IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 29, 2020, accepted July 13, 2020, date of publication July 22, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011289

A New Clinical Spectrum for the Assessment of Nonalcoholic Fatty Liver Disease Using Intelligent Methods

AMAN SINGH[1], PINKU NATH[2], VIVEK SINGHAL[3], DIVYA ANAND[1], KAVITA[1], (Member, IEEE), SAHIL VERMA[1], (Member, IEEE), AND TZUNG-PEI HONG[4], (Senior Member, IEEE)

[1]School of Computer Science and Engineering, Lovely Professional University, Phagwara 144001, India
[2]School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara 144001, India
[3]Department of Information Technology, JSS Academy of Technical Education, Noida 201301, India
[4]Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan

Corresponding author: Sahil Verma (sahilverma@ieee.org)

ABSTRACT Nonalcoholic Fatty Liver Disease (NAFLD) is the most common cause of chronic liver disease around the world. Remaining silent in the early stages makes its evaluation a challenge. Liver biopsy is still the gold standard method used to classify NAFLD stages but has important sample error issues and subjectivity in the interpretation. This research is an effort to overcome liver biopsy to a possible extent by forming a non-invasive clinical spectrum. This paper proposed an intelligent scheme using the forward algorithm, Viterbi algorithm, and Baum-welch algorithm for examining the disease, and a new clinical spectrum is introduced that incorporates most likely attributes associated with NAFLD stages. The experimental results verify that our method is efficient in distinguishing the credibility of an attribute being associated with a specific stage in case it is linked with more than one stage. Moreover, the proposed scheme can successfully estimate the likelihood of stage progression and supports medical knowledge more proficiently and realistically.

INDEX TERMS Nonalcoholic fatty liver disease, computational methods, forward-backward learning, intelligent systems, healthcare informatics.

I. INTRODUCTION

Intelligent method based frameworks have played a vital role in medicine. From statistical techniques to data mining algorithms to neural networks, all these have been widely deployed on medical datasets for evaluating the sickness. Due to increasing vagueness and complexities in the datasets, deriving intelligible information becomes a significant challenge for clinicians. This challenge could lead to an imprecise assessment of the disease, which would further guide inaccurate treatments to patients. So to avoid these uncertainties up to a feasible extent, medical professionals refer to the intelligent decision-making systems for a second thought on the interpretation of multifaceted datasets. Like for other health complications, intelligent methods have also been widely applied to diagnose and classify liver diseases (Singh and Pandey [1]).

The liver is the largest internal solid organ in a human body weighing about 3 pounds (Adams [52]). It performs several metabolic functions like metabolizing drugs, bile production, filtering blood, storing glucose, assisting in fat digestion, detoxifying harmful chemicals, and making proteins for blood plasma (Chuang [3]). Liver disease is defined as the improper functioning of the liver causes illness which further leads to serious health ramifications. General causes of the disease are genetic disorders, infected eatables, immoderate consumption of ethanol, a severe reaction to certain drugs, infections from bacteria, and excessive fat buildup in the body (Asrani *et al.* [53], Lonardo *et al.* [54], and Chuang [3]). Liver disorders are categorized into numerous types, out of which a few common ones include viral hepatitis, autoimmune hepatitis, neonatal hepatitis, fibrosis, cirrhosis, liver tumor, liver abscess, Wilson disease, alcoholic liver disease, and

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

non-alcoholic fatty liver disease (Adams [52]). Its presence around the world in diverse forms is increasing the mortality rate and making it a serious area of concern in the medical domain. The ability of the liver to function ordinarily even when partly damaged resists it from a timely diagnosis and also makes it more alarming because by then, the liver has suffered significant damage. This indicates that in time diagnosis is inevitable so that treatment can begin at the right stage and in a well-organized mode. During the assessment of sickness, selecting the features, evaluating the values, differentiating the dependent and independent predictors, and finding the co-relations between various attributes stretches the decision time of physicians. To solve these impediments mentioned and to reduce the cost, time, and effort needed, intelligent methods based on decision-making frameworks have been used.

A. STATUS OF NON-INVASIVE ASSESSMENT OF NAFLD

In literature, there are various studies on liver disease assessment using intelligent computing techniques. It was found that both individual and integrated intelligent techniques were widely used to develop the adaptable diagnostic frameworks for achieving better accuracies and decreasing the likelihood of occurrence of prediction errors (Singh [1]). Individually, artificial neural networks (ANNs) (Ansari *et al.* [5], Arsene and Lisboa [4], Babu and Suresh [7], Elizondo *et al.* [6], Hayashi *et al.* [8]), fuzzy logic (FL) (Ming *et al.* [9], and Obot and Udoh [10]), decision trees (Eastwood and Gabrys [11], Floares [12], and Yan *et al.* [13]), and in combination, artificial neural network-case-based reasoning (ANN-CBR) (Chuang [3], and Lin and Chuang [14]), data mining-fuzzy logic (DM-FL) (Luukka [15], and Torun and Tohumoğlu [16]), case-based reasoning-data mining (CBR-DM) (Lin [17]), artificial immune system-fuzzy logic (AIS-FL) (Mezyk and Unold [18], and Polat *et al.* [19]), artificial immune system-data mining-fuzzy logic (AIS-DM-FL) (Polat and Gunes [20], [21]), artificial neural network-data mining-genetic algorithm (ANN-DM-GA) (Cruz-Ramirez *et al.* [22]), were extensively applied to different types of liver disorders includes viral hepatitis, liver cancer, liver cirrhosis, liver fibrosis, hepatitis dataset, liver disorders dataset and hepatobiliary disorders dataset.

It is confirmed from the survey that the popularity and applicability of individual and integrated intelligent techniques to liver disorders are immense. Though, at the same time, it is important to note that researchers had shown negligible interest in applying these techniques for the non-invasive assessment of NAFLD, which is a majorly popular hepatic condition people are suffering from nowadays. NAFLD clinical syndrome is one of the most common causes of chronic liver disease around the world (Kalra *et al.* [23]). Its presence is estimated to be 9-32% in Indian population (Kalra *et al.* [23]), one-third of adults in the United States (Dunn *et al.* [24]), 20-30% in western countries (Paschos and Paletas [25]) and 10-24% worldwide (Obika and Noguchi [26]) with vast majority undiagnosed and are still increasing year-on-year. Patients with NAFLD are at a higher risk of death as compares to individuals from the general population (Lonardo *et al.* [55]). The pathogenesis of NAFLD excludes significant ethanol usage ($>$21 drinks per week for men and $>$14 drinks per week for women) and all other causes of the chronic liver disease because fatty infiltration can be caused through numerous conditions such as viral hepatitis B, C, infection and drug toxicity. However, the term NAFLD is predominantly associated with obesity, insulin resistance, type-2 diabetes, and metabolic risk factors (Nugent and Younossi [27]).

