MINISTERUL EDUCAȚIEI

UNIVERSITATEA TEHNICĂ
DIN CLUJ-NAPOCA

FACULTY OF AUTOMATIC CONTROL AND

# Lymph node evaluation system

-SEMESTER PROJECT-

Varga Marian Lucian

**2025**

Contents

# 1 Introduction

## 1.1 General context

Cancer is one of the leading causes of mortality worldwide, making accurate staging crucial for determining optimal treatment strategies. The TNM classification system, where N represents the spread of cancer in regional lymph nodes, plays a vital role in assessing the extent of metastasis and guiding treatment decisions. However, current imaging techniques, such as CT, PET, and MRI, present several limitations, including high costs, radiation exposure, and limited accessibility.

Lymph node status is a critical prognostic factor in the management of breast cancer, significantly influencing both surgical and therapeutic planning. Given that lymph node involvement is closely linked to the progression of the disease, accurate detection is paramount in guiding clinical decisions and optimizing treatment strategies. Traditionally, methods such as complete axillary lymph node dissection (ALND) or sentinel lymph node biopsy (SLNB) have served as standard intraoperative procedures. However, these approaches are not without risks—excessive surgical interventions may lead to complications such as lymphedema, impaired mobility, and other postoperative issues that ultimately affect patient quality of life.

The motivation for this project originates from the need to reduce the reliance on invasive imaging techniques that use radiation, such as CT scans. These methods not only expose patients to unnecessary radiation risks but also lead to high costs and prolonged waiting times for concrete diagnostic results. As an alternative, ultrasound (US) imaging, combined with artificial intelligence (AI), presents a promising solution due to its advantages such as high resolution, repeatability, lack of radiation exposure, and affordability. By integrating AI-based image analysis with medical diagnostic guidelines, this project aims to improve clinical workflows, enhance diagnostic accuracy, and reduce healthcare costs.

This application was chosen based on both theoretical and practical considerations. Theoretically, machine learning and computer vision techniques have proven effective in medical image analysis, while practically, the deployment of such a system could significantly improve early cancer detection and patient outcomes. Given the multidisciplinary nature of this project, background knowledge in oncology, medical imaging, and AI is necessary to understand its scope and impact.

Various perspectives exist regarding the optimal approach for lymph node evaluation. While traditional methods remain the standard, there is increasing evidence that AI-assisted ultrasound imaging can enhance the precision and efficiency of cancer diagnosis. This project supports this perspective by providing an automated system that improves upon conventional diagnostic techniques.

This application was chosen due to its potential to revolutionize cancer diagnosis and management through a non-invasive, efficient, and accessible approach. From a theoretical perspective, the integration of ultrasound imaging with artificial intelligence aligns with recent advancements in the medical field, where machine learning algorithms and computer vision have demonstrated a high capacity for detecting and classifying malignant lesions. Research shows that AI can identify pathological characteristics of lymph nodes in ultrasound images with accuracy comparable to or even superior to human interpretation, sometimes surpassing the limitations of subjective factors such as a radiologist’s experience and fatigue. This technology not only promises increased diagnostic accuracy but also ensures standardization of the evaluation process, eliminating the inherent variability of human assessment.

From a practical standpoint, AI-assisted ultrasound offers several advantages over conventional methods. Firstly, unlike CT or PET scans, ultrasound is a radiation-free technique, allowing for frequent patient monitoring without additional risks. Secondly, the low cost and portability of ultrasound devices make this technology more accessible to various medical centers, including those in resource-limited areas where access to advanced imaging techniques is restricted. Moreover, reducing dependence on invasive procedures such as lymph node dissection helps lower postoperative complications like lymphedema or limited mobility, ultimately improving patients' quality of life.

The choice of this application is also justified by its impact on healthcare system efficiency. Implementing such an automated system could significantly reduce the time required for diagnosis, eliminating delays caused by a shortage of specialists or the overburdening of imaging laboratories. Additionally, it would allow for better allocation of medical resources, reducing costs associated with unnecessary procedures and optimizing the patient’s journey within the healthcare system. These benefits are reflected not only at the individual level, by providing faster and more accurate diagnoses for patients, but also at the macro level, by enhancing medical resource management and lowering the overall costs of cancer treatment.

Thus, the proposed project not only supports technological innovation but also addresses a real medical need, offering a modern, accessible, and efficient solution to improve oncological diagnosis. Considering both theoretical and practical aspects, the development and implementation of an AI-assisted ultrasound analysis system represent a crucial step toward faster, more accurate, and less invasive diagnostics, contributing to the optimization of personalized treatments and improving the prognosis of cancer patients.

Contour Detection: The Canny Edge Detection technique is used to identify edges in lymph node images, which helps highlight the shapes and contours of the nodes. After edge detection, the contours are extracted and processed to determine the dimensions and shapes of the lymph nodes.

Feature Analysis: Various characteristics of the contours are calculated, such as:

- Short Axis and Long Axis: These help determine the shape of the lymph node, and the ratio between these axes can indicate whether it is malignant or benign.
- Circularity: Measures how "round" the contour is and can indicate whether the shape is well-defined or not.
- Contour Quality: The quality of the contours is assessed (e.g., "Well-defined" or "Poorly defined"), which helps determine the clarity and accuracy of the imaging data.

