Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Large annotated ultrasound dataset of non-alcoholic fatty liver from Saudi hospitals for analysis and applications

Fahad Alshagathrh[a], Mahmood Alzubaidi[a], Khalid Alswat[b,c], Ali Aldhebaib[c], Bushra Alahmadi[d], Meteb Alkubeyyer[e], Abdulaziz Alosaimi[f], Amani Alsadoon[b,c], Maram Alkhamash[b,c], Jens Schneider[a], Mowafa Househ[a,*]

[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*
[b] *Liver Disease Research Center, College of Medicine, King Saud University, Riyadh, Saudi Arabia*
[c] *Radiological Sciences Program, College of Applied Medical Sciences, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia*
[d] *Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia*
[e] *Radiology Department, King Saud University Medical City, Riyadh, Saudi Arabia*
[f] *Medical Imaging Department, King Abdulaziz Medical City, Riyadh, Saudi Arabia*

ARTICLE INFO

Article history:
Received 16 September 2024
Revised 9 December 2024
Accepted 24 December 2024
Available online 27 December 2024

Dataset link: Large Annotated Ultrasound Dataset of Non-Alcoholic Fatty Liver from Saudi Hospitals for Analysis and Applications (Original data)

Keywords:
Liver ultrasound imaging
Steatosis grading
Liver fibrosis Staging
Ultrasound dataset
Computer vision

ABSTRACT

This study presents a comprehensive ultrasound image dataset for Non-Alcoholic Fatty Liver Disease (NAFLD), addressing the critical need for standardized resources in AI-assisted diagnosis. The dataset comprises 10,352 high-resolution ultrasound images from 384 patients collected at King Saud University Medical City and National Guard Health Affairs in Saudi Arabia. Each image is meticulously annotated with NAFLD Activity Score (NAS) fibrosis staging and steatosis grading based on corresponding liver biopsy results. Unlike other datasets that rely on bounding boxes, we opted for full-image labelling based on biopsy findings, which link to histopathological results, ensuring more precise representation of liver conditions. Rigorous pre-processing ensures high-quality image preservation, including expert radiologist assessment, DICOM to PNG conversion, and standardization to 768 × 1024 pixels. This resource supports various com-

* Corresponding author.
E-mail address: mhouseh@hbku.edu.qa (M. Househ).

https://doi.org/10.1016/j.dib.2024.111266

puter vision tasks, enabling the development of AI algorithms for accurate NAFLD diagnosis and staging. A large, diverse, and well-annotated dataset like ours is essential for enhancing model performance and generalization, providing a valuable resource for researchers to develop robust AI models in medical imaging.

Specifications Table

Subject	Medical Imaging, Artificial Intelligence, and Computer Vision in Healthcare.
Specific subject area	Annotated ultrasound image dataset for liver steatosis and fibrosis staging in NAFLD patients.
Data format	Raw, Filtered, Processed
Type of data	Table, Image, Chart, Figure
Data collection	The data were collected from liver ultrasound examinations of 384 King Saud University Medical City and National Guard Health Affairs, Saudi Arabia patients. Ultrasound images were obtained using multiple devices used for ultrasound acquisition. The data includes manually cropped images and is pre-processed to remove patient-identifying information. DICOM images were converted to PNG format, standardized to 768 × 1024 pixels, and categorized based on liver biopsy results for fibrosis staging and steatosis grading. Exclusion criteria included suboptimal imaging quality and incomplete liver visualization.
Data source location	King Saud University Medical City (KSUMC) and National Guard Health Affairs (NGHA), Riyadh, Saudi Arabia.
Data accessibility	Repository name: Open Science Framework (OSF) Data identification number: DOI 10.17605/OSF.IO/C2YG8 Direct URL to data: https://doi.org/10.17605/OSF.IO/C2YG8 Instructions for accessing these data: The dataset requires approval through a formal agreement outlining permitted uses and confidentiality requirements. Researchers interested in using the dataset should submit a request through the Open Science Framework (OSF) repository. The data are restricted for non-commercial research purposes, and user agreements must be signed before access.
Related research article	None

1. Value of the Data

- Comprehensive Dataset for NAFLD Diagnosis: This comprehensive dataset includes ultrasound images from 384 patients, annotated with steatosis and fibrosis grades based on NAFLD Activity Scores (NAS), making it highly valuable for diagnostic research. This offers a valuable resource for researchers looking to develop and validate diagnostic tools for Non-Alcoholic Fatty Liver Disease (NAFLD) using medical imaging.
- Support for Machine Learning Model Development: Given the size and annotations of the dataset, it supports the development and validation of machine learning models aimed at automating the diagnosis of NAFLD. Researchers can leverage these images to create or enhance AI models focused on detecting liver disease and grading its severity.
- Broad Compatibility and Flexibility: The dataset is provided in PNG format, preserving ultrasound images' diagnostic integrity through lossless compression. Validation analyses confirmed that pixel-level features are retained during conversion from DICOM to PNG, ensuring the data's reliability for advanced imaging studies. Furthermore, PNG's compatibility with widely used tools like Python and MATLAB enhances its accessibility for diverse research workflows.

- Utility in Research Beyond Liver Disease: Although researchers originally developed the dataset for NAFLD studies, it can be repurposed for broader applications in medical imaging, including assessments of ultrasound image quality and advancements in computer vision algorithms for medical diagnostics. It serves as a valuable benchmark for evaluating the performance of various image-processing techniques.
- The dataset's comprehensive annotations and diverse distribution across different grades make it a valuable resource for developing advanced diagnostic tools and studying disease progression. Its structured format allows researchers to conduct reproducible analyses and explore novel medical imaging and computational research methodologies.

2. Background

Non-Alcoholic Fatty Liver Disease (NAFLD) is a prevalent global health concern affecting a significant portion of the population worldwide [1]. Clinicians commonly utilize ultrasound imaging as a non-invasive diagnostic tool for NAFLD [2]. However, interpreting ultrasound images is inherently subjective, leading to inconsistencies, particularly in assessing histological characteristics [3]. We have established a comprehensive, annotated dataset of NAFLD ultrasound images to improve diagnostic reliability and facilitate the development of artificial intelligence (AI) applications. This dataset addresses the limitations of existing NAFLD ultrasound datasets, which often lack adequate size, diversity, and comprehensive annotations [4,5].

The dataset consists of annotated ultrasound images collected from King Saud University Medical City and National Guard Health Affairs, linked to liver biopsy findings for precise staging of steatosis and fibrosis. It includes a diverse range of demographic and clinical variants of NAFLD. Researchers tagged every image according to the NAFLD Activity Score (NAS) for both Steatosis and Fibrosis stages, ensuring a uniform reference. This dataset provides a robust foundation for developing AI technologies aimed at improving the consistency and accuracy of NAFLD evaluation through ultrasound imaging.

The dataset enhances value by offering a more extensive, diversified, and meticulously annotated resource, facilitating the creation of more accurate diagnostic instruments.

3. Data Description

The dataset is publicly accessible via the OFS [6] repository at https://doi.org/10.17605/OSF.IO/C2YG8 and contains liver ultrasound images in standardized PNG format. Validation analyses confirmed that the PNG format preserves the diagnostic quality of images while ensuring compatibility with widely used analytical tools. Each image is standardized to dimensions of 768 × 1024 pixels. The dataset was developed for multi-class classification tasks related to Non-Alcoholic Fatty Liver Disease (NAFLD) and is structured to facilitate machine learning and computer vision research in this domain. Access to the dataset is restricted to ensure compliance with ethical guidelines and to protect patient confidentiality. Researchers seeking access must submit a formal request via the OSF platform detailing their intended use. After signing a confidentiality agreement outlining permitted uses and restrictions, access is granted upon approval. Commercial use of the dataset is strictly prohibited without explicit permission.

