



Published in final edited form as:

AI (Basel). 2022 September ; 3(3): 739–750. doi:10.3390/ai3030043.

Can Sequential Images from the Same Object Be Used for Training Machine Learning Models? A Case Study for Detecting Liver Disease by Ultrasound Radiomics

Laith R. Sultan^{1,*†}, Theodore W. Cary^{2,†}, Maryam Al-Hasani¹, Mrigendra B. Karmacharya², Santosh S. Venkatesh³, Charles-Antoine Assenmacher⁴, Enrico Radaelli⁴, Chandra M. Sehgal¹

¹Ultrasound Research Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

²Department of Radiology, Children Hospital of Philadelphia, Philadelphia, PA 19104, USA

³Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

⁴Department of Pathobiology, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

Machine learning for medical imaging not only requires sufficient amounts of data for training and testing but also that the data be independent. It is common to see highly interdependent data whenever there are inherent correlations between observations. This is especially to be expected for sequential imaging data taken from time series. In this study, we evaluate the use of statistical measures to test the independence of sequential ultrasound image data taken from the same case. A total of 1180 B-mode liver ultrasound images with 5903 regions of interests were analyzed. The ultrasound images were taken from two liver disease groups, fibrosis and steatosis, as well as normal cases. Computer-extracted texture features were then used to train a machine learning (ML) model for computer-aided diagnosis. The experiment resulted in high two-category diagnosis using logistic regression, with AUC of 0.928 and high performance of multicategory classification, using random forest ML, with AUC of 0.917. To evaluate the image region independence for machine learning, Jensen–Shannon (JS) divergence was used. JS distributions showed that images of normal liver were independent from each other, while the

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

*Correspondence: lsultan@pennmedicine.upenn.edu.

†These authors contributed equally to this work.

Author Contributions: Conceptualization, L.R.S., T.W.C., S.S.V. and C.M.S.; methodology, L.R.S., T.W.C. and C.M.S.; software, T.W.C.; validation, L.R.S., T.W.C. and C.M.S.; formal analysis, L.R.S., T.W.C., M.A.-H., C.-A.A., E.R. and M.B.K.; investigation, L.R.S. and T.W.C.; resources, C.M.S.; data curation, L.R.S. and T.W.C.; writing—original draft preparation, L.R.S. and T.W.C.; writing—review and editing, L.R.S., C.M.S. and T.W.C.; visualization, L.R.S.; supervision, C.M.S.; project administration, L.R.S., T.W.C. and M.B.K.; funding acquisition, C.M.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Institutional Animal Care and Use Committee and approved by the University of Pennsylvania (IACUC number 804998, approved on 26 September 2017).

Conflicts of Interest: The authors declare no conflict of interest.

images from the two disease pathologies were not independent. To guarantee the generalizability of machine learning models, and to prevent data leakage, multiple frames of image data acquired of the same object should be tested for independence before machine learning. Such tests can be applied to real-world medical image problems to determine if images from the same subject can be used for training.

Keywords

medical imaging; quantitative ultrasound; machine learning; independent data; liver disease

1. Introduction

Image data availability is vital for the implementation of machine learning (ML) methods in clinical settings [1,2]. Large datasets with high-quality images are essential for training, validation, and testing of algorithms for clinical applications. Typically, ML algorithm performance for predictive tasks increases with increased training data volume [3–5]. If numerous parameters are to be studied, there is a need to train ML models commensurately on abundant data to obtain a generalizable model. However, there is limited access to medical images, and the preparation of image data is a costly and time-intensive process [6]. Most health care systems are not adequately equipped to share large numbers of medical images [3]. Medical data are often stored in silos and are not available in enterprise-wide core clinical systems. These data silos prevent relevant data from being shared even between departments within the same institution due to the interoperability issues of the current institutional enterprise systems [7]. Image data therefore cannot be accessed easily by AI algorithm development for widespread clinical practice.

In addition to the data availability issues, in ML a common assumption is that the given data points are realizations of independent random variables [8]. However, this assumption is often violated when the data points are highly interdependent (e.g., when the data exhibit temporal or spatial correlations) [9]. Similar scenarios are typical situations in visual recognition and computational biology [10]. Dependent data arise whenever there are inherent correlations in between observations. This is to be expected for time series of imaging data, where we would intuitively expect that instances with similar time stamps have stronger dependencies than ones that are far away in time.

A common approach to bypass the problem of limited data is to use multiple images from the same subject as separate training instances for ML [11–14]. However, this approach raises the question of whether the data are independent. Depending on the independence assumptions of the learning algorithms, the performance of the resulting models trained and tested on the same patient(s) and same body region(s) might be inflated, and the models might not be generalizable to future images. In this situation, the predictive model developed using conventional ML algorithms could be biased, inaccurate, and tend to produce unsatisfactory classifiers. A common example where ML algorithms are well known to exhibit variations in prediction accuracy is when ML is provided with imbalanced training sets due to the imbalanced ratio of pathological and normal cases typically seen in

medical imaging [15]. Previous studies reported that a close-to-balanced training is required for best model performance, while data imbalance can have negative influence on the model performance.

In this study, we propose an approach to test the data independence for building a reliable diagnostic ML model using a liver disease data images set. For this purpose, we examined the independence of sequential ultrasound image frames acquired from the same cases of liver disease. We algorithmically extracted numerical liver texture features from the ultrasound images for machine learning. All these computer-generated features were used to train models. The independence between image region grayscale distributions were quantified by Jensen–Shannon (JS) divergence, a bounded symmetrization of the unbounded Kullback–Leibler (KL) divergence [16–19]. JS divergence was measured for B-mode ultrasound images acquired from images of three pathologies: normal cases, and then two groups of liver disease, namely steatosis (fatty liver) and fibrosis.

