Contrastive Learning Embedded Siamese Neural Network for the Assessment of Fatty Liver

Kumar Mohit
Department of Electronics and Communication Engineering
MNNIT Allahabad
Prayagraj, INDIA
kumarmohit@mnnit.ac.in

Ankit Shukla
Department of Electronics and Communication Engineering
MNNIT Allahabad
Prayagraj, INDIA
ankitshukla459@gmail.com

Rajeev Gupta
Department of Electronics and Communication Engineering
MNNIT Allahabad
Prayagraj, INDIA
rajeevg@mnnit.ac.in

Pramod Kumar Singh
Department of Radio-diagnosis Imaging
IMS-BHU
Varanasi, INDIA
pramodk.singh1@bhu.ac.in

Kushagra Agarwal
MD, Radiology
Kriti Scanning Centre
Prayagraj, INDIA
kritiscan@gmail.com

Basant Kumar
Department of Electronics and Communication Engineering
MNNIT Allahabad
Prayagraj, INDIA
singhbasant@mnnit.ac.in

***Abstract*— This paper presents an self-supervised Siamese neural network (SNN) for identification and classification of fatty liver severity. SNN is used for self-supervision tasks for being influenced from model optimization property of supervised and manual annotation property of unsupervised learning. This technique is based on contrastive learning of the joint embedding network which can learn more subtle representations from the medical images for classification task, with just one or few number of labelled images required from each class for training. The efficiency of the proposed model is validated on our dataset of liver ultrasound to classify them into three stages of the fatty liver disease and normal liver. A two-class classifier (normal/grade-I, normal/grade-II and normal/grade-III fatty liver) and four-class classifier (normal, grade-I, grade-II, grade-III fatty liver disease) were trained by minimizing contrastive loss to obtain classification accuracy of 98.91% and 96.84% respectively.**

Keywords*— Fatty liver, Contrastive learning, Deep learning, Siamese neural network, ResNet, Ultrasound images.

I. INTRODUCTION

Liver diseases are life threatening problems when not cured in time. The root of the prevention lies in the early identification of fatty liver for which computer-aided diagnosis (CAD) through deep learning techniques proved to be a success. This is the process of gradual accumulation of fat on liver, which if not cured in time may leads to chronic liver conditions of scarring the liver and liver damage, along with affecting cardiovascular conditions [1]. Ultrasonography (USG) is most preferred imaging technique for scanning soft tissues like liver, kidney, etc. Also, USG is cheapest, non-radiative and harmless imaging technique [2-3]. The principle behind liver identification is subjective and objective analysis. Subjective analysis is based on vascular visibility and echogenicity comparison with respect to right kidney. Healthy liver and kidney textures are isoechoic, while fatty liver tissues are hyperechoic causing blurring of intrahepatic veins and diaphragm [4]. Earlier objective analysis were based on CAD assessment for statistical calculations of hepatorenal sonographic index (HRI) [5-7], texture and morphological based features, etc. from manually selected ROI from the liver ultrasound images. Various feature extraction and classification techniques for characterization of liver tissues using machine learning is summarized in review paper of Bharti et. al. [8]. But observer dependent objective assessments are limited due to differ in the image acquisition settings and inter-observer variability. Also, manually selection of feature extraction methods were sort of setting gold standard for machine learning tasks and therefore classification accuracies were subjected to change with handling radiologists. Deep learning with convolutional neural network (DL-CNN) emerges as a tool for fully automatic approach towards improving diagnostic performances. Deep learning makes use of convolution filters to convolve with the input image matrix for automatic extraction of significant features and classification into different classes. The weights of pre-trained CNN can also be transferred to new model in deep learning algorithms just to train for new problems [9]. Byra et al. [10] makes use of transferring the learned model and introduced "transfer learning" concept into liver characterization study which proved to be a benchmark for many researchers. They modify their model with Inception-ResNet-v2 architecture and extracted hepatic features from liver ultrasound. Same work was extended by Reddy and his team in [11], where they incorporated pre-trained VGG-16 model for classifying between normal and fatty liver. Biswas et al. [12] worked on developing GPU based CNN framework for feature extraction, where they applied inception algorithm for feature selection and SoftMax classifier. Li et al. [13] applied scalable ResNet-18 neural network for classification of pre-processed B-mode liver ultrasound images into mild, moderate and severe steatosis.