The dataset, submitted to OSF as All_Labelled_Images.rar, is organized into two main subfolders: Steatosis and Fibrosis. Each contains Gray-Scale Images and Coloured Images, further divided by NAS classification. The 'Fibrosis' folder includes subfolders for No fibrosis, Mild fibrosis, Perisinusoidal fibrosis, Bridging fibrosis, and Cirrhosis. The 'Steatosis' folder is divided into Grade 0, Grade 1, Grade 2, and Grade 3.

The dataset comprises a description file that details the dataset's structure and the distribution of research completed throughout certain time intervals. However, no specific metadata file containing patient information, such as patient ID, study date, NAS scores, or biopsy results, is

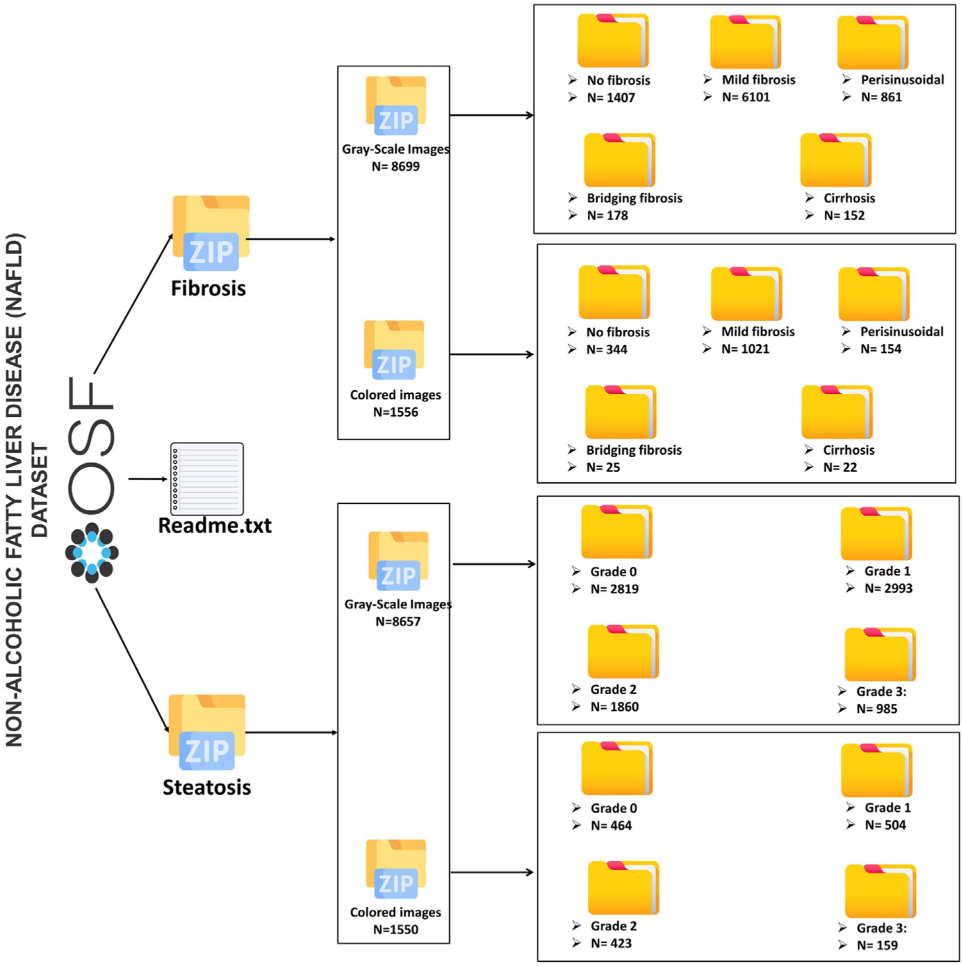

Fig. 1. Distribution of non-alcoholic fatty liver disease (NAFLD) ultrasound images across NAS fibrosis staging and NAS steatosis grading.

provided, as this information is confidential and already used to label the images directly. The attached file provides an overview of the dataset's structure, enabling scholars to comprehend the larger context of the image collection. This framework allows for focused analysis and model creation while preserving patient confidentiality.

All images are stored in PNG format to ensure compatibility with most image analysis software, and the standardized image dimensions ensure consistency across the dataset. Visual aids, such as Fig. 1, illustrate the distribution of images across fibrosis and steatosis stages, helping researchers navigate the dataset and select relevant subsets for analysis.

4. Materials and Methods

We generated a large and diverse ultrasound image dataset to address the scarcity of open data for analysing fatty liver disease [4,5,7]. Ultrasonography, the primary diagnostic tool with readily available data in other modalities (CT, MRI), was chosen for its accessibility [5,8].

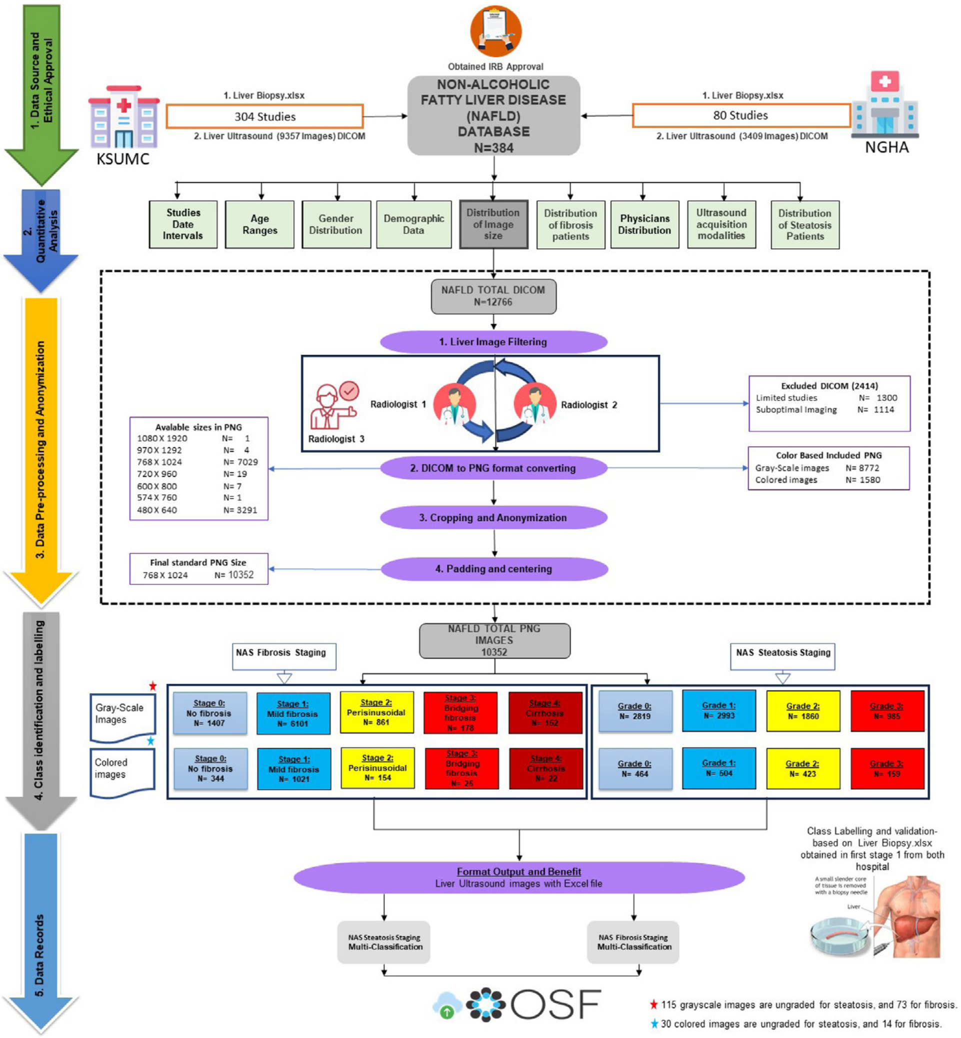

Fig. 2. Workflow diagram illustrating the data processing pipeline for NAFLD ultrasound images. The process begins with 12,766 DICOM files and progresses through stages of expert radiologist filtering, DICOM to PNG conversion, and image standardization.

