USLF-Net: A Task-Specific CNN for Liver Fibrosis Classification from Ultrasound with Mobile Clinical Integration

Tan Tai Phan[1], Thi Cam Hong Pham[1], Nguyen Minh Thao Phan[1,2], Cong-Tinh Dao[1,2], Hai Thanh Nguyen[1]
[1] Can Tho University, Can Tho City, Vietnam
[2] National Yang Ming Chiao Tung University, Hsinchu City, Taiwan
Correspondence: Hai Thanh Nguyen, nthai.cit@ctu.edu.vn
Communication: received 30 June 2025, revised 28 August 2025, accepted 01 October 2025
Digital Object Identifier: https://doi.org/10.32913/mic-ict-research.v2025.n3.1403

Abstract: Accurate liver fibrosis classification is crucial for making treatment decisions in patients with chronic liver disease. Liver biopsy and advanced noninvasive imaging modalities, including elastography, computed tomography (CT), and magnetic resonance imaging (MRI), remain challenges in resource-limited settings due to their high cost and specialized equipment requirements. Consequently, conventional B-mode ultrasound is widely available and cost-effective. Despite the considerable promise of deep learning in medical imaging, prior works often have high computational demands, which limit their practicality. In this study, we introduce USLF-Net, a novel UltraSound-Based Liver Fibrosis Classification using Convolutional Neural Network (CNN). Our model was validated using a comprehensive dataset comprising 6,323 ultrasound images, stratified across five fibrosis stages (F0-F4) according to the Meta-analysis of Histological Data in Viral Hepatitis (METAVIR) scoring system. USLF-Net achieved 97.64% classification accuracy, demonstrating superior performance compared to recent transfer learning approaches. We further developed a clinical application for automated liver fibrosis diagnosis to enhance diagnostic capabilities in healthcare settings.

Keywords: *Convolutional Neural Network, Diagnostic Systems, Liver Fibrosis Classification, Ultrasound Image*

I. Introduction

Chronic liver disease poses a substantial global health burden, with cirrhosis serving as the primary pathological mechanism leading to progression to hepatocellular carcinoma [1]. Multiple disease stages characterize liver cirrhosis, and end-stage cirrhosis is often associated with severe clinical complications, including portal hypertension, liver failure, and the risk of hepatocellular carcinoma [2]. Accurate automated classification of cirrhosis status is critical for determining treatment strategies, monitoring disease progression, and assessing response to therapy [3]. Recently, Deep Convolutional Neural Networks (DCNNs) have demonstrated promising performance in classifying cirrhosis stages according to the METAVIR scoring system using the ultrasound imaging [4]. The METAVIR classification system provides a standardized framework for assessing fibrosis, classifying disease progression into five distinct stages based on pathologic features.

There are several medical methods for diagnosing liver fibrosis, with liver biopsy being the gold standard for traditional diagnostic techniques. However, this method has inherent limitations, including invasiveness, inaccurate sampling, patient discomfort, high cost, and potential side effects [5]. In addition, advanced imaging modalities such as elastography, CT, and MRI offer high diagnostic accuracy for assessing liver fibrosis; however, their dependence on expensive equipment and specialized expertise limits their use in resource-poor settings where chronic liver disease (CLD) is most prevalent. Therefore, conventional B-mode ultrasound (US) is widely accessible, cost-effective, and noninvasive, making it a practical alternative for large-scale screening and routine clinical use [6]. Although the relatively low diagnostic sensitivity of B-mode ultrasound limits its effectiveness in accurately staging liver fibrosis, its advantages include real-time imaging capabilities, portability, and low operating costs, making it particularly suitable for both initial assessment and longitudinal follow-up in various healthcare settings.

Recent breakthroughs in deep learning, particularly convolutional neural networks (CNNs), offer a promising method for automating and directly extracting fibrosis features from B-mode ultrasound images. While CNNs applied to elastography data have yielded high performance,

studies such as Lee et al. [7] demonstrate that models trained on large-scale conventional ultrasound datasets can achieve accuracy levels comparable to or exceeding those of human radiologists. Given the widespread availability and affordability of conventional ultrasound systems, integrating AI-driven analysis into this modality represents a scalable and globally applicable strategy to democratize liver fibrosis assessment, especially in resource-constrained clinical settings.

Although CNNs have achieved breakthrough performance in medical imaging tasks, including mammography lesion classification [8], hepatocellular carcinoma detection [9], pneumonia diagnosis [10], and histopathological fibrosis quantification [11], their implementation for ultrasound-based liver fibrosis classification faces two fundamental challenges.

The first challenge to address is: *To what extent can deep convolutional architectures effectively discriminate subtle variations in ultrasound imaging while mitigating severe class imbalance in clinical fibrosis datasets?* Automated hepatic fibrosis classification presents a challenging fine-grained recognition problem due to progressive pathomorphological changes with minimal inter-class discriminative features. The pathophysiological progression (F0 → F4 with F0 representing No fibrosis, F1 representing Esophageal fibrosis, F2 representing Periportal fibrosis, F3 representing Septal fibrosis, F4 representing Cirrhosis) manifests as subtle echogenic modifications exhibiting substantial intra-class variance and inter-class similarity [12]. Ultrasound acquisition variability, operator-dependent protocols, and subjective ground-truth annotations compound this challenge. Additionally, natural fibrosis distribution demonstrates severe class imbalance, with early stages (F0-F1) significantly outnumbering advanced cases (F3-F4) due to screening protocols. This disparity in distribution creates biased predictions that favor common stages while failing to identify clinically critical cases of advanced fibrosis.

Another critical challenge is: *How can domain-specific architectural innovations overcome transfer learning limitations to achieve superior performance in ultrasound-based fibrosis classification while ensuring clinical deployment feasibility?* Conventional pretrained models present two key limitations when applied to ultrasound-based liver fibrosis assessment. First, their high computational demands render them impractical for real-time and resource-constrained clinical environments [13–15]. Second, their architectures, initially optimized for natural images, often fail to capture the spatial dependencies and multiscale textural features characteristic of ultrasound data. These challenges underscore the need for task-specific CNN architectures tailored to the unique properties of ultrasound imaging, while facilitating efficient clinical deployment.