2. Methods

2.1. Image Acquisition and Computerized Analysis

1180 B-mode ultrasound images acquired in vivo from rat livers were used for analysis. The images were taken from 3 different rat groups as follows: 450 images from fibrosis cases (rats $n = 6$), 450 images from steatosis cases (rats $n = 4$), and 280 from normal (rats $n = 4$). Four video clips of B-mode images were acquired from each rat in standard transverse and sagittal imaging planes of the right and left lobes of the liver. Each clip consisted of average of 25–35 images. Imaging presets (gain = 18 dB, high sensitivity, 100% power, transmit frequency 21 MHz, and high line density) and time compensation gain were optimized and standardized.

Five to six identical rectangular regions of interest (ROI) were placed manually on each image to ensure comprehensive inclusion of multiple representative parts of the liver parenchyma and exclusion of imaging artifacts such as acoustic shadowing, enhancement, or reverberation. A total of 5903 regions of interest were placed on the images.

A number of texture features were extracted from the ROIs, which include:

1. *First order histogram features*: including echo intensity, heterogeneity (regional variance between ROIs, internal heterogeneity (local variance within ROIs) [20]. Echo intensity and heterogeneity represent the mean and standard deviation of intensity within an ROI. Heterogeneity is the standard deviation of the echo intensity between the ROIs in all the planes measured throughout the liver.
2. *Run length features* include gray-level nonuniformity (GLNU) and run length nonuniformity (RLNU). These features represent the length of the run, usually the number of pixels for the horizontal or vertical scan direction, or the number of pixels multiplied by a diagonal direction [21].
3. *Entropy*: a gray level connectivity texture feature was also studied [21].

All image analysis was performed using a custom application written in the IDL (Interactive Data Language) programming language (version 8.5; Harris Geospatial, Broomfield, CO, USA) [21].

2.2. Feature Statistics and Machine Learning Diagnostic Models

The mean and standard error for the ultrasound texture features of the three different groups were compared by two-tailed paired Student's t -tests. $p < 0.05$ was considered significant. One-way analysis of variance (ANOVA) was used to compare the difference between the three study arms. Statistical analysis was performed using MedCalc (version 19.0.5, MedCalc Software Ltd., Ostend, Belgium).

Two classifiers were used for machine learning analysis. Random Forest [22] was used for multicategory classification, while logistic regression is used for the two groups' separation. Leave-one-out cross-validation approach (round-robin) was used for training and testing the data with both classifiers. Training and testing of data were performed using Weka software (version 3.8.5, University of Waikato, Hamilton, New Zealand) [23].

2.3. Intra- and Inter-Case Divergence Analysis

Jenson–Shannon (JS) divergence [16–19] was used to quantify the difference in grayscale distribution between two regions, for both intracase and intercase sampling. JS divergence offers an information-theoretic set-similarity measure that works naturally for pair-wise comparisons. To evaluate intra- and inter-divergence, we compared intracase to intercase pairs, calculating JS divergence for every pair. Intracase pairs were sampled for every possible time shift, and their divergence distributions were then tested against the divergence distribution of the intercase region pairs. The goal was to find the minimum time difference between image regions of the same case such that their divergences were distributed similarly to regions sample from completely different subjects. For each test at each time shift, the null hypothesis was that the distributions were different, so we performed a t -test for significant similarity or equivalence (not the more common Student's t -test for significant difference) [24].

For equivalence, to demonstrate a “lack of difference”: The t -test for equivalence, where δ depends on how much nonequivalence is acceptable in the research study. In our example, we were unwilling to accept more than 5% reduction in intra-divergence compared to inter-divergence, so we set δ equal to $0.05 M_1$ (where the 1 subscript indicates inter, and 2 subscript indicates intra):

$$t(df) = \frac{M_1 - M_2 - \delta}{\sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} \times \left[\frac{1}{n_1} \frac{1}{n_2} \right]}}$$

The denominator is simply the standard error; df is degrees of freedom or $n_1 + n_2 - 2$; and the M is mean. This was a one-sided test in this study, because we wanted to prove that intrasampled cases do not have significantly lower divergence than intercases. It is a

noninferiority test because we wanted to prove that choosing our samples from intracases performed no worse than choosing from our intercases.

3. Histopathologic Validation

Liver disease was confirmed by histopathological examination. The liver lobes were assessed in a blind fashion by a vet pathologist for fibrosis and lipidosis on gross pathology. Portions of liver were preserved in 10% phosphate-buffered formalin and transferred to 50% ethanol after 48 to 72 h and then embedded in paraffin and processed for histological examination with hematoxylin and eosin (H&E) and trichrome staining. Each histologic section was graded according to the METAVIR scoring system for fibrosis. Lipidosis was investigated in addition to presence of balloon cells, which is critical to finding fatty liver changes.

4. Results

4.1. The Classification Performance of Ultrasound Features

All features showed statistically significant differences between the three groups. Table 1 shows the difference in mean values of liver texture ultrasound features between the three groups

Logistic regression two-class analysis showed high performance. First, the model was able to detect the disease from normal cases with AUC 0.917 (Figure 1). Then, differentiation of the two liver disease groups, namely steatosis and fibrosis, showed a very high diagnostic performance with AUC of 0.928.

