

Machine Learning Technology for Evaluation of Liver Fibrosis, Inflammation Activity and Steatosis (LIVERFAST™)

Abhishek Aravind¹, Avinash G. Bahirvani², Ronald Quiambao¹, Teresa Gonzalo^{3*}

¹Fibronostics Pte Ltd., Singapore

²Sprim Pte Ltd., Singapore

³Fibronostics US, Inc., Orlando, FL, USA

Email: *teresa.gonzalo@fibronostics.com

How to cite this paper: Aravind, A., Bahirvani, A.G., Quiambao, R. and Gonzalo, T. (2020) Machine Learning Technology for Evaluation of Liver Fibrosis, Inflammation Activity and Steatosis (LIVERFAST™). *Journal of Intelligent Learning Systems and Applications*, 12, 31-49.

<https://doi.org/10.4236/jilsa.2020.122003>

Received: February 15, 2020

Accepted: April 7, 2020

Published: April 10, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Using the latest available artificial intelligence (AI) technology, an advanced algorithm LIVERFAST™ has been used to evaluate the diagnostic accuracy of machine learning (ML) biomarker algorithms to assess liver damage. Prevalence of NAFLD (Nonalcoholic fatty liver disease) and resulting NASH (non-alcoholic steatohepatitis) are constantly increasing worldwide, creating challenges for screening as the diagnosis for NASH requires invasive liver biopsy. Key issues in NAFLD patients are the differentiation of NASH from simple steatosis and identification of advanced hepatic fibrosis. In this prospective study, the staging of three different lesions of the liver to diagnose fatty liver was analyzed using a proprietary ML algorithm LIVERFAST™ developed with a database of 2862 unique medical assessments of biomarkers, where 1027 assessments were used to train the algorithm and 1835 constituted the validation set. Data of 13,068 patients who underwent the LIVERFAST™ test for evaluation of fatty liver disease were analysed. Data evaluation revealed 11% of the patients exhibited significant fibrosis with fibrosis scores 0.6 - 1.00. Approximately 7% of the population had severe hepatic inflammation. Steatosis was observed in most patients, 63%, whereas severe steatosis S3 was observed in 20%. Using modified SAF (Steatosis, Activity and Fibrosis) scores obtained using the LIVERFAST™ algorithm, NAFLD was detected in 13.41% of the patients ($S_x > 0, A_y < 2, F_z > 0$). Approximately 1.91% ($S_x > 0, A_y = 2, F_z > 0$) of the patients showed NAFLD or NASH scorings while 1.08% had confirmed NASH ($S_x > 0, A_y > 2, F_z = 1 - 2$) and 1.49% had advanced NASH ($S_x > 0, A_y > 2, F_z = 3 - 4$). The modified SAF scoring system generated by LIVERFAST™ provides a simple and convenient evaluation of NAFLD and NASH in a cohort of Southeast Asians. This system may lead to the use of

noninvasive liver tests in extended populations for more accurate diagnosis of liver pathology, prediction of clinical path of individuals at all stages of liver diseases, and provision of an efficient system for therapeutic interventions.

Keywords

Machine Learning (ML), Artificial Intelligence (AI), Neural Networks (NNs), Steatosis, Inflammation Activity, Fibrosis (SAF Score), Nonalcoholic Fatty Liver Disease (NAFLD), Non-Alcoholic Steatohepatitis (NASH)

1. Introduction

Artificial intelligence (AI), particularly deep learning algorithms, is gaining extensive attention for its excellent performance in liver disease-recognition tasks [1] [2] [3] [4]. Deep learning algorithms can automatically make a quantitative assessment of complex serum biomarkers results and achieve an increased accuracy for diagnosis with higher efficiency [1]. AI is widely used and getting increasingly popular in the medical diagnosing of the liver, including radiology, ultrasound, and nuclear medicine [5]. Further, AI can assist clinicians to make more accurate and reproductive liver disease diagnosis and also reduce the physicians' workload.

Global prevalence of nonalcoholic fatty liver disease (NAFLD) is approximately 25% [6], including an estimated 46% of middle-aged Americans [7], establishing this as the most common chronic liver disease. It is expected that 20% - 30% of those with newly detected NAFLD will have already progressed to NASH; and among these, 10% - 20% will progress to cirrhosis and/or hepatocellular carcinoma [8] [9] [10] [11]. Not surprisingly, nonalcoholic steatohepatitis (NASH) is currently the second leading cause of liver disease among those awaiting liver transplantation in the United States [8]. NAFLD is considered a hepatic manifestation of metabolic syndrome components such as obesity, type 2 diabetes (T2DM), dyslipidemia, and insulin resistance. Among patients with T2DM, the prevalence of NAFLD may be as high as 70% and is often associated with cardiovascular disease [12]. Nonetheless, current practice guidelines fail to support routine screening for NAFLD/NASH in patients with T2DM due in part to the difficulty in properly diagnosing the disease in high-risk groups [7].

Percutaneous liver biopsy remains the gold standard for making a precise diagnosis of NAFLD with specification categorization and is necessary to assess the histopathologic criteria essential to making a diagnosis of NASH [13] [14]. Biopsy allows for confirmation of steatosis as well as determining the degree of lobular inflammation, ballooning, and fibrosis. Commonly used scoring systems for evaluating the severity of NAFLD include the NAFLD Activity Score and SAF score [15]. The FLIP (Fatty Liver Inhibition of Progression) Pathology Consortium has expanded the scoring system for steatosis (S), inflammation activity (A), and fibrosis (F) as the SAF Score, based on partially standardized vis-

ual features of microscopic pathology while separately assessing “steatosis”, “activity” (the sum of hepatocyte ballooning and lobular inflammation), and “fibrosis” through semi-quantitative ordinal scales [16]. Used in combination, these three criteria provide a more accurate, more comprehensive, and less subjective description for the diagnosis of NASH than the previously employed NAFLD Activity Score (NAS) scoring system [17]. The use of liver biopsies, however, carries limitations including pain, risk of complications from invasive procedure, inter-observer variation, and applicability of sampling location and tissue volume; which also introduce heterogeneity of the pathologic signs [18] [19] [20] [21].

The enormity of the NAFLD population and the complications associated with their disorders, raise concerns about employing frequent biopsy assessment of individuals, making its use impractical as a screening tool and/or for serial evaluations in metabolically compromised patients [22]. Moreover, numerous clinical practice guidelines including AASLD, EASL-EASD-EASO, APASL, and WHO recommend non-invasive biomarker-based diagnostic modalities to diagnose liver diseases [16] [23] [24] [25] [26].

Accordingly, non-invasive diagnostics have been developed to screen for and monitor liver disease. Because imaging technologies such as ultrasonography, magnetic resonance imaging (MRI), transient elastography (TE), and computed tomography (CT) are expensive, they are generally impractical for most serial evaluations [27]. Besides high cost, the operator dependence, lower sensitivity and range, radiation exposure, and limited availability are some of the limitations of imaging-based diagnosis of liver damage [28].