To address these challenges, we introduce USLF-Net, a novel CNN architecture from scratch that effectively balances classification accuracy and computational efficiency, offering a practical and deployable solution for automated fibrosis staging using ultrasound imaging.

Our key contributions include:

- We implement USLF-Net, a novel CNN-based architecture for multi-class liver fibrosis staging using B-mode ultrasound images in accordance with the METAVIR scoring system.
- We develop a preprocessing and feature extraction pipeline that incorporates spatial regularization techniques, significantly enhancing the model's robustness.
- We conduct extensive experimental evaluations, comparing our model against several benchmark architectures, and demonstrate its superior performance in various evaluation metrics.
- We integrate the proposed model into a mobile diagnostic system, particularly advantageous in resource-constrained and remote clinical settings, providing an accessible and efficient solution for liver fibrosis diagnosis.

The subsequent sections of this paper are organized as follows. The most recent methods and uses of Deep Learning (DL) in predicting liver fibrosis levels are briefly reviewed in Section II. A DL model for liver fibrosis prediction is constructed as described in Section III. Section IV then presents the dataset description, experiment setup, comparison performance, and further analysis. The clinical application interface and system design are presented in Section V. Lastly, Section VI concludes the paper by providing a summary of the key conclusions and implications of the study.

II. Related Work

1. Machine Learning for Liver Fibrosis Classification

Traditional machine learning (ML) techniques were widely used for liver fibrosis classification. These methods relied on handcrafted features derived from ultrasound or elastography images, such as gray-level co-occurrence matrices (GLCMs), local binary patterns (LBPs), and intensity histograms. Classifiers such as support vector machines (SVMs), random forests (RFs), k-nearest neighbors (KNNs), and logistic regression were commonly used [6]. However, their effectiveness is limited by their limited feature representation capabilities and lack of adaptability across various imaging conditions and patient populations. Traditional ML models can also not capture hierarchical spatial patterns, which are essential for identifying subtle

parenchymal texture changes associated with fibrosis progression [16]. As a result, they have particular difficulty classifying adjacent fibrosis stages (e.g., Portal Fibrosis - F1 vs. Periportal Fibrosis - F2) [17], prompting a move toward more data-driven, adaptive methods.

2. Ultrasound Images Approaches for Liver Fibrosis Classification

Ultrasound-based automated liver fibrosis assessment has emerged as a significant research domain, with elastography modalities including shear wave elastography (SWE) [18–23] and transient elastography (TE) [24] demonstrating consistent efficacy in automated classification tasks [25]. Integrating convolutional neural networks (CNNs) with elastography data has substantially improved performance. Wang et al. [23] demonstrated that CNN models applied to 2D SWE images from patients with chronic hepatitis B outperformed conventional tools. Li et al. [26] developed a TE-based model exceeding standard liver stiffness measurement accuracy for detecting advanced fibrosis and cirrhosis. Xue et al. [20] proposed a multimodal CNN framework that combines grayscale and elastography images, outperforming both LSM and biological models. Similarly, studies [7, 27, 28] utilizing B-mode ultrasound without elastography have reported promising results in liver fibrosis staging. Lee et al. [7] used 13,608 Ultrasound images achieved 83.5% accuracy for METAVIR classification, with their CNN model outperforming five radiologists in identifying F4 fibrosis.

III. Methodology

1. Model Architecture

Fig. 1 illustrates our proposed model, USLF-Net, which is a novel hierarchical CNN for liver fibrosis classification using B-mode ultrasound images. It begins with a preprocessing step that resizes the input images to 224×224 pixels in a single-channel (grayscale) format, standardizing the data across varying acquisition conditions. The core architecture leverages a deep CNN framework for image classification, enabling the model to learn discriminative spatial features relevant to tissue differentiation. The proposed architecture has four convolutional layers designed for hierarchical feature extraction and classification of hepatic fibrosis stages. The initial convolutional layer employs 32 filters with 3×3 kernels, followed by batch normalization for training stability, max pooling for spatial dimensionality reduction, and spatial dropout with a rate of 0.2 to prevent overfitting. The subsequent second and third convolutional layers utilize 64 filters each with 3×3 kernels, maintaining the same regularization pipeline of batch normalization, max pooling, and spatial dropout at 0.2. The fourth convolutional layer expands the feature representation capacity through 128 filters with 3×3 kernels, incorporating batch normalization for gradient flow optimization. The feature outputs are flattened and passed through a fully connected dense layer comprising 256-512 neurons, which serves to integrate the extracted spatial features into high-level representations. The network terminates with a softmax output layer containing five neurons, corresponding to the classification of hepatic fibrosis stages: no fibrosis (F0), portal fibrosis (F1), periportal fibrosis with few septa (F2), septal fibrosis with numerous septa without cirrhosis (F3), and cirrhosis (F4), in accordance with the METAVIR histopathological scoring system.

Our model is trained using the Adam optimizer with an initial learning rate of 0.0001. To enhance convergence and adaptively adjust the learning rate, a `ReduceLROnPlateau` scheduler is integrated with a reduction factor of 0.2, a patience of 3 epochs, and a minimum learning rate of 1×10^{-6}. To prevent overfitting and promote generalization, the model incorporates L2 regularization with a weight decay coefficient of 1×10^{-4}, along with dropout mechanisms including both standard `Dropout` and `SpatialDropout2D`, applied with a rate of 0.2. The model is trained for 100 epochs with a batch size of 32 and utilizes the `sparse_categorical_crossentropy` loss function, appropriate for multi-class classification with integer-labeled data. The model parameters are selected based on various configurations, including learning rate and spatial dropout, as presented in Table II. The training procedure is summarized in Algorithm 1.