Random forest learning for multicategory classification also showed high performance in differentiation of the three groups (Figure 2). The model showed that the features can differentiate all three groups from each other with high diagnostic performance ranging from 0.854 to 0.917 with sensitivity up to 83.8 and specificity reaching 83.3.

4.2. Divergence Testing for Image Independence

Of the three tested liver pathologies, only normal cases demonstrated that intracase region divergence is statistically close to the intersampled case divergence (Table 2, Figure 3). In Figure 3C, we can observe that the mean divergence for intrasampling was almost the same as for intersampling.

On the other hand, steatosis and fibrosis cases failed the similarity test. Inter-divergence was significantly higher than intra-divergence (Table 2, Figures 4 and 5): regions sampled from different cases were more different than regions sampled from the same case on different frames within a video, at different time points.

The ultrasound images examples of the three liver pathologies, in Figures 3–5, demonstrate that although it is possible to visually distinguish between the images of the three groups, the differentiation of cases within a disease group is hard on different frames of the same cases or between cases.

5. Discussion

For successful application of ML methods in medical imaging research and deployment of high-performance generalizable models, there is a need for a sufficient number of training samples from large images databases. We used the common training practice of adding samples by training on a large number of sequential image frames taken from the same case. The results of ML models showed high classification performance for the three studied disease groups. However, when we tested the independence of the sequential images that were taken from the same case using JS divergence, of the three tested liver pathologies, only normal cases demonstrated statistically that intracase region divergence is close to the intersampled case divergence. This means that the normal cases diverged similarly between patients and within a patient, but for fibrosis and steatosis, samples within a patient were more similar to each other than samples from different patients. Two regions sampled from the same normal case were just as different as two regions sampled from completely different cases according to the t-test for equivalence, within 5%. Therefore, for normal cases we can reject the null hypothesis that there is a significant difference between inter- and intrasampling. In general, intra-divergence for all the three groups was small, but different. Fibrosis showed the highest JS divergence in comparison to other groups. This finding is expected as fibrosis is often associated with heterogenous tissue changes between different regions of the liver in comparison to more uniform changes seen in steatosis [25,26]. Yet, inter- and intracase JS fibrosis divergence are not statistically close to each other; therefore, we cannot claim that the intracase image frames are independent enough to train on as separate cases.

The intuition that divergence should increase with time-shift between samples proved to be incorrect, because any gradual divergence trend was overpowered by the cyclical effects of breathing motion. Looking at the graphs of divergence with increasing time change, it is evident that divergence between pairs of sampled regions is periodic, with a period of 35 frames or 4.5 s. This is believed to be due to breathing. A region sampled from a later frame will be most similar (less divergent) to a region sampled earlier at the same relative point in the respiratory cycle and least similar to a region sample out of phase in the cycle. For all three tested pathologies, but especially steatosis and fibrosis, divergence change within the breathing cycle was much stronger than divergence drift over the entire time-course of the study. In theory, if this divergence oscillated perfectly with constant-period and constant-amplitude cycles, to maximize sampling differences, maximally out-of-phase time points could be chosen within or even crossing multiple periods. For instance, if the reference interdivergence was 1, and the respiratory period was 1 s (1 Hz), then the peak-to-valley maximum divergence between frames within a cycle would be at half a cycle, or 0.5 s. If the divergence difference over this half-cycle was at least equal to our reference threshold of 1, then exactly two samples that were this half-period apart, or some integer multiple of this half period, could be taken for the *entire* time of the study to ensure that they were as different from each other as two samples from entirely different cases. All of our data show that, over the time of the study, the amplitude of the divergence oscillation within a cycle increases over time: in a later breathing cycle after an earlier reference breathing cycle, a sampled region will diverge more from the earlier reference sample even when both are still

in phase; and if the two samples are taken out of phase, they will also diverge much more over time. The strongest guarantee of sufficient divergence would be if the lower bound connecting the out-of-phase valleys on the increasing-oscillating curve increased by at least the reference intersampling divergence, but none of these studies showed such an effect, possibly because they were not long enough.

In this paper, we proposed the use of a statistical preliminary analysis to assess the quality of the imaging data before constructing an ML model. To our knowledge, the JS divergence test has not been evaluated to test the independence of image data. However, similar preliminary exploratory data analyses have been reported in literature for assessing the quality of data before going to ML models. One example of the exploratory analysis reported includes the use of Dynamic Time Warping (DTW) [27,28]. DTE is a measure of how similar two temporal sequences are in a time series analysis. DTW looks for the optimal alignment between the two series as opposed to looking at the Euclidean distance between two points at each time series. DTW was evaluated as a distance metric of fMRI time series with repeated measurements of an individual subject and showed that DTW analysis results in more stable connectivity patterns by reducing the within-subject variability and increasing robustness for preprocessing strategies.