B. TYPES OF NAFLD

The clinicopathologic spectrum of NAFLD encompasses hepatic steatosis, NASH, liver fibrosis, liver cirrhosis, and rarely liver cancer (Lonardo *et al.* [56], Dunn *et al.* [24], and Kanda *et al.* [57]). Unlike simple steatosis is generally non-progressive and has a relatively benign clinical course, NASH is progressive and can develop fibrosis, cirrhosis, and liver cancer. Steatosis is an accumulation of approximately 5% of fat in the hepatic parenchyma without inflammation, whereas NASH is distinguished by the presence of fat as well inflammation or ballooning degeneration of hepatocytes in addition to steatosis (Than and Newsome [29], and Yu *et al.* [28]). Liver fibrosis is also a histological transformation caused by inflammation in the liver that causes a disproportion between the collagen fiber synthesis and decomposition (Fitzpatrick and Dhawan [30]). It is a scarring process in which more than usual amounts of collagen fiber deposits in the liver cells but without the destruction of the lobular structure which specifically distinguishes it from cirrhosis. In cirrhosis, the architectural organization of the functional units of liver collapse and its fundamental structure is deformed. Liver cancer is a malignancy that usually originates from hepatocytes (European Association for the Study of the Liver, 2016; Azzam & Malnick, 2015; Byrne & Targher, 2015; Than & Newsome, 2015). The risk of developing primary liver cancer is increased in the background of advanced fibrosis and cirrhosis (Rinella [59] and Sanyal *et al.* [33]). In the end, all these nonalcoholic fatty liver disease complications lead to liver failure.

C. APPROACHES BASED ON THE HIDDEN MARKOV MODELS

Approaches based on the Hidden Markov models (HMMs) are especially known for describing unpredictable events or applications such as handwriting, part-of-speech tagging, musical score following, speech recognition, gesture recognition, partial discharges (Chen *et al.* [37], Gao *et al.* [38], Rabiner and Juang [34], Rabiner [35], and Tsakalidis *et al.* [36]). It also has extensive applicability in bioinformatics like genome sequence analysis, gene identifications, gene annotation, cancer detection through images, proteins, and DNA sequences (Krogh *et al.* [40] and Yoon [39]). Rastghalam and Pourghassem [41]

proposed a breast cancer detection framework in which HMM-based methodology was used as a fusion algorithm for decision-level fusion-based classification on thermography images. In multiple sequence alignment and phylogenetic analysis of protein and gene data, HMM approaches had provided compelling performance. For example, to align multiple genomic or proteomic sequences, a fuzzy-HMM-based framework was presented by Collyda *et al.* [42]. Besides, to perform a phylogenetic analysis of protein and gene data with the help of multiple sequence alignment generated, a fuzzy-HMM model was developed by Collyda *et al.* [43]. The identification of regions of the genome is a key element in genome sequence analysis. A computation-based discovery to recognize transcription factor binding sites using HMM techniques was proposed by Wu and Xie [44]. Takano *et al.* [45] presented an HMM-based framework for predicting human motions based on motion patterns. These motion patterns were used as parameters for HMM, and validation of the method was taken place on the motion dataset provided by CMU. Peng and Dong [46] used the HMM and grey model for modeling the time of states and developed a hybrid prognosis approach for age-dependent health prediction in terms of remaining valuable life of the assets. Complicated intersection designs and congested travel conditions lead to road accidents. For solving this problem, a traffic incident prediction system based on HMM was built by Zhou *et al.* [47] and Ling *et al.* [48] applied HMM techniques to predict the movement of a speaker's mouth from text inputs. Apart from the estimations, HMMs log-likelihood score was also used to rank the data and this ranking was useful in creating rules for making decisions (Rafiul Hassan *et al.* [49]).

Based on the comprehensive literature study, it was found that the scope of implementing HMM-based approaches in liver disorders and the evaluation of NAFLD using intelligent techniques have not been explored yet. Indeed, NAFLD stages are more commonly associated with obesity, insulin resistance, type-2 diabetes, metabolic syndrome, and a few elevated liver enzyme tests, but in actual situations, there are more insights into the pathogenesis of these hepatic conditions. Liver biopsy is still the gold standard method used to categorize NAFLD stages but has important sample error issues and subjectivity in the interpretation (Machado and Cortez-Pinto [50]). Moreover, it is an invasive method that can also raise the risk of complications if the mode of sampling is not appropriate. This study is an effort in that direction to overcome liver biopsy to a possible extent. The article accordingly applied HMM-based methods to evaluate health examination data of NAFLD patients. A new clinical spectrum is then formed, which incorporates a combination of non-invasive attributes including risk factors, symptoms, and biomarkers most likely associated with NAFLD stages. This proposed spectrum is expected to assist medical experts in performing a more proficient and inclusive assessment of NAFLD phases.

The rest of this paper is organized as follows: Section 2 presents material and methods containing a detailed description of the algorithms used. Section 3 covers the experimental results and discussion. Finally, conclusion and future work are drawn in Section 4.

II. MATERIAL AND METHODS

Based on the advice and assistance of clinicians and hepatology specialists, the data was collected and prepared to employ data collection tools and techniques, which include questionnaires, direct interviews, and assessment of existing medical records. Fig. 1 depicts the block diagram for the stepwise approach for the assessment of nonalcoholic fatty liver disease. The NAFLD spectrum was categorically divided into five stages including steatosis, NASH, fibrosis, cirrhosis, and liver cancer. Essential health examination data of 86 patients with nonalcoholic fatty liver disease and 22 healthy people were collected. Among 108 instances of data, 10 patients had steatosis, 24 patients had NASH, 22 patients had fibrosis, 24 patients had cirrhosis, 6 patients had liver cancer, and the remaining 22 were healthy individuals. The dataset incorporated the results of the medical expert's laboratory examination and health and lifestyle survey that include 62 relevant attributes, which are a combination of risk factors, symptoms, and biomarkers. These attributes were covering information about metabolic syndrome (MetS), hypertension (HTN), dyslipidemia (DLM), visceral obesity (VO), type-2 diabetes (T2DM), body mass index (BMI), gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), alanine aminotransaminase (ALT), aspartate aminotransferase/alanine aminotransaminase (AST/ALT ratio), fatigue (FT), malaise (ML), pain in upper quadrant (RUQ), adiponectin level (AL), tumor necrosis factor alpha (TNFα), triglycerides level (TG), hepatomegaly (HM), acanthosis nigricans (AN), total bilirubin (TBIL), alpha-2-macroglobulin (α2M), apolipoprotein A1 (Apo A-I), haptoglobin (HP), ferritin (FER), hyaluronic acid (HA), insulin resistance (IR), high-density lipoprotein cholesterol (HDL), hyperglycaemia (HG), platelets count (PC), albumin (ALB), TIMP metallopeptidase inhibitor 1 and TIMP metallopeptidase inhibitor 2 (TIMP-1 and TIMP-2), amino terminal type III procollagen peptide (P3NP), mitochondrial dysfunction (MD), increased free fatty acids (FFA), hepatocyte lipotoxicity (HL), fibrogenic response (FR), oxidative stress (OS), cytokeratin-18 (CK-18), C-reactive protein (CRP), hyperlipidemia (HLD), interleukin 6 and interleukin 8 (IL6 and IL8), lipid peroxidation (LOP), sleep apnea (AP-ne-ah), ammonia (H3N), variceal hemorrhage (VH), jaundice (JAUND), oedema (OD), asterixis (ASX), spider naevi (SN), hepatic encephalopathy (HE), esophageal varices (EV), gastrointestinal bleeding (GI bleed), thrombocytopenia (TTP), alpha-fetoprotein (AFP), des-carboxy prothrombin (DCP), glypican-3 (GPC3), osteopontin (OPN), lectin-bound AFP (AFP-L3), tyrosine kinase with Ig and EGF homology domains 2 (TIE-2), squamous cell carcinoma