Considering those complications and limitations, clinicians have turned to serum-based biomarkers and their associated investigational or commercial algorithms for the presence of advanced fibrosis in NAFLD (e.g. NAFLD fibrosis score, FIB-4 index, aspartate aminotransferase [AST] to platelet ratio index [APRI]), serum biomarkers (Enhanced Liver Fibrosis [ELF] panel, Fibrometer, FibroTest, and Hepascore) [16] [23] [24] [25] [26] [29] [30]. Concurrently, non-alcoholic fatty liver disease (NAFLD) now represents the most common cause of abnormal liver blood tests and chronic liver disease in the Western world [31].

Models based on machine learning (ML) algorithms (computer aided diagnosis based on artificial intelligence) have been shown to classify liver disease into distinct categories with ~80% accuracy [32] [33]. Biomarker-based diagnostic methods have been proven to fulfil these requirements for diagnosis [34]. Different stages of steatosis, inflammation, and fibrosis produce characteristic molecular changes (biomarkers) which can be detected in the serum and provide a snapshot of the liver disease stage. Therefore, assessment via algorithmic derivations of the same three components of liver disease assessed by biopsy (steatosis, inflammation activity, and fibrosis) can lead to a provisional diagnosis of NAFLD or NASH.

Using the latest available artificial intelligence (AI) technology, we have improved second generation advanced algorithm LIVERFAST™ to evaluate the di-

agnostic accuracy of ML biomarker algorithms to assess liver damage. Applying the diagnostic modality of FLIP pathology scoring system, the ML algorithm-based LIVERFASTM produced a modified SAF score to evaluate hepatic steatosis, necro-inflammatory activity, and fibrosis.

The updated algorithm has been trained to accommodate new data using AI, adding flexibility and improving from notoriously outdated linear regression models. The new ML algorithm improves accuracy of prediction of SAF score.

In this work there are three different networks, one for each lesion: fibrosis; inflammation activity; and steatosis. The three NNs shared common settings. They each used rectified linear unit as the activation function, *mean absolute error* as the loss function, Glorot uniform as the weight initialization methods, and Adam as the optimization technique. Other settings were determined through a cross-validation grid search (number of hidden layers, number of neurons in each layer, batch size, and Adam parameters).

The aim of this report is two-fold. First, to show utilization of ML to replicate the diagnostic accuracy of LIVERFASTM's latest generation of AI for algorithm using anthropometric and biomarker parameters to determine hepatic steatosis, necro-inflammatory activity, and fibrosis. Second, to show how the use of the existing SAF scoring system and its interpretation guided the creation of a new algorithmic tool, ML-based, providing accurate prediction of the SAF score determinations of NAFLD/NASH with a cost-efficient approach. Here, we have studied real-world data collected from a large number of patients whose blood samples were analysed using ML-based algorithm for the evaluation of fatty liver disease. Our results validate the applicability and accuracy of LIVERFASTM in a Southeast Asian population.

2. Methods

2.1. Healthcare Data

The data used in this research were obtained from the patients who were registered in ongoing clinical trials or through a diagnostic test for liver diseases recommended by treating physician during clinic visits. Data from all participating networks, provided by family physicians and other primary care providers were aggregated into a single database. An abstract overview of the dataset is given in **Table 1**. The database contains records of 13,068 patients from a three-year period ranging from 2016 to 2018, and every record includes various attributes regarding vital signs, diagnosis and demographics.

Patient data were collected from 16 sites across Asia including Hong Kong, Malaysia, Philippines, Singapore, Thailand and the United Arab Emirates. All participants provided written informed consent for the use of their data for research and analysis prior to blood sample collection.

2.2. LIVERFASTM Algorithm

The study aim was to facilitate physicians' and other healthcare professionals'

evaluation of liver disease, fibrosis, inflammation activity and steatosis in extended populations. This system may lead to the use of the ML algorithm LIVERFAST™ as a predictor of the clinical path of individuals at all stages of liver diseases, and provision of an efficient system for therapeutic interventions.

The ML-based diagnostic tool LIVERFAST™ was developed to diagnose liver damage that utilizes a combination of anthropometric and serum biomarkers to generate a report for healthcare professionals use. For an overview of the machine learning proposed workflow see **Figure 1**.

Table 1. Characteristics of the population in South East Asia. SD, standard deviation; BMI, body mass index.

Country	Number of Patients (%)	
	Number	%
Hong Kong	205	1.57
Malaysia	3055	23.38
Philippines	2435	18.63
Singapore	2884	22.07
Thailand	4481	34.29
United Arab Emirates	8	0.06
Total number of subjects	13068	

Demographics	Sample Size	
	%	Mean ± SD
Male	59.90	
Female	40.10	
Male age		49.71 ± 13.21
Female age		52.37 ± 13.29
BMI (kg/m ²)		27.22 ± 8.16

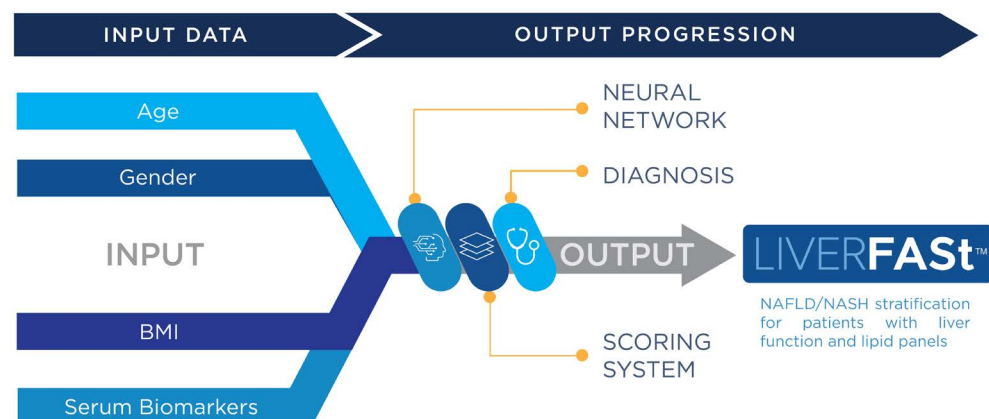


Figure 1. Machine Learning LIVERFAST™ algorithm: serum biomarkers, age, gender, and BMI as an input; and application of Neural Networks for final evaluation and liver disease scoring.

LIVERFAST™ is an artificial intelligence-based algorithm technology that uses a set of NNs, combined with a scaling and mathematical operation of input data, to generate continuous scores for three liver lesions. The required ML-based algorithm platform serum biomarkers are alpha-2-macroglobulin (a2M), apolipoprotein-A1 (ApoA1), haptoglobin, total bilirubin, gamma-glutamyl transferase (GGT), aspartate aminotransferase (AST), alanine aminotransferase (ALT), fasting cholesterol (total), fasting triglycerides and fasting glucose. Those individual serum biomarkers have been identified as appropriate biomarkers for liver disease evaluation [35]. In addition, patient anthropometric characteristics were included: age, gender, and height and weight for calculated BMI.

