

# 基於 NASRL 構建最佳化超參數之早期脂肪肝檢測模型

## Constructing an Optimized Hyperparameter Model for Early Fatty Liver Detection Based on NASRL

林泓錕 Hong-kun Lin<sup>1,2</sup> 王浩任 Hao-Jen Wang<sup>3</sup> 呂紹浦 Shao-Pu Lu<sup>1</sup> 李佳燕 Chia-Yen-Lee<sup>1\*</sup>

陳宗伯 Zong-Bo Chen<sup>4</sup> 陳錦得 Jin-De Chen<sup>5</sup> 楊昆澈 Kuen-Cheh Yang<sup>5</sup>

<sup>1</sup> 國立聯合大學, National United University

<sup>2</sup> 國立陽明交通大學, National Yang Ming Chiao Tung University

<sup>3</sup> 國立台灣大學, National Taiwan University

<sup>4</sup> 中國醫藥大學附設醫院, China Medical University Hospital

<sup>5</sup> 國立臺灣大學醫學院附設醫院北護分院, National Taiwan University Hospital Bei-Hu Branch

\* Corresponding Email: [leecyya@gmail.com](mailto:leecyya@gmail.com)

### 一、中文摘要

客觀且有效的早期脂肪肝電腦輔助檢測工具可以顯著縮短醫師的判讀時間，減少醫師間的判讀誤差，這對於早期脂肪肝疾病檢測至關重要。然而，開發這樣的工具面臨著一些挑戰。若要收集臨床上的黃金標準需要進行肝穿刺檢測，但該檢測具有侵入性而不受青睞，而非侵入式方法若要量測準確則需要使用非常昂貴的機器進行檢測、若使用超音波機器做為檢測工具則容易因肋骨、腸氣、超音波偽影、假影等影響導致畫面不清楚，易造成醫師誤判而給予錯誤的標籤，這些因素皆會造成資料收集困難，若當數據收集完成仍須面對資料不平衡或是缺乏多樣化的樣本問題，且不同的醫療機構使用不同的檢測設備和設置，圖像品質和格式可能有所不同。

在過往的相關研究，開發超音波脂肪肝影像分類模型時，使用同一受測者短時間多幀的肝腎影像作為訓練資料，與臨床上醫師須使用超音波檢測脂肪肝時需多個角度觀察多個肝臟特徵的方式相差甚遠，又或是使用只使用 S0 與 S3 的脂肪變性等級資料，而在模型選用方面使用在自然影像領域表現優秀的 SOTA 模型，然而這些模型是為了解決自然影像分類問題而開發，鮮少針對醫學影像問題進行模型架構設計。

為此本研究提出基於具備最適模型參數搜索能力之 Neural architecture search with reinforcement learning(NASRL)方法構建一最佳化的早期脂肪肝檢測深度學習架構，此架構可藉由輸入超音波 B-mode 肝實質影像即可獲得有效的脂肪肝判別結果。其效能經 10-Fold 交叉驗證後可分別達 Accuracy: 0.76±0.09、Sensitivity: 0.77±0.08、Specificity: 0.77±0.07，與過往相關研究常使用之 VGG19 模型效能分別僅為 Accuracy:0.69±0.13、Sensitivity:0.78±0.19、Specificity:0.72±0.05 相比有較好的效能表現。此外，本研究所提出模型所使用之參數量相比於 VGG19 模型減少了約 78%，具備模型訓練與推論效率上之優勢。

**關鍵詞：**非酒精性脂肪肝、FibroScan、深度學習、Neural Architecture Search、Reinforcement learning

### Abstract

An objective and effective computer-aided tool for early detection of fatty liver disease can significantly reduce the time required for physicians to interpret results and decrease inter-physician variability. This is crucial for the early detection of fatty liver disease. However,

developing such a tool faces several challenges. To collect clinical gold standard data, a liver biopsy is required, but this invasive procedure is not favored. On the other hand, non-invasive methods, if accuracy is required, non-invasive methods often necessitate the use of very expensive equipment. When using ultrasound machines as detection tools, the images can be unclear due to factors such as rib interference, bowel gas, ultrasound artifacts, and reverberation artifacts, which can lead to misdiagnosis and incorrect labeling by physicians. These factors all lead to difficulties in data collection. Even after data collection, issues such as data imbalance or a lack of diverse samples may arise. Additionally, different medical institutions use different equipment and settings, which can result in varying image quality and formats.