Collaborations with two hospitals in Riyadh and Jeddah, Saudi Arabia, facilitated data collection. Institutional Review Board (IRB) approval [9] and adherence to ethical protocols were strictly followed. This comprehensive dataset comprises medical ultrasound images paired with corresponding liver biopsy results, the gold standard for diagnosis, mitigating the inherent subjectivity and variability in ultrasound interpretations.

Ultrasound images and liver biopsy data were collected retrospectively from King Saud University Medical City (KSUMC) and National Guard Health Affairs (NGHA). At KSUMC, 312 patients met the eligibility criteria; at NGHA, 110 patients were included. Both institutions followed rigorous data collection protocols, ensuring data completeness and patient anonymity.

Table 1
Distribution of the included studies performed on NAFLD patients at KSUMC from 2009 to 2017 and at NGHA from 2009 to 2023.

Studies date intervals	No. of studies from KSUMC	No. Of studies from NGHA
Jan 2009–Dec 2010	3	3
Dec 2010–Dec 2011	39	0
Dec 2011–Dec 2012	90	4
Dec 2012–Dec 2013	89	0
Dec 2013–Dec 2014	39	6
Dec 2014–Dec 2015	28	0
Dec 2015–Dec 2017	16	1
Jan 2019–Dec 2020	0	9
Jan 2021–Dec 2022	0	43
Jan 2023–Dec 2023	0	14
Total	304	80

After data collection, we conducted rigorous curation and pre-processing to ensure completeness, quality, and reproducibility. Fig. 2 provides an overview of the workflow encompassing data acquisition, preprocessing, quality control, and labelling processes.

4.1. Data source

Data management procedures were conducted with precision. The dataset, consisting of ultrasound images and biopsy reports, was securely stored on the researcher's PC, with all analyses conducted using HBKU processing units to ensure data integrity and confidentiality. Regular team meetings were held to review data collection and management practices, ensuring the dataset's completeness, accuracy, and quality.

In preparing the dataset, a strong focus on quality and diversity was maintained. A sizable and diverse collection of annotated ultrasound images was curated, highlighting its relevance and broad applicability. Image pre-processing and region of interest (ROI) selection were emphasized, with standardized image sizes and automated ROI selection methods applied to enhance consistency.

The clinical validation phase evaluated the framework's applicability in a real-world setting. This phase, conducted under Institutional Review Board (IRB) guidelines (further detailed in the Ethical Statement), involved collecting images representing all disease stages and validating the framework's outcomes with scientific evidence. While biopsy result accuracy was not the primary focus, this phase reinforced the framework's reliability within a clinical context.

This section concludes by highlighting key aspects of dataset acquisition, management, preparation, ethics, and clinical validation. The methodology reflects scientific rigor and a commitment to ethical standards, reinforced by approvals from respected institutions. The study's processes were meticulously designed to adhere to IRB guidelines, ensuring responsible and transparent research practices.

4.2. Data collection

As previously stated, the data set was obtained from two hospitals and consists of ultrasound images of individuals with fatty liver, accompanied by biopsy findings. This section will discuss the collection of images and biopsy findings at both institutions. We shall represent the King Saud University Medical City as (KSUMC). The abbreviation for King Abdulaziz Medical City for the National Guard will be (NGHA).

Table 1 presents the distribution of ultrasound imaging sessions (studies) performed on 304 NAFLD patients at KSUMC from 2009 to 2017. As shown, the number of studies conducted each

Table 2
Age distribution in the study reflects typical NAFLD demographics, with middle-aged adults being the most prevalent.

Age ranges	No. of patients from KSUMC	No. Of patients from NGHA
0 to <20	11	8
20 to <40	160	25
40 to <60	120	14
≤60	13	17
Total	304	64

Table 3
Gender distribution in the study aligns with other research, showing a higher proportion of females with NAFLD.

Gender	No. of patients in KSUMC	No. of patients in NGHA
Male	85	36
Female	219	28
Total	304	64

year varied, with a peak of 90 in 2011–2012. This comprehensive dataset, covering a diverse range of patients and examinations, provides a valuable resource for researchers investigating the use of ultrasound in NAFLD diagnosis, staging, and treatment monitoring. The table also shows the distribution of ultrasound imaging sessions (studies) performed on patients at NGHA from January 2009 to December 2023. Image distribution across time points is clinically relevant to follow NAFLD disease development longitudinally. With images from several sessions over several years, this dataset allows researchers to analyse how NAFLD progresses and reacts to therapies, giving useful data for diagnostic model building and therapeutic monitoring. The number of studies conducted each year varied, with a peak of forty-three studies in January 2021 to December 2022. This dataset, spanning 15 years, offers a valuable resource for researchers investigating the use of ultrasound in NAFLD diagnosis, and staging.

The age distribution of the 304 NAFLD patients included in the study from KSUMC is shown in Table 2. Most patients were between 20 and 40 years old (52.6 %), followed by those between 40 and 60 years old (39.4 %). Younger and older age groups were less frequent, with only eleven patients (3.6 %) younger than 20 years and thirteen patients (4.3 %) older than 80 years. This distribution reflects the typical demographic of NAFLD, a disease that commonly affects middle-aged adults. Even though the variance can be seen in NAFLD patients in NGHA, the same observations can also be seen since the most affected group is middle-aged adults. It is essential to note the difference in the number of patients and studies between the two institutions. While KSUMC shows a one-to-one correspondence, some NGHA patients underwent multiple ultrasound imaging sessions during the study period.

The 304 NAFLD patients from KSUMC in this research were primarily female. Table 3 shows (72.3 %) female patients and (27.7 %) men. Other NAFLD investigations have shown a greater disease prevalence in women. Hormonal variations during menstruation and menopause may affect lipid metabolism and insulin sensitivity, explaining this discrepancy. Our dataset emphasizes age and gender, which are crucial to NAFLD diagnosis. Privacy concerns and the need to standardize ultrasound data capture across a homogenous group prevented the collection of ethnicity and pre-existing conditions. Future datasets should include more demographic characteristics to improve generalizability. Women's more significant obesity and metabolic syndrome rates may also explain the trend. More study is needed to understand this gender gap and its effects on NAFLD treatment. The table reveals a higher prevalence of male patients (36) with NAFLD compared to female patients (28) in NGHA, due to the military hospital's male workforce, where NAFLD is typically more prevalent in females.