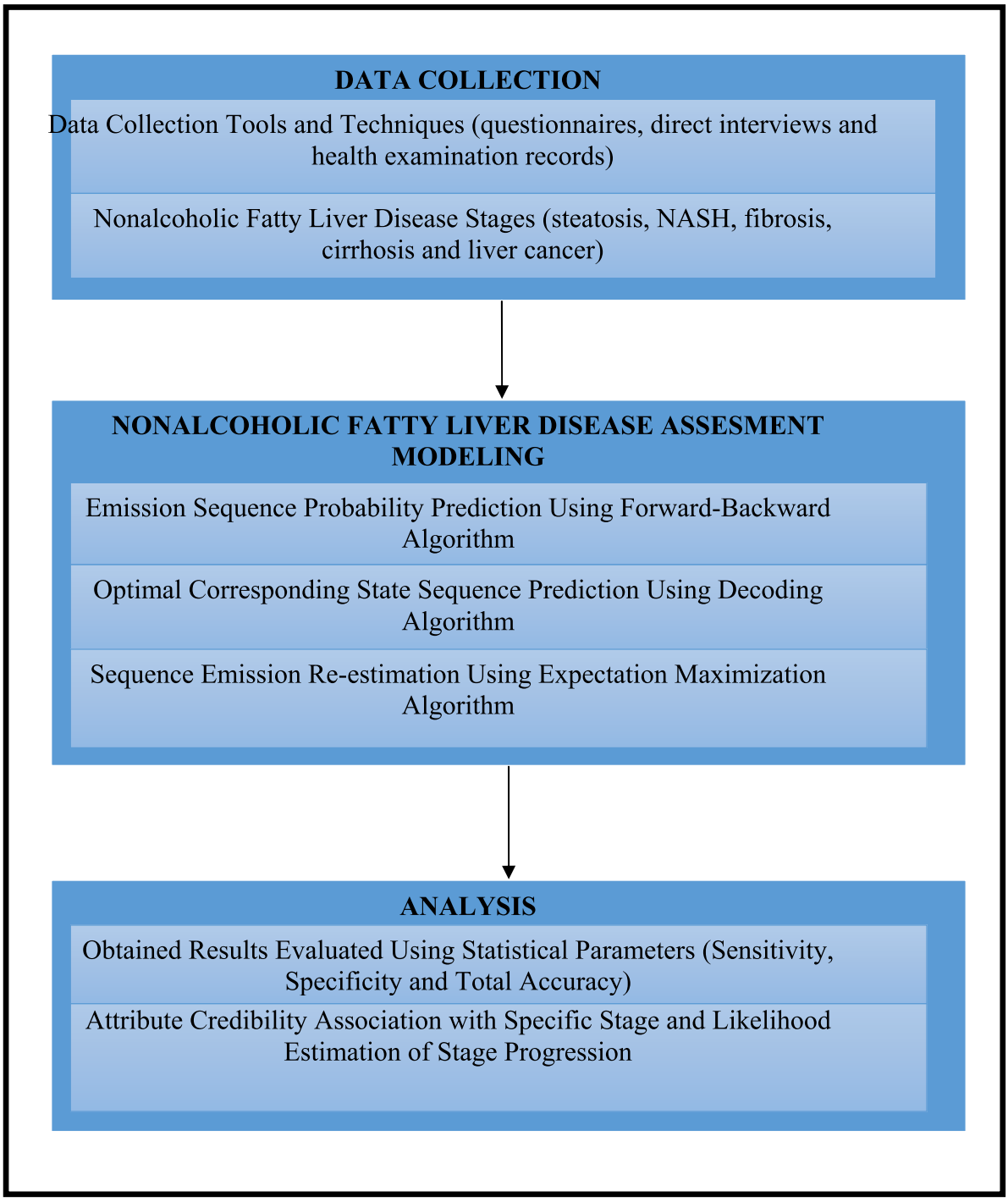

FIGURE 1. Block diagram for the stepwise approach for the assessment of nonalcoholic fatty liver disease.

antigen (SCCA), Golgi phosphoprotein 2 (GP73), alpha-l-fucosidase (Alpha_L_fucos) and human carbonyl reductase 2 (HCR2). Numerous attributes belong to more than one stage, making the assessment more challenging and ambiguous. Apart from the staging of NAFLD, the sound part of this study is to distinguish the credibility of an attribute being associated with a specific stage in case it is linked with more than one stage and to estimate the likelihood of stage progression.

The approaches based on the hidden Markov model are especially known for describing events or applications that are unpredictable due to the influence of random variables. Applicability of HMMs for different purposes proves its versatility in bioinformatics and other significant areas of research. The idea underlying to implement HMM-based methods is by considering three different canonical problems that are typically called evaluation, decoding, and learning.

In general, an HMM has numerous states, numerous distinct emission symbols per state, state transition probability distribution indicating the likelihood of transitioning from one state to another, emission symbol probability distribution that indicates how likely it is for a certain measure value to come from a given state and the initial state distribution for each state indicating how likely it is for a new input sequence to start in a given state. The notation used to represents these parameters is given below (Rabiner and Juang [34], Rabiner [35], and Stamp [51]):

A = the number of states,
B = the number of emission symbols,
T = the length of emission sequences,
$C = (c_1, c_2, \ldots, c_T)$, representing a current state sequence,
$D = \{d_1, d_2, \ldots, d_N\}$, representing distinct states in the model,
$U = (U_1, U_2, \ldots, U_T)$, representing an emission sequence,
$V = \{1, 2, \ldots, B\}$, representing a set of possible emissions; i.e. $U_i \in V$ for $i = 1, 2, \ldots T$,
$S = \{s_{ij}\}$, being an $A \times A$ matrix with $s_{ij} = P(d_j \text{ at } t+1 \mid d_i \text{ at } t)$, where the state transition probability s_{ij} is independent of the time t,
$E = \{e_j(k)\}$, being $A \times B$ with $e_j(k) = P(k \text{ at } t \mid d_j \text{ at } t)$, where the emission probability $e_j(k)$ is independent of time t, and
π = the initial state distribution.

An absolute requirement of an HMM thus includes S, E, and π. Therefore, the simple notation for HMM becomes $\lambda = (S, E, \pi)$.

Problem 1: Given the sequence of emissions $U = U_1, U_2, \ldots, U_T$, and the HMM $\lambda = (S, E, \pi)$, how do we efficiently compute $P(U|\lambda)$, the probability of emission sequence. This problem is called evaluation, and it is solved by either the forward algorithm or the backward algorithm, which measures how well a model matches the emission sequence. The forward-backward procedures (do not confuse them with the forward-backward algorithm) are explained below from which we can find the forward variable.

