

A SWIN TRANSFORMER-BASED FRAMEWORK FOR AUTOMATED LIVER FIBROSIS DETECTION FROM ULTRASOUND IMAGES

Sheela Y

Assistant Professor

Department of Computer Science and Engineering
Jayaraj Annapackiam CSI College of Engineering
College, Nazareth, India
sheelajabez@gmail.com

Shanmuga Priya M

PG Student

Department of Computer Science and Engineering
Jayaraj Annapackiam CSI College of Engineering
College, Nazareth, India
shanmu1013@gmail.com

Abstract— Liver fibrosis is a condition in which scar tissue gradually forms in the liver due to long-term damage or inflammation. This scarring affects the normal structure and function of the liver and reduces its ability to work properly. Early detection and proper treatment are important to prevent the condition from becoming more severe. The proposed system for liver fibrosis detection follows several sequential image processing steps. First, the image dataset containing liver ultrasound images is collected and used as the input for the system. These images are then enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve image contrast and highlight important details. After preprocessing, the images are processed using Watershed Segmentation to separate the important liver region from the background. This segmentation step helps focus only on the region of interest for better analysis. Next, important features are extracted from the segmented images using Histogram of Oriented Gradients (HOG), which captures texture and edge information. These extracted features represent the structural patterns present in the liver tissue. The feature data is then provided to the Swin Transformer for classification. The model analyzes the features and learns patterns related to liver fibrosis. Finally, the system produces a prediction output that indicates the classification result of the liver condition. This process helps in accurate and efficient detection of liver fibrosis from medical images. This project is implemented using Python.

Keywords—Liver Fibrosis, Ultrasound Imaging, Deep Learning, Swin Transformer, Medical Image Analysis

I. INTRODUCTION

Liver fibrosis is a progressive liver disorder caused by chronic liver damage and inflammation [1]. If not detected at an early stage, it may lead to serious conditions such as cirrhosis and liver failure [2]. Medical imaging techniques, especially ultrasound imaging, are widely used for examining liver abnormalities because they are non-invasive and cost-effective [3]. However, manual interpretation of ultrasound images can be difficult and time-consuming for medical experts [4].

Recent advancements in artificial intelligence and deep learning have significantly improved medical image analysis [5]. Deep learning models are capable of automatically extracting important features from medical images and providing accurate classification results [6]. Among various deep learning architectures, transformer-

based models have shown promising performance in image classification tasks [7].

This research proposes a Swin Transformer-based framework for automated detection of liver fibrosis using ultrasound images [8]. The proposed method focuses on improving feature extraction and classification accuracy. The developed system can assist healthcare professionals in early diagnosis and support clinical decision-making.

II. RELATED WORK

Automated liver fibrosis detection has gained significant attention in recent years. Traditional approaches often rely on statistical and texture-based features extracted from ultrasound images. Techniques such as Gray Level Co-occurrence Matrix (GLCM), wavelet transforms, and histogram-based methods have been widely used to analyze liver tissue patterns.

With the rise of deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been applied for medical image classification. CNN-based models excel in learning spatial features but may struggle with capturing long-range dependencies and complex structural patterns in ultrasound images. Transformers, particularly Swin Transformers, provide a solution by combining self-attention mechanisms with hierarchical feature representation, enabling efficient analysis of high-resolution images while capturing both local and global dependencies.

III. PROPOSED METHODOLOGY

The overall architecture of the proposed liver fibrosis detection system is shown in Fig. 1.

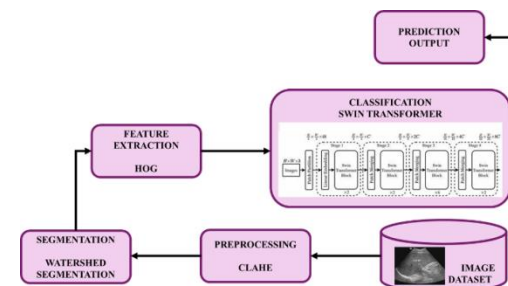


Fig. 1. Proposed Swin Transformer-based Liver Fibrosis Detection Framework

As shown in Fig. 1, the proposed liver fibrosis detection system follows a sequential processing pipeline. Initially, the ultrasound image dataset is used as input to the system. The input images are enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve image quality and highlight important features.