In previous related research, when developing ultrasound fatty liver imaging classification models, short-term multi-frame liver and kidney images of the same subject were used as training data. This approach differs from the clinical practice where doctors use ultrasound to detect fatty liver by observing multiple liver features from various angles. Some research uses only the steatosis grade data of S0 and S3. In terms of model selection, state-of-the-art (SOTA) models that perform well in the field of natural images are often used. However, these models are primarily designed for natural image classification and are rarely tailored to address medical imaging problems.

To address this, this study proposes a method based on Neural Architecture Search with Reinforcement Learning (NASRL) to construct an optimized deep learning architecture for early fatty liver detection. By inputting B-mode ultrasound images of the liver parenchyma, this architecture can effectively determine the presence of fatty liver. The performance, evaluated using 10-fold cross-validation, achieved an accuracy of 0.76±0.09, sensitivity of 0.77±0.08, and specificity of 0.77±0.07. In comparison, the commonly used VGG19 model in previous studies showed lower performance, with an accuracy of 0.69±0.13, sensitivity of 0.78±0.19, and specificity of 0.72±0.05. Additionally, the proposed model uses approximately 78% fewer parameters than the VGG19 model, offering advantages in training and inference efficiency.

**Keywords:** NAFLD、FibroScan、Deep learning、Neural Architecture Search、Reinforcement learning

## 二、緣由與目的

非酒精性脂肪肝 (Non-Alcoholic Fatty Liver Disease, NAFLD) 在台灣盛行率約 11.4%~41%[1]，而在 2023 年全球盛行率達到 32.4%[2]。早期診斷並治療將助於後續的康復流程使脂肪肝逆轉回正常肝臟，反之可能會發展為不可逆的肝硬化甚至肝癌影響生命安全。

臨床上有許多種對於脂肪肝的檢測方法，侵入式檢測為肝穿刺(Liver Biopsy)、血液檢查，而非侵入式檢測包含超音波(Ultrasound)、MRI 質子密度脂肪分數 (MRI-PDFF)、肝纖維化掃描(FibroScan)等技術。其中 FibroScan 是一個無創、即時、準確的儀器[3][4]，可用於診斷脂肪肝和評估肝纖維化程度，其診斷結果可以立即提供給醫生和患者，快速評估肝臟健康，然而 FibroScan 機台需要較昂貴的價格以及維護成本較難普及在每家醫院。多數人常使用臨床上已被廣泛應用的超音波檢測，醫師可以很直觀針對脂肪肝特徵去做診斷，然而其 B-Mode 影像複雜且易受肋骨、腸氣影響導致畫面不清楚，易造成醫師誤判，且醫師需要觀察多張影像，耗費人力與時間。

過往對於使用超音波影像進行脂肪肝檢測研究中，Michal 等人[5]使用經過 ImageNet 預訓練的模型 Inception-ResNet-v2 進行特徵提取，接著使用支援向量機(support vector machine, SVM)進行含有脂肪肝的圖像分類，最終效能表現(AUROC、Sensitivity、Specificity、Accuracy 分別為  $0.977 \pm 0.021$ 、1、0.882、0.963)，但該資料集為 55 個受試者，同一受測者短時間多幀的肝腎影像，與臨床上醫師須使用超音波檢測脂肪肝時需多個角度看多個肝臟特徵的方式相差甚遠，因此本研究進行收案時不僅採用多角度、多超音波機器，開發真正符合臨床應用模型。

Miriam 等人[6]的研究中，使用 U-Net 模型分割出超音波影像之肝實質的感興趣區域 (Region of interest, ROI)，並將 ROI 區域分為數個不重疊的區塊作為模型輸入，脂肪肝分級方面採用 FibroScan 所量測受控衰減參數作為肝臟脂肪變性分級之依據，並使用 DenseNet-121 模型進行脂肪肝檢測，取得了不錯的效能表現(AUROC、Sensitivity、Specificity 分別為 0.9、0.95、0.85)，然而，該研究僅定義 S0(CAP<238)作為正常，S3(CAP>290)作為有脂肪肝，並排除了 CAP 範圍介於 238 至 290 之間的 S1 和 S2 類別資料，這些類別在臨床上辨別難度較高對模型也較難分類，但是在臨床上同樣與 S0 與 S3 重要，因此即便獲得較好的檢測效能依然不符臨床所需。因此，為了使脂肪肝檢測模型更加符合實際臨床需求，本研究採用 S0~S3 所有脂肪肝嚴重程度級別資料並使用在醫學領域上常用的 VGG19 深度學習模型作為基線進行開發，收集從多個超音波影像機台掃描之脂肪肝變性影像，並使用可靠且不受醫師主觀影響的 FibroScan 作為脂肪變性分級標準，以檢測出高於 S0 之肝臟病變，盡可能幫助醫生找出有脂肪肝的患者之外，降低醫師自己觀察超音波影像時的主觀判斷誤差，以及不同醫師之間的觀察誤差及加速判讀速度，開發出客觀有效可以泛化應用於多個超音波機台早期脂肪肝檢測深度學習模型。本研究為國科會計畫，編號為【NSTC111-2628-E239-001-MY3】。