The given model is $\lambda = (S, E, \pi)$ and emission sequence is $U = (U_1, U_2 \ldots, U_T)$. To compute $P(U|\lambda)$ forward algorithm is used. In addition to the forward algorithm, we do have a naive approach for solving the evaluation problem, but even that we choose the former after executing both because the latter requires more multiplications in comparison with the former as explained below. Consider the state sequence $C = (c_1, c_2, \ldots, c_T)$. Now based on equation (2) we have

$$P(U \mid C, \lambda) = e_{c1}(U_1)\, e_{c2}(U_2) \ldots . e_{cT}(U_T) \tag{1}$$

and based on the initial state distribution and equation (1), it follows that

$$P(C \mid \lambda) = \pi_{c1} s_{c1,c2} s_{c2,c3} \ldots \ldots s_{cT-1,cT}.$$

Since

$$P(U, C \mid \lambda) = \frac{P(U \cap C \cap \lambda)}{P(\lambda)},$$

and

$$P(U \mid C, \lambda)\, P(C \mid \lambda) = \frac{P(U \cap C \cap \lambda)}{P(C \cap \lambda)} . \frac{P(C \cap \lambda)}{P(\lambda)} = \frac{P(U \cap C \cap \lambda)}{P(\lambda)},$$

we have:

$$P(U, C \mid \lambda) = P(U|C, \lambda)|P(C \mid \lambda) \tag{2}$$

Finally, we obtain all the possible state sequences as follows:

$$\begin{aligned} &P(U \mid \lambda) \\ &= \sum_C P(U, C \mid \lambda) \\ &= \sum_C P(U \mid C, \lambda)\, P(C|\lambda) \\ &= \sum_C \pi_{c1} e_{c1}(U_1)\, s_{c1,c2} e_{c2}(U_2) \ldots . s_{cT-1cT} e_{cT}(U_T). \end{aligned}$$

The computation mentioned above requires $2TA^T$ multiplications that decrease efficiency and makes it incompetent in comparison with the forward algorithm, which only requires A^2T multiplications. The forward algorithm (also known as α-pass) for finding $P(U|\lambda)$ is defined below.

For $t = 1, 2, \ldots, T$ and $i = 1, \ldots, A$, $\alpha_t(i)$ is the probability of a partial emission sequence up to a given time t. It is shown as follow:

$$\alpha_t(i) = P(U_1, U_2, \ldots . U_t, c_t = d_i \mid \lambda) \tag{3}$$

For computing $\alpha_t(i)$ defined in Eq. (3) recursively, the following steps are executed.

Step 1: Initialization

$$\alpha_1(i) = \pi_i e_i(U_1), \quad for\ i = 1, 2, \ldots, A;$$

Step 2: Induction:
For $i = 1, 2, \ldots, A$ and $t = 2, 3, \ldots, T$, compute:

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{A} \alpha_t(j)\, s_{ji}\right] e_i(U_{t+1}).$$

Step 3: Termination:

$$P(U \mid \lambda) = \sum_{i=1}^{A} \alpha_T(i).$$

The backward algorithm (also known as β-pass) is analogous to the forward algorithm for finding $P(U|\lambda)$ expect the fact that the later starts from the beginning and work forward towards the end and the former is vice-versa. For $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, A$, $\beta_t(i)$ is calculated as follows:

$$\beta_t(i) = P(U_{t+1}, U_{t+2}, \ldots . U_T \mid x_t = d_i, \lambda). \tag{4}$$

Similarly, for computing $\beta_t(i)$ defined in Eq. (4) recursively, the following steps are executed.

Initialization:

$$\beta_t(i) = 1, \quad for\ i = 1, 2, \ldots, A.$$

Induction:

For $t = T-1, \ldots, 1$ and $i = 1, 2, \ldots, A$, compute:

$$\beta_t(i) = \sum_{j=1}^{A} s_{ij} e_j(U_{t+1})\, \beta_{t+1}(j).$$

Termination:

$$P(U|\lambda) = \sum_{j=1}^{A} \sum_{i=1}^{A} \alpha_t(j) s_{ji} e_i(U_{t+1})\, \beta_{t+1}(j).$$

Problem 2: Given the HMM $\lambda = (S, E, \pi)$ and a sequence of emissions $U = U_1, U_2, \ldots, U_T$, find the optimal corresponding state sequence $D = d_1, d_2 \ldots d_T$. This problem is called decoding and is solved by the Viterbi algorithm, sometimes also called a decoding algorithm. The structure of this algorithm is quite similar to the forward algorithm.

For t = 1,2, …, T and $i = 1, 2, \ldots, A$, the optimal state sequence $\omega_t(i)$ would be

$$\omega_t(i) = P(c_t = d_i \mid U, \lambda).$$

As we have already calculated the probability measured by α_t(i) up to the time t and by β_t(i) after the time t, we get the following:

$$\omega_t(i) = \frac{\alpha_t(i)\, \beta_t(i)}{P(U|\lambda)},$$

where $P(U|\lambda)$ is obtained by summing $\alpha_t(j)$ over i. Now $\omega_t(i)$ defines that it is the maximum value for the most likely state at time t is the state d_t.

Problem 3: Given a sequence of emissions $U = U_1, U_2, \ldots, U_T$ and the dimensions A and B, re-estimate HMM $\lambda = (S, E, \pi)$ that maximizes the probability of U. This problem is called learning, and it is solved by training through the Baum-welch algorithm, which is a special case of the expectation-maximization algorithm. This algorithm adjusts the probabilities of the emissions until the training converges, and it works by assigning initial probabilities to all the parameters. To determine updated parameters for HMM, it uses both the forward and backward variables is also called the forward-backward algorithm.

For $t = 1, 2, \ldots, T-1$ and $i, j \in \{1, 2, \ldots, A\}$, $\omega_t(i, j)$ is defined as a probability of being in state d_i at time t and transition to state d_j at time $t+1$ as follows:

$$\omega_t(i, j) = P\left(c_t = d_i, c_{t+1} = d_j \mid U, \lambda\right).$$

In terms of $\alpha, \beta, S$and$E$, $\omega_t(i, j)$ represents the improved version of the estimated value and can be written as follows:

$$\omega_t(i, j) = \frac{\alpha_t(i)\, s_{ij} e_j(U_{t+1})\, \beta_{t+1}(j)}{P(U \mid \lambda)}.$$

For $t = 1, 2, \ldots, T-1$, the values of $\omega_t(i)$ and $\omega_t(i, j)$ are related by:

$$\omega_t(i) = \sum_{j=1}^{A} \omega_t(i, j).$$

Now, based on the above equations, the re-estimation of the HMM, $\lambda = (S, E, \pi)$ can be defined as follows. For $i = 1, 2, \ldots, A$,

$$\pi_i = \omega_1(i).$$

For $i = 1, 2, \ldots, A$ *and* $j = 1, 2, \ldots, A$, calculate s_{ij}, which is the probability of transition from state d_i to state d_j:

$$s_{ij} = \sum_{t=1}^{T-1} \omega_t(i, j) \Big/ \sum_{t=1}^{T-1} \omega_t(i). \tag{5}$$

For $j = 1, 2, \ldots, A$ and $k = 1, 2, \ldots, B$, given the model in state d_j, we can calculate the probability of observing symbol k as:

$$e_j(k) = \sum_{\substack{t \in \{1.2.....,T\} \\ U_t = k}} \omega_t(j) \Big/ \sum_{t=1}^{T} \gamma_t(j). \tag{6}$$

In Eq. (5), $\sum_{t=1}^{T-1} \omega_t(i, j)$ represents the probable number of transitions from state d_i to state d_j and $\sum_{t=1}^{T-1} \omega_t(i)$ is the probable number of transitions from d_i to any state.

