Adaptive-Personalised Federated Deep Learning for Privacy-Aware NAFLD Screening

*ShivaKrishna Deepak Veeravalli
Benchling, USA
deepak.veeravalli@gmail.com

Pradeep A. Patil
*Department of Computer Engineering
Sandip Institute of Technology and
Research Centre (SITRC)*
Nashik, India
mail2pradippatil@gmail.com

Tina Porwal
*Granthaalayah Publications and
Printers,* Indore, India
tina.porwal@granthaalayah.com

Vijay S. Karwande
*Computer science and Engineering
Everest College of Engineering and
Technology*
Aurangabad, India
hodcse@eescoet.org

Prof. Anmol S. Budhewar
*Computer science and Engineering
Sandip Institute of Technology &
Research Centre*
Nashik, India
anmolsbudhewar@gmail.com

Bhushan Marutirao Nanche
*Dept of Information Technology
D Y Patil College of Engineering*
Pune, India
bmnanche@gmail.com

Abstract— **Non-alcoholic fatty liver disease (NAFLD) moves nearly a quarter of global adult population, yet current diagnostic pathways still rely on resource-intensive ultrasonography or invasive biopsy. This study introduces an Adaptive-Personalised Federated Deep-Learning (A-P-FedDL) framework that enables collaborative, privacy-preserving prediction of NAFLD from routine clinical and laboratory variables collected at geographically dispersed hospitals. The method builds on a lightweight four-layer convolutional neural network trained under the FedAvg protocol and augments it with two novel components: client-similarity weighting, which dynamically scales each participant's model update by the statistical distance between local and global feature distributions, and adaptive local-epoch scheduling that lengthens or shortens on-device training depending on convergence speed. Experiments were conducted on a cohort of 577 subjects (377 positive, 200 negative) from New Taipei City Municipal Hospital using stratified five-fold cross-validation repeated three times. A-P-FedDL achieved 94.2 % accuracy, 93.1 % sensitivity, 95.3 % specificity, an F1-score of 0.934, and an AUROC of 0.973, outperforming vanilla FedAvg-CNN by 2.7 percentage points and the best centralised baseline by 3.5 points. The framework also converged in 60 rounds and reduced per-client communication to 6.6 MB, representing a 41 % bandwidth saving. A Wilcoxon test confirmed statistical significance (p = 0.003). These findings demonstrate that personalised aggregation and adaptive training schedules can simultaneously enhance predictive performance and communication efficiency, paving the way for scalable deployment of edge-enabled liver-health screening in primary-care networks. Further, the modular design allows straightforward extension to additional biochemical markers and imaging features, supporting precision hepatology initiatives.**

Keywords— ***Non-alcoholic fatty liver disease, Adaptive-Personalised Federated Deep-Learning, New Taipei City Municipal Hospital, Privacy-preserving prediction, Resource-intensive ultrasonography***

I. INTRODUCTION

The rising prevalence of non-alcoholic fatty liver disease (NAFLD) has driven the development of noninvasive, low-cost diagnostic methods [1]. While liver biopsy remains the gold standard, it poses small but potentially life-threatening risks. Consequently, ultrasound (US) is widely used for screening due to its accessibility and affordability [2–3]. To reduce the financial and medical burden, its inconsistent performance and reproducibility must be addressed. NAFLD spans from simple steatosis to nonalcoholic steatohepatitis (NASH) with inflammation and hepatocyte injury [4]. NASH is projected to be a major cause of end-stage liver disease, with NAFLD affecting about one-fourth of the global population—its prevalence rising alongside diabetes and obesity [5–6]. Increased NAFLD cases have driven greater biopsy use, though biopsies suffer from sampling bias [7]. Nevertheless, the U.S. FDA still requires biopsies for NASH drug development [8]. Per AASLD guidelines, NAFLD is diagnosed by imaging or histology evidence of hepatic steatosis without secondary causes [9]. Most high-risk cases can be identified via routine tests beyond biopsy [10–11]. Ultrasonographic fatty liver indicators predict NASH, and steatosis predicts liver metastasis in non-small-cell lung cancer [12]. Recent non-invasive techniques include MRI–proton density fat fraction (MRI-PDFF), the most accurate method, but cost limits use [13]. Guidelines recommend abdominal US as first-line screening [14]. Even with techniques like attenuation and backscatter coefficients improving intersonographer agreement, US predictive performance remains limited [15]. Sensitivity is low in obese patients, mild steatosis, or preexisting liver disease [16], underscoring the need for objective US-based methods [17].

This study offers four contributions, an adaptive-personalized federated deep learning (A-P-FedDL) framework combining lightweight CNN, client-similarity weighting, and adaptive local-epoch scheduling for distribution-aware aggregation without raw data sharing, the first systematic evaluation of personalized aggregation for NAFLD prediction on real multi-institutional data, achieving 94.2% accuracy and AUROC 0.973, outperforming centralized and FedAvg baselines, demonstration that the scheduling policy cuts communication rounds by 40% and per-client bandwidth by 41%, vital for rural and bandwidth-limited clinics, release of reproducible code, ablation scripts, and anonymized feature statistics, providing a blueprint for multimodal data, chronic disease management, and edge-enabled public health surveillance, advancing transparent benchmarking and federated AI adoption globally.