## 三、材料與方法

### 3.1.研究材料

本研究使用的超音波影像來自中國醫藥大學附設醫院和台大醫院北護分院，採用 LOGIQ S8、Toshiba Aplio300 和 Toshiba SSA-660A 設備拍攝。共 242 名受試者，除了拍攝超音波影像外，本研究使用 FibroScan 530 Compact 測量受控衰減參數(Controlled Attenuation Parameter, CAP)。實驗軟體方面，本研究在 Linux 系統下運行，使用 Python 3.8 和 Tensorflow 2.11.0 版本，並配備 Nvidia RTX 2080Ti GPU 進行 AI 運算。

本研究使用 310 筆具有肝實質影像的訓練資料如 Figure1 所示，並透過 FibroScan 測量的 CAP 值進行脂肪肝分類。本研究使用中國醫藥大學附設醫院和台大醫院北護分院的共同採用的 CAP 閾值，其中正常 (S0, CAP<238) 的有 96 筆，有脂肪肝 (S1~S3, CAP>=238) 的有 214 筆。

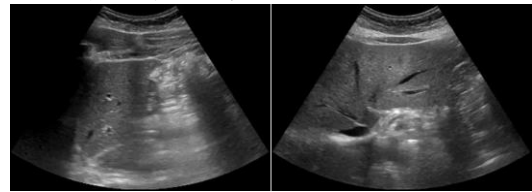


Figure 1.不同脂肪肝嚴重程度之超音波影像:  
左 S0(CAP:191)，右 S1~S3(CAP:302)

### 3.2.研究方法

本研究使用 VGG19 作為基底模型架構利用 NASRL 的方式開發出適合應用於超音波脂肪肝影像分類模型。

資料集依照 8:1:1 的比例分配用於早期脂肪肝預測模型的訓練、驗證與測試資料集，並通過 10Fold 進行交叉驗證。

#### ➤ 影像標準化

由於各醫院在拍攝超音波影像時使用的機台型號不同，影像的解析度和顏色等特徵也有所不同如 Figure2 所示。為了統一這些特徵，本研究對各機台獲得的 B-Mode 影像進行了前處理。首先將原始的 RGB 影像轉換為灰階影像，然後將影像大小統一調整為 224x224 像素。

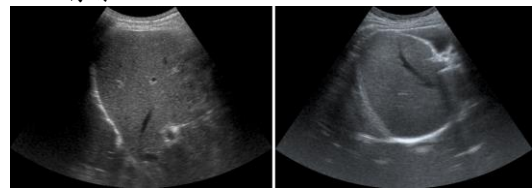


Figure 2.不同機台所拍攝出來的超音波影像  
左:Toshiba SSA-660A、右:Toshiba Aplio300

#### ➤ 資料平衡與資料增強

由於收集資料較為 S1~S3 類別因此需要對其進行資料平衡，避免模型訓練偏向於預測多數類別，忽略少數類別的情況發生，因此使用資料增強的方法降低資料不平衡的情況發生。

為了確保增強後的數據仍可代表真實世界的樣本分布，保持數據的真實性與可解釋性，本研究僅按照角度 90 度的倍數旋轉影像作為資料增強方法，解決資料不平衡問題。



### ➤ Visual Geometry Group 19 (VGG19)

VGG19 是一種基於卷積神經網路 (Convolutional Neural Network, CNN) 的模型如 Figure1 所示，其特點是具有較深的網路架構。通過多個卷積層 (Convolution Layer) 和池化 (Pooling) 操作，提取影像的低級特徵 (Low-Level Features) 如橫隔膜、肝臟邊界等，和高級特徵 (High-Level Features) 如特徵圖內每個特徵點之間的關係，從而學習到更多影像細節和結構訊息，這有助於模型更好地區分不同類別之間的差異。