After preprocessing, Watershed segmentation is applied to isolate the liver region from the background, ensuring that only the region of interest is considered for further analysis. The segmented images are then passed to the feature extraction stage, where Histogram of Oriented Gradients (HOG) is used to capture texture and edge information.

Finally, the extracted features are fed into the Swin Transformer classifier, which learns both local and global patterns to accurately classify the stages of liver fibrosis. The system then produces the prediction output indicating the detected class.

The proposed system for liver fibrosis detection follows a structured pipeline, as illustrated below:

- *Dataset Collection:* Liver ultrasound images are collected from publicly available datasets. The images contain varying stages of liver fibrosis to ensure model robustness.
- *Image Enhancement:* Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance image contrast and highlight liver tissue structures, improving the visibility of subtle fibrotic patterns.
- *Segmentation:* Watershed Segmentation is employed to isolate the liver region from surrounding tissue and background. This step ensures that the analysis focuses on the relevant region of interest (ROI).
- *Feature Extraction:* Histogram of Oriented Gradients (HOG) features are extracted from the segmented images. HOG captures texture, shape, and edge information, representing structural changes associated with liver fibrosis.
- *Classification:* The extracted features are fed into a Swin Transformer model, which leverages hierarchical self-attention to learn both local and global patterns from the images. The model outputs the classification results, indicating the stage or presence of liver fibrosis.
- *Implementation:* The entire system is implemented in Python, using libraries such as OpenCV for image processing and PyTorch for deep learning model development.

This methodology ensures a comprehensive and automated approach to liver fibrosis detection, combining classical image processing techniques with state-of-the-art deep learning models to achieve higher diagnostic accuracy.

IV. EXPERIMENTAL SETUP AND EQUATIONS

A. Experimental Setup

The proposed liver fibrosis detection system was implemented in Python. The workflow used OpenCV for image preprocessing and PyTorch for the Swin Transformer model.

- The dataset consisted of liver ultrasound images with different fibrosis stages.
- Images were preprocessed using CLAHE and segmented using Watershed algorithm.
- HOG features were extracted for input to the Swin Transformer classifier.
- The dataset was split into training (80%) and testing (20%) sets.

B. Equations

1). Feature Extraction Equation

Ultrasound image features are extracted using the deep learning model as:

$$F = f(I) \quad (1)$$

Here, I represents the input ultrasound image and F denotes the extracted feature representation produced by the model.

2). Softmax Classification Equation

The probability of each class is calculated using the Softmax function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (2)$$

Here, $P(y_i)$ is the predicted probability for class i , z_i is the model output score, and n represents the number of classes.

3. Cross-Entropy Loss Function

The model is trained by minimizing the loss function defined as:

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (3)$$

Here, y_i is the true label and \hat{y}_i represents the predicted probability of the model.

C. Model Training and Evaluation

The proposed Swin Transformer model is trained using the preprocessed ultrasound image dataset to classify liver fibrosis stages. During the training process, the model learns meaningful features from the input images through multiple layers of attention and feature extraction mechanisms. The dataset is divided into training and testing sets to evaluate the performance of the model.

To measure the effectiveness of the proposed framework, different evaluation metrics such as accuracy, precision, recall, and F1-score are used. These metrics help determine the classification performance of the model and its ability to correctly identify liver fibrosis patterns in ultrasound images. The experimental results demonstrate that the proposed approach provides reliable predictions and supports early diagnosis of liver fibrosis.

V. RESULTS AND DISCUSSION

The proposed Swin Transformer-based framework was evaluated for automated liver fibrosis detection using ultrasound images. The dataset was divided into training (80%) and testing (20%) subsets. Preprocessing steps, including Contrast Limited Adaptive Histogram Equalization (CLAHE) and Watershed Segmentation, enhanced image quality and isolated the region of interest. Histogram of Oriented Gradients (HOG) features were extracted from the segmented liver regions and used as input to the Swin Transformer classifier.