Although CNN provides early diagnosis solutions in liver and other medical problems, successful training for higher accuracy, sensitivity and specificity always requires very large amount of labelled (supervised) data; which is not always accessible. This generally limits the practical applications of the deep learning model designed for medical applications [14]. There always been a necessity of generalizing the unfamiliar categories without requiring to retrain always. Fei-Fei et al. [15] proposed 'one-shot' model of object categories which requires just one or few images from each category for learning features. They developed variational Bayesian probabilistic framework for one-shot image classification. Koch et al. [16] later developed CNN based self-supervised SNN for one-shot classification of images and set a benchmark for further studies. SNN was first proposed by Bromley et al. [17] for signature

verification based on similarity distance measure from twin network in the same feature space. These studies motivated medical scientists to develop SNN to overcome data limitations of supervised CNN.

Shorfuzzaman et al. [18] worked on the diagnosing of Covid-19 patients using n-shot SNN from chest X-ray images. They applied transfer learning by using pre-trained VGG-16 architecture for minimizing the contrastive loss. Mohit et al. [20] extended their work of [19] for pneumonia detection using transfer learned VGG-16 model to multiclass classification of Covid-19 cases using contrastive loss embedded SNN model for reducing number of false positives. In this proposed framework for fatty liver assessment, the highlights of our contributions are listed below.

(i) A new texture echogenicity based structural similarity model has been proposed in Fig. 1. which makes use of deep neural network built in Siamese architecture for binary-class and multi-class classification of liver ultrasound images.

(ii) Binary and multi-class classification are performed using same SNN.

(iii) Optimization has been performed which is based on minimizing the contrastive loss by training the model with contrastive learning.

(iv) Pre-trained ResNet-50 network is introduced as analogous networks for embedding and extracting the features, which helps in minimizing the contrastive loss.

The organization of rest of the paper follows proposed model in section 2, experimental setup in section 3, results in section 4 and conclusion in section 5.

II. Proposed Model

We have used SNN for fatty liver classification into normal, mild as Grade-I, moderate as Grade-II and severe as Grade-III fatty liver. Classification has been done through contrastive learning which is a self-supervised strategy. It acts as a bridge between supervised and unsupervised learning, requiring only one or few labelled data (n-shot learning) and ground truth for training data from each class [20]. Ground truth or labelled data assign the labels for rest of the images by generating pseudo-labels based on the similarities. For similarity calculations; two identical encoder networks (twin-network), shown in Fig. 1. are employed in SNN which work analogous with two different input images or data points. We have employed pre-trained modified ResNet-50 architecture as twin network for extracting low-dimensional representations in feature space and provide embeddings for both images. The residual network (ResNet) architecture uses skip connections to avoid network load and gradient vanishing while going deeper inside the network. Skipping the identity mappings boosts the training too [21]. Both networks in twin layout share common wights. These 'Siamese twins' use shared weights to discover similarities and differences among liver images. The residual part of the network helps in exploring enough features necessary for accurate training of our proposed model. For that, the representations from ResNet-50 passed to two-layer multilayer perceptron to get necessary representations which are processed in Siamese layer for distance measurement.

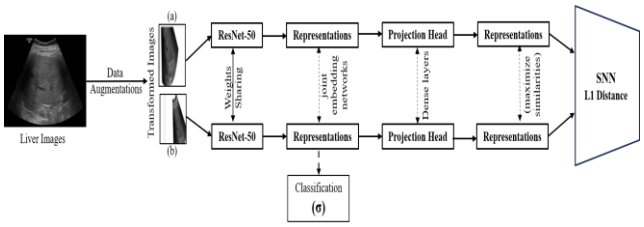

Fig. 1. Proposed SNN architecture for fatty liver characterization. The task is to maximize the similarity between representations from the same image.