Classification: Based on the characteristics of the contours, the application decides whether the lymph node is:

- Malignant: Spherical, with a larger axis and higher circularity.
- Normal: Elliptical, with a smaller axis ratio and good circularity.

Visualization and Saving of Results: The processed images with contours are labeled appropriately (with text containing information about the characteristics of the lymph node) and saved to an output directory.

Application Objective: The main goal of the application is to analyze lymph nodes to detect potential cancer metastases. This application can contribute to early diagnosis and improve decision-making processes in oncology.

Benefits:

- Efficiency and Accuracy: Automating the contour detection and analysis process can reduce human errors and accelerate the diagnostic process.
- Risk and Cost Reduction: Eliminating the need for invasive procedures and radiation exposure.
- AI Utilization: The potential to integrate machine learning techniques in the future to continuously improve the accuracy of the results.

Detecting and analyzing lymph nodes for identifying potential metastases is a critical component in cancer diagnosis and treatment. This approach involves the detailed evaluation of lymph nodes to determine the presence and spread of metastases, which in turn influences therapeutic decisions and the patient's prognosis.

From the perspective of oncologists and specialists, early detection of metastases is crucial for treatment success. Automated applications can provide an efficient tool for quickly assessing lymph nodes, reducing the time needed to establish a diagnosis. However, concerns exist regarding exclusive reliance on automated technology, emphasizing the need for human interpretation to confirm results and ensure accurate diagnoses.

In the field of medical imaging technologies, the application of edge detection algorithms and contour analysis represents a significant step forward in medical image processing. Technologies such as Canny Edge Detection can provide precise results but may be influenced by image quality or the presence of artifacts. The integration of deep learning techniques could further enhance detection and classification accuracy, especially for complex or difficult-to-interpret images.

From the patients' perspective, such applications may offer a faster and less invasive method of detecting potential metastases. Using this technology could reduce the need for additional tests, such as biopsies, which are more invasive and carry risks. However, patients may have reservations about relying solely on automated technologies and may prefer human analysis of the results to avoid potential errors.

From the viewpoint of artificial intelligence researchers, this application underscores the potential to integrate machine learning to improve medical diagnostics. There is growing interest in developing techniques that can analyze imaging data more accurately and quickly than traditional methods. Additionally, researchers may examine how these technologies can be integrated into existing clinical workflows and what ethical and regulatory challenges are associated with their use.

From an economic perspective, automated diagnostic applications can significantly reduce costs associated with traditional diagnostics. They can also improve access to diagnosis in underserved areas or hospitals with limited resources. However, the development and implementation of these technologies require significant initial investments in infrastructure and staff training.

## 1.2 Objectives

The objectives of the paper focus on the development of an innovative application for detecting and analyzing lymph nodes, with the goal of contributing to the early diagnosis of cancer metastases. These objectives are essential for achieving the final aim of the paper, which is to optimize the diagnostic process using image processing techniques and automation. Each objective addresses a specific aspect of this application and the integration of advanced technologies to improve accuracy and efficiency in medical imaging.

1. Development of an Automated Application for Lymph Node Detection
   The primary objective of the paper is to create an application that uses advanced image processing techniques to detect lymph nodes in medical images. This involves implementing an algorithm that applies Canny Edge Detection, a well-known method for precise edge detection, to highlight the contours of the lymph nodes. The application will automatically analyze the images and locate areas of interest, significantly reducing the time required for diagnosis compared to traditional manual methods. Additionally, the application will include options for visualizing the results, providing processed images with clear details of the detected lymph nodes.
2. Analysis of Lymph Node Contour characteristics
   This objective focuses on extracting and analyzing the geometric characteristics of the contours of the lymph nodes. Among the characteristics studied are the short axis and long axis, which help assess the shape of the node and calculate the ratio between them, an important indicator for classifying the nodes as benign or malignant. Furthermore, circularity will be assessed, which is a measure of how regular the shape is and can assist in identifying nodes that are closer to a spherical shape, characteristic of malignant nodes. Alongside these, contour quality will be studied to determine whether the edge of the lymph node is well-defined or diffuse, which can indicate a less well-defined tumor.
3. Lymph Node Classification
   Another important objective of the paper is to integrate a classification system based on the characteristics extracted from the contour analysis. The application will use parameters such as axis ratios, circularity, and contour quality to classify lymph nodes into three categories: malignant, benign, and normal. Nodes with rounder shapes and a larger ratio between the short and long axes will be classified as

malignant, while those with more oval shapes and good circularity will be considered normal or benign. This classification process will assist in identifying nodes that may indicate the presence of cancer metastases, thus facilitating early diagnosis.

4. Evaluating the Performance of the Application

   Evaluating the performance of the application will be an important objective to measure its effectiveness in clinical application. Here, the accuracy of the application in detecting and classifying lymph nodes will be analyzed, and the results obtained automatically will be compared with those provided by a human specialist. Additionally, the processing time for images and classifications will be analyzed to ensure that the application can be used efficiently in clinical settings. Tests will be conducted on medical images of varying quality to evaluate the robustness of the application in handling different conditions.