2. Network Structures

a) Convolutional Feature Extraction

The proposed architecture is designed to address the specific issues of liver fibrosis categorization, which necessitates a fine-grained distinction across five fibrosis phases (F0-F4). It uses a Conv2D hierarchy to extract multi-scale morphological characteristics from ultrasonographic inputs. These characteristics include modest textural alterations, changes in echogenicity, and parenchymal heterogeneities that develop gradually throughout the fibrosis process. The model can automatically capture these progression-related visual signals by using learnable convolutional kernels rather than handcrafted features. Batch normalization is applied after each convolutional layer to mitigate internal covariate shift and facilitate stable and rapid convergence.

b) Spatial Dimensionality Reduction and Feature Pooling

Max-pooling methods are used to downsample feature maps, retaining dominating activations while decreasing

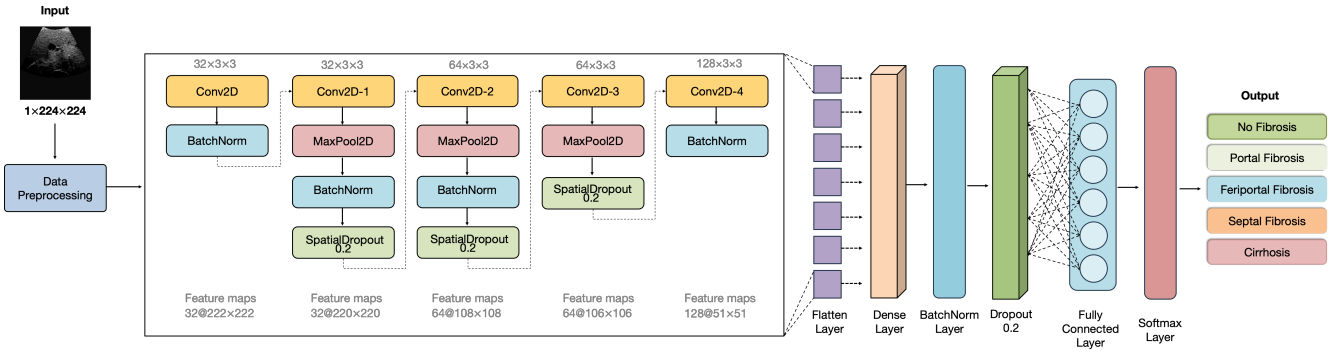

Figure 1. Our proposed model architecture (USLF-Net)

spatial resolution. This spatial abstraction not only reduces computing overhead but also facilitates the learning of translation-invariant features, which are crucial in ultrasound imaging, as anatomical landmarks can change location. By maintaining the most informative spatial information, the network retains the discriminative ability needed for stage-specific fibrosis detection.

c) Regularization and Generalization Strategy

To prevent overfitting, the model incorporates Spatial Dropout (SpatialDropout2D), which is especially important considering the significant intra-class similarity and inter-class heterogeneity among fibrosis phases. This approach deactivates whole feature maps at random during training, forcing the network to rely on scattered and redundant features. This enhances the model's robustness and generalizability, which is particularly crucial for detecting subtle differences between intermediate fibrosis phases (e.g., F2 vs. F3), as these are often visually ambiguous.

d) High-Level Representation and Classification

Following convolutional processing, the multidimensional feature maps undergo fully connected (dense) layers after being flattened into one-dimensional vectors. These layers pick up intricate, nonlinear mappings between fibrosis labels and retrieved characteristics. Instead of approaching the issue as binary or ordinal alone, the design allows for multi-class categorization over five different phases, enabling the model to represent progressive and hierarchical changes in the evolution of fibrosis.

3. Transfer Learning Models

To demonstrate the evolution of computer vision models, from traditional convolutional networks to state-of-the-art transformer-based approaches, we evaluate five representative deep learning architectures: VGGNet-16 [29], ResNet-50 [13], DenseNet-121 [15], EfficientNet-B0 [30], and Vision Transformer (ViT) [31] that provide diverse perspectives on architectural designs. Each model has its own unique characteristics and levels of complexity. VGGNet has a simple, sequential architecture with 3×3 filters and a fixed pooling layer, which is easy to implement but has a large number of parameters. ResNet uses residual blocks with skip connections, which helps overcome the gradient loss problem and supports efficient deep network training. DenseNet enhances the connectivity between layers by concatenation, promoting strong information flow but increasing memory costs. EfficientNet is designed based on the principle of compound scaling, which optimizes depth, width, and resolution simultaneously, thereby balancing performance and computational cost. Meanwhile, ViT employs a transformer architecture in image processing, utilizing a self-attention mechanism to model global relationships between image regions, which yields high efficiency in visual tasks but requires large training datasets and significant computational resources.

All models in this study used a pretrained version of ImageNet. The upper classification layer was removed, and a custom head was added. This head consists of a GlobalAveragePooling layer, a Dropout layer (with a dropout rate of 0.3-0.5, depending on the model), a Dense layer with 256 ReLU activations, and an output Softmax layer for multi-class classification. To address data imbalance, class weights were calculated based on the training set and applied to the training data. All models were compiled using the Adam optimizer, using the sparse categorical cross-entropy loss function, and evaluated using the accuracy metric. In **ResNet50**, the final 19 layers (approximately 10% of the 175 total) were fine-tuned with dropout rates of 0.4 and 0.3, using a learning rate of 1×10^{-6}. In **VGG16** and **DenseNet201**, the final four layers were fine-tuned with dropout rates of 0.4 and 0.3, and a learning rate of 1×10^{-5}. For **EfficientNetB0**, the final ten layers were fine-tuned with dropout rates of 0.5 and 0.4, along with a fully connected layer of 256 units incorporating L2 regularization, trained with a learning rate of 1×10^{-5}. Finally, the **ViT** employed the *vit-base-patch16-224* architecture,

Algorithm 1: Training USLF-Net for Liver Fibrosis Classification

Input: Training set $\mathcal{D}_{train} = \{(X_i, y_i)\}$;
Validation set $\mathcal{D}_{val}$;
Hyperparameters: epochs E=100, batch size B=32, learning rate α, patience p,;
Optimizer Adam$(\alpha, \beta_1, \beta_2)$
Output: Best parameters $\Theta^\star$ (minimizing validation loss)