One limitation of this study is that we did not test each of our measured radiomics features in this way but presented a method to perform it and tested it on a fundamental image property. Since most features depend somewhat on grayscale distribution, if ROIs between frames were not independent enough in those distributions, we did not look further. Grayscale histograms and first-order statistics are perhaps the simplest ways of characterizing image regions, so if those distributions are significantly more similar within a case than they are between cases, and more similar at certain regular time intervals within the clip, then many derived features may behave similarly. An image, though, is a spatial distribution of grayscale values, and it is very true that engineered features may only partly depend on the grayscale *value* distribution or may not depend on it at all. The methodology we presented, though, with enough samples to compare distributions (i.e., a radiomics image that maps the feature value to each pixel location) could test any quantitative feature to see if it was independent enough between frames to allow those frames to count as sufficiently independent images for that measurement. A more in-depth analysis of sampling divergence behavior over time is, however, beyond the scope of this research and would not be generalizable to other organs or modalities. The important result is that different pathologies might have different dependencies when sampling frames from video, since in this research one of three pathologies passed the independence test: sample regions between video frames of a normal case are as different from each other as samples from completely different cases. Future studies will also evaluate the effects of inter- and intrauser variability in ROI selection on the model performance. The study results are specific to ultrasound imaging, and findings could have been impacted by factors related to the choice of animal, tissue, ROI, and analysis. Future studies on a large scale are required for the proposed approach to be generalized to other imaging modalities and to be applied in human studies.

6. Conclusions

Whenever image frames taken from same case are used for training machine learning models, if that model assumes independence between images, an independence test should be performed. Not only might there be within-case image dependence, but that dependence also could be related to time separation of samples in a video study, so that enforcing a time interval between samples could meet the independence assumptions of the model. For our liver images, however, no such simple time interval could be found, because the periodic breathing motion was too strong an effect. However, for one of the three pathologies—the “normal” cases—image frames were as independent within a case as between cases. Such a result could be important for machine learning. In a clinical setting, it is not uncommon to acquire more disease images than normal controls, possibly creating an unbalanced imaging database. The result of this research on liver diseases suggests that it is acceptable to take many frames of the same normal case as independent training cases, demonstrating one possible application for thoroughly testing image frames for independence.

Funding:

The financial support from NIH grants CA204446-01, and EB022612-01, is gratefully acknowledged.

Data Availability Statement:

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the evolving nature of the project.

References

1. Brattain LJ; Telfer BA; Dhyani M; Grajo JR; Samir AE Machine learning for medical ultrasound: Status, methods, and future opportunities. *Abdom. Radiol* 2018, 43, 786–799.
2. Aung YYM; Wong DCS; Ting DSW The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. *Br. Med. Bull* 2021, 139, Idab016.
3. Willemink MJ; Koszek WA; Hardell C; Wu J; Fleischmann D; Harvey H; Folio LR; Summers RM; Rubin DL; Lungren MP Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020, 295, 4–15. [PubMed: 32068507]
4. Vabalas A; Gowen E; Poliakoff E; Casson AJ Machine learning algorithm validation with a limited sample size. *PLoS ONE* 2019, 14, e0224365. [PubMed: 31697686]
5. Teng D; Li F; Zhang W Using comprehensive machine-learning models to classify complex morphological characters. *Ecol. Evol* 2021, 11, 10421–10431. [PubMed: 34367585]
6. An C; Park YW; Ahn SS; Han K; Kim H; Lee SK Radiomics machine learning study with a small sample size: Single random training-test set split may lead to unreliable results. *PLoS ONE* 2021, 16, e0256152. [PubMed: 34383858]
7. Lehne M; Sass J; Essenwanger A; Schepers J; Thun S Why digital medicine depends on interoperability. *NPJ Digit. Med* 2019, 2, 79. [PubMed: 31453374]
8. Undar M; Krishnapuram B; Bi J; Rao RB Learning Classifiers When the Training Data Is Not IID. *IJCAI Int. Jt* 2007, 2007, 756–761.
9. Darrell T; Kloft M; Pontil M; Rätsch G; Rodner E Machine Learning with Interdependent and Non-identically Distributed Data (Dagstuhl Seminar 15152). *Schloss Dagstuhl-Leibniz-Zent. Fuer Inform* 2015, 5, 4.
10. Tillman R Structure learning with independent non-identically distributed data. *ACM Int. Conf. Proceeding Ser* 2009, 382, 131.