The training of the model is done to make network learn preciously to predict the dissimilarities between images of different classes, and similarities between images of same class. Shared weights are updated continuously through back-propagation during training of the contrastive loss function, so that feature space embeddings provided by the representations must have the minimum distance between similar pairs. The optimized weights helps in identifying the unknown images during downstream tasks, classification here, and predict the labels. Binary output is yielded to specify which instance pairs are similar (True or 1) and which are different (False or 0). In this way, model learn to make predictions by calculating the L1 distance between the corresponding output feature vectors; shown in Fig. 2. This figure describes parametric values of the representations at both end of twin network which are passed to flatten layer to get respective embeddings. Further L1 distance is calculated between both values to assign similarity score by sigmoid function in dense layer. Based on the probability of the similarity score, labels are assigned to unlabelled images in downstream tasks. Contrastive learning helps in enhancing the embedding quality of the feature extractor by contrasting sample pairs such that features belonging to the same instances are grouped together, and distance between instances of different class are increased.

III. Experimental Setup

A. Data Collection

Transabdominal USG is preferred imaging for visualizing hepatic echogenic pattern. We have collected a total of 749 B-mode liver ultrasound images (376 normal, 214 grade-I, 119 grade-II and 40 grade-III annotated by medical experts) recorded by scanning 112 patients of different age groups from Institute of Medical Sciences-Banaras Hindu University, Varanasi and Kriti scanning centre, Prayagraj, INDIA in portable network graphics (PNG).

B. Image Pre-processing

- *image annotation*: We have performed online cropping, shown in Fig. 3., from the video sequences to get rid of undesirable information from the ultrasound images and increase the length of dataset.
- *resizing*: Images had been resized to 224X224 array so that larger size can be handled easily and training happens smoothly. We have prepared the model with 'python 3.0' on 'google colab', so smaller size images are trained easily and memory is optimized.
- *rescaling*: Rescaling helps in making model computationally efficient by lowering the complexity and load of computations. In our work we have rescaled the distance by dividing each image pixels by 255 so that distance between data points is minimized and generalization error is avoided.

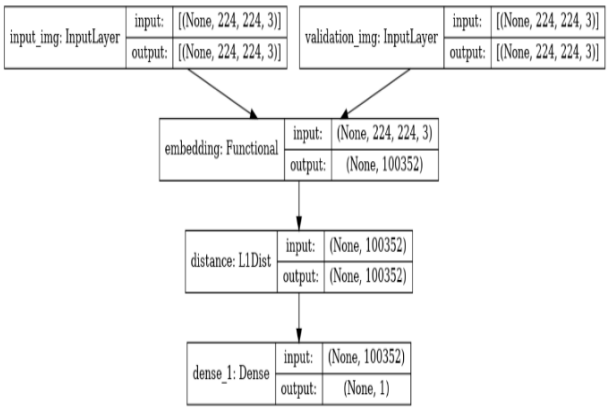

Fig. 2. SNN architecture for distance calculation.

- *data augmentation*: We have performed data augmentations for creating pseudo labels which supervised the model training for making predictions. We have performed rotation, shear, zoom, flip and shift as shown in Fig. 4. for creating positives from anchor images.

C. Training

Our model is prepared for multi-class classification from three binary classification models. The dataset is divided into anchor (reference images), positives (similar augmented images) and negatives (dissimilar images) dataset for training. Anchor dataset are made to compare with positives and negatives in dense layer to provide embeddings. Distance between anchor and negatives is maximized which roots for similarity check between same instances. Extracted features embeddings are given as image pairs of same classes are labelled 1 (True) i.e. to belong to positives, otherwise negatives. The labelling is used in flattened layer for calculation of distance between two feature vectors. The comparison is between normal-grade1, normal-grade2 and normal-grade3 set for classifying liver images into normal and grade-I, grade-II, grade-III fatty liver as shown in Fig. 5.