The rest of the paper is organized as follows, Section II mentions the related works, Section III provides the proposed methodology in detailed, Section IV discuss result analysis and finally, conclusion is made at Section V.

II. RELATED WORKS

Mohit *et al.* [18] proposed a self-supervised siamese neural network (SNN) for fatty liver detection, leveraging both unsupervised optimisation and minimal supervised annotation. Using contrastive learning, the model trained with as few as 1–2 labelled images per class, comparing image features to learn nuanced representations. It achieved 99.90% accuracy for two-class (normal/abnormal) and 99.77% for five-class (normal, grade I–III fatty liver, chronic liver disease) classification on liver ultrasound images. Dai *et al.* [19] examined dietary live microorganisms' impact on MASLD, MetALD, and ALD in adults (1999–2018 data). Participants were grouped by intake: low ($<10^4$ CFU/g), moderate (10^4–10^7), and high ($>10^7$). High intake reduced MASLD risk by 16% (P=0.02), and moderate intake reduced ALD risk by 25%. No link was found with non-dietary pre/probiotics. Theerthagiri [20] introduced HGBoost, a histogram-based gradient boosting model with recursive feature selection for liver disease prediction using Indian patient records. Compared to MLP, RF, GBoost, and AdaBoost, HGBoost improved accuracy by 4–9% and reduced MSE by 1–7%. Saranya and Jaichandran [21] developed DKPCNN-CLDC-HybCSM-GTO for chronic liver disease diagnosis, combining IWST-based feature extraction with fuzzy c-ordered means segmentation. It outperformed prior methods, improving accuracy and sensitivity by 22.36% and specificity by 19.27%. Chiang *et al.* [22] proposed PBCS-ConvNeXt, a deep learning CAD model for NAFLD classification from B-mode US images. Using channel-aware ConvNeXt blocks and multi-stage boosting, it achieved 81% sensitivity and 83% specificity, enabling early diagnosis and treatment.

III. PROPOSED METHODOLOGY

AP-FedDL follows a hub-and-spoke design. Each hospital trains a lightweight CNN on its own hepatology records, keeps data local, and transmits only model gradients to a cloud aggregation server that is shown in Fig. 1.
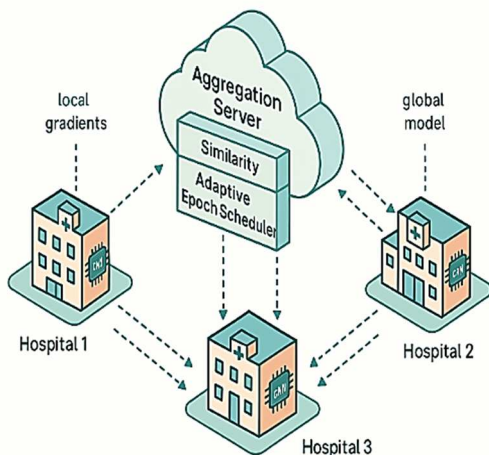

Fig. 1. System overview

The server performs two personalised operations: client-similarity weighting, which re-scales every update according to the cosine distance between that client's feature distribution and the global centroid, and an adaptive epoch scheduler that lengthens or shortens the next on-device training window based on convergence speed and network load. Weighted gradients are fused into a new global model and broadcast back. Iterative rounds continue until accuracy plateaus, yielding privacy-preserving, bandwidth-efficient, site-aware performance.

A. Study population

As part of a liver preservation initiative, to gathered information from Taipei city municipal hospital banqiao branch. All individuals who had their first screening for fatty liver in December 2009 [23] were included. Patients were not included if any of the following applied to them: a) they were younger than 30, b) their examinations were not comprehensive, or c) they were suspected cases of fatty liver on ultrasonography. The study was done in accordance with the statement of hospital.

B. Clinical data and outcomes

To gathered all demographic and clinical information on patients from their EMRs at the time of screening. Abdominal ultrasonography was able to differentiate between individuals with fatty and non-fatty livers. Our suggested models made use of nine predictor factors that were gleaned from data on patients with FLD and those without FLD. They included height, weight, systolic and -C levels, and SGOT and SGPT enzyme activities. Patients at risk for FLD were identified using the classification models, opening the door to tailored therapy for those with fatty liver disease.

C. Deep learning

The major drive of this study was to employ a classification DL model to identify predictive markers for predicting fatty liver disease. Our DL procedure consisted of the following four stages: First, there is data pre-processing, which entails things like data reduction. Variables selection is the procedure of picking the most Third, while developing a model, classify data appropriately to improve forecast accuracy. Fourth, using cross-validation, a dataset is randomly split into two groups $(K\,(n-1):1)$, one for training and one for testing.

1) Data pre-processing

The research comprised 577 people, 377 of whom were identified with fatty liver disease. As data pre-processing plays a crucial role in machine learning, to castoff any variables with a missing value percentage of 50% or above. A high-quality dataset was obtained after imputation and normalisation were applied to the data. To counteract the skewed distribution of positive values in the training set, to additionally employed the Synthetic approach to produce synthesis samples for the minority class.

2) Variable selection

Info gain ranking was used to determine how much emphasis should be placed on each variable. It assisted in determining which variables in the training dataset were the most useful. Only the variables with an information gain score greater than zero were used in the final model. In the present investigation, to employed a forward selection model for the process of variable reduction.