11. Kagadis GC; Drazinos P; Gatos I; Tsantis S; Papadimitroulas P; Spiliopoulos S; Karnabatidis D; Theotokas I; Zoumpoulis P; Hazle JD Deep learning networks on chronic liver disease assessment with fine-tuning of shear wave elastography image sequences. *Phys. Med. Biol* 2020, 65, 215027. [PubMed: 32998480]
12. Carrer L; Donini E; Marinelli D; Zanetti M; Mento F; Torri E; Smargiassi A; Inchingolo R; Soldati G; Demi L; et al. Automatic Pleural Line Extraction and COVID-19 Scoring From Lung Ultrasound Data. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 2020, 67, 2207–2217. [PubMed: 32746195]
13. Faes L; Wagner SK; Fu DJ; Liu X; Korot E; Ledsam JR; Back T; Chopra R; Pontikos N; Kern C; et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: A feasibility study. *Lancet Digit. Health* 2019, 1, e232–e242. [PubMed: 33323271]
14. Nirschl JJ; Janowczyk A; Peyster EG; Frank R; Margulies KB; Feldman MD; Madabhushi A A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue. *PLoS ONE* 2018, 13, e0192726. [PubMed: 29614076]
15. Qu W; Balki I; Mendez M; Valen J; Levman J; Tyrrell PN Assessing and mitigating the effects of class imbalance in machine learning with application to X-ray imaging. *Int. J. Comput. Assist. Radiol. Surg* 2020, 15, 2041–2048. [PubMed: 32965624]
16. Pheng HS; Shamsuddin SM; Leng WY Kullback Leibler divergence for image quantitative evaluation. *AIP Conf. Proc* 2016, 1750, 020003.
17. Nielsen F On the Jensen-Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy* 2019, 21, 485.
18. Okada K; Periaswamy S; Bi J Stratified regularity measures with Jensen-Shannon divergence. In *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
19. Ré MA; Azad RK Generalization of entropy based divergence measures for symbolic sequence analysis. *PLoS ONE* 2014, 9, e93532. [PubMed: 24728338]
20. D'Souza JC; Sultan LR; Hunt SJ; Schultz SM; Brice AK; Wood AKW; Sehgal CM B-mode ultrasound for the assessment of hepatic fibrosis: A quantitative multiparametric analysis for a radiomics approach. *Sci. Rep* 2019, 9, 8708. [PubMed: 31213661]
21. Sultan LR; Chen YT; Cary TW; Ashi K; Sehgal CM Quantitative pleural line characterization outperforms traditional lung texture ultrasound features in detection of COVID-19. *J. Am. Coll. Emerg. Physicians Open* 2021, 2, e12418. [PubMed: 33842925]
22. Tripathi A; Goswami T; Trivedi SK; Sharma RD A multi class random forest (MCRF) model for classification of small plant peptides. *Int. J. Inf. Manag. Data Insights* 2021, 1, 100029.
23. Frank E; Hall MA; Witten IH The WEKA Workbench. In *Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”*, 4th ed.; Morgan Kaufmann: Burlington, MA, USA, 2016.
24. Streiner DL Unicorns Do Exist: A Tutorial on ‘Proving’ the Null Hypothesis. *Can. J. Psychiatry* 2003, 48, 756–761. [PubMed: 14733457]
25. Chauhan A; Sultan LR; Furth EE; Jones LP; Khungar V; Sehgal CM Diagnostic accuracy of hepatorenal index in the detection and grading of hepatic steatosis. *J. Clin. Ultrasound* 2016, 44, 580–586. [PubMed: 27447717]
26. Layer G; Zuna I; Lorenz A; Zerban H; Haberkorn U; Bannasch P; van Kaick G; R ath U Computerized ultrasound B-scan texture analysis of experimental diffuse parenchymal liver disease: Correlation with histopathology and tissue composition. *J. Clin. Ultrasound* 1991, 19, 193–201. [PubMed: 1646222]
27. Meszlenyi RJ; Hermann P; Buza K; Gal V; Vidnyanszky Z Resting state fMRI functional connectivity analysis using dynamic time warping. *Front. Neurosci* 2017, 11, 75. [PubMed: 28261052]
28. Salvador S; Chan P FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intell. Data Anal* 2007, 11, 561–580.

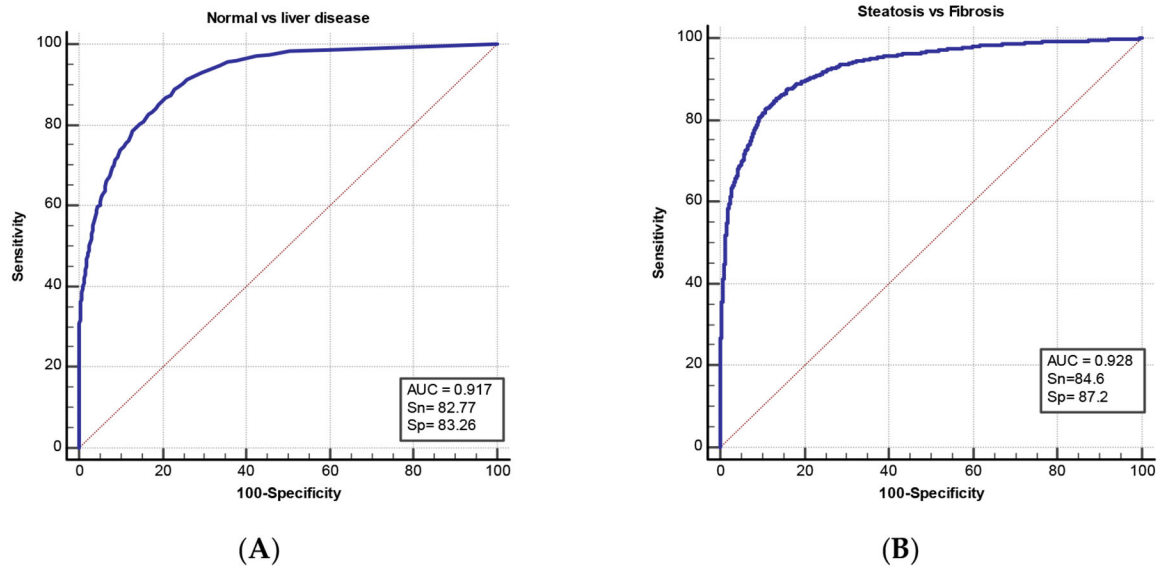


Figure 1. The diagnostic performance of quantitative liver texture ultrasound features using logistic regression (LR) machine learning for two-step pathology differentiation. (A) shows the diagnostic performance of ultrasound texture features in differentiating normal from liver disease including both fibrosis and steatosis cases, while (B) demonstrates the diagnostic performance of ultrasound features in differentiating steatosis from fibrosis cases. AUC refers to area under the curve, Sn: sensitivity, and Sp: specificity.