Finally, regarding the dataset gathered from KSUMC, Fig. 3 shows the dispersion of Saudi patient residences. Since the study's medical facility was in Riyadh, 75 % of patients were from

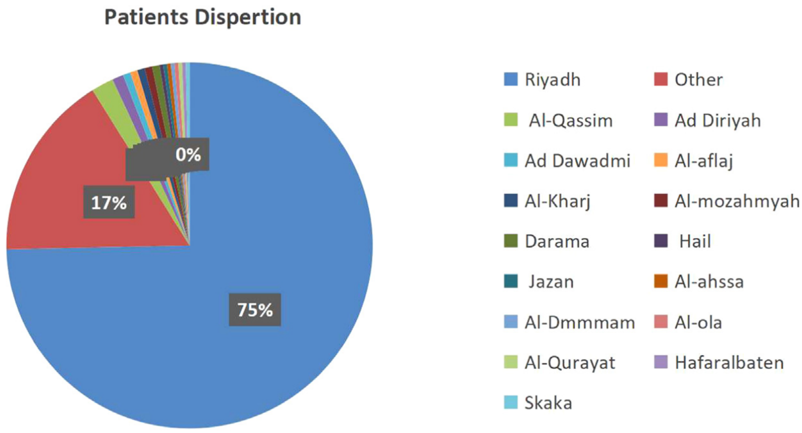

Fig. 3. Riyadh emerges as the primary residence for study participants (75 %).

there. Other prominent areas include Al-Qassim (16 %) and Ad Diriyah (7 %). Patients are referred to KSUMC in large metropolitan locations for expert NAFLD care in Saudi Arabia.

4.3. Biopsy acquisition and findings

The NAFLD Activity Score (NAS) assesses NAFLD severity and distinguishes between ordinary fatty liver and NASH. Steatosis, lobular inflammation, and hepatocyte ballooning are its primary targets. Based on the proportion of liver cells with fat, steatosis is rated 0–3, lobular inflammation 0–3, and hepatocyte ballooning 0–2. The fatty liver needs moderate steatosis (grade 1), 5 % or more significant fat build-up, whereas NASH requires a total NAS score of 5 or above. The minimum is mild steatosis and mild lobular inflammation or moderate hepatocyte ballooning. Thus, NAS assesses NAFLD severity and progression to NASH by measuring fat accumulation, inflammatory activity, and cellular harm.

Liver biopsy, while invasive, is universally acknowledged as the gold standard for the diagnosis and staging of liver disorders, including NAFLD. It offers comprehensive histological insights unattainable using non-invasive methods [10]. Nonetheless, possible sample mistakes and inter-observer heterogeneity provide difficulties to its definitive dependability [10]. To alleviate these issues, the research included rigorous procedural guidelines, including:

- Biopsy specimens were independently examined by three experienced hepatopathologists who were blinded to clinical information, therefore reducing bias.
- The NAFLD Activity Score (NAS) approach was uniformly used for the assessment of steatosis, inflammation, and fibrosis, guaranteeing standardized assessments [10].
- The inter-observer agreement was statistically assessed using Cohen's kappa, resulting in a high agreement level of 92.2 %, highlighting the reliability of the ground truth annotations [11].

The NAFLD Activity Score (NAS) is a powerful tool for diagnosing liver conditions and assessing their severity. This scoring system, encompassing the degree of steatosis, lobular inflammation, and hepatocyte ballooning, provides a nuanced picture of liver health beyond simply fat accumulation. In our study, we leverage the NAS advantage to accurately diagnose fatty liver and non-alcoholic steatohepatitis (NASH) in patients. Moreover, we utilize the NAS-based diagnoses to label corresponding ultrasound images, establishing a crucial link between histological evaluation and non-invasive imaging data. This linkage strengthens the dataset and paves the

Table 4
Distribution of steatosis patients and associated images based on the NAS scoring system.

NAS steatosis staging	No. of images from KSUMC	No. of patients from KSUMC	No. of images from NGHA	No. of patients from NGHA
<5 % (0)	3847	123	797	23
5–33 % (1)	3208	102	1129	27
>33–66 % (2)	1397	48	390	12
>66 % (3)	905	31	145	4
Total	9357	304	2461	66

way for developing robust machine-learning models capable of automated NAFLD diagnosis and condition monitoring based on readily available ultrasound images.

4.3.1. From KSUMC

All instances were liver biopsies done during cholecystectomy utilizing an 18 × 20 cm BARD Max-Core biopsy gun. To minimize the surgical effect on hepatic tissue, the biopsies were collected only from the right lobe before surgery to observe the biopsy site for bleeding. After extraction, the specimens were put in 10 % formalin, transported to the Histopathology Laboratory, embedded in paraffin blocks, and stained with hematoxylin-eosin and Masson's trichrome.

The histological NAFLD assessment score (NAS) was used to gather liver biopsy specimens from individuals with and without NAFLD. An expert hematopathologist analysed and interpreted all biopsies to guarantee histological integrity. Steatosis was graded as grade 0 (<5 %), grade 1 (5–33 %), grade 2 (>33–66 %), and grade 3 (>66 %). Normal steatosis was grade 0 and fibrosis stage 0–1, suggesting minor or moderate inflammation. At least 5 % steatotic hepatocytes, moderate lobular or portal inflammation, and stage 0–1 fibrosis indicated fatty liver. All liver samples met pathological analysis requirements with at least ten well-preserved portal tracts.

Hepatopathologists were blinded to clinical data throughout the procedure to guarantee fair interpretation. The investigations used the same steatotic hepatocyte grading, normal results, and fatty liver criteria. Histological diagnosis of non-alcoholic steatohepatitis (NASH) was based on the NAFLD Activity Score (5 or higher), steatosis grade, hepatocyte ballooning, and lobular inflammation with or without fibrosis. These stringent methods reinforce this study's conclusions' scientific rigor and dependability.

Table 4 details the distribution of NAS scores among the 304 KSUMC biopsy samples and their corresponding ultrasound images. As you can see, most patients (150, 49.3 %) presented with mild or moderate steatosis (NAS 1–2), represented by 4605 (49.2 %) of the total 9357 KSUMC ultrasound images in this category. This finding aligns with the overall dataset trend but leans even more towards earlier-stage NAFLD compared to other studies, further suggesting a potentially younger or less progressed NAFLD population at KSUMC. This warrants further investigation into the underlying factors contributing to this trend.

However, a significant 10.2 % of patients exhibited a NAS score of 3, indicative of borderline or definite NASH. This translates to 905 (9.7 %) of the KSUMC ultrasound images. This proportion is slightly higher than the overall dataset average and comparable to some more extensive studies, highlighting the presence of a notable subset at risk for NASH progression within the KSUMC cohort. This underscores the importance of accurate diagnosis and tailored intervention strategies for this population.

Analyzing the NAFLD fibrosis stage distribution based on the NAS scoring system (Table 5) revealed a predominance of mild stages. No fibrosis (stage 0) was observed in 4.6 % of patients. The majority (28 %) had perisinusoidal fibrosis (stage 1A), followed by perisinusoidal or periportal fibrosis (stage 1) in 6 %, perisinusoidal fibrosis (stage 1B) in 42 %, and portal/periportal fibrosis (stage 1C) in 5 %. Notably, only 5 % of patients exhibited advanced fibrosis stages, with bridging fibrosis (stage 3) present in 0.7 % and cirrhosis (stage 4) observed in 0.3 %. Interestingly,

Table 5
Distribution of fibrosis patients and associated images based on the NAS scoring system.

NAS fibrosis staging	No. of images from KSUMC	No. of patients from KSUMC	No. of images from NGHA	No. of patients from NGHA
None (0)	393	14	1426	42
Perisinusoidal or periportal (1)	7663	246	651	22
Perisinusoidal and portal/periportal (2)	1154	40	38	2
Bridging Fibrosis (3)	60	2	151	7
Cirrhosis (4)	40	1	146	5
Not Available	47	1	49	2
Total	9357	304	2461	80

while the number of images associated with each stage differs considerably (ranging from 393 for stage 0 to 4074 for stage 1B), this reflects the ease of identifying fibrosis at advanced stages through imaging techniques.

4.3.2. From NGHA

The retrospective analysis utilized liver biopsy specimens retrieved from the NGHAl central laboratory archives, spanning 2007 to 2023. Biopsies were acquired from patients undergoing a single-pass procedure targeting the left liver lobe with an 18-gauge needle. Following extraction, specimens were promptly immersed in 10 % formalin and transported to the Histopathology Laboratory for standard processing, including tissue embedding in paraffin blocks, sectioning, and staining with a suite of stains (hematoxylin-eosin, PAS, PASD, reticulin, VGM, and Masson's trichrome).