These features are applied in the ML algorithm meant to assess existence and the degree of severity of three liver lesions associated with NAFLD and NASH: fibrosis, inflammation, and steatosis as seen in **Table 2**.

The ML in LIVERFAST™ technology is comprised of two parts: 1) biomarker digital assays for three non-invasive diagnostic tests; and 2) software containing a proprietary algorithm to generate the ML-based biomarker digital assay scores from the serum biochemical markers, adjusted for patient demographics.

The serum biomarker assays are inputted into the LIVERFAST™ cloud-based physician portal in order to calculate the SAF scores using the ML technology. A liver evaluation based in SAF staging report is generated with all three non-invasive test scores, for the healthcare provider to use:

1) Fibrosis score to detect the degree of fibrosis. The result is provided as a score from 0 to 1, proportional to the severity of the fibrosis, with a conversion

Table 2. Features used in LIVERFAST™ algorithm to identify liver disease.

Feature	Liver Lesion			
	Fibrosis	Inflammation Activity	Steatosis	NASH
Age	x	x	x	x
Gender	x	x	x	x
BMI			x	x
Alpha-2-Macroglobulin	x	x	x	x
Haptoglobin	x	x	x	x
Apolipoprotein A1	x	x	x	x
Total Bilirubin	x	x	x	x
GGT	x	x	x	x
ALT (P5P)		x	x	x
AST (P5P)			x	x
Total Cholesterol			x	x
Triglyceride			x	x
Fasting Glucose			x	x

to the SAF scoring system (from F0 to F4). The five stages of histological scoring system are: F0 (no fibrosis), F1 (minimal fibrosis), F2 (moderate fibrosis), F3 (significant fibrosis), and F4 (severe fibrosis/cirrhosis).

2) Activity score to detect the degree of ballooning and lobular inflammation. The result is provided as a score of 0 to 1, proportional to the significance of the activity, with a conversion to the SAF scoring system (from A0 to A4). The five stages of histological scoring system are: A0 (no activity), A1 (minimal activity), A2 (moderate activity), A3 (significant activity), and A4 (severe activity).

3) Steatosis score to detect the degree of steatosis. The result is provided as a score from 0 to 1, proportional to the severity of steatosis, with a conversion to the SAF scoring system (from S0 to S3). The four stages of histological scoring system are: S0 (no steatosis), S1 (minimal steatosis), S2 (moderate steatosis), and S3 (severe steatosis)

2.3. LIVERFAST™ Algorithm Training of ML Technology with Neural Networks (NNs)

For further evaluation of the ML LIVERFAST™ algorithm, patients were enrolled into the database consistent with the treating physician's clinical suspicion of fatty liver disease. Their determination was based on presentation and risk factors for disease, plus the precondition of having had a complete record set already entered into the database. This research protocol received ethics approval from the research ethics board of IntegReview IRB to access de-identified medical records.

Data collected between January and November of 2016 were used to create a database consisting of 2862 medical assessments. From the database, 1027 assessments were used to develop the algorithm and subsequently 1835 assessments were applied for validation. The first round of algorithm training used 1027 of the 2862 medical records selected at random from throughout the full database. Each medical assessment included the thirteen combined anthropometric and blood biomarker features listed in Section 2.2.

Before training the NNs, standardization (a way to normalize each feature by removing its mean and dividing the result by its standard deviation) is invoked. It is particularly useful here since NNs rely on the gradient descent algorithm which converges much faster with scaled features.

In this work there are three different networks, one for each lesion: fibrosis; inflammation activity; and steatosis. The three NNs shared common settings. They each used rectified linear unit as the activation function, *mean absolute error* as the loss function, Glorot uniform as the weight initialization methods, and Adam as the optimization technique. Other settings were determined through a cross-validation grid search (number of hidden layers, number of neurons in each layer, batch size, and Adam parameters). The resulting network structure is shown in **Figure 2**.

The output dimensions applied to AI algorithm LIVERFAST were as follows:

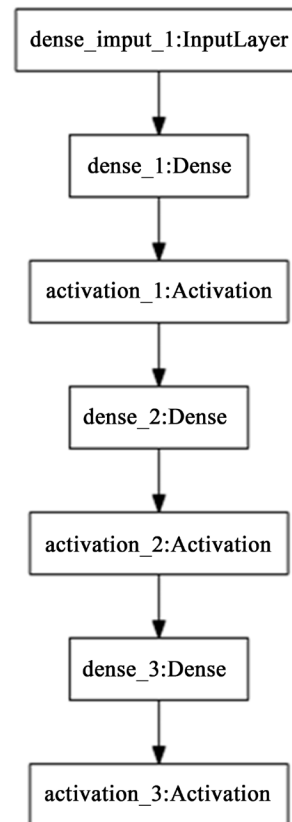


Figure 2. LIVERFAST™ Neural Network structure.

For Steatosis:

Dense_1_input: 12

Dense_1: 5

Dense_2: 45

Dense_3: 1

For Activity:

Dense_1_input: 8

Dense_1: 14

Dense_2: 12

Dense_3: 1

For Fibrosis:

Dense_1_input: 7

Dense_1: 12

Dense_2: 15

Dense_3: 1

In this study, three different NNs were trained, one for each LIVERFAST™ test. As described by **Table 2**, each NN includes a different set of features as input, which assess existence and the degree of severity of several liver pathologies, one for each of the lesions—fibrosis, activity and steatosis. Standardization of each feature must follow a normalization process and it is particularly useful here since NNs rely on the gradient descent algorithm which converges much

faster with scaled features.

Distribution of fibrosis, inflammation activity, or steatosis scores of the training and validation dataset are depicted in **Figure 3**. In this ML algorithm, the features are the independent variables, *i.e.*, biomarkers, age, gender and BMI. The outcome is the dependent variable for LIVERFAST fibrosis, inflammation activity, or steatosis score.

Supervised learning. Supervised learning is the primary modality of this work. The goal of supervised learning is to predict an output given a set of inputs. The model learns how to “make a decision” from examples that are known. A good supervised learning model is able to make acceptable predictions on new examples that were not involved in the original learning process. This ML strategy sought to predict numerical variables equivalent to the fibrosis, inflammation activity and steatosis according to the SAF scores and hence assign liver diagnosis and stage. The LIVERFAST™ ML algorithm creates its own versions of S, A, and F scores using patient age, gender, BMI and up to ten blood biomarkers. Subsequently, by retrieving the three separate ML-based test scores, the algorithm can approximate a patient’s biomarker derived (LIVERFAST™) SAF score— $S_xA_yF_z$ —and utilize that determination to assess the probable outcome of the FLIP Scoring System [17] for this combination (see **Table 3**).