Initialize Θ; best_val_loss $\leftarrow +\infty$; no_improve $\leftarrow 0$;
for $epoch \leftarrow 1$ **to** E **do**
  Shuffle $\mathcal{D}_{train}$ and split into mini-batches of size B;
  foreach *mini-batch* (X, y) **do**
    Preprocessing
      Normalize X to $[0, 1]$; Resize to 224×224
      Apply augmentation (random rotation, flip, contrast).
    USLF-Net Forward (per batch)
      $F_1 \leftarrow \mathrm{BN}(\mathrm{Conv}^{32}_{3\times3}(X))$
      $F_2 \leftarrow \mathrm{Drop}_{0.2}\big(\mathrm{BN}(\mathrm{MaxPool}(\mathrm{Conv}^{32}_{3\times3}(F_1)))\big)$
      $F_3 \leftarrow \mathrm{Drop}_{0.2}\big(\mathrm{BN}(\mathrm{MaxPool}(\mathrm{Conv}^{64}_{3\times3}(F_2)))\big)$
      $F_4 \leftarrow \mathrm{Drop}_{0.2}\big(\mathrm{MaxPool}(\mathrm{Conv}^{64}_{3\times3}(F_3))\big)$
      $F_5 \leftarrow \mathrm{BN}(\mathrm{Conv}^{128}_{3\times3}(F_4))$;
    Classification Head
      $v \leftarrow \mathrm{Flatten}(F_5)$
      $h \leftarrow \mathrm{Drop}_{0.2}(\mathrm{BN}(\mathrm{Dense}_{\mathrm{ReLU}}(v)))$
      $z \leftarrow \mathrm{Dense}_5(h)$
      $P \leftarrow \mathrm{Softmax}(z)$
    Update
      $\mathcal{L}_{train} \leftarrow \mathrm{CrossEntropy}(P, y)$
      Compute $\nabla_\Theta \mathcal{L}_{train}$ by backpropagation;
      Update $\Theta \leftarrow \mathrm{Adam}(\Theta, \nabla_\Theta \mathcal{L}_{train}; \alpha, \beta_1, \beta_2)$;
  end
  Compute validation loss $\mathcal{L}_{val}$ and accuracy on $\mathcal{D}_{val}$ (no augmentation)
  if $\mathcal{L}_{val}$ < best_val_loss **then**
    best_val_loss $\leftarrow \mathcal{L}_{val}$
    $\Theta^\star \leftarrow \Theta$
    no_improve $\leftarrow 0$;
  end
  else
    no_improve $\leftarrow$ no_improve +1
    if no_improve $\geq p$ **then**
      break ; `// early stopping`
    end
  end
end
return $\Theta^\star$;

where input images were partitioned into 16×16 patches and processed by a Transformer encoder with a learning rate of 1×10^{-5}.

IV. Experimental Results

1. Dataset and Metrics

We utilized the dataset from [32], which includes ultrasound (US) images from university hospitals, especially Seoul St. Mary's Hospital and Eunpyeong St. Mary's Hospital. The dataset includes images divided into five categories that indicate different phases of liver fibrosis: no fibrosis (F0), portal fibrosis (F1), periportal fibrosis (F2), septal fibrosis (F3), and cirrhosis (F4). These grades represent the evolution of liver fibrosis, from no fibrosis (F0) to severe cirrhosis (F4). Table I provides dataset details, whereas Fig. 2 illustrates the distribution of information across classes and Fig. 3 shows the experimental images of F0-F4 classes. Due to data imbalance between F0 and F2 classes, the model tends to favor the class with the larger sample size.

TABLE I
Data statistics used in this work.

Class	#Sample
F0 - No Fibrosis	2114
F1 - Portal Fibrosis	861
F2 - Periportal Fibrosis	793
F3 - Septal Fibrosis	857
F4 - Cirrhosis	1698
Total	6323

To address this issue, this study applied data augmentation to the class with the smaller sample size to increase the amount of training data and improve feature diversity. Specifically, we used image transformations such as rotation, flipping, translation, brightness/contrast changes, and the addition of Gaussian noise. These transformations preserved the pathological features of ultrasound images while creating more diverse data, helping to balance the training model and reduce overfitting.

For model training and evaluation, we divided the dataset into a training set (80%) and a testing set (20%) to ensure a robust model development process. To assess model performance, we employed standard classification metrics including accuracy, precision, recall, and F1-score, which collectively provide a comprehensive assessment of both overall and class-specific predictive capabilities.

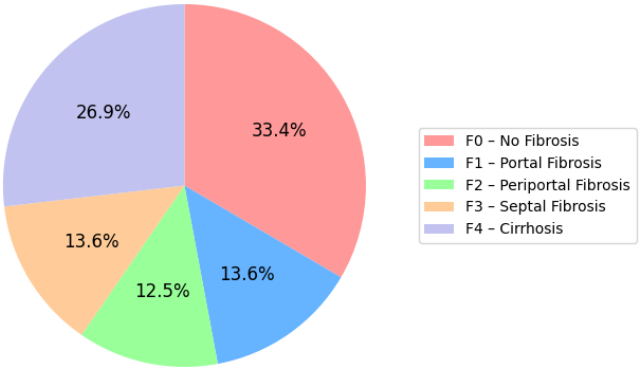

Figure 2. Distribution of liver fibrosis stages

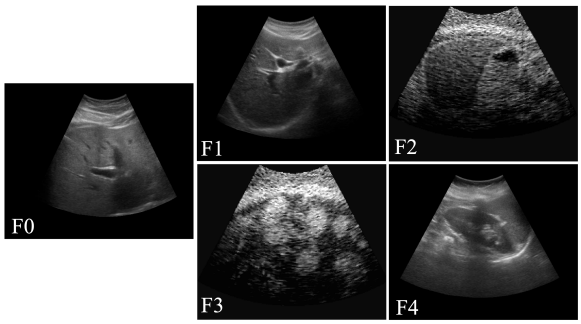

Figure 3. Illustrative B-mode ultrasound images showing liver fibrosis progression F0-F4 according to METAVIR classification.