In Eq. (6), $\sum_{\substack{t \in \{1.2.....,T\} \\ U_t = k}} \omega_t(j)$ is the number of times the model is in state d_j with emission k and $\sum_{t=1}^{T} \omega_t(j)$ is the number of times the model is in state d_j. As re-estimation is an iterative process, the algorithm would follow the mentioned steps.

Step 1: Initialize the HMM, $\lambda = (S, E, \pi)$;

Step 2: Calculate $\alpha_t(i)$, $\beta_t(i)$, $\omega_t(i, j)$, and $\omega_t(i)$;

Step 3: Re-estimate the HMM, $\lambda = (S, E, \pi)$;

Step 4: If the $P(U|\lambda)$ value increases, then repeat steps 2 and 3 until some convergence criteria are achieved and/or it reaches a maximum number of iterations.

III. EXPERIMENTAL RESULTS AND DISCUSSION

To demonstrate the effectiveness of the applied methods, we used the health examination dataset with a total of 108 instances. The NAFLD spectrum was categorically divided into five stages, including steatosis, NASH, fibrosis, cirrhosis, and liver cancer. Among 108 subjects, 10 patients had steatosis, 24 patients had NASH, 22 patients had fibrosis, 24 patients had cirrhosis, 6 patients had liver cancer, and the remaining 22 were healthy individuals. The dataset incorporated the results of medical experts' laboratory examinations and health and lifestyle surveys including the attributes which are a combination of risk factors, symptoms, and biomarkers. To validate the best performing one, we experimented with many different models as there is no detailed method or an obvious straightaway choice for determining the optimal model. It was necessary to initialize appropriate values during training as the performance of the model principally depends on it. The training was done with both Viterbi and Baum-welch algorithms, but the former was omitted and the later was considered based on the obtained outputs. Single sequences as well as multiple random sequences were given as inputs to train the model. Here, 20 multiple random sequences were finalized to enhance the recognition phase. More training samples iteratively learned values for

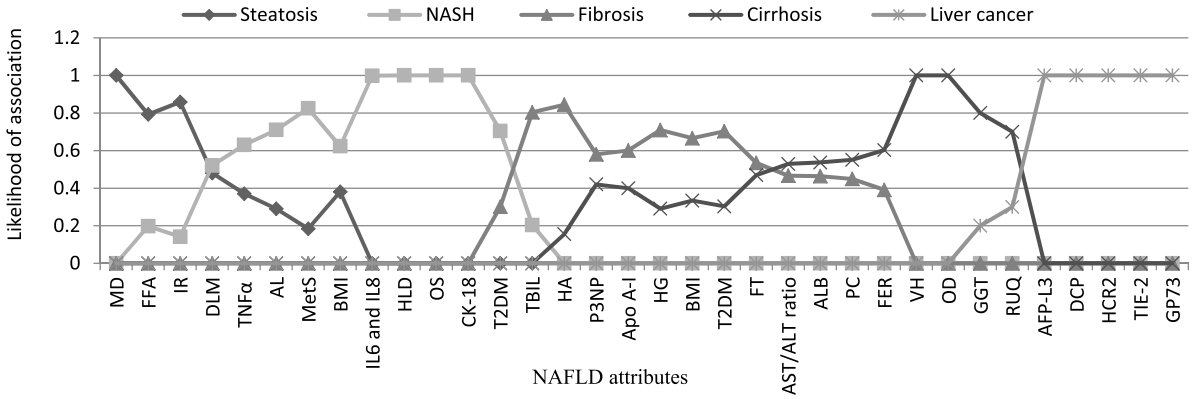

FIGURE 2. Likelihood of attributes association with NAFLD stages.

TABLE 1. Identification results of NAFLD stages.

	Steatosis	NASH	Fibrosis	Cirrhosis	Cancer	Healthy
Steatosis	7	0	0	0	0	3
NASH	2	22	0	0	0	0
Fibrosis	0	0	20	2	0	0
Cirrhosis	0	0	0	24	0	0
Cancer	0	0	0	0	6	0
Healthy	0	0	0	0	0	22

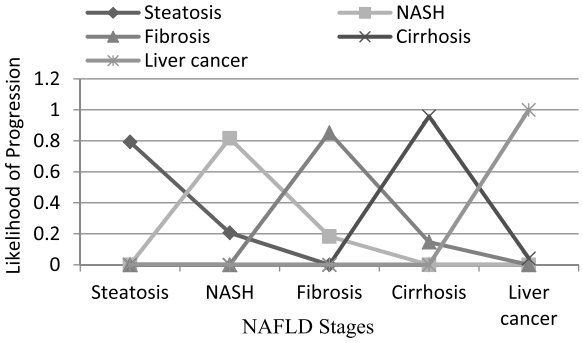

FIGURE 3. Likelihood of stage progression from one to another.

the model, and even if not exact it represented a good solution. Diverse values for tolerance levels and iteration limits were used in the iterative estimation process for testing convergence and checking the most advantageous maximum iteration limit. The most favorable tolerance level and maximum limit were set to 0.0001 and 70, respectively. If we further increased the tolerance, then it decreased the number of steps in the algorithm executed, which eventually reduced the quality of learning. Consequently, this showed that hidden Markov model-based methodology with fitting parameter estimation and training symbolizes a suitable approach for abstracting helpful information from the dataset.

Table 1 shows the identification results of the NAFLD stages. Based on the given matrix in Table 1, 3 patients with steatosis were wrongly identified as healthy, 2 patients with NASH were wrongly identified as patients with steatosis, and 2 patients with fibrosis were wrongly identified as patients suffering from cirrhosis. The rest 7 subjects with steatosis, 22 subjects with NASH, 24 subjects with cirrhosis, 6 subjects with liver cancer, and 22 healthy individuals were correctly recognized. Table 2 presents the achieved diagnostic rates evaluated using statistical parameters like sensitivity, specificity, and total accuracy. For pursuing better diagnostic performance, this study has also considered sensitivity as one of the most useful criteria to examine the evaluation as it measures the proportion of people rightly detected as being sick. 100% sensitivity indicated that the model recognizes all of the subjects who are sick to avoid delaying the treatment. The obtained results had a total accuracy rate of 93.5%, a sensitivity of 70% in steatosis, 91.7% in NASH, 90.9% in fibrosis, 100% in cirrhosis, 100% in liver cancer, and an overall specificity of 100% as all healthy people were correctly identified as healthy. The attained sensitivity was less in initial stages as compared to later stages indicating that the disease shows fewer warning signs while it starts developing and generates more comprehensible signs when progressing towards advanced stages.