D. Hyperparameters

Programming environment were tuned with the values provided in table 1 for learning the desired algorithm.

IV. Results

We have tested the Siamese network for calculations of accuracy and contrastive loss values with 10-shot learning. 'Contrastive loss' compares the extracted features from Siamese twins of two inputs for mapping vectors of similar inputs. These mapping helps network in contrastive learning.

TABLE I. HYPERPARAMETER TABLE

optimizer	Adam
learning rate	0.0001
activation function	ReLU, Sigmoid (last layer)
loss function	Contrastive loss
batch size	12
iterations	20

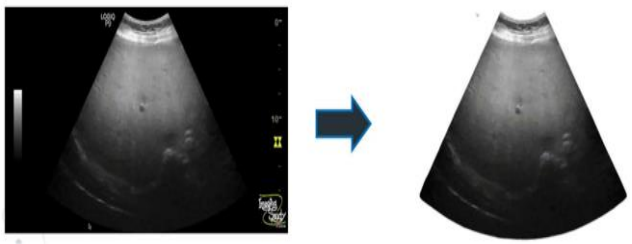

Fig. 3. Pre-processing for data annotations.

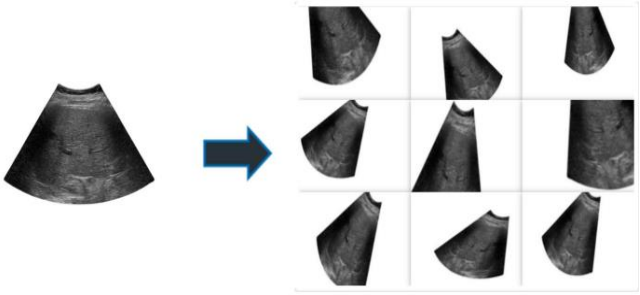

Fig. 4. Data augmented images used for training SNN.

If we consider 'y' to be ground truth relationship with image pairs, then contrastive loss we have calculated for a single pair using Eq. (1).

$$[(y-x) + (1-y) * max(margin-x, 0)] \quad (1)$$

where, square root of Manhattan distance between two image features is measured as 'x' and 'margin' denotes lower bound distance between dissimilar instances. So, 'y' here will be '1' for similar instances, and '0' for dissimilar. Fig. 6. shows the validation accuracy and loss graphs for binary and multi-class classifications. Fig. 6. (a) & (b) are accuracy and loss graphs respectively for multi-class classification, whereas Fig. 6. (c) & (d) for binary-class classification. The training graph reflects the prediction power of our model with minimum number of iterations. Loss function converges quickly through back-propagation and learn the necessary representations for classification. Accuracy is calculated by taking ratio of number of correct predictions to the total number of predictions. As we can see accuracy graph obtaining it's highest steady state within few epochs, which helped in predicting labels of the unlabeled liver images correctly in fig. 7 and 8.

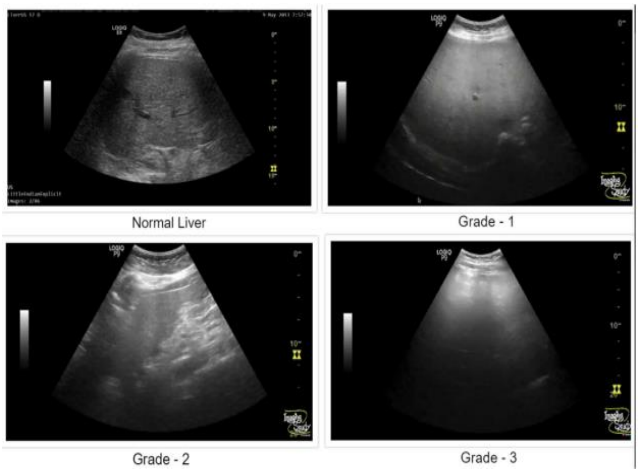

Fig. 5. Different classes of fatty liver based on echo discrepancy on B-mode liver USG images.