2. Experimental Setup

For model prediction, we developed using TensorFlow 2.16.2 within a GPU-enabled environment, allowing for accelerated training and inference. GPU support significantly reduces computational latency compared to CPU-only execution, facilitating efficient experimentation and deployment of deep learning models. This configuration is particularly suited for integrating AI functionalities into mobile platforms through lightweight, on-device inference.

For clinical application, we used a Pixel 3a Android emulator with an `arm64-v8a` architecture, ensuring compatibility with modern Android applications. The emulator features a 1080×2220 pixel display (393 × 808 dp) at 440 dpi, providing high-resolution output for accurate UI/UX evaluation. It operates on Android API level 36 (ext17), with 16 GB of allocated RAM and 880 MB of available storage within a 7.7 GB disk footprint. This environment supports responsive and reliable testing of mid-sized Android applications across various screen configurations.

3. Quantitative Evaluation and Comparative Analysis

a) Overall Performance Evaluation

Table II and Table III illustrate that our method significantly outperforms all baseline architectures on all evaluated metrics, where a higher value (denoted by ↑) indicates better performance. Our model, optimized with a learning rate of 0.0001 and spatial dropout of 0.2, achieves a classification accuracy of 97.64%, precision of 97.69%, recall of 97.64%, and F1-score of 97.65%.In contrast, classical models such as ResNet and DenseNet exhibit significantly lower performance, particularly in terms of recall and F1-score, which are crucial factors for capturing minority classes in imbalanced clinical datasets. EfficientNet performs relatively better than other baselines, achieving an F1-score of 77.67%, but is still far behind the proposed method. The results confirm the effectiveness of the task-specific architectural improvements incorporated into our design, as well as the spatial regularization adapted to the ultrasound image domain. Our proposed CNN-based system demonstrates robust performance in liver fibrosis classification from B-mode ultrasound, overcoming challenges such as subtle inter-class differences, intra-class variability, and class imbalance. Achieving high precision and recall across all METAVIR stages, the system provides reliable classification, especially for clinically critical cases.

TABLE II
THE PROPOSED MODEL PERFORMANCE ON VARIOUS SETTINGS (LR: LEARNING RATE, SD: SPATIAL DROPOUT)

LR	SD	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑
0.1	0.1	0.6115	0.6100	0.6115	0.5880
	0.2	0.5264	0.5472	0.5264	0.5009
	0.3	0.4129	0.3617	0.4129	0.3466
	0.4	0.4586	0.3203	0.4586	0.3615
0.01	0.1	0.9401	0.9435	0.9401	0.9401
	0.2	0.9551	0.9548	0.9551	0.9547
	0.3	0.9362	0.9366	0.9362	0.9344
	0.4	0.9448	0.9459	0.9448	0.9444
0.001	0.1	0.9622	0.9634	0.9622	0.9619
	0.2	0.9701	0.9698	0.9701	0.9698
	0.3	0.9630	0.9635	0.9630	0.9628
	0.4	0.9716	0.9716	0.9716	0.9715
0.0001	0.1	0.9748	0.9746	0.9748	0.9747
	0.2	**0.9764**	**0.9769**	**0.9746**	**0.9765**
	0.3	0.9732	0.9729	0.9732	0.9730
	0.4	0.9716	0.9715	0.9716	0.9715

b) Class-wise Performance Across Fibrosis Stages

Table IV provides a fine-grained, stage-specific evaluation based on the METAVIR scoring system (F0–F4). The proposed model demonstrates consistent superiority across all fibrosis stages, where a higher value (denoted by ↑) indicates better performance.

Notably, for the most clinically critical stages, F3 (septal fibrosis) and F4 (cirrhosis), our model achieves F1-scores of 0.9281 and 0.9913, respectively, highlighting its capacity

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑
VGGNet [29]	0.8731	0.8758	0.8731	0.8705
ResNet [13]	0.6147	0.6711	0.6147	0.6014
DenseNet [15]	0.6509	0.6828	0.6509	0.6406
EfficientNet [30]	0.7841	0.7956	0.7841	0.7767
ViT [31]	0.6320	0.7546	0.6320	0.6169
USLF-Net (Ours)	**0.9764**	**0.9769**	**0.9764**	**0.9765**

TABLE IV
CLASS-WISE PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	Class	Accuracy ↑	Precision	Recall ↑	F1-score ↑
VGGNet [29]	F0	0.9999	0.9998	0.9999	0.9999
	F1	0.6512	0.8814	0.6047	0.7172
	F2	0.7673	0.7083	0.7484	0.7278
	F3	0.7283	0.8540	0.6763	0.7548
	F4	0.9500	0.7886	0.9765	0.8725
ResNet [13]	F0	0.9624	0.8221	0.9459	0.8797
	F1	0.1517	0.5000	0.1313	0.0765
	F2	0.7610	0.2986	0.8302	0.4393
	F3	0.4566	0.5308	0.3988	0.4554
	F4	0.4235	0.7473	0.4000	0.5211
DenseNet [15]	F0	0.9859	0.7828	0.9835	0.8717
	F1	0.2849	0.6184	0.2733	0.3790
	F2	0.7358	0.3455	0.7736	0.4777
	F3	0.4451	0.5280	0.4913	0.5090
	F4	0.4824	0.8310	0.3471	0.4896
EfficientNet [30]	F0	0.9835	0.9649	0.9694	0.9671
	F1	0.5698	0.6707	0.6395	0.6548
	F2	0.7358	0.4842	0.7673	0.5937
	F3	0.3931	0.9273	0.2948	0.4474
	F4	0.8647	0.7817	0.8529	0.8185
ViT [31]	F0	0.9778	0.9953	0.9998	0.9976
	F1	0.2151	0.3866	0.9419	0.5482
	F2	0.1572	0.3721	0.3019	0.3333
	F3	0.9538	0.8361	0.2948	0.4359
	F4	0.4676	0.9828	0.6735	0.7993
USLF-Net (Ours)	**F0**	**0.9998**	**0.9998**	**0.9999**	**0.9998**
	F1	**0.9651**	**0.9543**	**0.9709**	**0.9625**
	F2	**0.9308**	**0.9074**	**0.9245**	**0.9159**
	F3	**0.8902**	**0.9627**	**0.8960**	**0.9281**
	F4	**0.9999**	**0.9827**	**0.9998**	**0.9913**

to accurately detect advanced fibrosis, which is essential for timely clinical intervention. In contrast, other models demonstrate uneven performance across classes. For example, ViT and ResNet show sharp performance degradation for early-stage fibrosis (F1) and intermediate stages (F2, F3), often with recall scores below 0.40, indicating unreliable sensitivity to those categories. Such inconsistencies can be detrimental in real-world clinical applications where early and accurate detection is paramount. Our model, by contrast, maintains a high recall across all stages, including a strong score for F0 and F4, underscoring its robustness and clinical utility.