The histological NAFLD Activity Score (NAS) served as the primary criterion for selecting biopsy specimens from both control individuals and those suspected of having NAFLD. Three expert GI and liver pathologists independently evaluated and interpreted all biopsies acquired during the study period. Steatosis severity was graded on a four-point scale: grade 0 (<5 % fat accumulation), grade 1 (5–33 % fat accumulation), grade 2 (>33–66 % fat accumulation), and grade 3 (>66 % fat accumulation). Normal steatosis was defined as grade 0 with fibrosis stage 0–1, indicating minimal or moderate inflammatory activity. Fatty liver was diagnosed in biopsies exhibiting at least 5 % steatotic hepatocytes, moderate lobular or portal inflammation, and fibrosis stage 0–1. To ensure adequate tissue quality for analysis, all included specimens contained at least ten well-preserved portal tracts. Finally, a histological diagnosis of non-alcoholic steatohepatitis (NASH) was established based on a NAFLD Activity Score of 5 or higher, considering the combined severity of steatosis, hepatocyte ballooning, and lobular inflammation with or without fibrosis.

Table 4 displays the allocation of ultrasound images and patients in various stages of liver steatosis according to the NAS Steatosis Staging, specifically for 66 NGHA biopsy samples. Stage 1 has the greatest quantity of photos (1129) and patients (27), suggesting that mild steatosis is the prevailing condition in this sample. Stage 0 consists of 797 images from twenty-three patients, demonstrating regular imaging even in cases with mild severity. Stage 2 and Stage 3 exhibit a reduced number of photographs, specifically 390 and 145, respectively, as well as a smaller number of patients, namely 12 and 4, respectively. This implies a less frequent incidence or an earlier intervention at a more severe level. The dataset consists of 2461 photos collected from sixty-six individuals, with an average of around thirty-seven photographs per patient, highlighting the importance of thorough surveillance. The distribution of this data highlights the high occurrence of mild steatosis and emphasizes the importance of using comprehensive imaging techniques to track the course of the disease. These findings offer significant information for both clinical and research purposes.

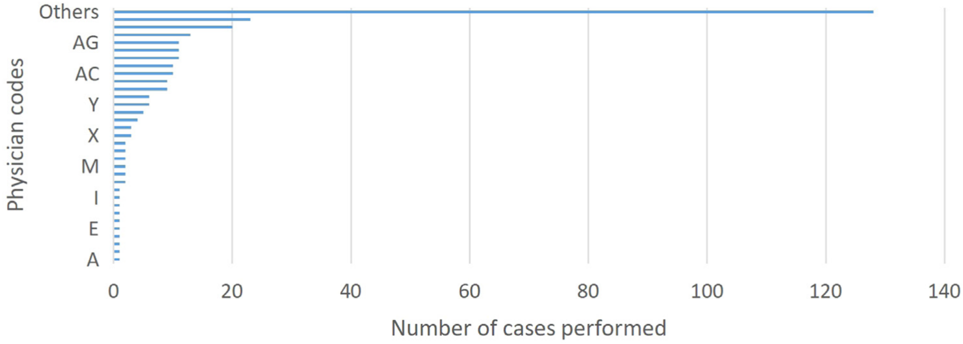

Fig. 4. Distribution of performing physicians and number of cases in KSUMC.

Table 6
Distribution of liver ultrasound acquisition modalities from both institutions used in the study.

Modality manufacturer	Model name	No. of studies from KSUMC	No. of studies from NGHA
GE Medical Systems	LOGIQ9	185	4
	LOGIQE9	0	4
	LOGIQE10s	0	1
Philips Medical Systems	EPIQ 7G	79	1
	EPIQ 5G	0	3
	CX50	0	1
	EPIQ Elite	0	4
	iU22	39	61
Hitachi Medical Co.	EUB-8500	1	0
SAMSUNG	MySono U6	0	1
Total		304	80

4.4. Ultrasound images acquisition and findings

In the pursuit of advancing diagnostic capabilities for Non-Alcoholic Fatty Liver Disease (NAFLD) at KSUMC and NGHA, our methodology centers around the meticulous data collection process facilitated by the GEIIS PACS system developed by GE Healthcare. This system serves as the primary repository for ultrasound medical images, providing a structured framework for of stored, managed, and distributed data within healthcare facilities.

4.4.1. From KSUMC

Ultrasound image acquisition involved a diverse group of physicians at KSUMC, as shown in Fig. 4. The data reveals that over thirty-three distinct physicians contributed to collecting liver ultrasound images in DICOM format. Among these physicians, some managed only a few cases, while others managed more cases. Notably, ten physicians were responsible for 42 % of the total cases acquired. This distribution suggests that while multiple physicians participated in the image acquisition process, a small subset contributed to a significant portion of the collected data.

Liver ultrasound images were acquired using a variety of devices from four different manufacturers: GE Medical Systems, Philips Medical Systems, Hitachi Medical Co, and Samsung. As shown in Table 6, a total of ten distinct ultrasound models were employed across these manufacturers. Notably, most patients underwent ultrasound examinations with GE Medical Systems devices, primarily using the LOGIQ9 model (49.2 %). Conversely, Philips Medical Systems' EPIQ 7G and iU22 were used for a smaller proportion of patients (20.8 %) and (26 %), respectively. However, Hitachi Medical Co.'s EUB-8500 model and Samsung's MySono U6 model were used only on one patient each. This distribution reflects the prevalence of GE Medical Systems equip-

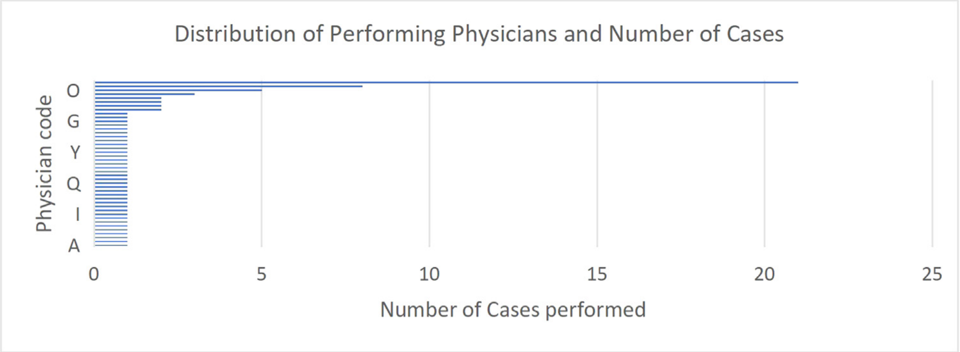

Fig. 5. Distribution of performing physicians and number of cases in NGHA.

Table 7
Distribution of liver ultrasound image dimensions and number of associated images.

Image size	No. of images from KSUMC	No. of images from NGHA
1080 × 1920	0	1
970 × 1292	0	4
768 × 1024	5568	2430
720 × 960	0	19
600 × 800	0	7

ment within the clinical setting where the data was collected. When interpreting ultrasound image findings, it is crucial to consider the potential influence of varying technical specifications and image acquisition protocols associated with different manufacturers and models.

4.4.2. From NGHA

Ultrasound image acquisition at NGHA also involved a diverse group of physicians, as shown in Fig. 5. The NGHA data reveals that a considerable number of physicians, represented by different codes, contributed to the collection of liver ultrasound images. Among these physicians, some managed only a few cases, while others managed more cases.