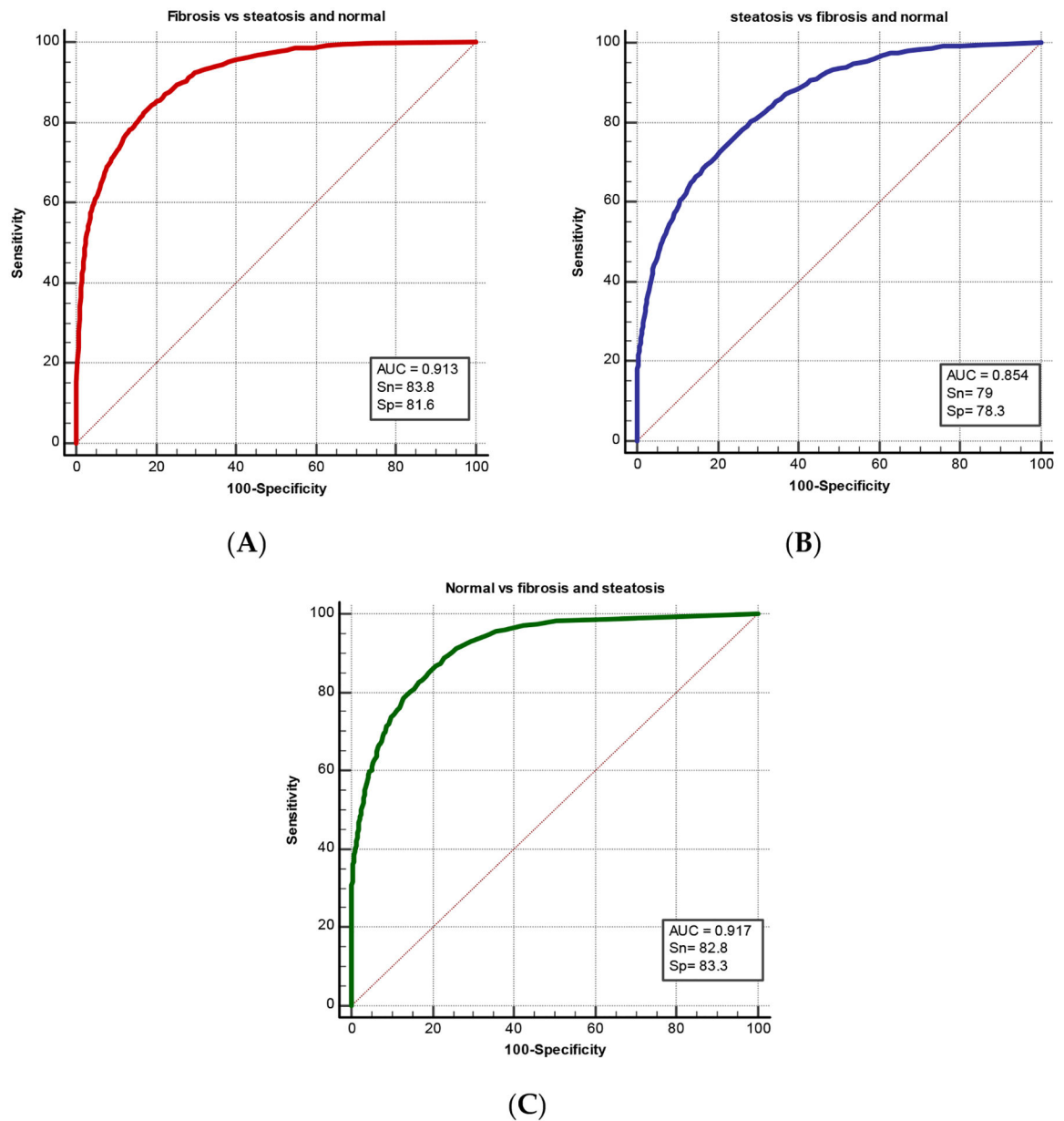
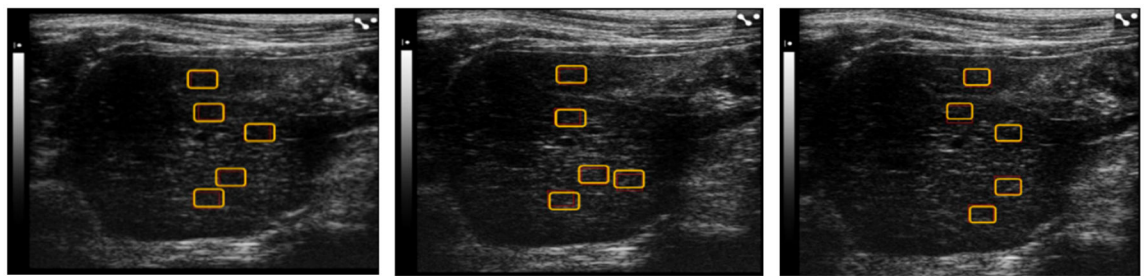
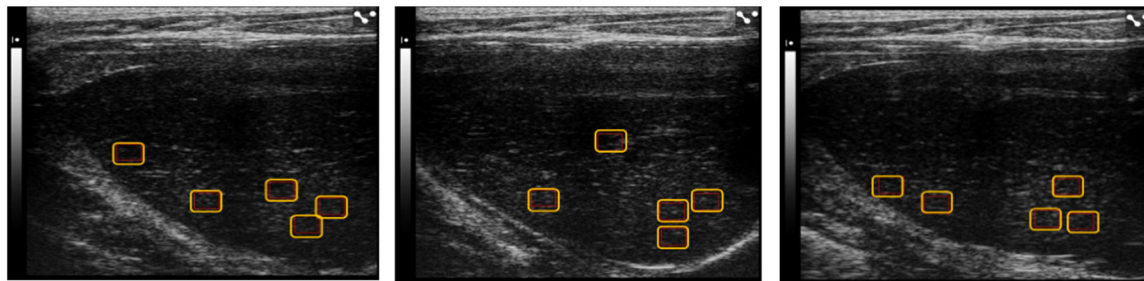