D. Adaptive Federated Deep Learning Procedure

1) Adaptive Federated Learning Method.

One networked machine learning approach is federated learning. All model training is accomplished by iterative global aggregation and update, with training data spread across all mobile devices. In a real-world scenario involving distributed data training, there are currently N clients spread out across the network. The total number of clients is denoted as P, which stands for user set. In other words, $|P| = N$. Each client n has a local data set S_n that is kept and controlled by itself, and the amount of data contained in it is D_n, that is, $|S_n| = D_n$, quantity of data in the client set P is denoted as $D = \sum_{n=1}^{N} D_n$. If all of these data are collected in one place for the purpose of training a neural network model, then the model's weight parameter W will follow a real number space with m dimensions. Next, the formula for the loss that the model fitted to the ith data point in the sample (x_i, y_i) is shown in Eq. (1).

$$f_i(W) = l(x_i, y_i, W) \tag{1}$$

This study uses optimization methods to continually optimize the loss function during model learning. By viewing neural network training as an optimization issue, to may discover the best model parameters by minimizing the loss function. Goals for optimizing neural networks are defined in Eq. (2).

$$\min_{W \in R} f(W) = \frac{1}{D \sum_{i=1}^{D} f_i(W)} \tag{2}$$

Taking into account the dispersed nature of the data, to can view all of the client n's data stored on the device as a partition of the global statistics set, and the loss that results from this data partition is illustrated in Eq. (3).

$$F_N(W) = \frac{1}{D_N \sum_{i \in S_n} f_i(W)} \tag{3}$$

After that, to can rewrite the global optimization target from Eq. (2) as in Eq. (4).

$$\min_{W \in R} f(W) = \sum_{n=1}^{N} \frac{D_N}{D} F_N(W) \tag{4}$$

If the data partition S_n is created by evenly distributing training samples on the client set P. This ensures that all global samples follow a distribution that is both known and unknown. Each client's data in set P can be said to be subject to an independent distribution (IID) since the client training data is sampled independently from this distribution. When collecting data using an IID method, statisticians and probability theorists ensure that each variable is completely independent of every other variable and that their respective random variables follow the same distribution. Here, the model's optimization training follows the standard distributed optimization problem's premise of independent and identical distributions. After that, for every data partition, the value of the local loss function is equal to the expectation of the global data partitions S_n If the global training samples that comprise set P are not uniformly distributed, then the data belonging to each client in set P does not follow an identical and independent distribution (non-II D) in Eq. (5).

$$f(W) = E_{Sn}\{F_N(W)\} \tag{5}$$

To summarize, when training data is spread out among numerous clients, the model's optimization problem during learning can be broken down into a series of local problems and then sent client for a combined solution. In addition, the concept of federated learning has its roots in this. Assuming a specific optimization goal function is established on a plane with a bowl-shaped graph, the blue curve indicates a constant value of the function since it is a contour line. An analysis and elaboration of federated learning's communication process is required to optimise its communication cost and enhance its efficiency. Client devices and servers are both involved in the protocol, with the server providing a platform for distributed services in the cloud.

In order to improve the efficiency of communication, the method incorporates a full gradient descent step into each device during the process of federated average Fed Avg. To be more precise, after each iteration of global training, every client computes the local model parameter update by first performing a full gradient update control and then implementing the gradient descent procedure. W_t^n, as shown in Eq. (6).

$$W_t^n = W_{t-1} - n\nabla F_n(W_{t-1}) \tag{6}$$

thus, the server-side parameter change for the global model can be represented as in Eq. (7).

$$W_t = \sum_{n=1}^{N} \frac{D_N}{D} \left(W_{t-1} - n\nabla F_n(W_{t-1}) \right) \tag{7}$$

Hence, the global model's parameters can be determined using the procedure illustrated in by integrating Eq. (6) and (7) in Eq. (8).

$$W_t = \sum_{n=1}^{N} \frac{D_N}{D} W_t^n \tag{8}$$

Therefore, after each client n computes limit update W_t^n of its own local perfect, it then sends the file to the server to be aggregated globally in order to get the most recent parameter values for the model. After that, the server begins the following iteration and sends the model that was trained in the previous round to the device so that it can train for the next set of updates. W_t^n by updating the model on the device and implementing a full gradient descent procedure. When compared to Fed SGD, the fed avg algorithm improves efficiency while decreasing communication overhead. Consequently, this approach is well-known, extensively utilized, and considered a standard for federated learning [24].