Fig. 2 and Fig. 3 instigate valuable discussion regarding the pathology of nonalcoholic fatty liver disease. Fig. 2 distinguishes the credibility of an attribute being associated with a specific stage in case it is linked with more than one stage. Fig. 3 gives us the estimation of the likelihood of stage progression from one to another. Based on Fig. 2, it was observed that the attributes including FFA, IR, DLM, TNFα, AL, MetS, BMI, T2DM, TBIL, HA, P3NP, Apo A-1, HG, FT, AST/ALT ratio, ALB, PC, FER, GGT, and RUQ were associated with more than one stage. FFA, IR, DLM, TNFα, AL, MetS, and BMI were linked with both steatosis and NASH; T2DM and TBIL were linked with both NASH and fibrosis; HA, P3NP, Apo A-1, HG, BMI,

TABLE 2. Diagnostic rates using statistical parameters.

	Sensitivity (Steatosis)	Sensitivity (NASH)	Sensitivity (Fibrosis)	Sensitivity (Cirrhosis)	Sensitivity (Cancer)	Specificity	Total accuracy
Proposed Model	70%	91.7%	90.9%	100%	100%	100%	93.5%

T2DM, FT, AST/ALT ratio, ALB, PC and FER were related with fibrosis and cirrhosis; GGT and RUQ were related with cirrhosis and liver cancer. In comparison, FFA and IR had a higher likelihood of association with steatosis than NASH; DLM, TNFα, AL, MetS, and BMI had a higher likelihood of association with NASH than steatosis; T2DM had a higher likelihood of association with NASH than fibrosis; TBIL had a higher likelihood of association with fibrosis than NASH; HA, P3NP, Apo A-1, HG, BMI, T2DM, and FT had a higher likelihood of association with fibrosis than cirrhosis; AST/ALT ratio, ALB, PC and FER had a higher likelihood of association with cirrhosis than fibrosis; GGT and RUQ had a higher likelihood of association with cirrhosis than liver cancer. MD, FFA, and IR showed the capability of becoming independent factors for evaluating steatosis; MetS, OS, CK-18, IL6, IL8, and HLD showed the capability of becoming independent factors for evaluating NASH; HA, HG, P3NP, and Apo A-1 showed the capability for fibrosis; AST/ALT ratio, PC, FER, VH, and OD showed the capability for cirrhosis; AFP-L3, DCP, HCR2, TIE-2, and GP73 showed the capability for liver cancer.

Individually, steatosis is related to MD, FFA, IR, DLM, TNFα, AL, MetS, and BMI out of which MD, FFA, IR, and DLM had more likelihood of association. NASH is related with FFA, IR, DLM, TNFα, AL, MetS, BMI, IL6 and IL8, HLD, OS, CK-18, T2DM, and TBIL out of which MetS, OS, CK-18, TNFα, T2DM, AL, IL6 and IL8, BMI and HLD had more likelihood of association. Fibrosis is linked with TBIL, T2DM, HA, P3NP, Apo A-1, HG, BMI, T2DM, FT, AST/ALT ratio, ALB, PC and FER out of which HA, HG, P3NP, Apo A-1, TBIL, AST/ALT ratio, ALB, T2DM, BMI, and FT had more likelihood of association. Cirrhosis is linked with HA, P3NP, Apo A-1, HG, BMI, T2DM, FT, AST/ALT ratio, ALB, PC, FER, VH, OD, GGT, and RUQ out of which AST/ALT ratio, PC, ALB, FER, VH, OD, GGT, FT, and RUQ had more likelihood of association. Liver cancer is related to GGT, RUQ, AFP-L3, DCP, HCR2, TIE-2, and GP73 out of which AFP-L3, DCP, HCR2, TIE-2, and GP73 had more likelihood of association.

Fig. 3 illustrates the estimation that 21% of patients had progressed from steatosis to NASH, 18% from NASH to fibrosis, 15% from fibrosis to cirrhosis, and only approx 4% had liver cancer from cirrhosis. Steatosis had the lowest and cirrhosis had the highest likelihood of staying in the same stage while progression which means patients rarely suffer from cancer if they already had cirrhosis. Indeed, the progression rate is slow but at the same time is alarming as the disease barely shows warning signs while developing in initial stages like steatosis and NASH which are reversible as compared to the later stages like fibrosis and cirrhosis which are almost irreversible. Fig. 2 helps discover the relations up to an appreciable extent by illustrating the variation in the likelihood of attribute association with NAFLD stages. Fig. 3 shows that NAFLD stages are progressive but has a very slow likelihood of transitions. It is revealed that after cirrhosis, fibrosis stood second in the likelihood of staying in the same stage. On relative comparison, it was estimated that steatosis had a 79% chance of retaining the same stage, NASH had 82%, fibrosis had 85%, and cirrhosis had 96%. It further signified that the likelihood of progression from steatosis to NASH and from NASH to fibrosis was higher than that from fibrosis to cirrhosis and cirrhosis to liver cancer. It was observed that MD, FFA, IR, DLM, TNFα, AL, MetS, BMI, IL6 and IL8, HLD, OS, CK-18, T2DM, TBIL, HA, P3NP, Apo A-1, HG, FT, AST/ALT ratio, ALB, PC, FER, VH, OD, GGT, RUQ, AFP-L3, DCP, HCR2, TIE-2, and GP73 played a vital role in the evaluation and became the part of the new clinical spectrum for assessing the nonalcoholic fatty liver disease.

The medical field is very dynamic, and the new finding comes regularly. The factual occurrence of NAFLD has been underrated even with a presence of 10-24% worldwide (Obika and Noguchi [26]) with the vast majority undiagnosed and is still increasing year-on-year. The search for an ideal non-invasive method to evaluate NAFLD has not been accomplished yet, and physicians broadly rely on ultrasound scans and liver biopsy. Liver biopsy is still the gold standard method used to categorize NAFLD stages but has important sample error issues and subjectivity in the interpretation. Furthermore, it is an invasive method that can also raise the risk of complications if the mode of sampling is not appropriate. The clinicopathologic spectrum of NAFLD encompasses hepatic steatosis, NASH, liver fibrosis, liver cirrhosis, and rarely liver cancer. It is principally associated with a few common blood enzymes tests and metabolic risk factors, but in actual, there are more insights into the attributes linked with these hepatic conditions. Parameters accountable for NAFLD development and progression keep updating. However, this study is an effort in that direction to overcome liver biopsy to a potential extent by forming a non-invasive clinical spectrum that incorporates most likely attributes linked with NAFLD. Undoubtedly, the diagnosis of a disease is one of the most crucial steps that mainly rely on a physician's clinical experience and analysis of health

examination data. This spectrum enhances the efficiency of NAFLD assessment and solves the information-overload problem by suggesting the most likely attributes of interest to clinicians. Still, some imperative attributes are fundamentally substantial in the evaluation of NAFLD stages but could not be included in the clinical spectrum due to any unobserved reason. The reliability and significance of an assessment test depend on the disease, the population where it is applied, and the change in the management induced by the test result. We can extend this work with well designed potential studies that scrutinize the long-term natural history of patients with validated biomarkers. Having a large dataset of patients who were followed for optimal periods can provide essential evidence to support medical experts and researchers, which may further lead to more generalized interpretations. Optimistically this work enhances the medical knowledge more sensibly and realistically and also assists physicians in evaluating NAFLD patients. In summary, the study is proved to be efficient as a medical support system for escalating the total accuracy, sensitivity, and specificity, for distinguishing the credibility of attributes association and for estimating stage progression of nonalcoholic fatty liver disease. Probably, NAFLD will soon become the general cause of hepatic disease worldwide and will surely keep busy the future researchers and hepatologists.