Table 4 displays the performance of ML-based algorithm with benchmark comparative tests. The maximum absolute error (MaxAE) and the mean absolute error (MAE) loss function was used to evaluate and tune the performance of the neural networks using machine learning.

Patient segmentation.

The association of an individual to be included in the machine learning

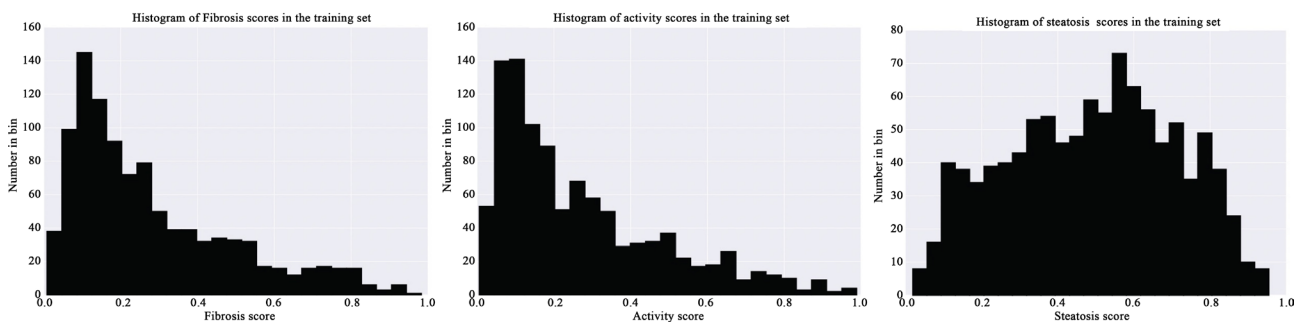


Figure 3. Histograms of Fibrosis, Inflammation Activity & Steatosis scores in the training and validation dataset (left to right).

Table 3. SAF score/diagnosis mapping: X, Y, and Z are integers as [0 - 3], [0 - 4] and [0 - 4] respectively [17].

Estimated SAF score	Diagnosis
$S_0A_yF_z$	No NAFLD
$S_{x>0}A_{y<2}F_z$	NAFLD
$S_{x>0}A_2F_z$	NAFLD or NASH
$S_{x>0}A_{y>2}F_z$	NASH

Table 4. Prediction of fibrosis score, activity score and steatosis score with the NN models using MAE (mean absolute error), MaxAE (maximum absolute error), R^2 (coefficient of determination) and CI (confidence interval).

Benchmark comparative test	Scoring Liver Lesion		
	Fibrosis score	Inflammation Activity Score	Steatosis Score
Mean Absolute Error	1.3E-3	3.4E-3	1.1E-3
Mean Absolute Error	3.2E-2	5.2E-2	2.4E-2
R^2	0.99992	0.99952	0.99991
CI (95%)	[1.2E-3, 1.4E-3]	[3.2E-3, 3.6E-3]	[1.0E-3, 1.2E-3]

algorithm categorization can then be evaluated using the below guidelines depicted in the following filters (1), (2), (3), (4) and (5):

$$\begin{aligned} &\text{Age, } 14 \leq \text{years} \leq 100; \\ &\text{Height, } 1.47 \leq \text{meters} \leq 2.0; \\ &\text{Weight, } 44 \leq \text{kg} \leq 122. \end{aligned} \quad (1)$$

$$\begin{aligned} &\text{ALT, } 1 \leq \text{IU/L} \leq 622; \\ &\text{AST, } 1 \leq \text{IU/L} \leq 1273; \\ &\text{Gamma GT, } 1 \leq \text{IU/L} \leq 2351. \end{aligned} \quad (2)$$

$$\begin{aligned} &\text{Alpha2 Macroglobulin, } 0.8 \leq \text{g/L} \leq 5.9; \\ &\text{Haptoglobin, } 0.08 \leq \text{g/L} \leq 3.2; \\ &\text{Apolipoprotein A1, } 0.56 \leq \text{g/L} \leq 2.5 \end{aligned} \quad (3)$$

$$\begin{aligned} &\text{Total cholesterol, } 2.26 \leq \text{mmol/L} \leq 8.43; \\ &\text{Triglycerides, } 0.38 \leq \text{mmol/L} \leq 7.35 \end{aligned} \quad (4)$$

$$\begin{aligned} &\text{Bilirubin, } 1 \leq \mu\text{mol/L} \leq 613; \\ &\text{Fasting glucose, } 3.04 \leq \text{mmol/L} \leq 13 \end{aligned} \quad (5)$$

For the evaluation of the Southeast Asian population, the score-stage conversions applied for LIVERFASt™ algorithm technology were the following:

Fibrosis score (x):

$$\begin{aligned} &\text{F0: } x \leq 0.27; \text{ F1: } 0.27 \leq x \leq 0.48; \text{ F2: } 0.48 \leq x \leq 0.58; \\ &\text{F3: } 0.58 \leq x \leq 0.74; \text{ and F4: } 0.74 \leq x. \end{aligned}$$

Inflammation Activity score (y):

$$\text{A0: } y \leq 0.29; \text{ A1: } 0.29 \leq y \leq 0.52; \text{ A2: } 0.52 \leq y \leq 0.62; \text{ and A3: } 0.62 \leq y.$$

Steatosis score (z):

$$\text{S0: } z \leq 0.37; \text{ S1: } 0.37 \leq z \leq 0.56; \text{ S2: } 0.56 \leq z \leq 0.6; \text{ and S3: } 0.68 \leq z.$$

Statistical Methods. In this study, two methods were used to evaluate the performance of the LIVERFASt™ platform for each of the three liver lesions: dichotomous and ordinal evaluation. According to the dichotomous evaluation method, the performance of the LIVERFASt platform was evaluated in its ability to classify between positive and negative disease states. For each of the 3 lesions,

this will be: $x(0)$ vs $x(>0)$ where “x”, “y” and “z” is each of the lesions of Fibrosis(F), Activity(A) and Steatosis(S), respectively. Using this dichotomy, we evaluated the LIVERFASTM technology on its specificity (Sp), sensitivity (Se), positive predictive value (PPV) and negative predictive value (NPV).

Following ordinal evaluation method [36], the evaluation of the performance of the LIVERFASTM platform in staging the condition for each of the three lesions, was determined as follows: Mean Squared Error (MSE) that provides a metric that penalizes larger errors; and MAE that provides a metric that reduces overall errors, and Adjusted MSE. The metrics were adjusted to penalize errors where the predicted stage is lower than the actual stage. This is done to account for the non-confirmatory indication of use of the LIVERFASTM platform for the conditions of NAFLD or NASH (penalizing errors that provide negative diagnosis for positive state patients more than positive diagnosis for negative state patients). The thresholds used were $MAE < 0.31^*$ and $MSE < 0.45^*$ (*referred to 20% 1 stage error, 4% 2 stage error and 1% 3 stage error).