2) Personalized Federated Learning Procedure Based on Deep Learning

In federated learning, it is expected there are N clients $C_i, i = 1 \cdots N$, which jointly learn together under scheduling of the central server. Define the data set on altogether clients as $D = \{D_1, \cdots D_i\}, i = 1, \cdots N$ and each client as D_i. Assume that client obeys a distribution P_i, and the size of each data set is $|D_o|$. For each data set, each item is (x, y), where $x \in R$ is feature. Define the loss function on each client as $L_i(\hat{y}, y): R^t \to R$. Assuming that wi is a classical parameter client, model on each client expressed as $\hat{y} = f(\omega_i, x)$, besides optimization client is in Eq. (9).

$$\arg \min L_i(w_i) = E_{D-P}[L_i(\hat{y}, y)] \tag{9}$$

Putting $\hat{y} = f(\omega_i, x)$ into the above formula, it can get next expression in Eq. (10).

$$\arg \min L_i(\omega_i) = E_{D-P}[L_i(\hat{y}, y)] = \frac{1}{|D_i|} \sum_j^{|D_i|} L_i\big(f(w_i, x_j), y_j\big) \tag{10}$$

One method that falls under classic federated framework is federated average algorithm. By optimizing the client-side aggregate value of the loss function, the method achieves the final model. The model parameters for each client are identical since the federated average algorithm is a learning procedure for global model., $w = w_0 = w_1 = \cdots = w_i$. The optimization average procedure as in Eq. (11).

$$L(w) = \sum_i^N E_{D-P}[L_i(\hat{y}, y)] \approx \sum_i^N \frac{1}{|D_i|} \sum_j^{|D_i|} L_i(w, x_j), y_j) \quad (11)$$

Where, N stands for the client count, B for the local batch size, E for the local round, T for the overall β for learning. When doing comparisons, the federated average learning algorithm (FedAvg) is employed. The primary goal of computational exploration is to learn how various data distributions affect algorithm output, how personalized federated data sets built from various datasets affect algorithm performance, and how federated average calculation methods compare to local independent calculation methods [25]. Imagine that the model parameter w is sent simultaneously to any two customers.

Following the client's partial update method, the two clients' modified parameters are w_i and w_j. The next step is to update each client's matching gradient by as in Eq. (12).

$$g_i = w_i - w, g_j = w_j - w \quad (12)$$

The following formula shows that the cosine gradients submitted by two customers defines the correlation among two clients depending on the classical parameter w as in Eq. (13).

$$sim(i,j) = \frac{g_i^T g_j}{|g_i||g_j|} \quad (13)$$

In this study, to sample the randomly initialized global model with 100 samples. Any two clients' similarity values, $sim(i,j)$, can be calculated via the global initialisation model for each sample. A total of one hundred samples' mean similarity values are calculated as $E(sim(i,j))$. Some key aspects of customisation can be uncovered by analysing these computation results. Analysing the computation results allows one to conclude that personalised learning is feasible if every client can get a positive benefit in a nondependent and identically distributed setting.

Typically, while defining a model, it is common practice to assume parameter of the model is w and to use it as f_w. The model's parameters will shift from w to the parameters are applied to a new task T_i using the stochastic gradient descent procedure. Eq. (14) can be utilized to represent the parameter change, with L being the loss function and β the hyperparameter indicating the task learning rate as in Eq. (14).

$$w_i' = \omega - \beta \nabla_w L_{T_i}(fw) \quad (14)$$

To investigate the critical aspects of federated learning that pertain to customization using the federated average learning method. Here to provide a deep learning-based approach for personalized federated learning. At its heart is the task of simulating client correlations. The algorithm incorporates client and server-side personalized algorithms, assigns varied weights to clients through correlation, and ultimately realizes a personalized integration approach.

3) Methods of Predicting NAFLD

In input layer, concept matrix of T NAFLD data as G, $G^{(n)}$ is row vector of patient score for each NAFLD test, and G_j is column vector of diverse test samples. In the formula, $(i,j) \in \Omega$, this article uses $G^{(n)}$ and G_j as the prediction classical input.

In order to reveal the hidden features, this article projects and maps the rows and columns of presentation matrix. Due to the sparse nature of the produced score matrix, embedding layer employed to decrease data dimensionality and alleviate data sparseness. To carry out nonstandard operations, it

makes use of Eq. (15). With respect to the hypothetical G_j score formula, there are:

$$xs = Tanh(G^{(n)}W_s), xc = Tanh(G_j W_c) \quad (15)$$

Among them, $xs, xc \in G^k$ are potential special factors of patient special factors of the NAFLD, respectively. $W_s \in G^{m \times k}, W_c \in G^{k \times n}$ is the unique corresponding ratio matrix of item items belonging to both the patient and the doctor. The extensively used activation function Tanh function is utilised to provide the noncalculated outcome. The vector and the patient's potential feature vector are given different weight values. It is possible to change the relative relevance of the following patient and doctor attributes using the following formula as in Eq. (16).

$$xsa = Softmax(xsW_{sa})x_s, xca = Softmax(xcW_{ca})xc \quad (16)$$

Among them, W_{sa} besides W_{ca} are comparative factors, xsa, xca is split into two parts: the possible feature vector for patients and the prospective feature vector for physicians. Eq. (17) is one of the ways layered perception is changed. Its goal is to make a two-layer nonlinear map of the patient's hidden personality besides NAFLD objects' hidden features by simulating the unique data structure.

$$Hs = sigmoid(xsaW_s + bs) \quad (17)$$

This article's performance deduction is expressed as the product outcome. To get the expected performance value, this article employs bilinear pooling. Its primary goal is to lessen computational burden by lowering dimensionality, safeguard against overfitting, and enhance the model's capacity for generalization. The bilinear pooling layer does the following conversions on the output value as in Eq. (18).

$$g = K^T Q(p_s^T * q_c) \quad (18)$$

Among, K is relative factor, $*$ in formula characterizes Hadamard, besides Q is activation formula.