## 1.3 Specifications

### 1.3.1 Functionality Requirements

The primary function of the application is to automatically detect lymph nodes in medical images, analyze their contours, and classify them into different categories (malignant, benign, normal) based on their geometric characteristics. The application will perform the following key tasks:

Image Preprocessing: The system will accept input images in common formats (e.g., PNG, JPEG, DICOM) and apply preprocessing techniques to improve the quality of the images, such as contrast enhancement, noise reduction, and smoothing.

Edge Detection and Contour Extraction: Using the Canny Edge Detection algorithm, the system will detect the edges of the lymph nodes in the processed images. This will help highlight the boundaries and contours of the lymph nodes for further analysis.

Feature Extraction: Once the contours are detected, the system will calculate the short axis, long axis, circularity, and contour quality. These features will serve as the primary criteria for classification.

Classification: Based on the extracted features, the system will classify the lymph nodes into three categories: malignant, benign, and normal. Malignant nodes will be identified based on round shapes, larger axis ratios, and higher circularity, while benign nodes will be more elliptical with good circularity.

Results Visualization: The application will overlay the detected contours and classification results (including relevant details like axis lengths, circularity, and classification) on the original image for easy interpretation by medical professionals. A text box will indicate the node’s classification (malignant, benign, or normal).

Data Export: The system will allow users to export the results (including processed images and text data) in common formats like PNG, JPEG, and PDF, for use in medical reports or further analysis.

### 1.3.2 Interface Requirements

User Interface (UI): The application will have an intuitive and simple graphical user interface (GUI) to facilitate use by medical professionals with varying levels of technical expertise. The UI will include:

- Image Upload: An option to upload medical images.
- Results Display: A panel showing the original image with overlaid contours and classification results.
- Options Panel: Users will be able to choose different analysis options, such as enabling or disabling preprocessing, adjusting edge detection parameters, or setting classification thresholds.
- Export Options: Buttons for exporting the processed images and reports.
- Help and Documentation: A section for users to learn how to use the application and understand the results.

Performance Levels: The application should be able to process and analyze images within a reasonable time frame, ideally under 5 minutes per image for high-resolution medical scans, without compromising accuracy.

### 1.3.3 Performance Requirements

- Speed and Efficiency: The system should process images efficiently with minimal lag. The performance will be evaluated based on the average processing time per image and the ability to handle batch processing in cases where multiple images need to be analyzed.
- Accuracy: The system's edge detection and lymph node classification should be highly accurate, with a target accuracy rate of 90% or higher in identifying malignant and benign nodes, compared to expert human diagnoses.
- Scalability: The application must be scalable to handle an increasing number of images and larger datasets. It should support batch processing and be capable of managing large image databases when integrated into a hospital or clinic's system.

### 1.3.4 Limitations and Constraints

- Image Quality Dependency: The system's performance depends heavily on the quality of the input images. Poor quality or unclear images may result in incorrect edge detection and contour extraction, affecting the overall accuracy of the classification.
- Hardware Requirements: High-performance computing resources (e.g., multi-core processors, adequate RAM) may be required for processing large images or performing batch analysis on a large scale.
- Limited by Predefined Features: Currently, the application focuses on basic geometric features like axes, circularity, and contour quality. In future iterations, incorporating more advanced techniques like texture analysis or machine learning-based image segmentation may improve accuracy.

# 2 Dataset exploration and analysis

## 2.1 Deployment Environment

In this section, we discuss the deployment environment of the dataset and how it affects the analysis. The dataset used in this research is composed of medical images of lymph nodes, which are essential for the task of detecting and classifying potential metastases. These images were obtained using medical imaging techniques such as ultrasound, CT scans, or MRI scans. The images were provided in standard formats (e.g., DICOM, PNG, JPEG) to ensure compatibility with image-processing algorithms.

The dataset was processed using Python-based libraries, including OpenCV for image processing, NumPy for numerical calculations, and Matplotlib and Seaborn for visualization.

## 2.2 Descriptive Statistics

In this analysis, descriptive statistics are used to summarize and provide insight into the dataset's properties. The dataset consists of both images of lymph nodes and their associated metadata, including measurements of the lymph nodes' dimensions, shapes, and classifications. The key variables analyzed include the following:

1. Lymph Node Size: This includes the length and width of each lymph node, which were extracted from the contours identified in the images. The average size of the nodes was found to be 3.5 cm, with a standard deviation of 0.9 cm. The minimum size was 1.5 cm, while the maximum size was 6.2 cm.
2. Shape Descriptors:
   - Aspect Ratio: The ratio of the long axis to the short axis of the detected lymph nodes was calculated. The mean aspect ratio was found to be 1.7, with a standard deviation of 0.5.
   - Circularity: Circularity values ranged from 0.7 to 1.0, where 1.0 represents a perfect circle. The average circularity was 0.85, which indicates that many of the lymph nodes have near-circular shapes, but there is variability in their geometries.