3. Results

A multi-country clinical dataset of 13,068 individuals over a period of 3 years is incorporated in this study. The degree of distribution of each country and demographics figures are given in **Table 1**, as well as gender specifics and average BMI.

For the evaluation of liver disease, fibrosis, inflammation activity and steatosis in this extended clinical dataset, a ML LIVERFASTM algorithm tool (see **Figure 1**) was applied based in the features depicted in **Table 2**, a2M, ApoA1, haptoglobin, total bilirubin, GGT, AST, ALT, total cholesterol, triglycerides and fasting glucose, as well as gender, age and BMI. For further evaluation of the AI-based algorithm, a database consisting of 2862 unique medical assessments of biomarkers and biopsy reports were trained and validated.

The training dataset ($n = 1027$) consisted of 60% males, while the validation dataset ($n = 1835$) had a greater proportion of males (77%). The training set was also slightly older (~51 years to 45 years old), and with a slightly lower mean BMI. The medical records contained liver staging scores generated by the AP-HP set of algorithms [37]. These were used as the benchmark against which the neural network outputs were evaluated. Per the histograms in **Figure 3**, most of the patients included in the training set were in early stages of fibrosis and inflammatory activity at the time of assessment. Among the patients with steatosis, the repartition of scores is more balanced throughout the full scoring range. Training and validation datasets displayed roughly similar but not identical distributions in terms of features as well as in terms of outcome in scoring test.

NASH is diagnosed based on an overall assessment by a pathologist using scoring systems such as the steatosis, activity, and fibrosis (SAF) score, which evaluates of the presence and extent of each individual component of steatosis, inflammation, and ballooning. NASH can be diagnosed using a validated algo-

rithm based on the SAF scoring system [17]. The correlation between the estimated SAF score and the NAFLD/NASH diagnosis that LIVERFASTM provides is shown in **Table 3**.

As seen in **Table 4**, the MAE function evaluated the machine learning performance of the NNs the using machine learning as well as the MaxAE of the AI-based algorithm test compared to benchmark comparative tests. For the three NNs of the LIVERFASTM ML algorithm, the order of magnitude of the MAE was $1E-3$, compared to benchmark algorithm that obtained a precision of $1E-2$ [37]. Aside from this, cross-validated grid search was used to optimize the hyperparameters which played a role in the performance of the resulting neural network. The R^2 values for all three networks are very close to 1, with 0.99992 for fibrosis, 0.99952 for activity and 0.99991 for steatosis which shows a very high level of linear correlation between the scores generated by the neural networks and those generated by the benchmark comparative tests.

3.1. Determination of Fibrosis, Inflammation and Steatosis in the Liver

Based on the patient's age, gender, a2M, ApoA1, haptoglobin, bilirubin, ALT and GGT levels LIVERFASTM generated fibrosis score of the patients. As shown in **Figure 3**, data evaluation revealed 11% of the patients exhibited significant fibrosis with fibrosis scores 0.6 - 1.00 (**Figure 4(a)**) while most of the patients (59.17%) did not show fibrosis.

The ML-based algorithm generated the inflammation score using the total cholesterol value added to the features utilized for fibrosis determination at the scale of 0.00 - 1.00. Approximately 7% of the population had A4 stage severe hepatic inflammation (**Figure 4(b)**). A majority of the patients (76.07%) did not have elevated levels of biomarkers for hepatic inflammation.

The ML-based algorithm also determined degree of hepatic steatosis by utilizing biomarkers utilized for inflammation activity determination adding AST, fasting glucose, triglycerides and BMI, on a scale of 0 to 1.0. Steatosis was observed in most patients (63%) whereas severe steatosis S3 was observed in 20% (**Figure 4(c)**).

3.2. Assessment of NAFLD and NASH Based on LIVERFASTM Algorithm

Based on ML-based algorithm, patients were evaluated for NAFLD and NASH using modified SAF scores indicating the degree of steatosis, inflammatory activity, and fibrosis in the liver of the subjects. LIVERFASTM classified patients:

- NAFLD only when steatosis score equated 0; inflammation activity score was less than 2; and the fibrosis score equated 0.
- NAFLD or initial NASH was predicted when steatosis score equated 0; inflammation activity score equated 2; and fibrosis score equated 0.
- Moderate NASH was predicted when steatosis score was more than 0, inflammation activity score was more than 2 and fibrosis score was either 1 or 2.