IV. CONCLUSION AND FUTURE WORK

Determining the natural history and associated features of NAFLD are vitally important due to its high presence in the community.

The main contribution of this study is:

- To discover the most likely attributes associated with NAFLD stages given the health examination data.
- To efficiently triumph over the scantiness and helps in achieving efficient evaluation.
- To decrease the likelihood of occurrence of identification errors, to reduce the cost, time, and effort needed.
- To distinguish the credibility of attributes related to the various hepatic conditions.
- To estimate the likelihood of stage progression and finally to form a new clinical spectrum incorporates symptoms, risk factors, and biomarkers that would effectively and efficiently assist physicians in the assessment of NAFLD using HMM-based methods.

The approaches based on the hidden Markov model are especially known for describing the events or applications that are unpredictable due to the influence of random variables. The applicability of HMMs for different purposes proves its versatility in bioinformatics and other important areas of research.

This study has also provided a foundation for future work. The medical field is exceptionally vibrant and dynamic as a variety of findings are discovered regularly. Still, some imperative attributes are fundamentally substantial in the evaluation of the NAFLD stages but could not be included in the clinical spectrum due to any unobserved reason. An enhanced clinical spectrum can be formed by adding new findings, which can lead to more generalized interpretations. Besides, the reliability and significance of an assessment test generally depend on the disease, the population where it is applied, and the change in the management induced by the test result.

We can also extend the work in the prognosis of NAFLD. As development procedures of the disease are very unpredictable, it makes prognosis a null hypothetical concept formed based on past clinical experience and the patterns observed from the collected datasets on a timely basis. Prognosis can be done to estimate when the current patient stage will transit to another. Yes, it will be impossible to predict the exact moment, but still, it would be feasible to give a reasonable estimation for the progression. The limitation in prognosis will be the identification of the first stage as the prognosis procedure needs some information from the patient before any prediction would be made. The first stage would be steatosis, NASH, fibrosis, or cirrhosis. With the supervision of medical experts, we would recommend a prognosis procedure based on the achieved results. For the last decade, physicians have been excitingly involved in understanding the pathology of NAFLD as it holds remarkable research opportunities to refine and diversify the results like studying unidentified biomarkers accountable for evolution, recognizing other diseases co-existing with these hepatic conditions, and investigating the unknown transitional period between the phases. Probably, NAFLD will soon become the general cause of hepatic disease worldwide and will surely keep busy the future researchers and hepatologists.

REFERENCES

[1] A. Singh and B. Pandey, "Intelligent techniques and applications in liver disorders: A survey," *Int. J. Biomed. Eng. Technol.*, vol. 16, no. 1, pp. 27–70, 2014.
[2] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9661–9668, Aug. 2012.
[3] C.-L. Chuang, "Case-based reasoning support for liver disease diagnosis," *Artif. Intell. Med.*, vol. 53, no. 1, pp. 15–23, Sep. 2011.
[4] C. T. C. Arsene and P. J. Lisboa, "Bayesian neural network applied in medical survival analysis of primary biliary cirrhosis," in *Proc. UKSim 14th Int. Conf. Comput. Modeling Simulation*, Mar. 2012, pp. 81–85.
[5] S. Ansari, I. Shafi, A. Ansari, J. Ahmad, and S. I. Shah, "Diagnosis of liver disease induced by hepatitis virus using artificial neural networks," in *Proc. IEEE 14th Int. Multitopic Conf.*, Dec. 2011, pp. 8–12.
[6] D. A. Elizondo, R. Birkenhead, M. Gamez, N. Garcia, and E. Alfaro, "Linear separability and classification complexity," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7796–7807, Jul. 2012.
[7] G. S. Babu and S. Suresh, "Meta-cognitive RBF network and its projection based learning algorithm for classification problems," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 654–666, Jan. 2013.
[8] Y. Hayashi, R. Setiono, and K. Yoshida, "A comparison between two neural network rule extraction techniques for the diagnosis of hepatobiliary disorders," *Artif. Intell. Med.*, vol. 20, no. 3, pp. 205–216, Nov. 2000.
[9] L. K. Ming, L. C. Kiong, and L. W. Soong, "Autonomous and deterministic supervised fuzzy clustering with data imputation capabilities," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 1117–1125, Jan. 2011.
[10] O. U. Obot and S. S. Udoh, "A framework for fuzzy diagnosis of hepatitis," in *Proc. World Congr. Inf. Commun. Technol.*, Dec. 2011, pp. 439–443.

[11] M. Eastwood and B. Gabrys, "Generalised bottom-up pruning: A model level combination of decision trees," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9150–9158, Aug. 2012.

[12] A. G. Floares, "Chronic hepatitis C and B based on i-BiopsyTM," Tech. Rep., 2009, pp. 855–860.

[13] W. Yan, M. Lizhuang, L. Xiaowei, and L. Ping, "Correlation between child-pugh degree and the four examinations of traditional chinese medicine (TCM) with liver cirrhosis," in *Proc. Int. Conf. Biomed. Eng. Informat.*, May 2008, pp. 858–862.

[14] R.-H. Lin and C.-L. Chuang, "A hybrid diagnosis model for determining the types of the liver disease," *Comput. Biol. Med.*, vol. 40, no. 7, pp. 665–670, Jul. 2010.

[15] P. Luukka, "Classification based on fuzzy robust PCA algorithms and similarity classifier," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7463–7468, May 2009.

[16] Y. Torun and G. Tohumoğlu, "Designing simulated annealing and subtractive clustering based fuzzy classifier," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2193–2201, Mar. 2011.

[17] R.-H. Lin, "An intelligent model for liver disease diagnosis," *Artif. Intell. Med.*, vol. 47, no. 1, pp. 53–62, Sep. 2009.

[18] E. Mężyk and O. Unold, "Mining fuzzy rules using an artificial immune system with fuzzy partition learning," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1965–1974, Mar. 2011.

[19] K. Polat, S. Şahan, H. Kodaz, and S. Güneş, "Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 172–183, Jan. 2007.

[20] K. Polat and S. Güneş, "Medical decision support system based on artificial immune recognition immune system (AIRS), fuzzy weighted pre-processing and feature selection," *Expert Syst. Appl.*, vol. 33, no. 2, pp. 484–490, Aug. 2007.