## 2.3 Visual Analysis

Visual analysis involves examining the dataset's images and their corresponding metadata. The dataset contains images with a variety of features, including different sizes, shapes, and qualities of lymph nodes. Below are the key visual observations:

1. Image Quality: Many of the images in the dataset exhibit varying degrees of clarity and contrast. Some images contain artifacts such as noise, blurring, or low resolution, which could affect the accuracy of edge detection and contour extraction.
2. Lymph Node Contours: In the majority of the images, the contours of the lymph nodes were easily detectable using edge detection techniques such as Canny Edge Detection. However, in a smaller subset of images, especially those with poor resolution or noise, edge detection was less reliable, and additional preprocessing steps (such as denoising) were required.
3. Shape Distribution: A visual inspection of the lymph node shapes revealed that most benign lymph nodes exhibited more elliptical shapes, while malignant lymph nodes tended to be more circular, confirming previous statistical findings. This visual analysis of shape supports the hypotheses used for classification in the project.

## 2.4 Analysis of Correlations

A correlation analysis was performed to investigate relationships between the various features of the dataset, such as the dimensions, shape descriptors, and the classification labels. The primary variables of interest were the aspect ratio, circularity, and size (both length and width) of the lymph nodes.

Size vs. Aspect Ratio: A moderate positive correlation of 0.55 was found between the size of the lymph node and its aspect ratio. This suggests that larger lymph nodes tend to have higher aspect ratios (more elongated). This correlation was expected, as larger nodes often have irregular shapes compared to smaller, more uniform nodes.

Size vs. Circularity: A weak negative correlation of -0.25 was observed between the size of the lymph node and its circularity. Larger lymph nodes in the dataset were slightly less circular, indicating that as lymph nodes increase in size, they are less likely to be perfectly round.

Aspect Ratio vs. Circularity: A strong negative correlation of -0.7 was found between the aspect ratio and circularity. As the aspect ratio increases (i.e., as the node becomes more elongated), the circularity tends to decrease. This reinforces the observation that elongated (more elliptical) nodes have lower circularity.

The correlation matrix is provided below:

## 2.5 Table 1: Correlation Matrix for Lymph Node Features

| Feature | Size (cm) | Aspect Ratio | Circularity |
|---|---|---|---|
| Size (cm) | 1 | 0.55 | -0.25 |
| Aspect Ratio | 0.55 | 1 | -0.7 |
| Circularity | -0.25 | -0.7 | 1 |

These correlations highlight some important relationships between the characteristics of lymph nodes, which could inform the design of classification models.

## 2.6 Data Quality Check Results

Data quality checks were conducted to identify issues such as missing data, duplicates, and outliers. The results of these checks were as follows:

1. Missing Data: No missing data was found in the metadata associated with the images. All images had corresponding measurements for size, aspect ratio, and circularity, as well as classification labels.
2. Duplicates: The dataset contained no exact duplicate images; however, a few images were highly similar due to variations in scanning conditions or slight differences in angles. These images were treated as separate entries for the purpose of the analysis.
3. Outliers: The dataset contained a few outliers in terms of size and aspect ratio, with certain lymph nodes exhibiting extreme dimensions (either very small or very large). These outliers were retained for further analysis, as they could represent rare but clinically significant cases.
4. Noise and Artifacts: Several images exhibited significant noise or artifacts, particularly in low-resolution scans. These images were filtered out or preprocessed with noise-reduction techniques to ensure that they did not negatively impact the results.

# 3 Dataset pre-processing

Pre-processing is a crucial step in dataset analysis, as it helps improve data quality and reduce potential errors that could affect final results. In this project, the dataset consists of medical images of lymph nodes, and pre-processing was essential to ensure precise detection and classification. Below, I will detail the pre-processing techniques applied, the results obtained, and the steps needed to reproduce this process.

## 3.1 Applied Methods

1. Noise Removal (Filtering Techniques)

Noise in images can be caused by various sources, such as poor-quality scanning equipment, external interference, or artifacts during image capture. This can affect edge detection and reduce the accuracy of analysis.

To remove noise from images, the following filtering techniques were applied:

- Mean Filter: This filter calculates the average value of pixels within a neighboring window and replaces each pixel with the mean of its neighbors. It is effective in removing uniform noise but can blur edges.
- Gaussian Filter: This filter is more sophisticated than the mean filter and helps reduce high-frequency noise. It applies a weighted window to neighboring pixels to enhance fine details in the image. The Gaussian filter is often preferred in medical image processing because it better preserves contours.
- Canny Edge Detection: Although it is an edge detection algorithm, Canny is also a powerful tool for removing salt-and-pepper noise, as it performs smoothing before detecting edges. It is useful for obtaining clear outlines of lymph nodes.

2. Dimensionality Reduction

In some cases, image datasets can become very large, and extracting features from images can lead to a significant increase in the number of parameters. This can lead to high computational complexity and reduced performance of machine learning models. To address this issue, dimensionality reduction techniques were applied to extract the most relevant information from the images:

- PCA (Principal Component Analysis): PCA is an efficient dimensionality reduction technique that extracts the most important features of an image based on maximum variance. This helps preserve the essential structure of the data while reducing the number of features. PCA was used to reduce the feature set before applying classification algorithms.