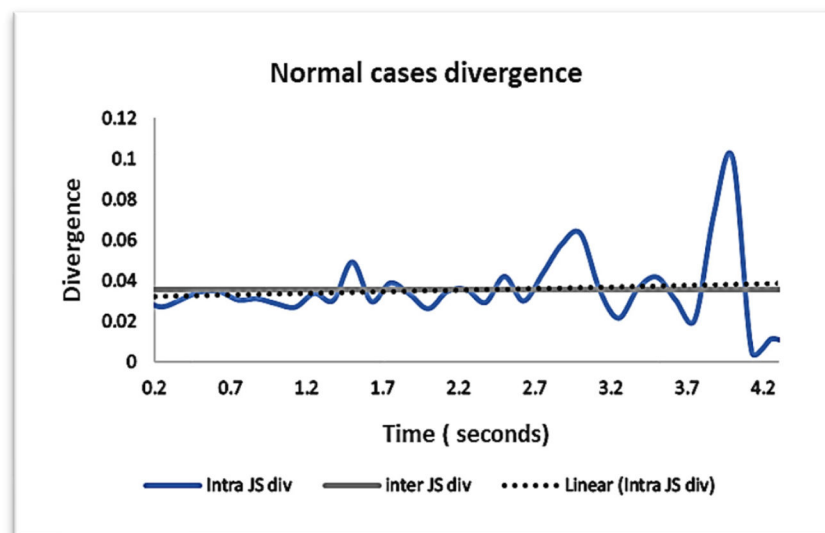
Figure 2. ROC curves of the diagnostic performance of quantitative liver texture ultrasound features using random forest machine learning for multicategory classification. **(A)** shows the diagnostic performance of ultrasound features in differentiating fibrosis from the two remaining groups: steatosis and normal. **(B)** demonstrates diagnostic performance in differentiating steatosis group from fibrosis and normal groups. **(C)** shows the performance of normal cases versus liver disease: fibrosis and steatosis. AUC refers to area under the curve, Sn: sensitivity, and Sp: specificity.



(A)



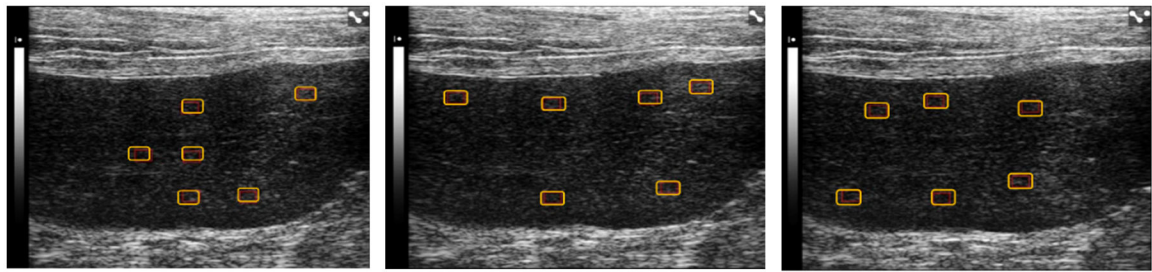
(B)



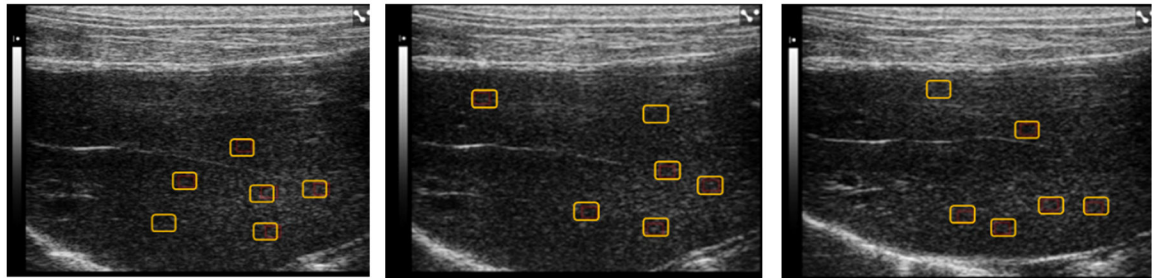
(C)

Figure 3.

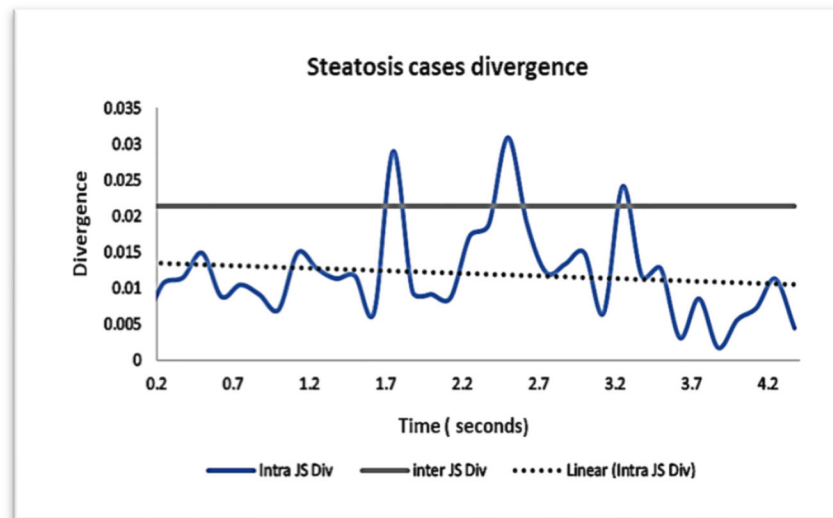
Panels (A,B) show examples of B-mode ultrasound liver images taken from two normal cases. Panel (A) shows three sequential from the same case with region of interests (ROIs) for quantitative analysis. Panel (B) shows three sequential images from a second normal case. Panel (C) shows the intra- and intercase JS divergence for cases in general. Intra- and intercase divergence for normal cases are close to each other, indicating that intrasampled cases may be just as independent inter sampling. Five to six regions of interests are placed (red rectangular boxes) on each image for quantitative analysis.