4. Further Analysis

Fig. 4 illustrates the performance of the USLF-Net for cirrhosis classification, highlighting both accuracy and error over 100 epochs. The left graph shows that the model rapidly achieves high accuracy on the training dataset, with stable validation accuracy, suggesting strong learning on the training data while maintaining consistent performance on unseen data. The right graph shows a significant reduction in training loss, while the validation loss remains slightly higher, indicating a minor gap in the model's ability to generalize to unknown data.

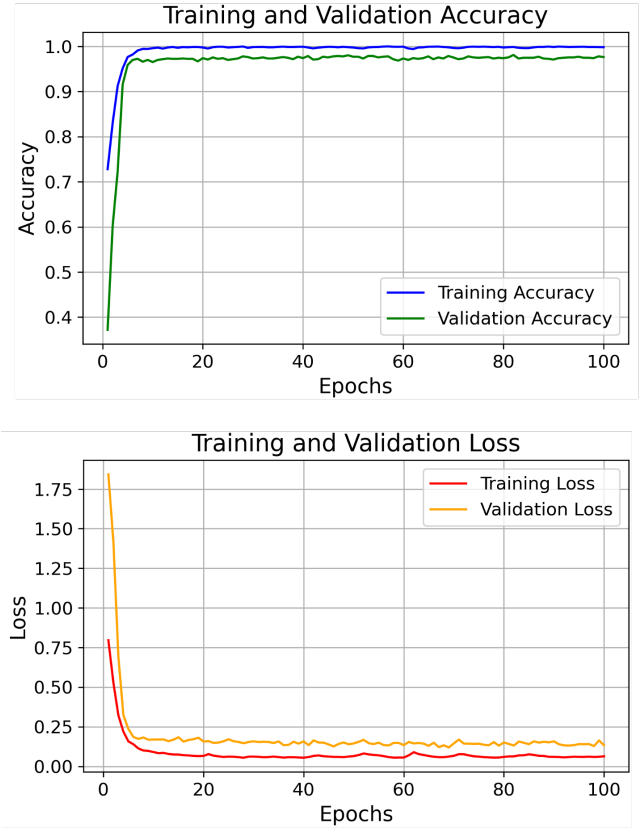

Figure 4. Our Model Analysis of Loss and Accuracy Across Epochs

Fig. 5 presents the normalized confusion matrix for the proposed model, which classifies cirrhosis into five distinct stages (F0–F4). The results indicate that the model demonstrates high accuracy across all classes, with particularly notable performance in the F0 and F4 categories, both achieving an absolute accuracy of 1.00. The accuracy for the other classes, namely F1, F2, and F3, is 0.95, 0.89, and 0.94, respectively, suggesting a strong classification capability. However, minor misclassifications occur between adjacent stages of cirrhosis (e.g., F2 is occasionally misclassified as F1 or F3). Overall, the model exhibits strong performance in classifying the entire dataset.

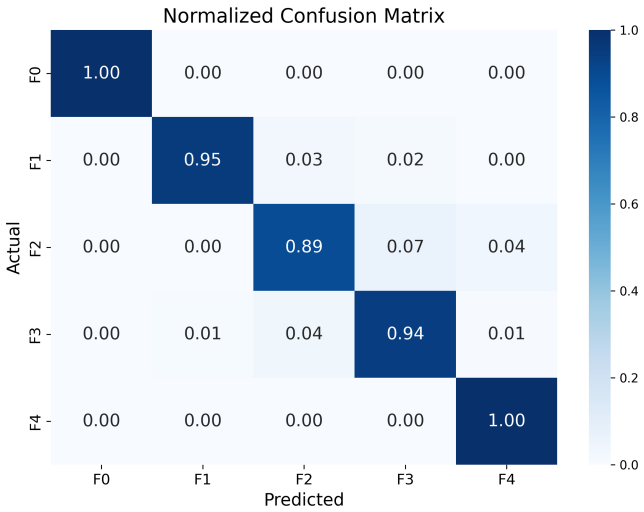

Figure 5. USLF-Net Model Accuracy Using Confusion Matrix

V. System Design and Clinical Application

1. System Design

Our system was developed to assist physicians in diagnosing liver fibrosis from 2D ultrasound images and managing patient care information (as shown in Fig. 6). The user interface was designed as a mobile application

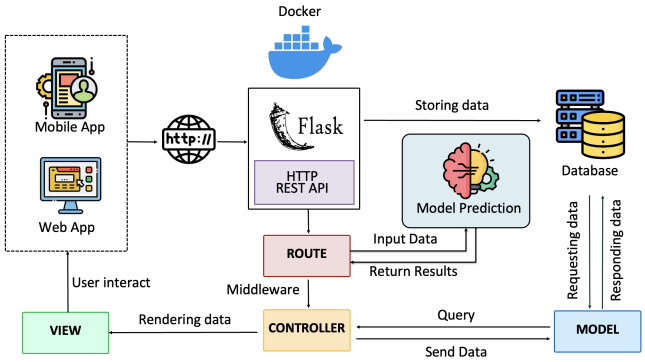

Figure 6. Our System Design for Liver Fibrosis Classification

utilizing Kotlin Jetpack Compose, allowing physicians to submit ultrasound images, access examination histories, and receive diagnostic results quickly on mobile devices. The backend was developed using Python Flask and implemented as a RESTful API, serving to process client requests and perform inference with pre-trained deep learning models. The backend architecture adheres to the Model-View-Controller (MVC) pattern, ensuring clear separation between logic processing components, API route management, and database interactions.

