average gray level of its neighboring pixels. For each pixel:

- Compute the average intensity of neighboring pixels.
- Calculate the difference between the center pixel and the neighborhood average.
- These differences are accumulated to create the NGTDM matrix.

NGTDM Features

Five features are derived:

Feature	Description
Coarseness	Measures texture smoothness
Contrast	Measures intensity variation
Busyness	Measures rapid intensity changes
Complexity	Measures structural complexity
Strength	Measures pattern clarity

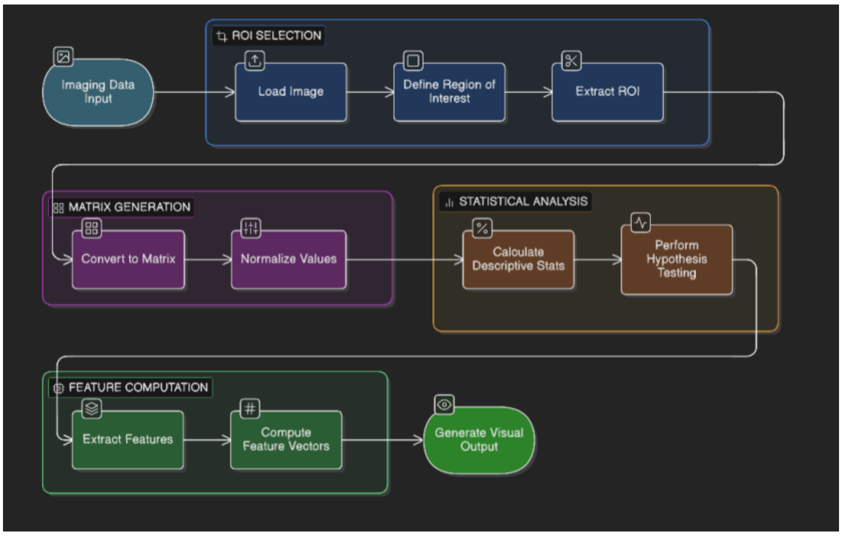


Figure 3 Flowchart of Texture Feature Extraction