CHAPTER I
INTRODUCTION

1.1 Background

Chronic liver disease remains a major health problem worldwide and in Indonesia, as it can lead to long-term impairment of liver function and even death. [1], [2]. The main causes include chronic hepatitis infection and fat accumulation in the liver, known as non-alcoholic fatty liver disease (NAFLD) [2], [3]. In Indonesia, the 2023 Health Survey reported that the prevalence of hepatitis B decreased from 7.1% in 2013 to 2.4% in 2023. However, the disease burden remains substantial, as the same survey recorded approximately 6.7 million people infected with hepatitis B and 2.5 million people infected with hepatitis C. Therefore, hepatitis continues to be a major focus of national health programs. Indonesia has also been reported to rank fourth in Southeast Asia in terms of liver disease incidence and mortality [1]. At the global and regional levels, meta-analyses and reports from Asia estimate that NAFLD affects approximately one-quarter to one-third of adults. It is more common among individuals with obesity, diabetes, or metabolic syndrome, and its prevalence continues to increase alongside changes in diet and lifestyle [2], [3]. Both NAFLD and chronic hepatitis can progress to liver fibrosis, which is commonly assessed using the F0-F4 staging scale. Each increase in fibrosis stage is associated with a higher risk of severe complications, such as gastrointestinal bleeding and liver cancer. Therefore, accurate fibrosis staging is essential for treatment selection, follow-up scheduling, and referral prioritization [4], [5].

Clinically, liver biopsy is still considered the gold standard for assessing the degree of liver damage. The collected tissue sample is examined under a microscope, mapped to the METAVIR F0-F4 scale, and often used as the ground truth in imaging and machine learning studies [4]. However, biopsy is an invasive procedure that carries risks of pain and bleeding. It is also impractical for repeated assessment and is prone to sampling error because it represents only a small portion of the liver tissue [4]. Various non-invasive examinations, such as elastography and biomarker panels, have been developed. However, their use in primary healthcare settings remains limited due to cost, equipment availability, and the need for trained personnel [2]. In contrast, B-mode ultrasonography (US) is relatively inexpensive,

radiation-free, widely available, and therefore commonly used for screening and monitoring chronic liver disease [1]. Several image analysis studies have shown that brightness and texture patterns in liver ultrasound images change according to the degree of fibrosis and may be utilized for automated staging. Therefore, a more objective and standardized ultrasound image analysis system is needed to reduce the subjectivity of visual assessment among examiners [6].

The development of deep learning has created significant opportunities for automating liver image analysis, including F0-F4 fibrosis staging. Zhu et al. (2021) demonstrated that convolutional neural networks could classify fibrosis stages in patients with chronic hepatitis B using apparent diffusion coefficient (ADC)-based MRI images with high accuracy, indicating that image patterns can be effectively mapped to METAVIR stages [4]. In ultrasound imaging, Joo et al. (2023) and Park et al. (2024) reported that various CNN and Vision Transformer architectures were able to distinguish mild, moderate, and severe fibrosis, as well as the five METAVIR classes, with high accuracy and AUC values, approaching or even exceeding the performance of experienced assessors [5], [6]. In Indonesia, Febryanto and Tahyudin (2025) found that CNN produced the most consistent results compared with LSTM and FNN for liver fibrosis diagnosis based on medical images, further supporting the relevance of deep learning approaches in the local context [7]. Beyond B-mode ultrasound, Lu et al. (2025) and Chiang et al. (2025) showed that the ConvNeXt family is effective for classifying fibrosis, compensated advanced chronic liver disease (cACLD), and NAFLD using 2D shear wave elastography and abdominal ultrasound images, with stable performance across different healthcare centers [8], [9].

However, most of these studies still modeled F0-F4 as nominal multiclass labels without considering the ordered relationship between fibrosis stages [4]. In fact, the METAVIR scale is ordinal, namely F0 < F1 < F2 < F3 < F4, where each increase in stage reflects a greater degree of fibrosis and a higher risk of complications [10]. In nominal classification, each class is treated as an independent category, so the hierarchical relationship between fibrosis stages is not explicitly modeled [4], [11]. This condition is less ideal for fibrosis staging, because prediction errors between adjacent classes have a different clinical meaning from

errors that jump across distant stages. To address this issue, modern ordinal regression approaches such as CORAL model classes through a series of ordered decision thresholds. This allows the natural order of F0-F4 to be learned more explicitly and evaluated using distance-aware metrics, such as Mean Absolute Error (MAE) and quadratic weighted kappa (QWK) [11]. Saito et al. (2021) showed that this type of ordinal formulation in an ultrasound-based fibrosis classification system produced better MAE and QWK values and reduced the frequency of large prediction errors compared with nominal classification approaches [10].

From an architectural perspective, many ultrasound image-based fibrosis staging studies still rely on conventional CNNs with ImageNet-based transfer learning and have not fully exploited the distinctive characteristics of liver ultrasound texture, which typically presents low contrast and granular speckle patterns [5]. ConvNeXt V2 is a modern ConvNet family that incorporates design principles from Vision Transformers and masked autoencoders into a fully convolutional framework, while introducing Global Response Normalization (GRN) to improve training stability [12]. Woo et al. (2023) demonstrated that ConvNeXt V2 pretrained using a Fully Convolutional Masked Autoencoder (FCMAE) produces strong and efficient visual representations, including in the lightweight ConvNeXt V2 Tiny variant, which requires fewer parameters and lower computational cost. In the context of relatively limited liver ultrasound datasets, the Tiny variant provides a balanced trade-off between model capacity and the risk of overfitting, while still leveraging visual representations learned from FCMAE pretraining and the GRN module [12]. The findings from DLRE-X and PBCS-ConvNeXt also indicate that the ConvNeXt family is capable of capturing complex liver texture patterns in both 2D shear wave elastography and abdominal ultrasound images [8], [9]. Therefore, ConvNeXt V2 Tiny was selected as the backbone for learning representations from liver ultrasound images and was combined with an ordinal approach to better align with the ordered nature of the F0-F4 fibrosis labels [11].