(A)



(B)



(C)

Figure 4.

Panels (A,B) show examples of B-mode ultrasound liver images from two steatosis cases. Panel (A) shows three sequential images from the same case with region of interests (ROIs) for quantitative analysis. Panel (B) shows three sequential images from a second steatosis case. Panel (C) displays the intra- and intercase JS divergence for steatosis cases in general. Intra- and intercase divergence for steatosis cases is far apart, indicating that we cannot claim independence of intrasampling in these cases. Five to six regions of interests are placed (red rectangular boxes) on each image for quantitative analysis.

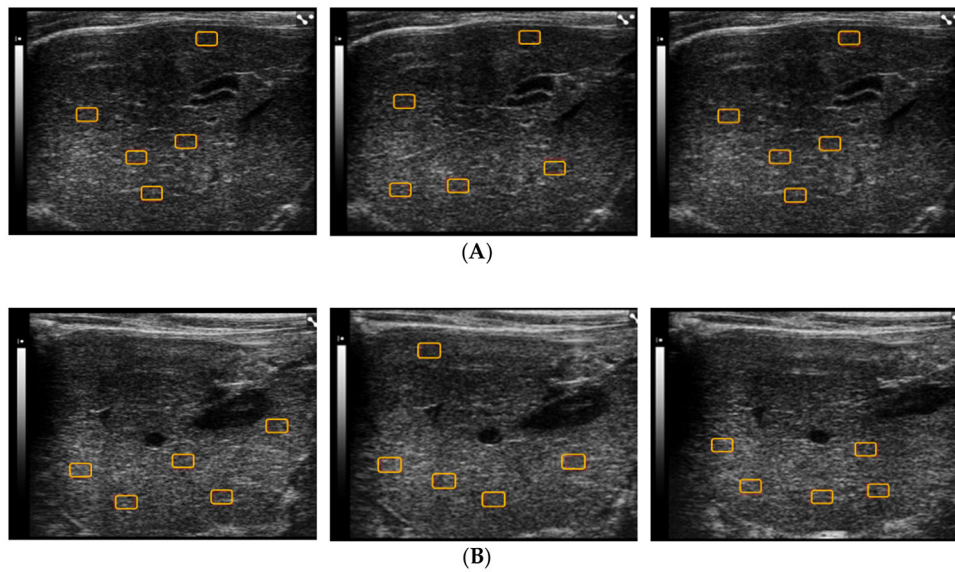


Figure 5. Panels (A,B) show examples of B-mode ultrasound liver images from two fibrosis cases. Panel (A) shows three sequential images from the same case with region of interests (ROIs) for quantitative analysis. Panel (B) shows three sequential images from another steatosis case. Panel (C) shows the intra- and intercase JS divergence for fibrosis cases in general. Intra- and intercase divergence for fibrosis cases is far apart, indicating that we cannot claim independence of intrasampling in these cases. Five to six regions of interests are placed (red rectangular boxes) on each image for quantitative analysis.

Table 1.

Mean \pm standard errors of liver texture ultrasound features studied for the three liver pathologies. Two-sided *t*-test *p*-values are shown for each group against the other two groups. $p < 0.05$ is considered significant.

Quantitative Ultrasound Features	Steatosis	Fibrosis	Normal	<i>p</i> -Value: Fibrosis vs. Steatosis	<i>p</i> -Value: Fibrosis vs. Normal	<i>p</i> -Value: Steatosis vs. Normal
Echo intensity	34.7 \pm 12.0	55.9 \pm 16.3	25.4 \pm 13.6	0.00	0.00	0.00
Heterogeneity	14.6 \pm 3.7	20.5 \pm 3.9	12.7 \pm 4.7	0.00	0.00	0.0
Internal Heterogeneity	12.0 \pm 1.2	16.3 \pm 2.4	13.2 \pm 1.8	0.00	0.00	00
GLNU	0.3 \pm 0.1	0.3 \pm 0.0	0.4 \pm 0.2	0.00	4.37×10^{-52}	0.00
RLNU	0.2 \pm 0.0	0.2 \pm 0.0	0.3 \pm 0.1	0.00	1.7×10^{-68}	2.914×10^{-72}
Entropy	3.5 \pm 0.2	4.5 \pm 0.18	3.1 \pm 0.5	0.00	2.4×10^{-116}	2.041×10^{-123}

Table 2.

Mean and standard deviation of intra- and inter-case JS divergence values for the three pathology groups.

JS Divergence	Normal	Fibrosis	Steatosis
Intra	0.01 ± 0.03	0.05 ± 0.05	0.01 ± 0.05
Inter	0.01 ± 0.03	0.07 ± 0.08	0.03 ± 0.05

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript