Technology of Hierarchical Classification Applied to Ensemble Methods for Liver Fibrosis Staging in Ultrasound

Vitalii Babenko
Department of Biomedical Cybernetics
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
vbabenko2191@gmail.com

Ievgen Nastenko
Department of Biomedical Cybernetics
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
nastenko.e@gmail.com

Volodymyr Solodushenko
Department of General and Pediatric Radiology
Institute of Nuclear Medicine and Diagonstic Radiology
Kyiv, Ukraine
solod-vova@ukr.net

Boris Tarasyuk
Department of General and Pediatric Radiology
Institute of Nuclear Medicine and Diagonstic Radiology
Kyiv, Ukraine
btarasyuk13@gmail.com

Volodymyr Pavlov
Department of Biomedical Cybernetics
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
pavlov.vladimir264@gmail.com

Olha Averianova
Department of Biomedical Cybernetics
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
averianova.olga@lll.kpi.ua

2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT) | 979-8-3315-4263-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/CSIT65290.2024.10982652

***Abstract*—Automated liver fibrosis staging is crucial for timely diagnosis and treatment. This paper presents a novel Technology of Hierarchical Classification (Tech-HC) applied to ensemble methods for accurate staging from ultrasound images. Using a dataset of 806 manually segmented regions of interest from 426 ultrasound images, we trained and evaluated Random Forest, XGBoost, LightGBM, and the "Random Forest of Optimal Complexity Trees" (RFOCT) models. LightGBM achieved 82% classification accuracy for F0-1 vs. F2-4, 86% for F0-2 vs. F3-4, and 96% for F0-3 vs. F4, while RFOCT achieved 77% accuracy for F0 vs. F1-4. Integrating these models into our hierarchical classification framework resulted in 99% accuracy for all patients. This technology has the potential to significantly improve diagnostic accuracy and efficiency in liver fibrosis assessment, particularly in resource-constrained settings.**

Keywords—ensemble methods, liver fibrosis, machine learning, medical image analysis, ultrasound imaging.

I. Introduction

Liver diseases represent a significant global health challenge, contributing to substantial morbidity and mortality rates worldwide. Recent estimates indicate that chronic liver diseases globally affect over 1.5 billion individuals, with liver fibrosis being a critical stage in the progression of various hepatic disorders. These disorders include hepatitis B, hepatitis C, non-alcoholic fatty liver disease (NAFLD), and alcoholic liver disease (ALD) [1]. Liver fibrosis, characterized by the excessive accumulation of extracellular matrix proteins, plays a critical role in the development of cirrhosis and hepatocellular carcinoma. These two conditions, collectively, are responsible for over two million deaths annually [2].

Accurate staging of liver fibrosis is essential for effective management and treatment of chronic liver diseases. Early detection and precise staging enable timely therapeutic interventions, which can slow disease progression and potentially prevent irreversible liver damage. Traditionally, liver biopsy has been considered the gold standard for fibrosis assessment. However, this procedure is invasive, costly, and carries the risk of potential complications [3]. This has led to a growing demand for non-invasive imaging techniques, particularly ultrasound, that can reliably stage liver fibrosis.

II. Problem Formulation and Related Works

Building upon the introduction's overview the global burden of liver disease and the need for accurate fibrosis staging, this section delves deeper into the specific challenges addressed by this research. Despite the advancements in non-invasive techniques like ultrasound, the subjectivity of image interpretation and the requirement for specialized expertise hinder widespread adoption and limit access, particularly in resource-constrained settings.

The application of machine learning models has yielded promising results in diagnosing and staging liver fibrosis using various imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound [4-6]. Notably, a recent study [7] demonstrated improvements in classification accuracy by employing a novel colorization technique for ultrasound images. This method achieved an 87% accuracy and an AUC of 0.94, surpassing traditional grayscale image analysis.

A study [8] investigated the effectiveness of combining pre-trained deep convolutional neural networks (CNNs) with a hybrid classifier for classifying liver conditions. The researchers evaluated various CNN architectures, including ResNet50, ResNeXt, and AlexNet, to classify liver images into three categories: normal, hepatitis, and cirrhosis. Their findings indicate that the hybrid classifier significantly improved the multi-classification accuracy to 86.4%, highlighting its potential for clinical applications where accurate differentiation between liver diseases is crucial.

Researchers in [9] developed a CNN model incorporating multi-scale features and attention mechanisms to accurately classify liver fibrosis stages from ultrasound images. This innovative approach achieved a 95.66% accuracy, significantly surpassing existing methods.

This research addresses the need for an automated and accurate method to stage liver fibrosis using ultrasound images. Our goal is to leverage the capabilities of machine learning to enhance the diagnostic accuracy and efficiency of ultrasound-based assessments. The work builds upon our

979-8-3315-4263-4/24/$31.00 ©2024 IEEE

previous research [10-13], which focused on differentiating between normal and pathological liver conditions.

III. Methods

A. Data

This study utilizes a valuable dataset provided by specialists from the State Institution "Institute of Nuclear Medicine and Radiation Diagnostics of the National Academy of Medical Sciences of Ukraine."

The dataset, collected between November 2015 and September 2023, encompasses 119 pediatric patients aged 3 to 18 years (mean age 10.5 ± 2.7 years). A control group of 63 patients exhibited no clinical or biochemical indications of liver damage. The remaining 56 patients presented with diffuse liver disease, with fibrosis stages determined according to the METAVIR scoring system. Specifically, 7 patients were classified as having mild liver fibrosis (F1), 17 as moderate fibrosis (F2), 21 as severe fibrosis (F3), and 11 as cirrhosis (F4).

Specialists at the Institute performed liver ultrasound examinations on each patient by using specialized ultrasound machines capable of producing grayscale images. Distinct types of transducers were employed, including:

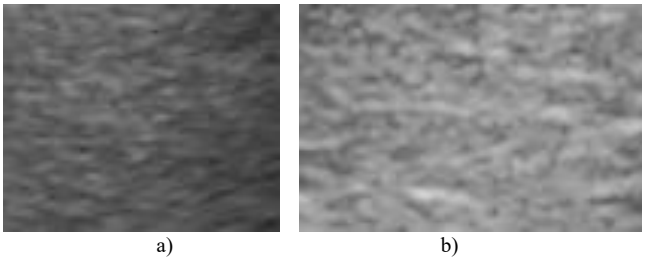
- Convex transducers (3-5 MHz), which offering a wider field of view and deeper tissue penetration.
- Linear transducers (8-10 MHz), which providing higher resolution images for detailed visualization of superficial structures.
- Enhanced mode linear transducers, which enable clearer visualization of liver tissue characteristics.

The dataset consists of 426 ultrasound images. Images have dimensions of 800x608 pixels (4:3 aspect ratio) and stored in the lossless BMP format. This format preserves the original image resolution without compression, ensuring the retention of all details and nuances within the ultrasound images, thus preventing any distortion or loss of information.

To focus the analysis on diagnostically relevant areas, specialists manually segmented regions of interest (ROIs) on each ultrasound image. Each image contained between one and five ROIs, which were extracted and used as the primary objects of study. Examples of these extracted ROIs are illustrated in Fig. 1.

The 806 extracted ROIs, with average dimensions of 88.6x113.5 pixels, were stored in the PNG format. Unlike BMP, PNG employs lossless compression, allowing for efficient file size reduction without compromising image quality.

Table 1 details the distribution of ROIs across the different fibrosis stages. Thus, a sample of ROIs was formed, representing the first stage of the proposed pipeline.



a) b)

Fig. 1. Examples of extracted ROIs:
a) normal liver; b) liver with pathology

TABLE I. Statistics of ROIs Regarding Liver Fibrosis Stages

Liver Fibrosis Stage	ROIs Amount
Normal liver condition	
F0	480 (59.5%)
Pathological liver condition	
F1	44 (5.5%)
F2	103 (12.8%)
F3	129 (16%)
F4	50 (6.2%)
Total	**806**

B. Image analysis pipeline for automated liver diagnosis

1) Histogram processing of images

Due to variations in ultrasound acquisition settings, the pixel intensity histograms of the liver images exhibited significant heterogeneity. This variability could potentially impact the reliability of data analysis. To mitigate this, histogram processing techniques were applied to normalize the pixel histograms within the defined ROIs. This standardization of image characteristics enhances the quality and consistency of the comparative analysis.

In digital image processing, histogram processing techniques are commonly used to optimize image contrast [10-13]. These techniques include normalization (scaling pixel intensities to a specific range), equalization (distributing pixel intensities more evenly across the histogram), and differentiation (enhancing edges and details by calculating pixel intensity differences).

2) Feature Engineering

Texture analysis, an image processing technique that examines the structural properties of an image (e.g., uniformity, roughness, granularity), can differentiate between normal and pathological liver states. Normal liver images typically exhibit darker shades of gray and a uniform, granular texture, while pathological states often present with brighter regions and a non-uniform texture, as illustrated in Fig. 1.

Texture analysis quantifies visual differences in images using computer algorithms that analyze pixel intensity distributions and their spatial relationships. Common methods for texture analysis include:

- First-Order Texture Statistics that analyzing the distribution of individual pixel intensities.
- Gray-Level Co-occurrence Matrix (GLCM) that examining the relationships between pairs of pixels at specific distances and angles.
- Gray-Level Run-Length Matrix (GLRLM) that analyzing the lengths of consecutive pixels with the same intensity.

These methods have been previously explored in our research [10, 13].

3) Feature Selection

Feature selection is performed using the correlation-based feature selection method [14]. This method prioritizes feature subsets that exhibit high correlation with the target variable while minimizing intercorrelation among the features themselves. This approach aims to mitigate multicollinearity that can arise from high interdependence between features and negatively impact model performance.

The central idea is to identify a subset of k independent features that optimize a given criterion (1) for evaluating feature subsets S:

$$S_k = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (1)$$

where: $\overline{r_{cf}}$ – the average of the absolute values of the correlations of all independent features with the dependent variable, $\overline{r_{ff}}$ – the average of the absolute values of the correlations among all features within the subset.

4) Development of Analytical Models

This stage involves a complex Technology of Hierarchical Classification (Tech-HC), as depicted in Fig. 2. The algorithm operates through a multi-level hierarchical structure:

- Level 1: the five stages of liver fibrosis (F0-F4) are decomposed into four binary classification tasks: "F0 vs. F1-4," "F0-1 vs. F2-4," "F0-2 vs. F3-4," and "F0-3 vs. F4." This approach, like that used in [12], simplifies the initial classification problem.
- Level 2: a meta-classifier integrates the results from the four binary classifiers to determine the specific fibrosis stage for each ROI.
- Level 3: a final meta-classifier aggregates data from the ROIs associated with each subject to make the final determination of the liver fibrosis stage for that individual.

The initial stage of the Tech-HC involved evaluating four ensemble machine learning methods: Random Forest, XGBoost, LightGBM, and Random Forest of Optimal Complexity Trees (RFOCT), a novel classification algorithm proposed in our previous work [13]. These methods were selected based on their proven effectiveness and reliability in previous research, as demonstrated in our prior studies [10-13]. Furthermore, ensemble methods were prioritized over deep learning approaches due to their superior performance with limited datasets, reduced computational requirements, and inherent interpretability, which are crucial factors in the clinical context of this research.

To ensure the accuracy and objectivity of evaluating the selected ensemble methods for the multi-classification task, the following conditions were established:

- 9:1 ratio was used to divide the dataset into training and testing sets.
- 10-fold cross-validation was employed for hyperparameter optimization of the modeling algorithms.
- The Synthetic Minority Oversampling Technique (SMOTE) was implemented to address class imbalance within the dataset.

IV. Results

Table 2 presents the performance metrics (accuracy, sensitivity, and specificity) for each binary classification scenario analyzed.

A comparison of the ensemble methods revealed that LightGBM consistently outperformed other models for the tasks "F0-1 vs. F2-4," "F0-2 vs. F3-4," and "F0-3 vs. F4." However, RFOCT demonstrated superior performance for the task "F0 vs. F1-4."

A multinomial logistic regression meta-classifier was employed to integrate the results from the individual classification models, enabling the prediction of specific fibrosis stages for each object. Table 3 presents the performance of this approach, evaluated using multi-classification metrics such as sensitivity, precision, accuracy, and F1-score. Notably, the probabilities of classes "1-4," "2-4," "3-4," and "4," as calculated by the binary classification models, served as input features for the meta-classifier.

This approach achieved a 69% accuracy on the test set, with sensitivity scores of 0.73 for F0, 0.33 for F1, 0.38 for F2, 0.86 for F3, and 0.75 for F4. This performance surpassed that of the standard one-vs-all multi-classification scheme, as detailed in Table 4.

A majority voting procedure, applied to the multi-classification results for each ROI within a subject, yielded the results presented in Table 5. This approach achieved a 99% accuracy on the cohort of patients, with sensitivity scores of 1.0 for F0, 1.0 for F1, 0.97 for F2, 0.98 for F3, and 1.0 for F4.