TABLE II. PERFORMANCE COMPARISON TABLE

Authors	Diagnostic Outcome	Embedding Model	Imaging Modality	Accuracy/AUC
Zeng et al. [25]	Diabetic Retinopathy	Inception-V3	Fundus images	AUC = 0.95 (Binary-class), Kappa = 0.83 (Multi-class)
Hou et al. [23]	Covid-19	ResNet-50	CT images	98.18% (Binary-class), 96.73% (Multi-class)
Li et al. [24]	Covid-19	ResNet-50	CT images	AUC = 0.96 (Multi-class)
Chikontwe et al. [21]	Covid-19	CNN (DA-CMIL)	CT images	98.60% (Binary-class)
Lu et al. [22]	Breast Cancer	ResNet-50	Histology images	AUC= 0.968 (Binary-class)
Mohit et al. [20]	Covid-19	Inception	CT images	AUC = 0.97 (Binary-class)
Proposed model	Fatty liver	ResNet-50	USG images	**98.91%** (Binary-class), **96.84%** (Multi-class)

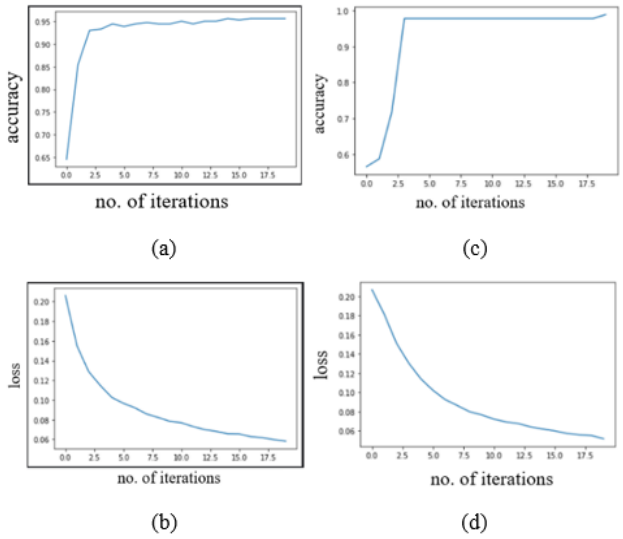

Fig. 6. Resultant graph of SNN architecture for accuracy in (a) & (c) and loss in (b) & (d) of multi-class and binary-class classification of fatty liver diseases respectively. Plot is made against 20 epochs.

We have resultant anchor and predicted images of normal liver, and positive, negatives for binary and multi-class classification in Fig. 7. & Fig. 8. respectively. This is a prototype of how unlabeled images are being compared with images of a class made anchor images, and getting labelled. Here normal cases are shown as 'anchor', similar cases as 'positives' and rest cases as 'negatives' for binary classification. Similarly, multi-class classification gets labeled by considering one by one each class in abnormal cases as 'anchor' and rest as 'positives-negatives'. Few works had been done in medical diagnostic field till date using self-supervised learning which we are comparing with our work in table 2. based on accuracy and loss calculations. With good amount of non-invasive dataset, our model is showing promising results, compared to others, with same encoder network as others used. However, fatty liver identification using SNN is less explored research area and we have focussed on accurately classifying the liver images, inspired from other studies but with better results. We succeeded in achieving an overall multi-class testing accuracy of 96.84% and binary-class accuracy of 98.91%. Loss parameters for multi-class was 0.0581, and 0.0517 for binary-class. We had also experimented our model for new CNN in twin architecture in place of pre-trained ResNet-50, but achieved only 96.82% and 84.54% testing accuracy for binary and multi-class respectively.