At NGHA, radiologists acquired liver ultrasound images using a variety of devices from three different manufacturers: GE Medical Systems, Philips Medical Systems, and Samsung. As shown in Table 6, radiologists employed a total of nine distinct ultrasound models across these manufacturers. Notably, Philips Medical Systems devices were used, with the iU22 model accounting for sixty-one studies. Radiologists also used GE Medical Systems' LOGIQ9 and LOGIQE9 models in four studies each, while Samsung's MySono U6 model was employed in one study. This distribution highlights the varied equipment usage within NGHA's clinical setting.

4.4.2.1. Comparative analysis of ultrasound image acquisition at KSUMC and NGHA. The collected liver ultrasound images from both KSUMC and NGHA exhibited various dimensions, as detailed in Table 7. The two most prevalent image sizes were 768 × 1024 pixels and 480 × 640 pixels, representing 67.7 % and 32 % of the total images at KSUMC, respectively. At NGHA, the image sizes included 1080 × 1920 pixels, 970 × 1292 pixels, 768 × 1024 pixels, 720 × 960 pixels, and 600 × 800 pixels, with 768 × 1024 pixels being the most common (2430 images). While these dimensions may reflect specific acquisition preferences or device capabilities, it is important to acknowledge that healthcare providers and institutions do not universally enforce standardized image dimensions in liver ultrasound imaging. Variations can occur based on specific equipment, settings, and clinical requirements.

Notably, only one image at KSUMC was found to have a unique dimension of 574 × 760 pixels, and another image had a unique dimension of 1080 × 1920 pixels. For NGHA, the unique dimensions were more varied, including multiple images across varied sizes. For a more precise understanding of the potential factors influencing image dimensions in this study, further investigation into the specific devices used, acquisition protocols employed, and relevant recent literature on standard liver ultrasound protocols is recommended. This information will provide valuable context for interpreting the acquired data and ensure accurate conclusions about the potential implications of image size variations.

The technical specifications of the stored images adhere to the Digital Imaging and Communications in Medicine (DICOM) format, ensuring a standardized and interoperable representation across diverse medical systems. Each pixel sample within the images is encoded with eight bits, with the highest bit at position seven, a crucial detail for accurate interpretation. Furthermore, the categorization of images into "Ultrasound Image Storage" and "Secondary Capture Image Storage" provides valuable insights into the original, unaltered images and their modified counterparts, enriching the dataset for nuanced diagnostic analysis.

Ensuring data integrity and preserving image quality are paramount considerations in our methodology. The decision to eschew lossy compression in image storage, coupled with the meticulous representation of 921,600-pixel elements per image, reflects our commitment to maintaining fidelity in ultrasound data. This approach establishes a robust foundation for accurate diagnostic interpretation, reinforcing our dedication to advancing healthcare outcomes in NAFLD diagnosis at KSUMC City and NGHA.

4.5. Data pre-processing and anonymization

To prepare the images for analysis, 12,766 DICOM files were processed through a systematic pipeline to ensure quality and privacy. This included converting DICOM images to PNG format using lossless compression, followed by anonymization and standardization to 768 × 1024 pixels using padding to preserve pixel integrity. As illustrated in Fig. 2.

Three expert radiologists, each with over ten years of experience in ultrasound imaging, participated in the initial screening process. Radiologists 1 and 2 independently examined each DICOM file using MicroDicom DICOM Viewer [12], evaluating two key criteria:

1. Image completeness: Ensuring the liver image was clear and fully captured.
2. Image quality: Assessing whether the image quality was sufficient for further analysis.

To ensure reliability and transparency, we calculated the inter-rater agreement using Cohen's kappa coefficient [11]. This statistic is appropriate for measuring agreement between two raters on categorical decisions (inclusion/exclusion). Any conflicts in image inclusion were resolved by Radiologist 3.

Following the expert assessment, 2414 DICOM files were excluded:

1. Limited studies (incomplete liver visualization): 1300 images
2. Suboptimal imaging (poor image quality): 1114 images

The remaining DICOM files were converted to PNG format using MicroDicom DICOM Viewer [12], carefully preserve the photometric interpretation of each image. During this conversion process, we also categorized the images based on their colour properties. Of the converted images, 8772 were grayscale, and 1580 were colour images. The conversion resulted in PNG images with a range of dimensions. Most of the images, 7029 in total, were converted to 768 × 1024 pixels. The second most common size was 480 × 640 pixels, accounting for 3291 images. A small number of images were converted to other dimensions: nineteen images at 720 × 960 pixels, seven images at 600 × 800 pixels, four images at 970 × 1292 pixels, and single instances of images at 1080 × 1920 pixels and 574 × 760 pixels. This variation in image sizes and colour properties necessitated further standardization in subsequent processing steps to ensure consistency across the dataset.

ImageJ software [13] was employed to manually crop the images, removing patient-identifying information and selecting the region of interest (ROI) containing the liver. We developed a Python script to add padding and centre the images to standardize image dimensions while maintaining image quality. The final standard size of 768 × 1024 pixels was chosen, as it represented the dimensions of approximately 70 % of the processed images.

It is important to note that we deliberately avoided resizing techniques to maintain image quality. Resizing, particularly downscaling, can lead to loss of fine details and introduce artifacts, potentially compromising the diagnostic value of ultrasound images. By using padding instead, we preserved the original pixel information while achieving a uniform image size across the dataset.

The pre-processing and anonymization steps resulted in a final dataset of 10,352 standardized PNG images, each with dimensions of 768 × 1024 pixels. In the Technical Validation section, we will elaborate on the methods used to validate the radiologists' image selection process and confirm that image quality was preserved throughout the pre-processing pipeline, from initial DICOM files to the final padded and centred PNG images.

Next, to ensure the reliability and quality of our dataset, we conducted rigorous technical validation processes focusing on two key aspects: the accuracy of image selection and the preservation of image quality throughout the processing pipeline.

4.6. Validation of image selection

The selection of appropriate fatty liver DICOM images was crucial to the integrity of our dataset. To validate this process, we employed a robust inter-rater reliability assessment between Radiologist 1 and Radiologist 2 using Cohen's kappa coefficient [11]. This statistical measure is particularly suitable for evaluating agreement on categorical decisions such as image inclusion or exclusion.

$$K = (po - pe)/(1 - pe)$$

where:

- *po: Observed agreement, the proportion of cases where the two raters agree.*
- *pe: Expected agreement, the proportion of agreement that would occur by chance.*

Cohen's kappa quantifies the level of agreement beyond what is expected by random chance, with values closer to 1 indicating stronger agreement.

The resulting Cohen's kappa coefficient showed an agreement of 92.2 % between Radiologist 1 and Radiologist 2. This prominent level of agreement translates to consensus on approximately 9544 images out of the total 10,352 in the final dataset, indicating strong inter-rater reliability. For the remaining images where disagreement occurred, Radiologist 3 made the final inclusion/exclusion decision. This process ensured that only the most appropriate and high-quality liver ultrasound images were included in the dataset, enhancing its overall reliability for future research applications.

4.7. Validation of image quality preservation

To ascertain that the image processing procedures, namely the transformation from DICOM to PNG and the ensuing standardization to 768 × 1024 pixels, did not modify the intrinsic image attributes, we performed an exhaustive image intensity analysis [14]. We used histogram charting to evaluate pixel intensity distributions between sample DICOM images from the original dataset and their respective processed PNG images after padding and centring. This method is crucial for demonstrating the preservation of image information throughout our processing pipeline.

Histogram analysis is widely accepted for validating image fidelity as it captures the distribution of pixel intensities across the image, which directly correlates with critical diagnostic features, such as texture and contrast. By demonstrating consistency in histograms before and after conversion, we ensure that the preprocessing pipeline maintains diagnostic information essential for both clinical evaluation and AI-based analysis [4,14].