Patient data and diagnostic results are stored in a MySQL database management system, with DataGrip employed as the management and querying tool during development and testing phases. The entire backend system, including the AI model, is containerized and deployed through Docker, ensuring stability, scalability, and high reusability across diverse environments. The ultrasound image analysis results from the model are transmitted back to the mobile application, enabling physicians to make accurate and timely clinical decisions. The implementation of automated ultrasound fibrosis staging integrated with mobile diagnostic technology represents a significant advancement in hepatology practice, enabling early detection, treatment monitoring, and improved patient outcomes through enhanced diagnostic precision, real-time clinical decision support, and accessibility in diverse healthcare settings.

2. Clinical Application

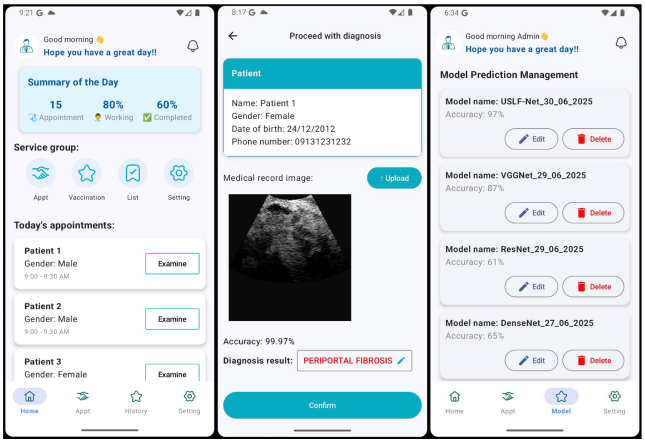

Figure 7. Physician-Centric Mobile Interface for Liver Fibrosis Diagnosis

Our mobile application is designed to support the diagnosis of clinical liver fibrosis (as shown in Fig. 7). After logging in, the physician user can manage their appointments and pending tasks. The patient list interface allows doctors to easily access patient records, including demographic information, medical history, and previous test results. In the diagnosis interface, physicians can upload ultrasound images directly from the device at each appointment. Our system then returns the prediction results from the best model and displays them along with the patient's data. Our system enables physicians to review and edit the model output before submitting the final diagnosis. In the right-hand interface, we provide the admin with the ability to manage trained models. Our application can increase the reliability of diagnosis, especially in real-time medical situations.

VI. Conclusion

This study introduces the USLF-Net model for automated liver fibrosis classification, achieving 97.64% accuracy

and outperforming existing transfer learning methods. The model demonstrates robust performance across all fibrosis stages, with particular accuracy in identifying F1 (no fibrosis) and F4 (cirrhosis), highlighting its clinical utility in detecting advanced fibrosis and accurately classifying cirrhosis. The model enables real-time, on-site assessment of liver fibrosis when integrated into a mobile diagnostic system, addressing significant healthcare accessibility challenges, particularly in resource-limited regions. In future work, we will focus on enhancing the prediction model using interpretability techniques, such as bottleneck analysis, Grad-CAM, and LIME, to improve model transparency, boost physician reliability, and further evaluate the clinical impact and cost-effectiveness of this mobile-enhanced diagnostic system.

REFERENCES

[1] N. A. Terrault, A. S. Lok, B. J. McMahon, K.-M. Chang, J. P. Hwang, M. M. Jonas, R. S. Brown Jr, N. H. Bzowej, and J. B. Wong, "Update on prevention, diagnosis, and treatment of chronic hepatitis b: Aasld 2018 hepatitis b guidance," *Hepatology*, vol. 67, no. 4, pp. 1560–1599, 2018.

[2] D. Schuppan and N. H. Afdhal, "Liver cirrhosis," *The Lancet*, vol. 371, no. 9615, pp. 838–851, 2008.

[3] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *Journal of hepatology*, vol. 70, no. 1, pp. 151–171, 2019.

[4] P. Bedossa and T. Poynard, "An algorithm for the grading of activity in chronic hepatitis c," *Hepatology*, vol. 24, no. 2, pp. 289–293, 1996.

[5] E. B. Tapper and N. D. Parikh, "Mortality due to cirrhosis and liver cancer in the united states, 1999-2016: observational study," *bmj*, vol. 362, 2018.

[6] S. S.-D. Xu, C.-C. Chang, C.-T. Su, and P. Q. Phu, "Classification of liver diseases based on ultrasound image texture features," *Applied Sciences*, vol. 9, no. 2, p. 342, 2019.

[7] J. H. Lee, I. Joo, T. W. Kang, Y. H. Paik, D. H. Sinn, S. Y. Ha, K. Kim, C. Choi, G. Lee, J. Yi *et al.*, "Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network," *European radiology*, vol. 30, pp. 1264–1273, 2020.

[8] F. Gao, T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, and B. Patel, "Sd-cnn: A shallow-deep cnn for improved breast cancer diagnosis," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 53–62, 2018.

[9] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. D. Pande, and S. H. Ahammad, "A densenet cnn-based liver lesion prediction and classification for future medical diagnosis," *Scientific African*, vol. 20, p. e01629, 2023.

[10] T. M. N. Phan and H. T. Nguyen, "Clinical decision support systems for pneumonia diagnosis using gradient-weighted class activation mapping and convolutional neural networks," in *Soft Computing: Biomedical and Related Applications*. Springer, 2021, pp. 81–92.

[11] C.-C. Wu, W.-C. Yeh, W.-D. Hsu, M. M. Islam, P. A. A. Nguyen, T. N. Poly, Y.-C. Wang, H.-C. Yang, and Y.-C. J. Li, "Prediction of fatty liver disease using machine learning algorithms," *Computer methods and programs in biomedicine*, vol. 170, pp. 23–29, 2019.