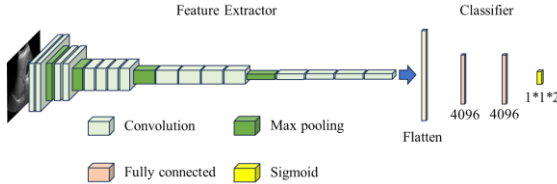


Figure 1. VGG19 架構圖

### ➤ Neural Architecture Search

神經網路搜索 (Neural Architecture Search, NAS) 在近年來逐漸受到重視[7]。神經網路在各種領域的任務中，如車牌辨識、醫學影像辨識和聲音訊號辨識等，已經展示了卓越的性能，作為一種非常強大且靈活的模型。然而，儘管神經網路取得了諸多成功，其設計過程卻相當複雜且耗時，需要大量時間和豐富的專家知識來手動構建模型和調整參數。

目前，大多數著名的神經網路 SOTA (State-Of-The-Art) 模型，如 AlexNet、VGG、GoogLeNet、ResNet 和 EfficientNet 等，主要針對自然影像辨識問題而設計，針對醫學影像辨識任務的開發相對較少。NAS 提供了一種自動化的途徑，能夠為特定任務自動尋找最佳的神經網路架構。這種方法有效解決了設計神經網路所需的專業知識問題，使非專業研究人員也可搜索出好的神經網路架構，此外顯著減少模型設計過程中所需的時間成本，使其能夠更加高效地應對各類任務。

NAS 可區分為三個階段[8]: 1. Search Space、2. Search Strategy、3. Performance Estimation Strategy 彼此之間的關係如 Figure2 所示，Search Space  $A$  定義為可搜索的 Operations 如: Convolution、Pooling、Kernel Size、Filter 等，透過 Search Strategy 選擇適當的 Operations 組合成模型架構  $A$ ，在評估階段使用 Performance Estimation Strategy 將評估結果回傳給 Search Strategy 更新 Search Strategy 參數產生新的模型架構，經疊代多次後搜索出最佳架構。

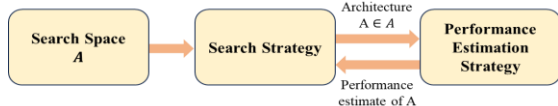


Figure 2. Neural Architecture Search 架構圖

### ➤ Search Space

在 NAS 任務中，定義一個良好的搜尋空間至關重要。如果搜尋空間較小，搜尋過程可能會較早收斂，進而找到局部最優的網路架構，而擴大搜尋空間可以增加找到真正新穎的神經網路模型的機會，但是，如果一味的擴大 Search Space 反而對 NAS 搜索最佳架構有害無益[9]，因此本計畫根據先前 SOTA 模型所使用 operations 作為 Search Space 利用這些先驗知識減少 NAS 搜索出最佳模型的難度。

本計畫使用 VGG19 作為 NAS 模型基底架構，採用 operations 有: Convolution\_kernel size: {1,3,5,7}、Convolution\_kernel\_size:{64,128,512}、總共 16 層 Convolution Layer，共有約  $10^{18}$  種組合，倘若直接進行隨機搜索找出最佳組合將會消耗大量的運算成本，因此選擇合適的搜索策略至關重要。

### ➤ Search Strategy

本研究所採用的搜索策略為 Reinforcement Learning 是由 Zoph 等人[10]所提出之方法，透過 Controller 循環神經網路 (Recurrent neural network, RNN) 搜索 operations 並組成子神經網路如 Figure3 所示，預測每一層的 Filter 數量、卷積核大小並重複。每個預測都由 Softmax 分類器執行，然後輸入到下一個時間步作為輸入。

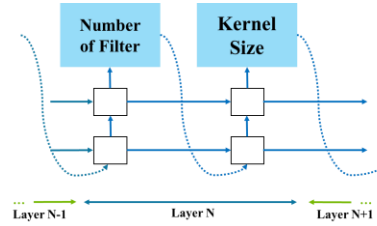


Figure 3. Controller RNN 採樣之子神經網路圖

其中 RNN 所採用的 operation 定義為 Action。Action 為經 RNN 運算後各個 operations 的權重，並經過 Softmax 運算後提取具有最大權重的 operation，整個過程如 Figure4 所示。