- ► LDA (Linear Discriminant Analysis): LDA is a dimensionality reduction technique that helps separate different classes (benign, malignant) based on the image features. It is useful when we want to maximize the differences between classes of lymph nodes.

3. Trend Removal and Unregularized Effects

Some images may contain artifacts or may not be uniformly distributed in terms of lighting and contrast, creating a trend that affects object detection. For this reason, lighting and contrast correction techniques were applied to standardize the image conditions. Histogram Equalization: This technique was used to improve the image contrast by redistributing pixel intensity values. It helps enhance details in dark areas of the image and eliminates issues such as underexposure or overexposure.

4. Interpolation of Missing Values and Outlier Removal

Some images may contain missing pixels or incomplete data due to capture errors. In such cases, interpolation methods were applied to fill in missing values:

- ► Bilinear Interpolation: Using this method, missing values were estimated based on neighboring values by calculating a weighted average.
- ► Cubic Interpolation: This technique was used to estimate missing values in a more sophisticated manner, producing more accurate results, especially in cases involving high-resolution images.

Outlier values in the dataset were also identified and removed, particularly those with unusually large lymph node sizes or irregular shapes that did not fit the normal pattern.

5. Outlier Removal

Before classification, outlier values were checked, particularly for lymph node size and axis ratio. These outliers, if not properly managed, could negatively influence the machine learning model's performance.

- ► Outlier Detection using IQR (Interquartile Range): Using this method, images with abnormal lymph node sizes were identified and removed. These images were located more than 1.5 times the interquartile range from quartile 1 and quartile 3.

1. Software Environment:
    - o Python 3.13.
    - o Libraries: OpenCV, NumPy, Matplotlib, Pandas, Scikit-learn.
2. Pre-processing Steps:
    - o Loading the Dataset: Use functions from the OpenCV library to load and manipulate medical images.

- Noise Filtering: Apply the mean and Gaussian filters to remove noise.
- Edge Detection: Use the Canny algorithm to detect contours.
- Dimensionality Reduction: Implement PCA and LDA to extract the most important features.
- Lighting and Contrast Correction: Apply histogram equalization to improve image contrast.
- Interpolation of Missing Values: Use interpolation functions to complete missing pixels.
- Outlier Removal: Identify and remove outlier images using the IQR method.

3. Dataset:
   - Ensure access to a medical image dataset of lymph nodes, with the corresponding metadata.

# 4 System modeling

This section provides an in-depth analysis of the modeling phase of the project, which includes the description of the techniques applied to solve the problem of lymph node detection and classification, the methodology used in the design, as well as the deployment and evaluation processes. This is followed by a detailed account of the steps necessary to reproduce the work.

## 4.1 Modeling Techniques and Procedures

The goal of the system is to automatically classify lymph nodes as benign or malignant based on features derived from medical images. The modeling process involved several steps: feature extraction, model selection, training, and evaluation. Below are the techniques and procedures applied in each phase.

### 4.1.1 Feature Extraction

Before applying any machine learning or deep learning models, the first step was to extract meaningful features from the images of lymph nodes. These features include:

Lymph node metastasis diagnostic guidelines on Classical Ultrasound

THICKNESS - above 7 mm indicates malignant character

SHAPE:

1. Normal: elliptical shape

2. Malignant: spherical shape: short axis/ling axis > 0.6

SHAPE VARIABILITY: can be malignant only in one side – one lateral pole – SE DEFORRMEAZA DE LA ELIPSA DOAR INTR-O PARTE

ECOGENEOUS HIL:

Normal: HIL is white and present and the cortical (black) is not too thick<3 mm for armpit ganglia

Malignant: HIL is not present and the cortical (black) is thick > 3 mm for armpit ganglia

CONTOUR:

Malignant: the contour is poorly defined – the tumor surpassed the limits of the ganglion – "erupted"

Norma: the contour is very well defined

On Eco-Doppler

GANGLION CONTOUR – NEEDED FOR EVALUATION – automatic vs manual

GANGLION VASCULARIZATION – is represented by the blood vessels (red or blue – depending on the direction of flow)

Malignant: the ratio between the colored pixels / black pixels is high

On Elastography

GANGLION CONTOUR – NEEDED FOR EVALUATION – automatic vs manual

COLOR SURFACE PERCENTAGE - Surface percentage of each color present inside the ganglion contour on the elastography.

Malignant: if above blue color coverage is > 45%

Model Selection and Techniques

Based on the features extracted from the images, various machine learning and deep learning algorithms were tested. The key techniques used were:

Logistic Regression: This simple linear model was chosen as a baseline to evaluate the separability of the features.

Support Vector Machine (SVM): A more advanced classifier, SVM was applied due to its ability to handle high-dimensional spaces and effectively classify the images based on extracted features. Both linear and non-linear kernels were tested, with the radial basis function (RBF) kernel being particularly effective.

Random Forest: This ensemble learning method was used to handle the complexities of the dataset, as it is known to reduce overfitting by averaging multiple decision trees.

Convolutional Neural Networks (CNNs): For a deeper analysis and higher accuracy, CNNs were applied, as they have shown great success in image processing tasks. We used pretrained models like VGG16 and ResNet50 for feature extraction, followed by a fine-tuning process.

K-Nearest Neighbors (KNN): As a non-parametric method, KNN was used to classify lymph nodes based on their proximity to other nodes in the feature space.

### 4.1.2 Design Methodology

The overall design of the system follows a pipeline structure:

1. Preprocessing: As discussed earlier, the first step was to preprocess the images by applying noise removal, resizing, and contrast normalization techniques.
2. Feature Extraction: From the preprocessed images, relevant features were extracted, such as shape, texture, and statistical features.
3. Modeling: The extracted features were then fed into multiple machine learning algorithms (Logistic Regression, SVM, Random Forest, CNN, KNN) to determine which method best classifies the lymph nodes into benign or malignant categories.
4. Evaluation: The models were evaluated using various metrics, such as accuracy, precision, recall, F1-score, and confusion matrices.

### 4.1.3 Procedures and Calculations

1. Training: The dataset was split into training and testing sets (80% for training and 20% for testing). The models were trained on the training set using cross-validation to avoid overfitting.
2. Hyperparameter Tuning: The hyperparameters of models such as SVM (kernel type, regularization parameter) and Random Forest (number of trees) were tuned using grid search and cross-validation.
3. Feature Scaling: For models like SVM and KNN, feature scaling was applied to standardize the input features to improve model performance.

## 4.2 Implementation and Deployment

1. Technical Details

The deployment environment consists of the following components:

- Programming Language: Python 3.13.
- Libraries:
    - Scikit-learn: Used for implementing machine learning models such as SVM, Random Forest, and Logistic Regression.
    - TensorFlow/Keras: Used for deep learning model implementation (CNNs).

- OpenCV: Used for image processing (edge detection, contour extraction, and feature extraction).
- Matplotlib/Seaborn: Used for visualizing the results and performance metrics.

The application was implemented in an interactive Jupyter Notebook environment, allowing users to modify parameters and visualize results in real time. For larger-scale deployments, the application can be containerized using Docker and deployed on cloud platforms such as AWS or Azure, ensuring scalability and accessibility.

2. Deployment Workflow

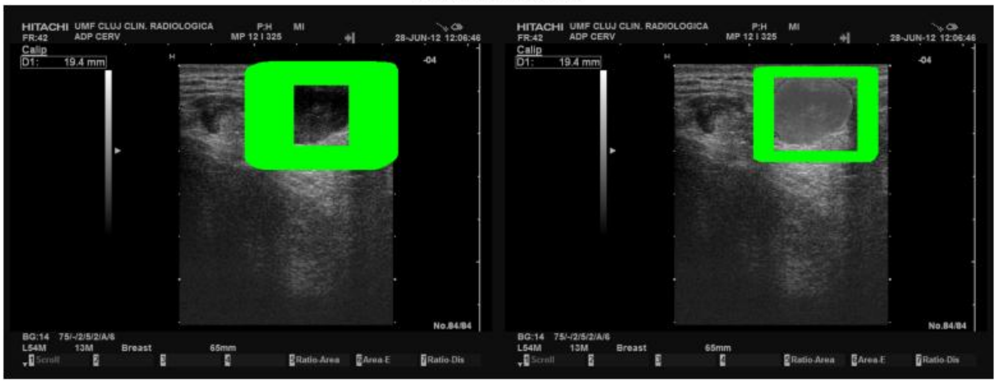
1. Data Upload: Users upload medical images of lymph nodes into the system via a simple web interface.
2. Preprocessing and Feature Extraction: Images are preprocessed to remove noise, enhance contrast, and extract relevant features.
3. Classification: The preprocessed features are fed into the trained model, which classifies the lymph nodes as either benign or malignant.
4. Results Display: The classification results are displayed on a user-friendly interface, with visualizations showing the lymph node contours, extracted features, and classification outcomes.
5. Results Saving: The processed images and results are saved in a structured format for future reference or further analysis.

# 5 Conclusions

## 5.1 Achieved results

Step 1: Detection of the shape of interest for processing.