[12] W.-C. Yeh, S.-W. Huang, and P.-C. Li, "Liver fibrosis grade classification with b-mode ultrasound," *Ultrasound in medicine & biology*, vol. 29, no. 9, pp. 1229–1235, 2003.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[16] A. Q. Md, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease," *Biomedicines*, vol. 11, no. 2, p. 581, 2023.

[17] U. R. Acharya, O. Faust, F. Molinari, S. V. Sree, S. P. Junnarkar, and V. Sudarshan, "Ultrasound-based tissue characterization and classification of fatty liver disease: A screening and diagnostic paradigm," *Knowledge-Based Systems*, vol. 75, pp. 66–77, 2015.

[18] G. C. Kagadis, P. Drazinos, I. Gatos, S. Tsantis, P. Papadimitroulas, S. Spiliopoulos, D. Karnabatidis, I. Theotokas, P. Zoumpoulis, and J. D. Hazle, "Deep learning networks on chronic liver disease assessment with fine-tuning of shear wave elastography image sequences," *Physics in Medicine & Biology*, vol. 65, no. 21, p. 215027, 2020.

[19] B. C. Byram and N. V. Ruiter, "Medical imaging 2019: Ultrasonic imaging and tomography," in *Society of Photo-Optical Instrumentation Engineers (SPIE)*

Conference Series, vol. 10955, 2019.

[20] L.-Y. Xue, Z.-Y. Jiang, T.-T. Fu, Q.-M. Wang, Y.-L. Zhu, M. Dai, W.-P. Wang, J.-H. Yu, and H. Ding, "Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis," *European radiology*, vol. 30, pp. 2973–2983, 2020.

[21] I. Gatos, S. Tsantis, S. Spiliopoulos, D. Karnabatidis, I. Theotokas, P. Zoumpoulis, T. Loupas, J. D. Hazle, and G. C. Kagadis, "Temporal stability assessment in shear wave elasticity images validated by deep learning neural network for chronic liver disease fibrosis stage assessment," *Medical physics*, vol. 46, no. 5, pp. 2298–2309, 2019.

[22] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, and A. E. Samir, "Objective liver fibrosis estimation from shear wave elastography," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1–5.

[23] K. Wang, X. Lu, H. Zhou, Y. Gao, J. Zheng, M. Tong, C. Wu, C. Liu, L. Huang, T. Jiang *et al.*, "Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis b: a prospective multicentre study," *Gut*, vol. 68, no. 4, pp. 729–741, 2019.

[24] K. Patel and G. Sebastiani, "Limitations of non-invasive tests for assessment of liver fibrosis," *JHEP Reports*, vol. 2, no. 2, p. 100067, 2020.

[25] R. Anteby, E. Klang, N. Horesh, I. Nachmany, O. Shimon, Y. Barash, U. Kopylov, and S. Soffer, "Deep learning for noninvasive liver fibrosis classification: A systematic review," *Liver International*, vol. 41, no. 10, pp. 2269–2278, 2021.

[26] Y. Li, Q. He, and J. Luo, "A deep learning trial on transient elastography for assessment of liver fibrosis," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[27] L. Dandan, M. Huanhuan, L. Xiang, J. Yu, J. Jing, and S. Yi, "Classification of diffuse liver diseases based on ultrasound images with multimodal features," in *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2019, pp. 1–5.

[28] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, "Liver fibrosis classification based on transfer learning and fcnet for ultrasound images," *Ieee Access*, vol. 5, pp. 5804–5810, 2017.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[32] Y. Joo, H.-C. Park, O.-J. Lee, C. Yoon, M. H. Choi, and C. Choi, "Classification of liver fibrosis from heterogeneous ultrasound image," *IEEE Access*, vol. 11, pp. 9920–9930, 2023.

Tan Tai Phan is a Main Lecturer at College of Information and Communication Technology, Can Tho University, Vietnam. He obtained his B.Sc. in Informatics from Can Tho University in 1996 and his M.Sc. in Information Systems Development from HAN University, the Netherlands, in 2004. His research focuses on information systems, machine learning, artificial intelligence, and their applications in healthcare, agriculture, and social management.
Email: pttai@ctu.edu.vn

Thi Cam Hong Pham received her degree in Information Systems from Can Tho University, Vietnam. Her research interests include machine learning and the application of artificial intelligence in healthcare.
Email: phamthicamhong0301@gmail.com

Nguyen Minh Thao Phan is a lecturer in the Department of Information Systems at the College of Information and Communication Technology, Can Tho University, Vietnam. She earned her B.Sc. in Information Systems from Can Tho University in 2021 and her M.Sc. in Electrical Engineering and Computer Science from National Yang Ming Chiao Tung University (NYCU), Taiwan, in 2024, where she is now pursuing a Ph.D. Her research interests include deep learning, natural language processing, recommender systems, and medical image processing.
Email: pnmthao@ctu.edu.vn

Cong-Tinh Dao is a lecturer in the Department of Computer Science at the College of Information and Communication Technology, Can Tho University, Vietnam. He received his B.Sc. in Computer Science from Can Tho University in 2021 and his M.Sc. in Electrical Engineering and Computer Science from National Yang Ming Chiao Tung University (NYCU), Taiwan, in 2024, where he is pursuing a Ph.D. His research interests include deep reinforcement learning, large language models, and biomedical & healthcare informatics, with a focus on developing intelligent systems that address real-world challenges in healthcare and AI applications.
Email: dctinh@ctu.edu.vn

Hai Thanh Nguyen is an Associate Professor at the College of Information and Communication Technology, Can Tho University, Vietnam. He received his M.S. in Computer Science and Engineering from National Chiao Tung University (now National Yang Ming Chiao Tung University), Taiwan, in 2014, and his Ph.D. in Computer Science from Sorbonne University, France, in 2018. His research interests include technology-driven applications in healthcare systems, smart agriculture, and smart education, with a focus on computer vision, machine learning, and data analysis.
Email: nthai.cit@ctu.edu.vn