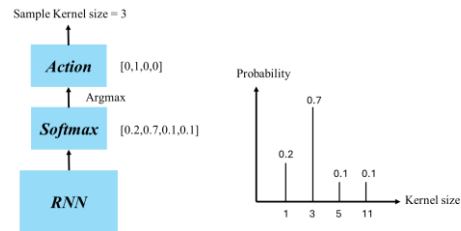


Figure 4. 舉例 Controller RNN 採樣子神經網路之 Kernel size 過程

當模型收斂時該子網路將最優 Validation Accuracy 作為獎勵訊號 (Reward Signal, R)，透過強化學習來訓練 Controller，換句話說，為了找到最佳架構需要 Controller 取得最大期望獎勵訊號  $J(\theta_c)$  如式 1 所示，其中  $\theta_c$  為 Controller 之參數，用於搜索架構，而  $a_{1:T}$  為每個 RNN 輸出所組成之 Action 列表。

$$J(\theta_c) = E_{p(a_{1:T}; \theta_c)}[R] \quad \text{式(1)}$$

由於 R 無法描述與  $\theta_c$  之關係因此 R 無法對  $\theta_c$  透過微分進行倒傳遞更新  $\theta_c$ ，因此 Zoph 等人使用 reinforce rule from Williams (1992)[11] 更新  $\theta_c$  如式 2 所示。

$$\nabla_{\theta_c} J(\theta_c) = \sum_{t=1}^T E_{p(a_{1:T}; \theta_c)} [\nabla_{\theta_c} \log P(a_t | a_{(t-1):1}; \theta_c) R] \quad \text{式(2)}$$

進一步根據經驗法則將式 2 近似為式 3，其中  $m$  代表 NAS 搜索出之不同架構數量，第  $k$  個神經網路架構在訓練資料集上進行訓練後所達到的 Validation Accuracy 為  $R_k$  由於式 3 更新梯度是無偏估計且具有較高的變異數，為了使估計更加精確，可將式 3 進一步修改為式 4，其中  $b$  為先前架構所得 Validation Accuracy 的指數移動平均值。

$$\frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla_{\theta_c} \log P(a_t | a_{(t-1):1}; \theta_c) R_k \quad \text{式(3)}$$

$$\frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla \theta_c \log P(a_t | a_{(t-1):1}; \theta_c) (R_k - b) \text{式(4)}$$

#### ➤ Performance Estimation Strategy

本研究將子神經網路最佳的 Validation Accuracy 作為 Reward Signal 更新  $\theta_c$ ，並存取 Validation loss 之權重作為挑選架構之依據，並進行 10-Fold 的交叉驗證，紀錄每層 Layer 之 Operation，最終產生一個最佳模型。

#### ➤ 模型評估標準

本研究採用了常見的分類模型評估標準，包括 Accuracy、Sensitivity、Specificity、Parameters、FPS 來進行評估模型的效能表現，其中 Accuracy 計算所有樣本的正确預測率、Sensitivity 為模型正确檢測出脂肪肝的能力，Specificity 為模型正确檢測非脂肪肝能力，Parameters 為評估模型參數使用效率、Frame per second(FPS)評估模型在測試階段每秒可以預測多少張影像，當 FPS 越高越能實現模型在臨床中進行 real time 的檢測。

#### 四、結果與討論

VGG19 雖然架構簡單、在不管是在自然影像或是醫學影像上進行分類任務皆取得不錯的成績，更有許多研究使用 VGG19 進行脂肪肝分類模型開發並取得不錯表現[12]，因此將 VGG19 作為本研究之基線模型進行開發與比較。