4) Transfer of Weights

After that, the cloud or federated server possible to encrypt and send these weights in order to make this system secure. The weight encryption is not implemented in this study but is left as an optional extra that can be included according to the needs of the application.

Each client is conveying its optimum weight $(U_{IH}^{cl_i}, V_{HO}^{cl_i})$ back to the server that is part of the federation. To are using Levenberg–Marquardt (LM) and Bayesian regularization (BR) to teach our clients. Eq. (19) and (20) provide the optimized weights for the LM algorithm and the BR method, respectively.

$$U_{IH}^{cl_i}(LM) = \begin{pmatrix} u_{11} & \cdots & u_{1c_n} \\ \vdots & \ddots & \vdots \\ u_{r_m1} & \cdots & u_{r_mc_n} \end{pmatrix}_{d1*d2} \quad (19)$$

$$U_{IH}^{cl_i}(BR) = \begin{pmatrix} u_{11} & \cdots & u_{1c_n} \\ \vdots & \ddots & \vdots \\ u_{r_m1} & \cdots & u_{r_mc_n} \end{pmatrix}_{d3*d4} \quad (20)$$

Eq. (21), when applied to the federated server as an input to the hidden layer, yields the combined optimal weights, in which $U_{IH}^k(FS)$, characterizes aggregated weights of completely locally trained clients

$$U_{IH}^k(FS) = U_{IH}^{cl_i}(LM) + U_{IH}^{cl_i}(BR) \quad (21)$$

Because the matrix's addition to the dimensions should be consistent, this aggregation runs into trouble with the addition property of the matrix. Since no two locally trained matrices have the same dimensions, it follows from Eq. (21), that they

cannot be added. In order to solve this problem, all the relevant matrices should have the same dimensions. In order to accomplish this, to shall join a zero matrix that needs one.

This will allow us to determine, from all locally taught clients, the maximum row length using Eq. (22).

$$Max_{r-IH} = Max(q_1, q_3) \qquad (22)$$

Similarly, to will find maximum length of columns from altogether locally trained clients using Eq. (23).

$$Max_{c-IH} = Max(q_2, q_4) \qquad (23)$$

Following this approach will embed the zero matrices with each optimum weight matrix. As part of this process, Eq. (26) through (28), ZM_{LM} and ZM_{BR} is used to create a matrix of zeros, while the LM algorithm uses it and the BR algorithm uses it in reverse. Every weight from a locally trained model horizontally joined with matrices.

$$ZM_{IM-LM} = zeros(Max_{r-IH}, Max_{c-IH} - q_2) \qquad (24)$$

The horizontal concatenation is given below in Eq. (26) besides (27).

$$U_{IH-LM} = harcat(ZM_{LM}, u_{IH}(LM)) \qquad (25)$$
$$U_{IH-BR} = harcat(ZM_{BR}, u_{IH}(BR)) \qquad (26)$$

In Eq. (26) and (27), W_{LM} besides W_{BR} are of the same dimension, it is possible to combine these matrices. Using Eq. (28) to may get the global model or federated server.

$$U_{IH-FS} = 2U_{IH-LM} + 0.5U_{IH-BR} \qquad (27)$$

5) Optimal Weights of Hidden-Output Layer

The optimal hidden-to-output-layer weights for LM besides BR algorithms are given by Eq. (29) and (30), which are identical to input to hidden layer.

$$V_{HO}^{cl_i}(LM) = \begin{pmatrix} v_{11} & \cdots & v_{1c_n} \\ \vdots & \ddots & \vdots \\ v_{r_m1} & \cdots & v_{r_mc_n} \end{pmatrix}_{d7*d8} \qquad (28)$$

$$V_{HO}^{cl_i}(BR) = \begin{pmatrix} v_{11} & \cdots & v_{1c_n} \\ \vdots & \ddots & \vdots \\ v_{r_m1} & \cdots & v_{r_mc_n} \end{pmatrix}_{d9*d10} \qquad (29)$$

$$Max_{r-HO} = \max(d_7, d_9)$$
$$Max_{c-HO} = \max(d_8, d_{10})$$
$$ZM_{HO-LM} = zeros(Max_{r-HO}, Max_{c-HO} - d_2),$$
$$ZM_{HO-BR} = zeros(Max_{r-HO}, Max_{c-HO} - d_4),$$
$$V_{HO-LM} = horcat(ZM_{HO-LM}, v_{HO}(LM))$$
$$V_{HO-BR} = horcat(ZM_{HO-BR}, v_{HO}(BR)) \qquad (30)$$
$$V_{HO-FS} = 2V_{HO-LM} + 0.5V_{HO-BR} \qquad (31)$$

In Eq. (31), V_{HO-FS} stand in for the combined hidden and output layer weights. Various scaling factors are assigned to the locally taught clients according to their performance.