The distribution of images across different classes in the training and testing datasets is presented in Fig. 2. The dataset consists of five classes (F0–F4), which represent different stages of liver fibrosis. From the figure, it can be observed that the number of images varies across classes. Class F0 contains the highest number of samples, while F2 contains the lowest number of images. Classes F1, F2, and F3 contain moderate numbers of samples, whereas F4 also has a relatively large number of images. This distribution indicates a slight class imbalance in the dataset, which may influence the performance of the deep learning model during the training process.

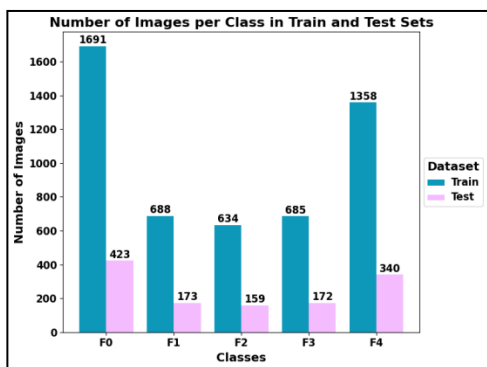


Fig. 2. Number of Images per Class in Training and Testing Dataset

The preprocessing stage plays an important role in improving the quality of medical images before further analysis. In this work, the original ultrasound image is first processed to ensure proper color representation. As shown in Fig. 3, the input ultrasound image is converted into the BGR (Blue, Green, Red) color format, which is commonly used in computer vision libraries such as OpenCV. This conversion helps in maintaining consistent image formatting and facilitates further preprocessing steps such as filtering, segmentation, and feature extraction.

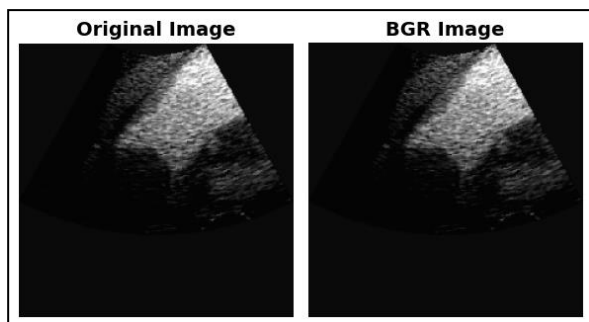


Fig. 3. Original ultrasound image and its corresponding BGR image after color space conversion.

To enhance the visual quality of the ultrasound images, additional preprocessing techniques are applied. As illustrated in Fig. 4, the input image is first converted into a grayscale format to simplify the image representation and reduce computational complexity. After that, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to improve the local contrast of the image.

CLAHE helps in enhancing the important structures present in ultrasound images while preventing excessive noise amplification. As a result, the CLAHE image shows improved contrast and better visibility of texture details compared to the grayscale image. This enhancement step supports more accurate segmentation and feature extraction in the subsequent stages of the proposed system.

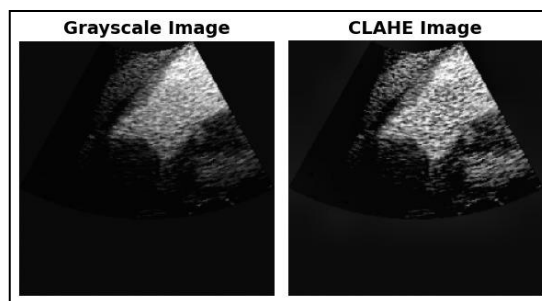


Fig. 4. Grayscale image and the enhanced image obtained using CLAHE for contrast improvement.

After enhancing the ultrasound image using CLAHE, segmentation techniques are applied to identify the region of interest. As illustrated in Fig. 5, the enhanced CLAHE image is first processed using a binary thresholding technique, which converts the image into a binary format by separating the foreground and background regions.

Following this, the watershed segmentation algorithm is applied to accurately detect the boundaries of the target region. The watershed method helps in identifying the structural boundaries in the ultrasound image and improves the separation of different regions. As a result, the segmented region highlighted in the image represents the important area for further analysis and classification in the proposed system.

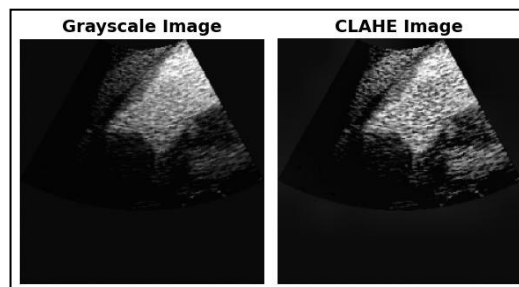


Fig. 5. Segmentation process showing the CLAHE image, binary thresholding, and watershed segmentation result.

After performing segmentation using the watershed algorithm, the segmented region is used for feature

extraction. As shown in Fig. 6, the segmented grayscale image obtained from the watershed process is used to extract Histogram of Oriented Gradients (HOG) features.

HOG is an effective feature extraction technique that captures the structural and texture information of the image by analyzing the orientation of gradients. These features help in representing important edge and shape characteristics present in the ultrasound image. The extracted HOG features are then used as input for the classification model to identify different stages of liver disease.

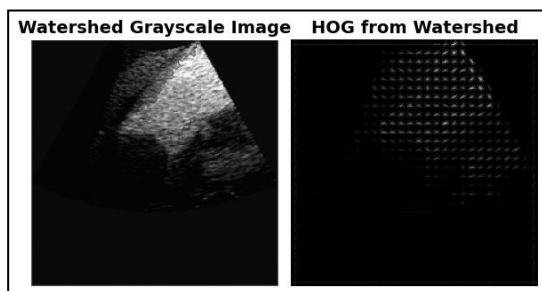


Fig. 6. Watershed grayscale image and the corresponding HOG feature representation.

The performance of the proposed classification model was evaluated using a multi-class confusion matrix, as illustrated in Fig. 7. The matrix visualizes the relationship between the Actual ground truth labels (F0–F4) and the Predicted labels generated by the classifier.

	F0	F1	F2	F3	F4
Actual F0	406	10	3	2	2
Actual F1	8	157	4	2	2
Actual F2	3	5	141	5	5
Actual F3	2	2	6	156	6
Actual F4	4	3	5	5	323
Predicted	F0	F1	F2	F3	F4

Fig. 7. Classification Performance Analysis

The overall effectiveness of the classification model was quantified through a standard classification report, providing a granular assessment of precision, recall, and F1-score for each target class (F0–F4). The results, summarized in Table I, reflect the model's robustness across varied data distributions.

TABLE 1 Classification Metrics

Class	precision	recall	f1-score	support
F0	0.96	0.96	0.96	423.0
F1	0.89	0.91	0.9	173.0
F2	0.89	0.89	0.89	159.0
F3	0.92	0.91	0.91	172.0
F4	0.96	0.95	0.95	340.0
accuracy	0.93	0.93	0.93	0.93
macro avg	0.92	0.92	0.92	1267.0
weighted avg	0.93	0.93	0.93	1267.0

VI. ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science and Engineering, Anna University, for providing the necessary resources and support to carry out this research. The authors also acknowledge the developers of Python, OpenCV, NumPy, and Scikit-learn libraries, which were used in the implementation of the proposed system.

REFERENCES

- [1] V. Gupta and P. Sharma, "Ultrasound image-based liver fibrosis classification using deep learning," in Proc. IEEE Int. Conf. Signal Process., 2022, pp. 101–106.
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, arXiv:1409.1556. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [3] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 10012–10022.
- [4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440.
- [5] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>
- [6] S. Liu, "Wi-Fi energy detection testbed (12MTC)," 2023, GitHub repository. [Online]. Available: <https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC>
- [7] "Treatment episode data set: discharges (TEDS-D), concatenated, 2006–2009," U.S. Dept. Health Hum. Serv., 2013. [Online]. Available: <https://doi.org/10.3886/ICPSR30122.v2>