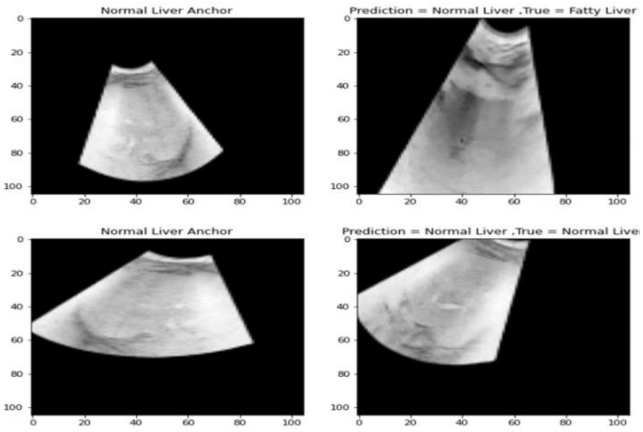

Fig. 7. Predictions made on test data for binary-class classification.

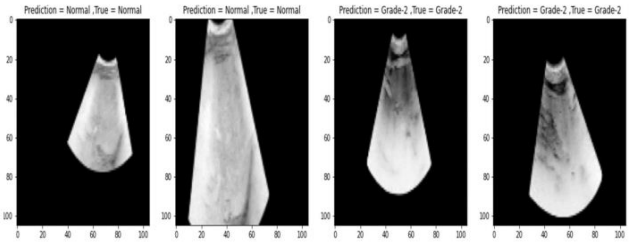

Fig. 8. Predictions made on test data for multi-class classification.

V. CONCLUSION

We proposed a novel method for fatty liver characterization from USG images using ResNet-50 embedded self-supervised SNN, so automatic liver disease detections algorithms need not depends on large amount of annotated data. This technique was based on contrastive learning of the joint embedding network which could learn more subtle representations from the medical images for classification task, with just one or few number of labelled images required from each class for training. We validated the efficiency of the proposed model on our dataset of liver ultrasound to classify them into different classes; normal, grade-I, grade-II and grade-III fatty liver disease. Proposed model trained with contrastive loss showed a reliable classification accuracy for automatic early and accurate

detections to be used as a second opinion by the radiologists; however, we will optimize the model with triplet loss as a future work for getting more accuracy, sensitivity and specificity.

ACKNOWLEDGMENT

The authors would like to thank Dr. Gulab Soni, MD, DM, Panna Neuro Care, Prayagraj, India for his support in completion of the study by providing necessary inputs wherever required.

REFERENCES

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

[2] K. Mohit, R. Gupta and B. Kumar, "A survey study of diseases diagnosed through imaging methodology using ultrasonography" in *Advances in VLSI, Communication, and Signal Processing. Lecture Notes in Electrical Engineering*, vol. 683, D. Harvey, H. Kar, S. Verma, V. Bhadauria, Eds. Singapore: Springer, 2021, 689-703. doi:10.1007/978-981-15-6840-4_57.

[3] Mohit Kumar, Gupta Rajeev and Kumar Basant, "A Survey on the Machine Learning Techniques for Automated Diagnosis from Ultrasound Images," Current Medical Imaging 2024; 20(). doi:10.2174/1573405620666230529112655.

[4] Mohit Kumar, Gupta Rajeev and Kumar Basant, "Computer-aided Diagnosis of Various Diseases Using Ultrasonography Images," Current Medical Imaging 2024; 20(). doi:10.2174/1573405619666230306101012.

[5] M. Webb et al., "Diagnostic value of a computerized hepatorenal index for sonographic quantification of liver steatosis," *AJR Am. J. Roentgenol.*, vol. 192, no. 4, pp. 909-914, 2009. doi:10.2214/AJR.07.4016.

[6] M. F. Xia et al., "Standardized ultrasound hepatic/renal ratio and hepatic attenuation rate to quantify liver fat content: An improvement method," *Obesity (Silver Spring)*, vol. 20, no. 2, pp. 444-452, 2012. doi:10.1038/oby.2011.302.

[7] A. Chauhan et al., "Diagnostic accuracy of hepatorenal index in the detection and grading of hepatic steatosis," *J. Clin. Ultrasound*, vol. 44, no. 9, pp. 580-586, 2016. doi:10.1002/jcu.22382.