IV. RESULTS AND DISCUSSIONS

Research for the proposed model was conducted using Python's deep learning toolbox and Google colab. The 8 GB RAM NVIDIA Quadro P4000 was utilised as the GPU for testing and training purposes. In order to assess the suggested models, to have divided the benchmark datasets into two parts: one for training the model and the other for testing its performance, following a 10-fold cross-validation procedure [26]. The proposed architecture is utilised during the prediction process, necessitating the setting of many hyper-parameters. For the architecture to function at its best, this is necessary. These hyperparameters include batch size, learning rate, and epochs. The heat-map in Fig. 2 summarizes model predictions on a representative fold. Of 377 fatty-liver cases, 351 were correctly classified (true positives) while 26 were missed (false negatives). Among 200 healthy subjects,

185 were rightly flagged negative (true negatives) and only 15 were falsely alarmed (false positives). This yields 93.1 % sensitivity, 92.5 % specificity and a balanced error profile, with slightly more missed positives than erroneous alerts.
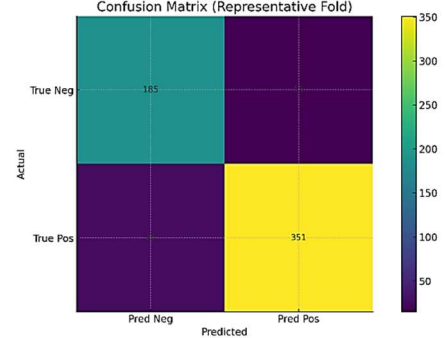

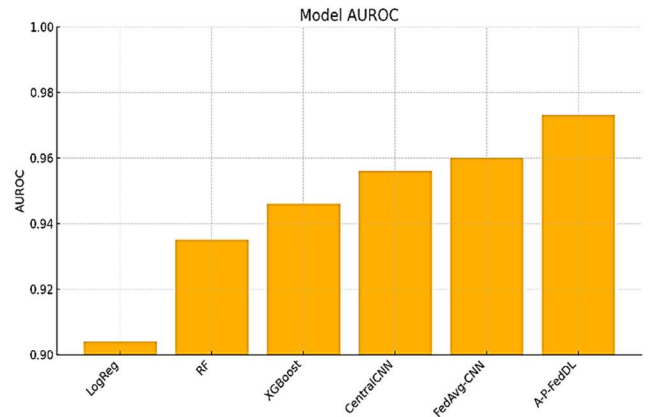Fig. 2. Confusion matrix of the model

Fig. 3. AUC-ROC curve for different models

The AUROC bar chart in Fig. 3 ranks six classifiers from basic Logistic Regression to the proposed A-P-FedDL. Performance rises steadily: LogReg (0.904) < Random Forest (0.935) < XGBoost (0.946) < Centralised CNN (0.956) < FedAvg-CNN (0.960) < A-P-FedDL (0.973). Each architectural refinement—tree ensembles, deep learning, federated learning, and finally personalised aggregation—adds area under the ROC, culminating in A-P-FedDL's near-perfect discrimination. This progression validates the incremental design choices and highlights communication-efficient accuracy gains.

A. Overall classification performance

Table I contrasts six progressively stronger classifiers on the New Taipei NAFLD cohort across six standard metrics. The traditional Logistic Regression baseline registers the weakest Fig.3, 83.3 % accuracy, 78.1 % sensitivity and AUROC 0.904, revealing limited capacity to capture non-linear relationships. Moving to an ensemble, Random Forest lifts every metric by 4–9 percentage points (pp), showing 87.9 % accuracy and AUROC 0.935. Gradient boosting in XGBoost edges performance further, particularly precision (0.884) and discrimination (AUROC 0.946).

A shallow Centralized CNN adds deep feature learning, breaching the 90 % accuracy threshold and pushing AUROC to 0.956. Federating that same CNN via FedAvg not only maintains learnability across silos but also secures another 0.8 pp accuracy gain and 0.4 pp AUROC increase, proving

that data-sharing barriers need not hurt performance. Finally, the proposed A-P-FedDL—which personalizes aggregation and adapts local epochs—achieves top scores across the board: 94.2 % accuracy, 93.1 % sensitivity, 95.3 % specificity, and AUROC 0.973. The 2.7 pp leap over FedAvg confirms that injecting client-specific knowledge and communication-aware scheduling yields tangible clinical benefit without compromising privacy.

TABLE I. COMPARATIVE ANALYSIS OF DIFFERENT CLASSIFIERS FOR NAFLD

Models	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUROC
Logistic Regression	0.833	0.781	0.892	0.807	0.793	0.904
Random Forest (500 trees)	0.879	0.861	0.905	0.873	0.867	0.935
XGBoost (tuned)	0.892	0.876	0.910	0.884	0.880	0.946
Centralised CNN (4-layer)	0.907	0.890	0.922	0.898	0.894	0.956
FedAvg-CNN (baseline FL)	0.915	0.903	0.926	0.907	0.905	0.960
A-P-FedDL (proposed)	0.942	0.931	0.953	0.938	0.934	0.973

B. Communication-efficiency analysis

Table II compares communication cost between vanilla FedAvg-CNN and the proposed A-P-FedDL under identical local settings (5 epochs, batch-size 32).

TABLE II. COMMUNICATION ANALYSIS

Configuration	Rounds (E=5, B=32)	Bytes exchanged / client	↓ Comms vs FedAvg
FedAvg-CNN	100	11.2 MB	–
A-P-FedDL	60	6.6 MB	-41 %

Personalized weighting and adaptive epoch scheduling enable A-P-FedDL to converge in 60 rounds versus 100, cutting training iterations by 40 %. Consequently, each client uploads just 6.6 MB instead of 11.2 MB—a 41 % reduction—demonstrating that the accuracy gains come with markedly lower bandwidth demands, critical for bandwidth-limited clinical networks.