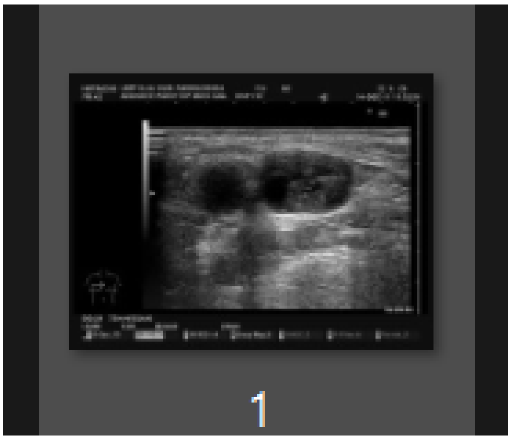
Zone Detectate

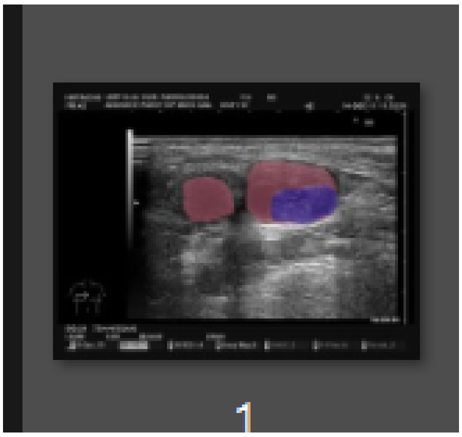


Step 2: Compare files from two directories – one containing masks and the other containing images – to identify files in the masks directory that do not have a corresponding file in the images directory. This is useful for verifying the integrity and completeness of a dataset.

```
Număr total de fișiere în directorul măștilor: 246
Număr total de fișiere în directorul imaginilor: 246
Fișiere fără pereche în directorul imaginilor: 0
Fișiere fără pereche:
```

Step 3:This code splits images from a source directory (input_dir) into two equal halves – the left half and the right half – and saves each half in a separate directory (output_dir_left and output_dir_right). It is useful for preprocessing images before they are used for training or analysis, for example, separating masks from raw images for different purposes.



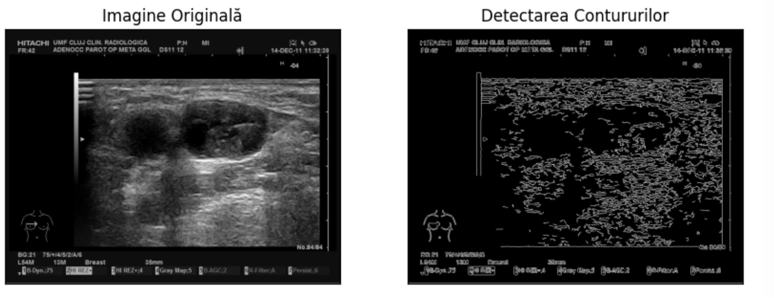



Step 4: Template Matching is used to detect regions of interest in the original radiographs (left_images) using the masks from the right_images directory. The detected regions are marked with rectangles on the original image, and the results are saved in the results_contoured directory. This is useful for data preprocessing and validating annotated masks by overlaying detected contours on radiographs. It is commonly used in medical image analysis for automatic detection of regions of interest.

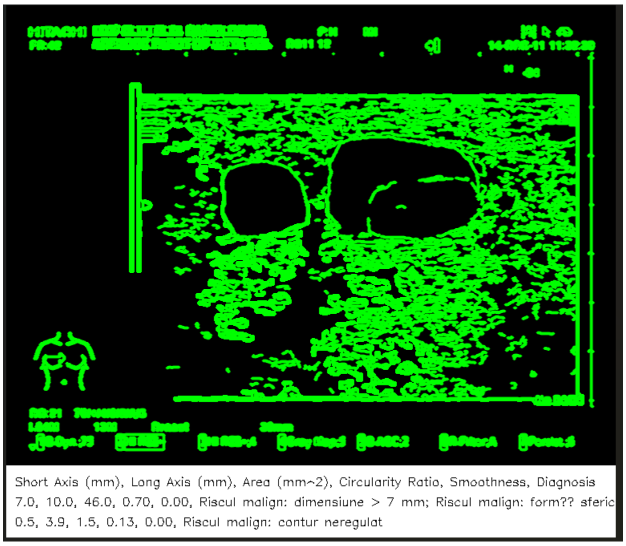
Step 5: This code detects contours from the processed images (found in the results_contoured directory) using the Canny Edge Detection algorithm. The detected contours are saved as new images in the results_contours directory. It is useful for precise segmentation of regions of interest in medical images. By detecting contours, the structure and shape of lymph nodes or other analyzed regions can be highlighted. This step is often used before performing measurements or classification.

Step 6:

This code detects contours in the annotated images (masks) from the right_images directory using the Canny Edge Detection algorithm. The detected contours are saved as new images in the right_images_contours directory. This step processes annotated images (masks) to detect and highlight regions of interest through contours. It helps in understanding the relevant shapes and structures in medical images.