Fig. 6 illustrates the histogram comparisons [14], showing intensity distributions for images before and after preprocessing and standardization. Specifically, 1.A and 2.A presents the intensity histograms for two random DICOM images before and after preprocessing, respectively. 1.B and 2.B show the corresponding images for these histograms. The consistency in these histograms confirms that our processing steps, including format conversion, padding, and centring, did not introduce significant alterations to the pixel intensity values.

To further validate the preservation of image quality, we calculated similarity metrics between the histograms of original DICOM images and their PNG counterparts. The results showed a Pearson correlation coefficient of 0.998, indicating near-perfect alignment in pixel intensity distributions, and confirming that no significant information loss occurred during the conversion [14].

The consistent pixel intensity distributions observed ensure that clinically relevant features, such as liver texture, echogenicity, and contrast gradients, are preserved. These features are critical for the accurate staging and grading of NAFLD and remain intact through the conversion process [5,10].

This validation is particularly important for several reasons:

1. Data Integrity: It ensures that the essential diagnostic information contained in the pixel intensities remains unchanged, preserving the clinical relevance of the images.
2. Reproducibility: Researchers using this dataset can be confident that the images accurately represent the original ultrasound data, facilitating reproducible studies.
3. Standardization Without Compromise: It demonstrates our success in achieving a standardized image format and size without sacrificing image quality or altering pixel intensity distributions.
4. Applicability for Advanced Analysis: The preserved intensity information makes the dataset suitable for advanced image analysis techniques, including machine learning and deep learning applications in liver disease diagnosis.

By rigorously validating both the image selection process and the preservation of image quality, we have ensured that our dataset meets high standards of reliability and accuracy. This technical validation strengthens the dataset's value for future research in hepatology, radiology, and artificial intelligence in medical imaging.

4.8. Class identification and labelling

To organize and categorize the liver ultrasound images, we implemented a systematic labelling process using Python scripts. This process integrated clinical data from Liver Biopsy.xlsx files provided by participating hospitals, allowing us to accurately classify each image according to both NAS (NAFLD Activity Score) Fibrosis Staging and NAS Steatosis Grading.

The labelling process resulted in the following classification structure:

NAS Fibrosis Staging:

- Stage 0 (No fibrosis): 1751 images (1407 Grayscale, 344 coloured)
- Stage 1 (Mild fibrosis): 7122 images (6101 Grayscale, 1021 coloured)
- Stage 2 (Perisinusoidal fibrosis): 1015 images (861 Grayscale, 154 coloured)
- Stage 3 (Bridging fibrosis): 203 images (178 Grayscale, 25 coloured)
- Stage 4 (Cirrhosis): 174 images (152 Grayscale, 22 coloured)

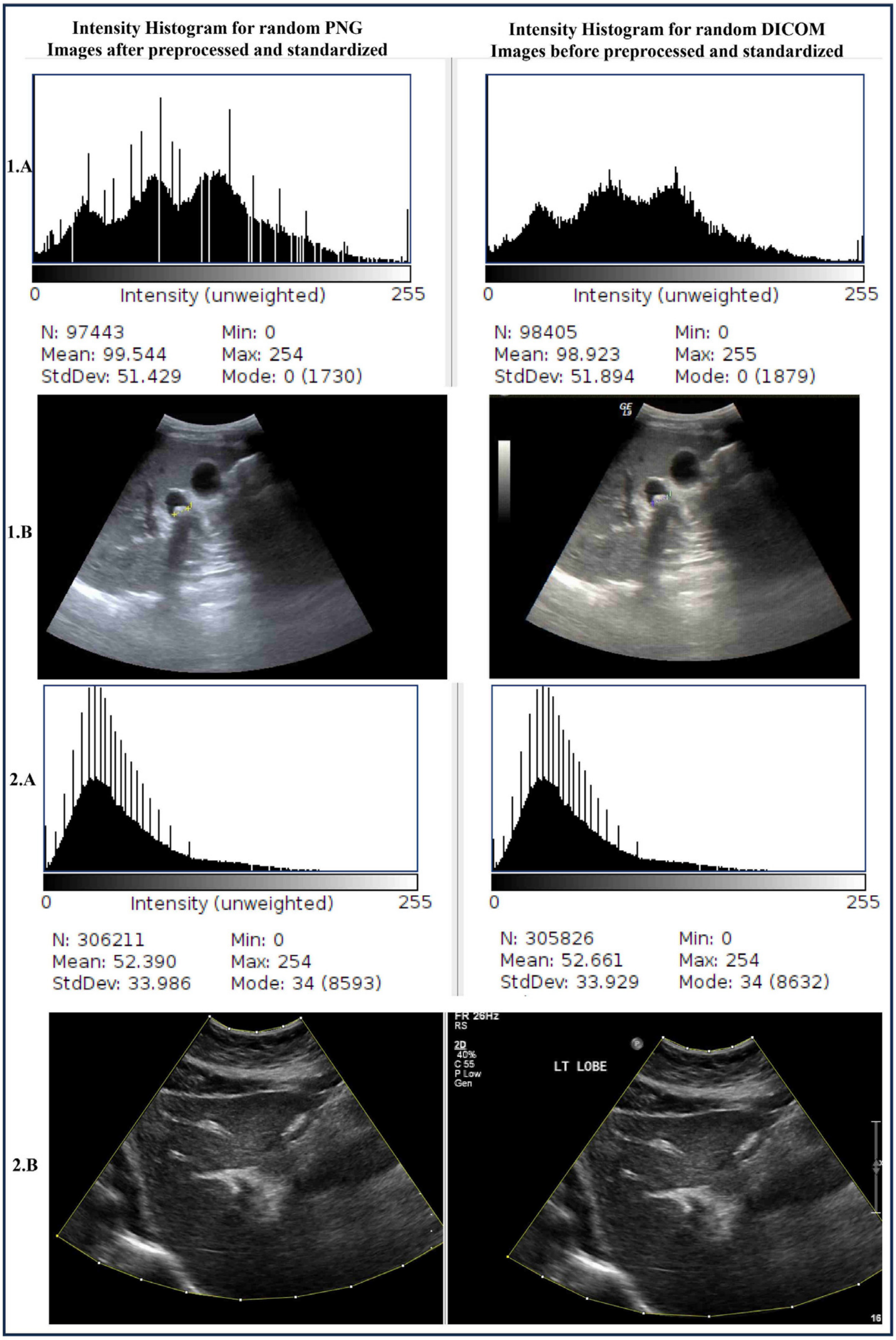

Fig. 6. This figure shows intensity histograms and corresponding ultrasound images before and after preprocessing. Panels 1.A and 2.A compare histograms for random DICOM images before (right) and after (left) preprocessing, demonstrating minimal changes in pix pixel intensity. Panels 1.B and 2.B display the corresponding images, confirming quality preservation.

NAS Steatosis Grading:

- Grade 0: 3283 images (2819 Grayscale, 464 coloured)
- Grade 1: 2597 images (2093 Grayscale, 504 coloured)
- Grade 2: 2283 images (1860 Grayscale, 423 coloured)
- Grade 3: 1144 images (985 Grayscale, 159 coloured)

It is important to note that a small subset of images remained ungraded in our dataset. Specifically, 145 images (115 grayscale and 30 coloureds) lack steatosis grading, while 87 images (73 grayscale and 14 coloureds) are ungraded for fibrosis. These ungraded images represent cases where definitive classification was not possible based on the available clinical data. Their inclusion in the dataset maintains transparency and may provide opportunities for future refinement of grading processes or specialized studies on challenging cases in NAFLD assessment.