在早期脂肪肝檢測模型，使用 B-Mode 影像作為輸入資料，VGG19 模型的效能表現在 Accuracy、Sensitivity、Specificity、Parameters、FPS 分別為  $0.72 \pm 0.06$ 、 $0.69 \pm 0.13$ 、 $0.78 \pm 0.19$ 、139,577,282、38。為進一步提升模型的效能，透過使用 NAS 方法，改善模型結構，效能表現在 Accuracy、Sensitivity、Specificity、Parameters、FPS 分別為  $0.77 \pm 0.07$ 、 $0.76 \pm 0.1$ 、 $0.77 \pm 0.08$ 、31,164,674、41，在 Specificity 方面不僅維持預測非脂肪肝的效能，並降低標準差，模型預測非脂肪肝能力更加穩定，此外在 Accuracy、Sensitivity 表現上也有所提升，Parameters 相較於 VGG19 顯著減少 78%，不僅降低計算資源需求，更提升模型 FPS，整體效能表現如 Table1 所示。研究結果顯示透過 NAS 方法搜索出之架構相較於 VGG19 透過降低參數量的方式有助於提升模型泛化能力，並降低過擬合的狀況，提高了模型在預測有無脂肪肝時的平衡性，降低使模型偏好預測某一個類別。以上實驗結果表明，NAS 方法是一種有效的模型優化策略，能夠在應用於超音波早期脂肪肝檢測任務中，提高模型的效能。

Table 1. 模型預測結果

	Sensitivity	Specificity	Accuracy	Parameters	FPS
VGG19	$0.69 \pm 0.13$	$0.78 \pm 0.19$	$0.72 \pm 0.06$	139,577,282	38
NASRL	$0.76 \pm 0.10$	$0.77 \pm 0.08$	$0.77 \pm 0.07$	31,164,674	41

#### 五、結論

為了客觀、準確、有效地輔助醫師診斷，本研究提出可泛用於多個超音波機器的早期脂肪肝檢測模型，透過 NAS 方法搜索出之架構不僅克服訓練時資料不平衡問題，在 Accuracy、Sensitivity、Specificity、Parameters、FPS 分別達到了  $0.77 \pm 0.07$ 、 $0.76 \pm 0.1$ 、 $0.77 \pm 0.08$ 、31,164,674、41，研究結果顯示使用 NAS 方法改良 VGG19 在早期脂肪肝分類任務上取得一定的效能提升，在未來期望持續收集來自不同超音波機器所拍攝影像以及不同醫生所拍攝之超音波肝實質影

像進一步提升模型泛化程度，而在模型方面將持續開發 NAS 應用於脂肪肝分級模型。

#### 六、致謝

本研究感謝國科會計畫【NSTC111-2628-E239-001-MY3】的支持。特別感謝中國醫藥大學附設醫院及國立臺灣大學醫學院附設醫院北護分院在資料收集方面的協助。

#### 七、參考文獻

- [1] C. S.Hsu et al, "Non-alcoholic fatty liver disease: An emerging liver disease in Taiwan," J. Formos. Med. Assoc., vol.111, no.10, pp.527-535, Oct.2012,
- [2] Devarbhavi, H., Asrani, S. K., Arab, J. P., Nartey, Y. A., Pose, E., & Kamath, P. S. (2023). Global burden of liver disease: 2023 update. *Journal of hepatology*, 79(2), 516-537
- [3] P. J.Eddowes et al., "Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease," *Gastroenterology*, vol. 156, no. 6, pp. 1717-1730, May2019
- [4] Eddowes, P. J., Sasso, M., Allison, M., Tsochatzis, E., Anstee, Q. M., Sheridan, D., ... & Newsome, P. N. (2019). Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology*, 156(6), 1717-1730.
- [5] Byra, M., Styczynski, G., Szmigielski, C., Kalinowski, P., Michałowski, Ł., Paluszkiwicz, R., ... & Nowicki, A. (2018). Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *International journal of computer assisted radiology and surgery*, 13, 1895-1903.
- [6] M.N.Ibrahim et al., "Automated fatty liver disease detection in point-of-care ultrasound B-mode images" *J.Med.Imaging*, vol.10, no.3, p.34505, 2023
- [7] Difan Deng and Marius Lindauer. Literature list on Neural Architecture Search, 2021.
- [8] Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural architecture search: A survey. *Journal of Machine Learning Research*, 20(55), 1-21.
- [9] Ci, Y., Lin, C., Sun, M., Chen, B., Zhang, H., & Ouyang, W. (2021). Evolving search space for neural architecture search. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6659-6669).
- [10] Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.
- [11] Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8, 229-256.
- [12] Zamanian, H., Shalbaf, A., Zali, M. R., Khalaj, A. R., Dehghan, P., Tabesh, M., ... & Acharya, U. R. (2023). Application of artificial intelligence techniques for non-alcoholic fatty liver disease diagnosis: A systematic review (2005-2023). *Computer Methods and Programs in Biomedicine*, 107932.