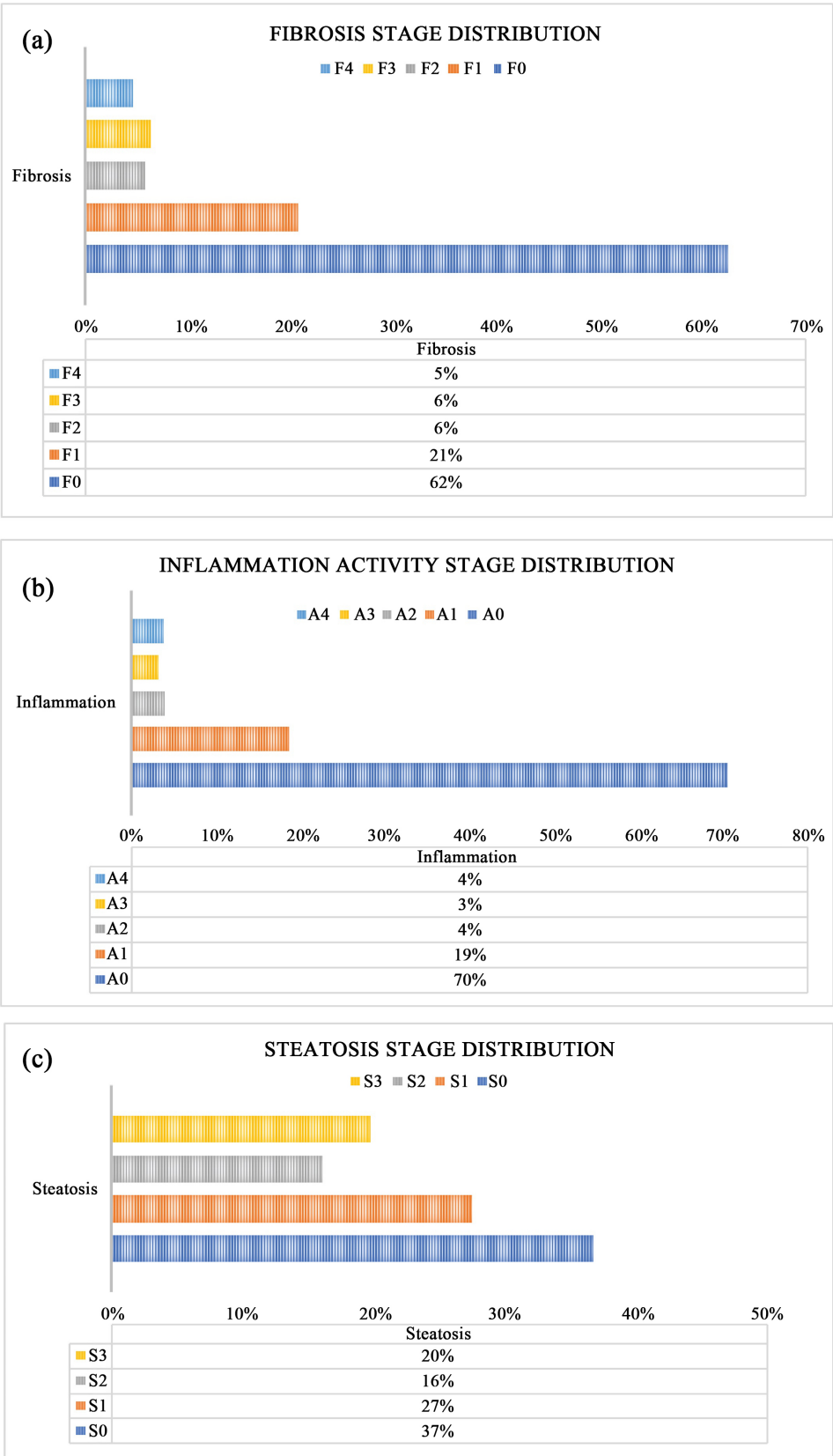


Figure 4. Machine learning LIVERFAST™ algorithm applied to South East Asian population. (a) Fibrosis; (b) Inflammation activity; (c) Steatosis.

- Advanced NASH was predicted when steatosis score was more than 0, inflammation activity score was more than 2 and fibrosis score was 3 or 4.

As shown in **Table 5**, using modified SAF (scores obtained using machine learning LIVERFAST™ algorithm, NAFLD was detected in 13.41% of the patients ($Sx > 0, Ay < 2, Fz > 0$). Approximately 1.91% ($Sx > 0, Ay = 2, Fz > 0$) of the patients showed NAFLD or NASH scorings while 1.08% had confirmed NASH ($Sx > 0, Ay > 2, Fz = 1 - 2$) and 1.49% had advanced NASH ($Sx > 0, Ay > 2, Fz = 3 - 4$).

4. Discussion

As discussed previously, NAFLD may progress to NASH and subsequently cirrhosis and HCC. In addition, NASH is associated with increased mortality compared with the general population [38]. If diagnosed early, weight loss and lifestyle modification may improve liver histology and prevent further damage [38]. Thus, it is essential to diagnose NAFLD and NASH to prevent progression to cirrhosis or HCC. Biopsy examination has been the gold standard for evaluation of liver health and disease diagnosis and remains the primary standard to which all other methods for liver evaluation are compared (e.g. multiple imaging techniques; serology).

As the prevalence of NAFLD increases worldwide, a critical need has arisen for reliable tools that are non-invasive, safe, quick, inexpensive, and suitable to evaluate patients with metabolic complications or provide sensible review of treatment progress for individual patients or intervention trials [39]. Liver biopsy, though compelling, carries an element of risk and is subject to varying interpretations and inter-observer discordance [20]. Due to the limitations of these conventional methods, the necessity for a non-invasive, highly sensitive, specific, easy and readily available and cost-effective method to diagnose fatty liver disease is warranted [25].

Table 5. Evaluation of NAFLD and NASH based on LIVERFAST™ algorithm.

Diagnosis	Scoring liver lesion according to LIVERFAST™ algorithm		
	Diagnostic criteria	Description of diagnosis	% Patients
NAFLD only	$Sx > 0, Ay < 2, Fz > 0$	Patients with NAFLD	13.41%
NAFLD or Initial NASH	$Sx > 0, Ay = 2, Fz > 0$	Patients with inflammation activity $Ay = 2$, possible diagnosis of either NAFLD or NASH	1.91%
Moderate NASH	$Sx > 0, Ay > 2, Fz = 1 - 2$	Patients scoring activity $Ay > 2$, indicated NASH condition, not including patients with advanced NASH	1.08%
Advanced NASH	$Sx > 0, Ay > 2, Fz = 3 - 4$	Patients with advanced NASH, <i>i.e.</i> NASH with advanced Fibrosis $Fz = 3 - 4$	1.49%

In this study, we analysed the real-world data of biomarker-based diagnosis of non-alcoholic fatty liver disease in 13,068 subjects from Southeast Asia. The evaluation was made by AI biomarker-based algorithm LIVERFASTTM which was developed using ML second generation logarithms.

We have taken quantitative scores defined by the FLIP consortium where NASH is diagnosed based on an overall assessment by a pathologist using validated algorithm scoring systems such as the steatosis, activity, and fibrosis (SAF) score [17]. The correlation between the estimated SAF score and the NAFLD/NASH diagnosis that LIVERFASTTM is using is shown in **Table 3**. Approximately 13.41% of the patient population was diagnosed with NAFLD or NASH.

Determination of disease severity is a challenging element of the diagnostic workup of patients with NAFLD. The goal here is to identify patients with more advanced disease at increased risk for advancing further to irreversible disease with associated morbidity and mortality. From the population analysed, 1.08% had confirmed NASH ($S_x > 0$, $A_y > 2$, $F_z = 1 - 2$) and 1.49% had advanced NASH.

A challenge with NNs is that they have a very high number of hyperparameters to tune. Therefore, finding the optimal set of parameters for a given problem can be long and cumbersome. In this work there are three different problems, one for each lesion: fibrosis; inflammation activity; and steatosis. The three NNs shared common settings. They each used rectified linear unit as the activation function, mean absolute error as the loss function, Glorot uniform as the weight initialization methods, and Adam as the optimization technique. Other settings were determined through a cross-validation grid search with number of hidden layers, number of neurons in each layer, batch size, and Adam parameters. Here, the tool used for this process was Keras, The Python Deep Learning library [40].

Standard metrics used to describe the accuracy of regression models have been computed to assess the performance of the new models. MAE, MaxAE and coefficient of determination (R^2) are computed over validation sets of anthropometric and laboratory assessments that were not used during the training phase. The values (**Table 4**) for these three metrics are evidence of the ability of the new models to generalize readily to unknown data, thereby validating their high predictive power over the most significant range of medical relevance. Having developed the tools to reach this milestone, it is now possible for researchers to provide significant improvements to create the second-generation assessment tools that the clinical community seeks for even more useful non-invasive liver diagnostics. One limitation of the study is the discrepancies in gender, age and to a lesser extent, BMI between the training data set of medical assessments and the validation dataset. The algorithm is being upgraded as more patient medical assessments are applied to the current model, additional clinical studies are in progress to that end.