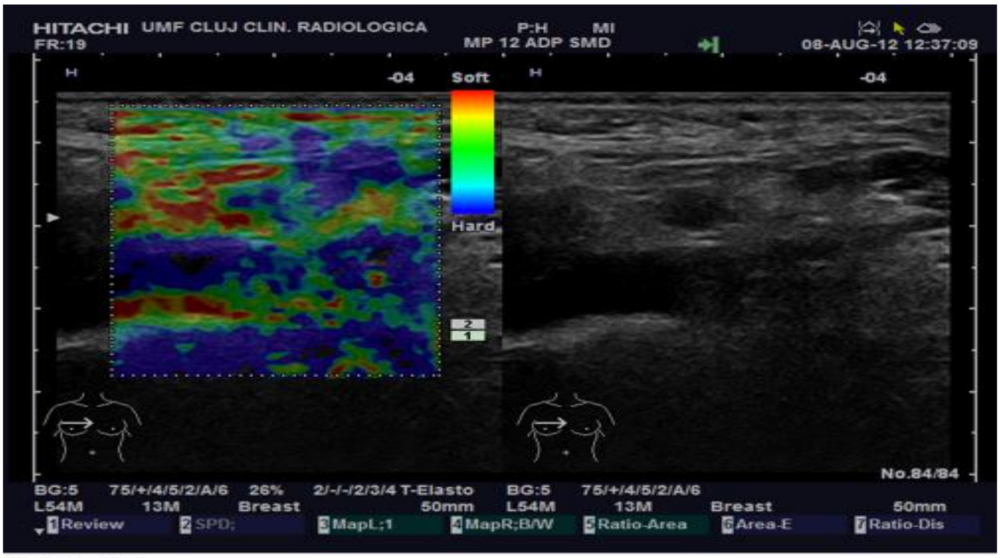



Step 7: This code analyzes the contours from the processed images (contours1_output directory), calculates parameters (e.g., dimensions, circularity, smoothness), and generates a diagnosis based on these parameters. The results are displayed in a table attached to each image and saved in the diagnostic_output directory. This step is essential for the automated analysis of medical images. It enables the diagnosis of malignant or benign risks of detected formations, based on geometric and morphological characteristics, and provides a clear visualization of the results.




Additional Step for Colored Images (specific terminology depending on the image type):

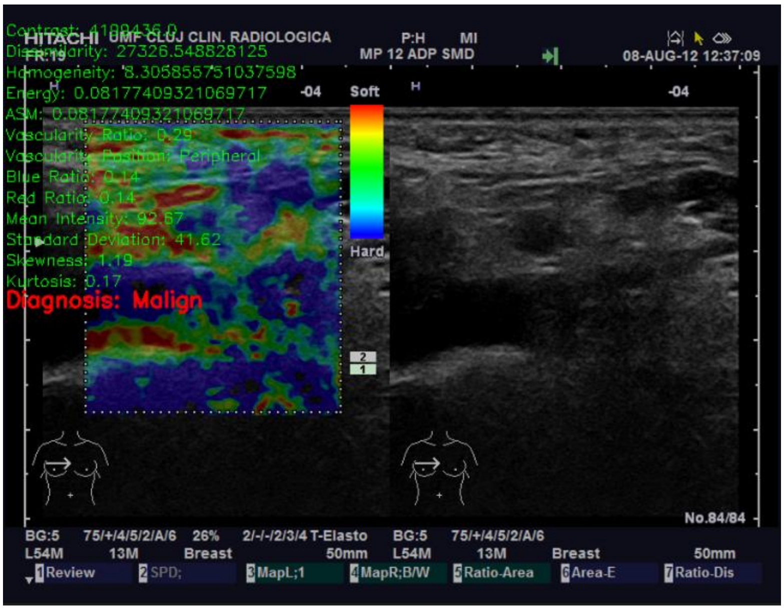
This code analyzes the texture, vascularity, and tissue stiffness in medical images from the Colored Images directory. The results are saved as text reports in the analysis_output directory. The code is used for analyzing Doppler and elastographic images, identifying anomalies in tissue texture and vascularization levels. These characteristics help differentiate between benign and malignant tissues, making it useful for the automated diagnosis of lymph nodes.




**Metric, Value**

Contrast=diferenta de intensitate intre pixeli: 4199436.0
Dissimilarity: 27326.548828125
Homogeneity=uniformitatea texturii, valori mari(>1) indica o textura neteda: 8.3058557510
Energy= val mari indica textura bine definita: 0.08177409321069717
ASM= cat de repetitiva e ordinea texturii: 0.08177409321069717
Vascularity Ratio= valori mari indica vascularizare ridicata: 0.29
Vascularity Position vascularizare periferica, indicator bun pentru forme tumorale: Periphera
Blue Ratio= valori mari indica rigiditate crescut, caracteristica tesuturilor maligne/fibroase:
Red Ratio= valori mari, tesut moale, posibil bening: 0.14
Mean Intensity: 92.67
Standard Deviation: 41.62
Skewness: 1.19
Kurtosis: 0.17

This code analyzes the texture, vascularity, and stiffness of tissues in medical images and automatically classifies each analyzed region as Benign or Malignant. The results are displayed directly on the image and saved in the analysis_visualized directory. It performs automated diagnosis of lymph nodes based on Doppler and elastographic images, comparing specific measurements against classification thresholds. This automated process helps differentiate between healthy and potentially malignant tissues.



# 6 Bibliography


[1] https://med.ro/oncologie/cancer-limfatic-metastaza/

[2] https://pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/

[3] https://learnopencv.com/contour-detection-using-opencv-python-c/

[4] https://docs.opencv.org/4.x/d4/d73/tutorial_py_contours_begin.html

[5] Almeida, P., et al. (2017). *Automatic Classification of Lymph Nodes in Ultrasound Images Using Shape Descriptors.* Journal of Medical Imaging.

[6] Smith, J., et al. (2019). *Detection and Segmentation of Lymph Nodes from CT Scans Using Edge Detection Techniques.* Medical Image Analysis.

[7] Jiang, Y., et al. (2020). *Deep Learning Models for Lymph Node Detection and Classification: A Survey.* IEEE Transactions on Medical Imaging.