In addition to architectural considerations and label ordinality, the model's probability outputs also need to be considered so that predictions can be interpreted more informatively. This study aims not only to distinguish each fibrosis stage from

F0 to F4, but also to evaluate derived endpoints such as significant fibrosis, advanced fibrosis (F≥3), and cirrhosis. These endpoints can help assess model performance at clinically interpretable fibrosis thresholds, without positioning the model as a standalone clinical diagnostic tool [4], [5]. On the other hand, many studies have reported that modern deep learning networks tend to be poorly calibrated, meaning that the probability outputs produced by the model do not always align with the actual frequency of events [13]. Sambyal et al. (2023) showed that even models with high accuracy may suffer from miscalibration when optimized solely using cross-entropy. Meanwhile, Balanya et al. (2024) proposed Adaptive Temperature Scaling as a more stable post-hoc calibration method under limited-data conditions and demonstrated improved calibration quality across various architectures [13], [14]. Therefore, probability calibration is relevant to ensure that the risk scores generated by the model can be interpreted more cautiously, especially when the model outputs are presented in the form of probabilities and risk categories.

Based on these gaps, this study develops an ordinal classification model for F0–F4 liver fibrosis from B-mode ultrasound images using the public "Liver Histopathology (Fibrosis) Ultrasound Images" dataset as the primary data source [15]. Prior to training, the dataset was audited and deduplicated to reduce the risk of image duplication that could affect the reliability of the evaluation. A pretrained ConvNeXt V2 Tiny backbone was used as the feature extractor to capture global and local texture representations in ultrasound images. The Tiny variant was selected because it offers a balanced compromise between model capacity and the risk of overfitting on a medium-sized dataset [12]. On top of this backbone, the CORAL approach was applied to model F0–F4 staging as an ordinal output, allowing the ordered structure of the METAVIR scale to be learned more explicitly than in a nominal softmax classification baseline [11], [12]. This study also compares the nominal baseline and the ordinal approach as initial experimental scenarios before evaluating the main model using 5-fold cross-validation, out-of-fold predictions, and a holdout set. In addition to F0-F4 prediction, the model's ordinal probabilities were used to evaluate derived endpoints, including F≥2, F≥3, and F=4. Meanwhile, the advanced fibrosis binary head was used as the source of

calibrated risk probability through temperature scaling and was complemented with uncertainty estimation [13], [14]. Model performance was evaluated using ordinal metrics such as MAE and QWK, discrimination metrics such as AUC, and calibration metrics. Therefore, the findings of this study can serve as a basis for developing a more structured and transparent liver ultrasound image analysis system [5], [9].

1.2 Problem Statement

1. How can an ordinal classification model for ultrasound image-based liver fibrosis be designed using ConvNeXt V2 Tiny and the CORAL approach to predict fibrosis stages F0-F4?
2. How does the CORAL-based ordinal approach affect the prediction of F0–F4 liver fibrosis stages compared with a nominal softmax classification baseline?
3. How does the main model perform in staging F0–F4 liver fibrosis and distinguishing between non-advanced fibrosis and advanced fibrosis (F≥3) on the ultrasound dataset used in this study?
4. How can the model outputs be implemented in a simple website so that ultrasound image prediction results can be displayed in a structured format, including the fibrosis class, class probabilities, and advanced/non-advanced fibrosis status?

1.3 Research Objectives

The objective of this study is to develop and evaluate a deep learning model based on ConvNeXt V2 Tiny with the CORAL approach for ordinal classification of F0–F4 liver fibrosis from B-mode ultrasound images. This study also aims to assess the effect of the ordinal approach compared with the nominal softmax classification baseline, particularly in preserving the ordered structure of fibrosis stages from F0 to F4. In addition, the main model is evaluated to distinguish between non-advanced fibrosis (F0-F2) and advanced fibrosis (F3–F4) as the primary binary interpretation focus.

In addition to model evaluation, this study aims to develop a simple website interface prototype to present the model outputs in a more structured manner. The displayed outputs include the input image, F0-F4 fibrosis class prediction,

probability distribution, and advanced/non-advanced fibrosis status. This website is not intended as a standalone diagnostic system, but as a research implementation that makes the model inference results easier to understand.

1.4 Research Contributions

This study is expected to contribute to the development of deep learning-based liver ultrasound image analysis methods, particularly for fibrosis classification with an ordered stage structure. By using the CORAL ordinal approach, this study can serve as a reference for developing models that not only treat fibrosis classes as separate labels, but also consider the severity order from F0 to F4. The findings of this study may also provide a foundation for future research involving larger datasets, external validation, or patient-level data.

From an implementation perspective, this study produces a FibrosisRisk website prototype to present the model prediction results in a concise and informative manner. The interface demonstrates how the model can display the predicted class, class probabilities, and advanced/non-advanced fibrosis status for new ultrasound images. Therefore, the contribution of this study is not limited to model performance evaluation, but also includes the design of a more readable output format, while still positioning the system as a research prototype rather than a substitute for medical examination.

1.5 Research Scope and Limitations

1. This study used the fibrosis labels F0, F1, F2, F3, and F4 according to the information provided in the dataset. Label reassessment by medical experts was not performed because this study focused on model development and evaluation based on the dataset used. Therefore, label quality follows the characteristics of the available data.
2. This study only used liver ultrasound images as model inputs. Other supporting information, such as patient data, laboratory results, or medical descriptions, was not used during training or evaluation.