Furthermore, replacing or adding biomarkers to those described here may provide solutions that improve upon the first-generation accuracy. With the use

of advanced ML techniques, investigators may also target the removal of specific, less significant biomarkers used in the first generation. Thus, while improving sensitivity and specificity, the new diagnostic tools might also increase cost-effectiveness (assuming equal or better accuracy and/or the cumulated expense of deleted biomarkers is greater than the expense of any new biomarkers). A thorough analysis for the selection of innovative and more accurate biometrics would be an obvious first step for such work. Ideally the pursuit of multiple putative modifications would include the realization of additional clinical studies which would directly provide pathology assessments from high quality biopsy tissue and an abundance of potential biomarkers.

The most immediate utilization of the algorithmic assessments described here is the application of these tools by any licensed physician or health care professional. At the same time, the primary focus and ultimate goal of this work is the development of advanced algorithms that will more accurately diagnose whole liver pathology; predict the clinical path of individuals at all stages of the NAFLD spectrum; and provide an efficient and improved system with which to examine new and critically needed therapeutic interventions that mitigate or reverse NAFLD progression.

5. Conclusions

The modified SAF scoring system generated by LIVERFASTTM provided a simple and convenient diagnosis of NAFLD and NASH and staging of the three liver lesions as shown in a cohort of South East Asia.

The use of noninvasive liver tests in extended populations provides an accurate diagnosis of liver pathology, prediction of clinical path of individuals at all stages of liver diseases, and an efficient system for therapeutic interventions. Non-invasive diagnostic tools such as LIVERFASTTM are easy to perform, less expensive, and readily available and aid to the early diagnosis and better prognosis in patients with NAFLD and NASH.

In accord with the 2016 EASL-EASO-EADO Guidelines, use of noninvasive serum biomarkers should aim to: 1) in primary care settings, identify the risk of NAFLD among individuals with increased metabolic risk; 2) in secondary and tertiary care settings, identify those with worse prognosis, e.g. severe NASH; 3) monitor disease progression; 4) predict response to therapeutic interventions. Achieving these objectives could reduce the need for liver biopsy.

The applicability of LIVERFASTTM extends beyond hepatologists and includes primary care providers, as well as endocrinologists, diabetologists and other medical disciplines that manage and monitor fatty liver, liver fibrosis and liver activity. LIVERFASTTM test has a potential role worldwide in the clinical care settings as screening for NAFLD and NASH population at risk.

Acknowledgements

The authors wish to thank Imtiaz Alam, MD that provided significant back-

ground on the liver disorders addressed by this algorithmic and contributed to revision of the final version of the manuscript. Nelly Conus, PhD contributed to revision of the final version of the manuscript. Hawley K. Linke, PhD provided inputs for important intellectual content for results interpretation and contributed to revision of the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wei, W., Wu, X., Zhou, J., Sun, Y., Kong, Y. and Yang, X. (2019) Noninvasive Evaluation of Liver Fibrosis Reverse Using Artificial Neural Network Model for Chronic Hepatitis B Patients. *Computational and Mathematical Methods in Medicine*, **2019**, 7239780-7239788. <https://doi.org/10.1155/2019/7239780>
- [2] Chang, N.-W., *et al.* (2017) Biomarker Identification of Hepatocellular Carcinoma Using a Methodical Literature Mining Strategy. *Database (Oxford)*, **2017**, bax082. <https://doi.org/10.1093/database/bax082>
- [3] Li, B., *et al.* (2017) Artificial Neural Network Models for Early Diagnosis of Hepatocellular Carcinoma Using Serum Levels of α -Fetoprotein, α -Fetoprotein-L3, Des- γ -Carboxy prothrombin and Golgi Protein 73. *Oncotarget*, **8**, 80521-80530. <https://doi.org/10.18632/oncotarget.19298>
- [4] Choi, K.J., *et al.* (2018) Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent-Enhanced CT Images in the Liver. *Radiology*, **289**, 688-697. <https://doi.org/10.1148/radiol.2018180763>
- [5] Huang, Q., Zhang, F. and Li, X. (2018) Machine Learning in Ultrasound Computer-Aided Diagnostic Systems: A Survey. *BioMed Research International*, **2018**, Article ID: 5137904. <https://www.hindawi.com/journals/bmri/2018/5137904/>
<https://doi.org/10.1155/2018/5137904>
- [6] Younossi, Z.M., Koenig, A.B., Abdelatif, D., Fazel, Y., Henry, L. and Wymer, M. (2016) Global Epidemiology of Nonalcoholic Fatty Liver Disease-Meta-Analytic Assessment of Prevalence, Incidence and Outcomes. *Hepatology*, **64**, 73-84. <https://doi.org/10.1002/hep.28431>
- [7] Chalasani, N., *et al.* (2012) The Diagnosis and Management of Non-Alcoholic Fatty Liver Disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology and the American Gastroenterological Association. *Hepatology*, **55**, 2005-2023. <https://doi.org/10.1002/hep.25762>
- [8] Wong, R.J., *et al.* (2015) Nonalcoholic Steatohepatitis Is the Second Leading Etiology of Liver Disease among Adults Awaiting Liver Transplantation in the United States. *Gastroenterology*, **148**, 547-555. <https://doi.org/10.1053/j.gastro.2014.11.039>
- [9] Ascha, M.S., Hanouneh, I.A., Lopez, R., Tamimi, T.A.-R., Feldstein, A.F. and Zein, N.N. (2010) The Incidence and Risk Factors of Hepatocellular Carcinoma in Patients with Nonalcoholic Steatohepatitis. *Hepatology*, **51**, 1972-1978. <https://doi.org/10.1002/hep.23527>
- [10] Bhala, N., *et al.* (2011) The Natural History of Nonalcoholic Fatty Liver Disease with Advanced Fibrosis or Cirrhosis: An International Collaborative Study. *Hepatology*, **54**, 1208-1216. <https://doi.org/10.1002/hep.24491>