[8] P. Bharti, D. Mittal and R. Ananthasivan, "Computer-aided characterization and diagnosis of diffuse liver diseases based on ultrasound imaging: A Review," *Ultrason. Imaging*, vol. 39, no. 1, pp. 33-61, 2017. doi:10.1177/0161734616639875.

[9] M. Singh, B. Kumar and D. Agrawal, "Good view frames from ultrasonography (USG) video containing ONS diameter using state-of-the-art deep learning architectures," *Med. Biol. Eng. Comput.*, vol. 60, no. 12, pp. 3397-3417, 2022. doi:10.1007/s11517-022-02680-3.

[10] M. Byra et al., "Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 12, pp. 1895-1903, 2018. doi:10.1007/s11548-018-1843-2.

[11] D. S. Reddy, R. Bharath and P. Rajalakshmi, "A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging" in *Proc. 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom)*, Ostrava, Czech Republic, Sept. 17-20 2018, pp. 1-5. doi:10.1109/HealthCom.2018.8531118.

[12] M. Biswas et al., "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm," *Comput. Methods Programs Biomed.*, vol. 155, pp. 165-177, 2018. doi:10.1016/j.cmpb.2017.12.016.

[13] B. Li et al., "Accurate and generalizable quantitative scoring of liver steatosis from ultrasound images *via* scalable deep learning," *World J. Gastroenterol.*, vol. 28, no. 22, pp. 2494-2508, 2022 Jun. 14. doi:10.3748/wjg.v28.i22.2494.

[14] H. C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016. doi:10.1109/TMI.2016.2528162.

[15] L. Fei-Fei, R. Fergus and P. Perona, "One-shot learning of object categories" in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594-611, Apr. 2006. doi:10.1109/TPAMI.2006.79.

[16] G. Koch, R. Zemel and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition" in *Proc. ICML 2015 – The 32nd international conference on machine learning*, deep learning workshop, vol. 2, 2015.

[17] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger and R. Shah, "Signature verification using a 'Siamese' time delay neural network," *Adv. Neural Inf. Process. Syst.*, vol. 6, pp. 737-744, 1994.

[18] M. Shorfuzzaman and M. S. Hossain, "MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients," *Pattern Recognit.*, vol. 113, p. 107700, 2021. doi:10.1016/j.patcog.2020.107700.

[19] A. Awasthi, K. Mohit, R. Gupta, and B. Kumar, "Data collection Website with brain tumor and pneumonia detection" in *Recent Trends in Electronics and Communication. Lecture Notes in Electrical Engineering*, vol. 777, A. Dhawan, V. S. Tripathi, K. V. Arya and K. Naik, Eds. Singapore: Springer, 2022, 1149-1156. doi:10.1007/978-981-16-2761-3_99.

[20] K. Mohit, R. Gupta and B. Kumar, "Self-supervised contrastive learning for Covid-19 classification from computed tomography images," 9th Uttar Pradesh Sect. International Conference *on* Electrical*, Electronics and* Computer Engineering *(UPCON)*, Prayagraj, India, vol. 2022. IEEE, *2022*, pp. 1-5. doi:10.1109/UPCON56432.2022.9986406.

[21] P. Chikontwe et al., "Dual attention multiple instance learning with unsupervised complementary loss for COVID-19 screening," *Med. Image Anal.*, vol. 72, p. 102105, 2021. doi:10.1016/j.media.2021.102105.

[22] L. Ming, R. Chen, J. Wang, D. Dillon and F. Mahmood, "Semi-supervised Histology Classification Using Deep Multiple Instance Learning and Contrastive Predictive Coding", 2019. doi: 10.48550/arXiv.1910.10825.

[23] J. Hou et al., "Periphery-aware COVID-19 diagnosis with contrastive representation enhancement," *Pattern Recognit.*, vol. 118, p. 108005, 2021. doi:10.1016/j.patcog.2021.108005.

[24] L. Li et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65-E71, 2020. doi:10.1148/radiol.2020200905.

[25] X. Zeng, H. Chen, Y. Luo and W. Ye, "Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744-30753, 2019. doi:10.1109/ACCESS.2019.2903171.