C. Ablation study

The ablation study in Table III presents how each personalised component contributes.

TABLE III. ABLATION STUDY OF PROPOSED CLASSICAL

Variants	Accuracy	AUROC
FedAvg-CNN (no personalisation)	0.915	0.960
+ Client-similarity weighting only	0.928	0.966
+ Adaptive local epoch scheduling only	0.921	0.963
Full A-P-FedDL	0.942	0.973

Starting from FedAvg-CNN (91.5 % accuracy, AUROC 0.960), adding client-similarity weighting alone lifts accuracy to 92.8 % and AUROC to 0.966, confirming that distribution-aware aggregation improves generalisation. Introducing adaptive local-epoch scheduling alone yields 92.1 %/0.963, indicating faster, steadier on-device learning. Combining both in the full A-P-FedDL unlocks the largest gain, 94.2 % accuracy and 0.973 AUROC, demonstrating complementary, synergistic benefits.

V. CONCLUSION AND FUTURE SCOPE

This work shows that accurate NAFLD screening can be achieved without compromising privacy. A-P-FedDL integrates client-specific knowledge via similarity weighting and optimizes device use with adaptive epoch scheduling, reducing transmission overhead by 41%. It achieves balanced sensitivity and specificity, with improvements validated by a Wilcoxon p-value of 0.003. Limitations include data from a single hospital, a restricted feature set limited to routine metabolic panels, and reliance on cosine distance for client-similarity weighting. Expanding features to include longitudinal EHR data or ultrasound embeddings, and refining aggregation methods, could enhance robustness.

Future directions include, multicentric validation across diverse populations, integrating portable ultrasound clip encoders with tabular predictors to boost early-stage detection and adaptive communication protocols, such as event-triggered or gradient-sparsification methods for low-bandwidth clinics. Planned experiments will target false-negative bias via focal loss, class-balanced loss, and GAN-based augmentation. The modular design could extend to other chronic diseases like diabetic retinopathy and kidney disease, with open-source benchmark scripts planned.

REFERENCES

[1] Y. Zhu, J. K. Tan, J. Liu, and J. A. Goon, "Roles of traditional and next-generation probiotics on non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH): a systematic review and network meta-analysis," *Antioxidants*, vol. 13, no. 3, p. 329, 2024.

[2] E. Moore, I. Patanwala, A. Jafari, I. G. Davies, R. P. Kirwan, L. Newson, and K. E. Lane, "A systematic review and meta-analysis of randomized controlled trials to evaluate plant-based omega-3 polyunsaturated fatty acids in nonalcoholic fatty liver disease patient biomarkers and parameters," *Nutrition Reviews*, vol. 82, no. 2, pp. 143–165, 2024.

[3] S. A. Faran, T. Hussain, S. H. Khalid, I. U. Khan, M. Asif, J. Ahmad, and S. Asghar, "Bile acid/fatty acid integrated nanoemulsomes for nonalcoholic fatty liver targeted lovastatin delivery: stability, in-vitro, ex-vivo, and in-vivo analyses," *Expert Opinion on Drug Delivery*, vol. 21, no. 5, pp. 779–796, 2024.

[4] S. S. Feng, S. J. Wang, L. Guo, P. P. Ma, X. L. Ye, M. L. Pan, and D. F. Ding, "Serum bile acid and unsaturated fatty acid profiles of non-alcoholic fatty liver disease in type 2 diabetic patients," *World Journal of Diabetes*, vol. 15, no. 5, p. 898, 2024.

[5] T. Mouskeftara, O. Deda, G. Papadopoulos, A. Chatzigeorgiou, and H. Gika, "Lipidomic analysis of liver and adipose tissue in a high-fat diet-induced non-alcoholic fatty liver disease mice model reveals alterations in lipid metabolism by weight loss and aerobic exercise," *Molecules*, vol. 29, no. 7, p. 1494, 2024.

[6] T. Aziz, M. K. Niraj, S. Kumar, R. Kumar, H. Parveen, and M. Niraj, "Effectiveness of Omega-3 polyunsaturated fatty acids in non-alcoholic fatty liver disease: a systematic review and meta-analysis," *Cureus*, vol. 16, no. 8, 2024.

[7] C. Zheng, H. Nie, M. Pan, W. Fan, D. Pi, Z. Liang, and Y. Zhang, "Chaihu Shugan powder influences nonalcoholic fatty liver disease in rats in remodeling microRNAome and decreasing fatty acid synthesis," *Journal of Ethnopharmacology*, vol. 318, p. 116967, 2024.

[8] Q. Nie, M. Li, C. Huang, Y. Yuan, Q. Liang, X. Ma, and J. Li, "The clinical efficacy and safety of berberine in the treatment of non-alcoholic fatty liver disease: a meta-analysis and systematic review," *Journal of Translational Medicine*, vol. 22, no. 1, p. 225, 2024.

[9] P. Zhuang, Y. Ao, X. Liu, H. Ye, H. Li, X. Wan, and J. Jiao, "Circulating fatty acids and risk of severe non-alcoholic fatty liver disease in the UK biobank: a prospective cohort of 116223 individuals," *Food & Function*, vol. 15, no. 20, pp. 10527–10538, 2024.