This structured labelling approach enables researchers to easily select subsets of images based on specific fibrosis stages or steatosis grades, facilitating targeted analyses and machine learning model development for various aspects of NAFLD progression.

Limitations

The dataset outlined in this paper has limitations that may influence its generalizability and reliability. A significant drawback is the dataset's tilt towards middle-aged persons, which constrains its application to other age groups and may hinder its use for comprehensive demographic study. The use of diverse ultrasound apparatus across several institutions, including many manufacturers and models, may have resulted in discrepancies in image quality and acquisition techniques. Despite attempts to standardize image processing, heterogeneity in source equipment may affect dataset homogeneity. Additionally, the observer-dependency of ultrasound imaging, where the quality of the interpretation can vary based on the operator's expertise and experience, is another challenge. This subjectivity could influence the consistency in identifying and grading steatosis or fibrosis in NAFLD diagnosis. A segment of the gathered images was omitted owing to inadequate liver visualization or poor image quality, potentially limiting the dataset's variety. Notwithstanding these constraints, the dataset continues to be significant and beneficial for the progression of AI-assisted diagnosis of NAFLD.

Ethics Statement

This research included human participants and was executed in complete compliance with the Declaration of Helsinki, Good Clinical Practice standards, and local ethical regulations. Ethical oversight was conducted by the Institutional Review Boards (IRBs) of King Saud University Medical City (KSUMC), which granted approval number E-23-7672 for the study entitled "Develop Deep Learning Solutions for Detecting and Staging Fatty Liver using Ultrasound Images" and the King Abdullah International Medical Research Center (KAIMRC) at National Guard Health Affairs (NGHA), under approval number SP23R/010/01. The inclusion criteria guaranteed the dataset's relevance to Non-Alcoholic Fatty Liver Disease (NAFLD) patients, while exclusion criteria were implemented to exclude people with other liver disorders.

Informed permission was secured from all participants in compliance with local laws and regulations, safeguarding patient confidentiality, anonymity, and autonomy. The data gathering rigorously followed ethical rules, with the dataset, including anonymized ultrasound pictures and biopsy results, securely saved on the researcher's computer. The researchers adhered to stringent data management and security policies throughout the study, guaranteeing the confidentiality of the data and compliance with institutional standards.

Declaration of generative AI and AI-assisted technologies in the writing process

While preparing this work, the authors utilized QuillBot and Grammarly to improve the manuscript's language. After using these tools, the authors thoroughly reviewed and edited the content as needed and assumed full responsibility for the publication's final content.

Data Availability

Large Annotated Ultrasound Dataset of Non-Alcoholic Fatty Liver from Saudi Hospitals for Analysis and Applications (Original data) (Open Science framework or (OSF)).

CRediT Author Statement

Fahad Alshagathrh: Conceptualization, Methodology, Data curation, Formal analysis, Project administration, Writing – original draft; **Mahmood Alzubaidi:** Data curation, Writing – original draft; **Khalid Alswat:** Supervision, Project administration, Resources, Writing – review & editing; **Ali Aldhebaib:** Validation, Writing – review & editing; **Bushra Alahmadi:** Data curation, Writing – review & editing; **Meteb Alkubeyyer:** Validation, Writing – review & editing; **Abdulaziz Alosaimi:** Validation, Writing – review & editing; **Amani Alsadoon:** Project administration, Writing – review & editing; **Maram Alkhamash:** Data curation, Writing – review & editing; **Jens Schneider:** Supervision, Writing – review & editing; **Mowafa Househ:** Supervision, Writing – review & editing.

Acknowledgements

This study was made possible with the support of the Liver Disease Research Center, Department of Medicine, College of Medicine, King Saud University, Saudi Arabia, Riyadh. The center is funded by the National Plan for Science, Technology, and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, grant Number 08-MED512-02.

Special thanks to National Guard Health Affairs/Ministry of National Guard, Riyadh, Saudi Arabia, for providing us with histopathological and radiological data related to Non-Alcoholic Fatty Liver Disease (NAFLD) between 2009 and 2023.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Alswat, A.A. Aljumah, F.M. Sanai, F. Abaalkhail, M. Alghamdi, W.K. Al Hamoudi, A. Al Khathlan, H. Al Quraishi, A. Al Rifai, M. Al Zaabi, M.A. Babatin, C. Estes, A. Hashim, H. Razavi, Nonalcoholic fatty liver disease burden—Saudi Arabia and United Arab Emirates, 2017–2030, Saudi J. Gastroenterol. 24 (2018) https://journals.lww.com/sjga/fulltext/2018/24040/nonalcoholic_fatty_liver_disease_burden___saudi.3.aspx.

[2] S. Saadeh, Z.M. Younossi, E.M. Remer, T. Gramlich, J.P. Ong, M. Hurley, K.D. Mullen, J.N. Cooper, M.J. Sheridan, The utility of radiological imaging in nonalcoholic fatty liver disease, Gastroenterology 123 (2002) 745–750, doi:10.1053/gast.2002.35354.

[3] S. Ballestri, A. Mantovani, C.D. Byrne, A. Lonardo, G. Targher, Diagnostic accuracy of ultrasonography for the detection of hepatic steatosis: an updated meta-analysis of observational studies, Metab. Target Organ Damage 1 (2021) 7, doi:10.20517/mtod.2021.05.

[4] F.M. Alshagathrh, M.S. Househ, Artificial intelligence for detecting and quantifying fatty liver in ultrasound images: a systematic review, Bioengineering 9 (2022), doi:10.3390/bioengineering9120748.

[5] G. Petzold, Role of ultrasound methods for the assessment of NAFLD, J. Clin. Med. 11 (2022), doi:10.3390/jcm11154581.
[6] I. Sullivan, A. DeHaven, D. Mellor, Open and reproducible research on open science framework, Curr. Protoc. Essent. Lab. Tech. 18 (2019), doi:10.1002/CPET.32.
[7] P.C. Eng, R. Forlano, T. Tan, P. Manousou, W.S. Dhillo, C. Izzi-Engbeaya, Non-alcoholic fatty liver disease in women—Current knowledge and emerging concepts, JHEP Rep. 5 (2023) 100835, doi:10.1016/j.jhepr.2023.100835.
[8] M. Alzubaidi, M. Agus, K. Alyafei, K.A. Althelaya, U. Shah, A. Abd-Alrazaq, M. Anbar, M. Makhlouf, M. Househ, Toward deep observation: a systematic survey on artificial intelligence techniques to monitor fetus via ultrasound images, iScience 25 (2022), doi:10.1016/j.isci.2022.104713.
[9] R.J. Amdur, C. Biddle, Institutional review board approval and publication of human research results, JAMA 277 (1997) 909–914, doi:10.1001/jama.1997.03540350059034.
[10] I.L.K. Nalbantoglu, E.M. Brunt, Role of liver biopsy in nonalcoholic fatty liver disease, World J. Gastroenterol. 20 (2014) 9026–9037, doi:10.3748/wjg.v20.i27.9026.
[11] R. Chow, N. Chiu, E. Bruera, M. Krishnan, L. Chiu, H. Lam, C. DeAngelis, N. Pulenzas, S. Vuong, E. Chow, Inter-rater reliability in performance status assessment among health care professionals: a systematic review, Ann. Palliat. Med. 5 (2016) 83–92, doi:10.21037/apm.2016.03.02.
[12] MicroDicom, MicroDicom DICOM Viewer, (n.d.).
[13] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675, doi:10.1038/nmeth.2089.
[14] J.F. Peters, Visualising pixel intensity distributions, in: Found. Comput. Vis. Comput. Geom. Vis. Image Struct. Object Shape Detect, Springer International Publishing, Cham, 2017, pp. 125–144, doi:10.1007/978-3-319-52483-2_3.