- [11] Wong, V.W.-S., *et al.* (2010) Disease Progression of Non-Alcoholic Fatty Liver Disease: A Prospective Study with Paired Liver Biopsies at 3 Years. *Gut*, **59**, 969-974. <https://doi.org/10.1136/gut.2009.205088>
- [12] Targher, G., *et al.* (2007) Prevalence of Nonalcoholic Fatty Liver Disease and Its Association with Cardiovascular Disease among Type 2 Diabetic Patients. *Diabetes Care*, **30**, 1212-1218. <https://doi.org/10.2337/dc06-2247>
- [13] Chalasani, N., *et al.* (2018) The Diagnosis and Management of Nonalcoholic Fatty Liver Disease: Practice Guidance from the American Association for the Study of Liver Diseases. *Hepatology*, **67**, 328-357. <https://doi.org/10.1002/hep.29367>
- [14] European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD) and European Association for the Study of Obesity (EASO) (2016) EASL-EASD-EASO Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease. *Journal of Hepatology*, **64**, 1388-1402. <https://doi.org/10.1016/j.jhep.2015.11.004>
- [15] Kleiner, D. E., *et al.* (2005) Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease. *Hepatology*, **41**, 1313-1321. <https://doi.org/10.1002/hep.20701>
- [16] Shiha, G., *et al.* (2009) Liver Fibrosis: Consensus Recommendations of the Asian Pacific Association for the Study of the Liver (APASL). *Hepatology International*, **3**, 323-333. <https://doi.org/10.1007/s12072-008-9114-x>
- [17] Bedossa, P., *et al.* (2012) Histopathological Algorithm and Scoring System for Evaluation of Liver Lesions in Morbidly Obese Patients. *Hepatology*, **56**, 1751-1759. <https://doi.org/10.1002/hep.25889>
- [18] Sumida, Y., Nakajima, A. and Itoh, Y. (2014) Limitations of Liver Biopsy and Non-Invasive Diagnostic Tests for the Diagnosis of Nonalcoholic Fatty Liver Disease/Nonalcoholic Steatohepatitis. *World Journal of Gastroenterology*, **20**, 475-485. <https://doi.org/10.3748/wjg.v20.i2.475>
- [19] Janiec, D.J., Jacobson, E.R., Freeth, A., Spaulding, L. and Blaszyk, H. (2005) Histologic Variation of Grade and Stage of Non-Alcoholic Fatty Liver Disease in Liver Biopsies. *Obesity Surgery*, **15**, 497-501. <https://doi.org/10.1381/0960892053723268>
- [20] Ratziu, V., *et al.* (2005) Sampling Variability of Liver Biopsy in Nonalcoholic Fatty Liver Disease. *Gastroenterology*, **128**, 1898-1906. <https://doi.org/10.1053/j.gastro.2005.03.084>
- [21] Rousset, M.-C., *et al.* (2005) Sources of Variability in Histological Scoring of Chronic Viral Hepatitis. *Hepatology*, **41**, 257-264. <https://doi.org/10.1002/hep.20535>
- [22] Pandeyarajan, V., Gish, R.G., Alkhouri, N. and Noureddin, M. (2019) Screening for Nonalcoholic Fatty Liver Disease in the Primary Care Clinic. *Gastroenterology & Hepatology*, **15**, 357-365.
- [23] WHO (2014) Guidelines for the Screening, Care and Treatment of Persons with Hepatitis C Infection.
- [24] Sarin, S.K., *et al.* (2016) Asian-Pacific Clinical Practice Guidelines on the Management of Hepatitis B: A 2015 Update. *Hepatology International*, **10**, 1-98. <https://doi.org/10.1007/s12072-015-9675-4>
- [25] Afdhal, N., Bedossa, P., Friedrich-Rust, M., Han, K.-H. and Pinzani, M. (2015) EASL-ALEH Clinical Practice Guidelines: Non-Invasive Tests for Evaluation of Liver Disease Severity and Prognosis. *Journal of Hepatology*, **63**, 237-264. <https://doi.org/10.1016/j.jhep.2015.04.006>

- [26] AASLD and IDSA (2014) Recommendations for Testing, Managing and Treating Hepatitis C. *Aasld*, 1-51.
- [27] Schwenzer, N.F., Springer, F., Schraml, C., Stefan, N., Machann, J. and Schick, F. (2009) Non-Invasive Assessment and Quantification of Liver Steatosis by Ultrasound, Computed Tomography and Magnetic Resonance. *Journal of Hepatology*, **51**, 433-445. <https://doi.org/10.1016/j.jhep.2009.05.023>
- [28] Calès, P., et al. (2008) Reproducibility of Blood Tests of Liver Fibrosis in Clinical Practice. *Clinical Biochemistry*, **41**, 10-18. <https://doi.org/10.1016/j.clinbiochem.2007.08.009>
- [29] Angulo, P., et al. (2007) The NAFLD Fibrosis Score: A Noninvasive System that Identifies Liver Fibrosis in Patients with NAFLD. *Hepatology*, **45**, 846-854. <https://doi.org/10.1002/hep.21496>
- [30] Vallet-Pichard, A., et al. (2007) FIB-4: An Inexpensive and Accurate Marker of Fibrosis in HCV Infection Comparison with Liver Biopsy and Fibrotest. *Hepatology*, **46**, 32-36. <https://doi.org/10.1002/hep.21669>
- [31] Clark, J.M. (2006) The Epidemiology of Nonalcoholic Fatty Liver Disease in Adults. *Journal of Clinical Gastroenterology*, **40**, S5-S10.
- [32] Fatima, M. and Pasha, M. (2017) Survey of Machine Learning Algorithms for Disease Diagnostic. *Journal of Intelligent Learning Systems and Applications*, **9**, 1-16. <https://doi.org/10.4236/jilsa.2017.91001>
- [33] Vijayarani, D.S. and Dhayanand, M.S. (2020) Liver Disease Prediction Using SVM and Naïve Bayes Algorithms.
- [34] Hadizadeh, F., Faghihimani, E. and Adibi, P. (2017) Nonalcoholic Fatty Liver Disease: Diagnostic Biomarkers. *World Journal of Gastrointestinal Pathophysiology*, **8**, 11-26. <https://doi.org/10.4291/wjgp.v8.i2.11>
- [35] Neuman, M.G., Cohen, L.B. and Nanau, R.M. (2014) Biomarkers in Nonalcoholic Fatty Liver Disease. *Canadian Journal of Gastroenterology and Hepatology*, **28**, 607-618. <https://doi.org/10.1155/2014/757929>
- [36] Gaudette, L. and Japkowicz, N. (2009) Evaluation Methods for Ordinal Classification. In: *Advances in Artificial Intelligence*, Springer, Berlin, Heidelberg, 207-210. https://doi.org/10.1007/978-3-642-01818-3_25
- [37] Munteanu, M., Ratziu, V., Morra, R., Messous, D., Imbert-Bismut, F. and Poynard, T. (2008) Noninvasive Biomarkers for the Screening of Fibrosis, Steatosis and Steatohepatitis in Patients with Metabolic Risk Factors: Fibro Test-Fibro Max Experience. *Journal of Gastrointestinal and Liver Diseases*, **17**, 187-191.
- [38] Marchesini, G., Roden, M. and Vettor, R. (2017) Response to: Comment to EASL-EASD-EASO Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease. *Journal of Hepatology*, **66**, 466-467. <https://doi.org/10.1016/j.jhep.2016.11.002>
- [39] Tapper, E.B., Hunink, M.G.M., Afdhal, N.H., Lai, M. and Sengupta, N. (2016) Cost-Effectiveness Analysis: Risk Stratification of Nonalcoholic Fatty Liver Disease (NAFLD) by the Primary Care Physician Using the NAFLD Fibrosis Score. *PLoS ONE*, **11**, e0147237. <https://doi.org/10.1371/journal.pone.0147237>
- [40] ASCL.Net-Keras: The Python Deep Learning Library. <https://ascl.net/1806.022>