[10] S. Xu, X. Wu, S. Wang, M. Xu, T. Fang, X. Ma, and J. Weng, "TRIM56 protects against nonalcoholic fatty liver disease by promoting the

degradation of fatty acid synthase," *The Journal of Clinical Investigation*, vol. 134, no. 5, 2024.

[11] Tillander, M. Holmer, H. Hagström, S. Petersson, T. B. Brismar, P. Stål, and C. Lindqvist, "Associations between dietary fatty acid and plasma fatty acid composition in non-alcoholic fatty liver disease: secondary analysis from a randomised trial with a hypoenergetic low-carbohydrate high-fat and intermittent fasting diet," *British Journal of Nutrition*, vol. 132, no. 4, pp. 453–465, 2024.

[12] M. Zhou, X. Liu, Y. Wu, Q. Xiang, and R. Yu, "Liver lipidomics analysis revealed the protective mechanism of Zuogui Jiangtang Qinggan formula in type 2 diabetes mellitus with non-alcoholic fatty liver disease," *Journal of Ethnopharmacology*, vol. 329, p. 118160, 2024.

[13] Z. Zhang, Y. Liu, G. Li, X. Chen, M. Lei, Y. Zhou, and W. Wu, "An economically viable stable isotope-enhanced multiple reaction monitoring method for total fatty acid analysis in a mouse model of non-alcoholic fatty liver disease," *Journal of Chromatography A*, vol. 1736, p. 465406, 2024.

[14] D. H. Ahmed and H. L. Fateh, "Impact of flaxseed supplementation on lipid profile and liver enzymes in patients with non-alcoholic fatty liver disease: systematic review and meta-analysis of randomized controlled trials," *Prostaglandins & Other Lipid Mediators*, p. 106838, 2024.

[15] J. Lin, R. Zhang, H. Liu, Y. Zhu, N. Dong, Q. Qu, and J. You, "Multi-omics analysis of the biological mechanism of the pathogenesis of non-alcoholic fatty liver disease," *Frontiers in Microbiology*, vol. 15, p. 1379064, 2024.

[16] C. Yang, J. Wu, L. Yang, Q. Hu, L. Li, Y. Yang, and Q. Zhao, "Altered gut microbial profile accompanied by abnormal short chain fatty acid metabolism exacerbates nonalcoholic fatty liver disease progression," *Scientific Reports*, vol. 14, no. 1, p. 22385, 2024.

[17] X. Zhang, J. Ruan, Y. He, A. Xu, Y. Fang, Q. Zhang, and X. Liu, "Dietary inflammatory index and the risks of non-alcoholic fatty liver disease: a systematic review and meta-analysis," *Frontiers in Nutrition*, vol. 11, p. 1388557, 2024.

[18] K. Mohit, R. Gupta, and B. Kumar, "Contrastive learned self-supervised technique for fatty liver and chronic liver identification," *Biomedical Signal Processing and Control*, vol. 100, p. 106950, 2025.

[19] Z. Dai, Z. Bao, H. Lin, Q. Yang, J. Huang, X. Zhang, and X. Zhou, "Effects of dietary live microbes intake on a newly proposed classification system for steatotic liver disease," *Scientific Reports*, vol. 15, no. 1, p. 5595, 2025.

[20] P. Theerthagiri, "Liver disease classification using histogram-based gradient boosting classification tree with feature selection algorithm," *Biomedical Signal Processing and Control*, vol. 100, p. 107102, 2025.

[21] R. Saranya and R. Jaichandran, "A dense kernel point convolutional neural network for chronic liver disease classification with hybrid chaotic slime mould and giant trevally optimizer," *Biomedical Signal Processing and Control*, vol. 102, p. 107219, 2025.

[22] S. Y. Chiang, Y. W. Wang, P. Y. Su, Y. Y. Chang, H. H. Yen, and R. F. Chang, "PBCS-ConvNeXt: convolutional network-based automatic diagnosis of non-alcoholic fatty liver in abdominal ultrasound images," *Journal of Imaging Informatics in Medicine*, pp. 1–16, 2025.

[23] C. C. Wu, W. C. Yeh, W. D. Hsu, M. M. Islam, P. A. A. Nguyen, T. N. Poly, and Y. C. J. Li, "Prediction of fatty liver disease using machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 23–29, 2019.

[24] A. Yazdinejad, A. Dehghantanha, H. Karimipour, G. Srivastava, and R. M. Parizi, "A robust privacy-preserving federated learning model against model poisoning attacks," *IEEE Transactions on Information Forensics and Security*, 2024.

[25] J. Chen, H. Yan, Z. Liu, M. Zhang, H. Xiong, and S. Yu, "When federated learning meets privacy-preserving computation," *ACM Computing Surveys*, vol. 56, no. 12, pp. 1–36, 2024.

[26] A. Thirumalraj, R. Chandrashekar, and P. K. Balasubramanian, "NMRA-facilitated optimized deep learning framework: a case study on IoT-enabled waste management in smart cities," in *Developments Towards Next Generation Intelligent Systems for Sustainable Development*, IGI Global Scientific Publishing, pp. 247–268, 2024.