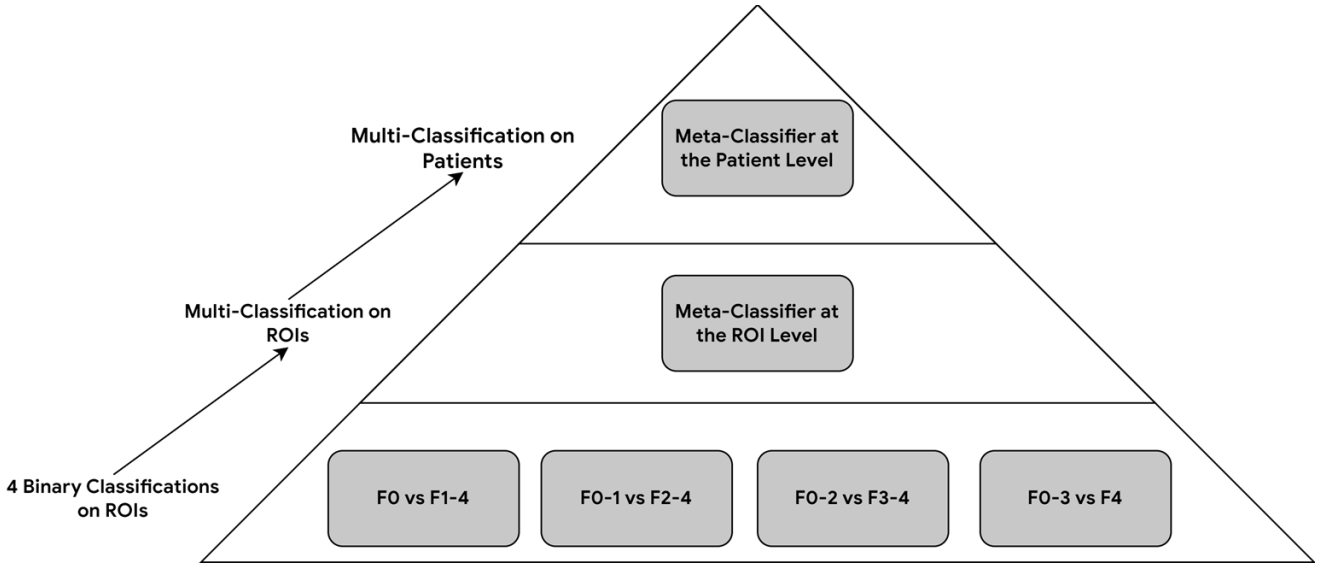


Fig. 2. Technology of Hierarchical Classification

TABLE II. RESULTS OF THE FIRST STAGE OF THE TECH-HC

Task	Level 1 Model	Performance Metrics on Test Set		
		Accuracy	*Sensitivity*	*Specificity*
F0 vs F1-4	RFOCT	0.77	0.75	0.78
F0-1 vs F2-4	LightGBM	0.82	0.81	0.82
F0-2 vs F3-4	LightGBM	0.86	0.83	0.87
F0-3 vs F4	LightGBM	0.96	0.75	0.97

TABLE III. RESULTS OF THE SECOND STAGE OF THE TECH-HC

Fibrosis Class	Performance Metrics on Test Set			
	Sensitivity	*Precision*	*Accuracy*	*F1-score*
F0	0.73	0.82	–	0.77
F1	0.33	0.40		0.36
F2	0.38	0.38		0.38
F3	0.86	0.63		0.73
F4	0.75	0.60		0.67
Total	**–**	**–**	**0.69**	**0.69**

TABLE IV. RESULTS OF THE STANDARD ONE-VS-ALL MULTI-CLASSIFICATION SCHEME

Fibrosis Class	Performance Metrics on Test Set			
	Sensitivity	*Precision*	*Accuracy*	*F1-score*
F0	0.76	0.86	–	0.80
F1	0.50	0.50		0.50
F2	0.50	0.24		0.32
F3	0.50	0.58		0.54
F4	0.50	0.67		0.57
Total	**–**	**–**	**0.65**	**0.68**

TABLE V. RESULTS OF THE THIRD STAGE OF THE TECH-HC

Fibrosis Class	Performance Metrics on Patients Cohort			
	Sensitivity	*Precision*	*Accuracy*	*F1-score*
F0	1.00	1.00	–	1.00
F1	1.00	1.00		1.00
F2	1.00	0.94		0.97
F3	0.95	0.91		0.98
F4	1.00	1.00		1.00
Total	**–**	**–**	**0.99**	**0.99**

V. CONCLUSIONS

This paper proposes a Technology of Hierarchical Classification (Tech-HC). This algorithm leverages the results of group classification of regions of interest (ROIs) as meta-features for a first-level meta-classifier to determine the class of each ROI. A second-level meta-classifier then integrates these ROI-level classifications to produce a final class assessment for the entire subject.

The Tech-HC not only identifies the specific stage of fibrosis but also adapts to individual patient characteristics. At the ROI level, the algorithm achieved a 69% accuracy on the test set, with sensitivity scores ranging from 0.33 to 0.86 for individual fibrosis stages. Furthermore, at the subject level, the algorithm achieved a 99% accuracy, correctly identifying the fibrosis stage in 118 out of 119 cases.

In conclusion, the results presented in this study demonstrate the potential of the proposed Tech-HC algorithm to improve the diagnosis of liver fibrosis stages. This technology offers a foundation for developing decision support systems based on non-invasive ultrasound examinations.

REFERENCES

[1] G. A. Roth et al., "Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the global burden of disease study 2017," The Lancet, vol. 392, no. 10159, pp. 1736–1788, 2018. doi: 10.1016/s0140-6736(18)32203-7

[2] H. B. El-Serag, "Epidemiology of Viral Hepatitis and hepatocellular carcinoma," Gastroenterology, vol. 142, no. 6, 2012. doi: 10.1053/j.gastro.2011.12.061

[3] D. C. Rockey et al., "Liver biopsy," Hepatology, vol. 49, no. 3, pp. 1017–1044, 2008. doi: 10.1002/hep.22742

[4] K. Yasaka et al., "Deep learning for staging liver fibrosis on CT: A pilot study," European Radiology, vol. 28, no. 11, pp. 4578–4585, May 2018. doi: 10.1007/s00330-018-5499-7

[5] S. J. Hectors et al., "Fully automated prediction of liver fibrosis using deep learning analysis of gadoxetic acid–enhanced MRI," European Radiology, vol. 31, no. 6, pp. 3805–3814, Nov. 2020. doi: 10.1007/s00330-020-07475-4

[6] Y. H. Kim, "Artificial Intelligence in medical ultrasonography: Driving on an unpaved road," Ultrasonography, vol. 40, no. 3, pp. 313–317, Jul. 2021. doi: 10.14366/usg.21031

[7] S. Hirata et al., "Convolutional neural network classification of ultrasound images by liver fibrosis stages based on echo-envelope statistics," Front. Phys., vol. 11, p. 1164622, 2023. doi: 10.3389/fphy.2023.1164622

[8] P. Pasyar et al., "Hybrid classification of diffuse liver diseases in ultrasound images using deep convolutional neural networks," Informatics in Medicine Unlocked, vol. 22, p. 100496, 2021. doi: 10.1016/j.imu.2020.100496

[9] X. Feng, X. Chen, C. Dong, Y. Liu, Z. Liu, R. Ding, and Q. Huang, "Multi-scale information with attention integration for classification of liver fibrosis in B-mode US image," Computer Methods and Programs in Biomedicine, vol. 215, p. 106598, 2022. doi: 10.1016/j.cmpb.2021.106598

[10] I. Nastenko et al., "Liver Pathological States Identification in Diffuse Diseases with Self-Organization Models Based on Ultrasound Images Texture Features," in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), Zbarazh, Ukraine, 2020, pp. 21–25. doi: 10.1109/CSIT49958.2020.9321999

[11] D. Hrishko et al., "Optimal complexity structures of ultrasound image models in Diagnostic Decision Making System," in 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), Sep. 2021. doi: 10.1109/csit52700.2021.9648686

[12] E. Nastenko et al., "Analysis of ultrasound images of chronic diffuse liver diseases using texture parameters of gray gradations," Radiation Diagnostics, Radiation Therapy, vol. 13, no. 3, p. 7, 2022. doi: 10.37336/2707-0700-2022-3-1

[13] V. Babenko et al., "Classification of Pathologies on Medical Images Using the Algorithm of Random Forest of Optimal-Complexity Trees," Cybernetics and Systems Analysis, 2023. doi: 10.1007/s10559-023-00569-z

[14] N. Gopika and A. M. Kowshalaya, "Correlation based feature selection algorithm for Machine Learning," in 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Oct. 2018. doi: 10.1109/cesys.2018.8723