[21] K. Polat and S. Güneş, "A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS.," *Comput. Methods Programs Biomed.*, vol. 88, no. 2, pp. 74–164, 2007.

[22] M. Cruz-Ramírez, C. Hervás-Martínez, J. C. Fernández, J. Briceño, and M. de la Mata, "Multi-objective evolutionary algorithm for donor-recipient decision system in liver transplants," *Eur. J. Oper. Res.*, vol. 222, no. 2, pp. 317–327, Oct. 2012.

[23] S. Kalra, M. Vithalani, G. Gulati, C. M. Kulkarni, Y. Kadam, J. Pallivathukkal, B. Das, R. Sahay, and K. D. Modi, "Study of prevalence of nonalcoholic fatty liver disease (NAFLD) in type 2 diabetes patients in India (SPRINT)," *J. Assoc. Physicians India*, vol. 61, no. 7, pp. 448–453, 2013.

[24] W. Dunn, R. Xu, D. L. Wingard, C. Rogers, P. Angulo, Z. M. Younossi, and J. B. Schwimmer, "Suspected nonalcoholic fatty liver disease and mortality risk in a population-based cohort study," *Amer. J. Gastroenterology*, vol. 103, no. 9, pp. 2263–2271, Sep. 2008.

[25] P. Paschos and K. Paletas, "Non-alcoholic fatty liver disease and metabolic syndrome," *Hippokratia*, vol. 13, no. 1, pp. 9–19, 2009.

[26] M. Obika and H. Noguchi, "Diagnosis and evaluation of nonalcoholic fatty liver disease.," *Express Diabetes Res.*, vol. 2012, Jan. 2012, Art. no. 145754.

[27] C. Nugent and Z. M. Younossi, "Evaluation and management of obesity-related nonalcoholic fatty liver disease," *Nature Clin. Pract. Gastroenterology Hepatology*, vol. 4, no. 8, pp. 432–441, Aug. 2007.

[28] E. L. Yu, J. B. Schwimmer, and J. E. Lavine, "Non-alcoholic fatty liver disease: Epidemiology, pathophysiology, diagnosis and treatment," *Paediatrics Child Health*, vol. 20, no. 1, pp. 26–29, Jan. 2010.

[29] N. N. Than and P. N. Newsome, "A concise review of non-alcoholic fatty liver disease," *Atherosclerosis*, vol. 239, no. 1, pp. 192–202, Mar. 2015.

[30] E. Fitzpatrick and A. Dhawan, "Noninvasive biomarkers in non-alcoholic fatty liver disease: Current status and a glimpse of the future," *World J. Gastroenterol.*, vol. 20, no. 31, pp. 10851–10863, Aug. 2014.

[31] C. D. Byrne and G. Targher, "NAFLD: A multisystem disease," *J. Hepatol.*, vol. 62, no. 1S, pp. S47–S64, Apr. 2015.

[32] H. Azzam and S. Malnick, "Non-alcoholic fatty liver disease—the heart of the matter," *World J. Hepatol.*, vol. 7, no. 10, pp. 1369–1376, Jun. 2015.

[33] A. J. Sanyal, C. Banas, C. Sargeant, V. A. Luketic, R. K. Sterling, R. T. Stravitz, M. L. Shiffman, D. Heuman, A. Coterrell, R. A. Fisher, M. J. Contos, and A. S. Mills, "Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C," *Hepatology*, vol. 43, no. 4, pp. 682–689, Apr. 2006.

[34] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[35] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[36] S. Tsakalidis, V. Digalakis, and L. Neumeyer, "Efficient speech recognition using subvector quantization and discrete-mixture HMMs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. ICASSP*, Mar. 1999, pp. 569–572.

[37] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image Vis. Comput.*, vol. 21, no. 8, pp. 745–758, Aug. 2003.

[38] W. Gao, G. Fang, D. Zhao, and Y. Chen, "A chinese sign language recognition system based on SOFM/SRN/HMM," *Pattern Recognit.*, vol. 37, no. 12, pp. 2389–2402, Dec. 2004.

[39] B.-J. Yoon, "Hidden Markov models and their applications in biological sequence analysis," *Curr. Genomics*, vol. 10, no. 6, pp. 402–415, Sep. 2009.

[40] A. Krogh, M. Brown, I. S. Mian, K. Sjölander, and D. Haussler, "Hidden Markov models in computational biology. Applications to protein modeling," *J. Mol. Biol.*, vol. 235, no. 5, pp. 1501–1531, Feb. 1994.

[41] R. Rastghalam and H. Pourghassem, "Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images," *Pattern Recognit.*, vol. 51, pp. 176–186, Mar. 2016.

[42] C. Collyda, S. Diplaris, P. A. Mitkas, N. Maglaveras, and C. Pappas, "Fuzzy Hidden Markov models: A new approach in multiple sequence alignment.," *Stud. Health Technol. Inform.*, vol. 124, pp. 99–104, Jan. 2006.

[43] C. Collyda, S. Diplaris, P. Mitkas, N. Maglaveras, and C. Pappas, "Enhancing the quality of phylogenetic analysis using fuzzy hidden Markov model alignments.," *Stud. Health Technol. Inform.*, vol. 129, no. 2, pp. 1245–1249, Jan. 2007.

[44] J. Wu and J. Xie, "Computation-based discovery of cis-regulatory modules by hidden Markov model," *J. Comput. Biol.*, vol. 15, no. 3, pp. 279–290, Apr. 2008.

[45] W. Takano, H. Imagawa, and Y. Nakamura, "Spatio-temporal structure of human motion primitives and its application to motion prediction," *Robot. Auto. Syst.*, vol. 75, pp. 288–296, Jan. 2016.

[46] Y. Peng and M. Dong, "A hybrid approach of HMM and grey model for age-dependent health prediction of engineering assets," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12946–12953, Sep. 2011.

[47] J. Zhou, L. Cheng, L. Zhou, and Z. Chu, "Traffic incident prediction on intersections based on HMM," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 13, no. 6, pp. 52–59, Dec. 2013.

[48] Z.-H. Ling, K. Richmond, and J. Yamagishi, "An analysis of HMM-based prediction of articulatory movements," *Speech Commun.*, vol. 52, no. 10, pp. 834–846, Oct. 2010.

[49] M. Rafiul Hassan, B. Nath, M. Kirley, and J. Kamruzzaman, "A hybrid of multiobjective evolutionary algorithm and HMM-fuzzy model for time series prediction," *Neurocomputing*, vol. 81, pp. 1–11, Apr. 2012.

[50] M. V. Machado and H. Cortez-pinto, "Review Non-invasive diagnosis of non-alcoholic fatty liver disease. A critical appraisal," *J. Hepatol.*, vol. 58, no. 5, pp. 1007–1019, 2013.

[51] M. Stamp, "A revealing introduction to hidden Markov models," Dept. Comput. Sci., San Jose State Univ., San Jose, CA, USA, Tech. Rep., 2004, pp. 1–20.

[52] D. H. Adams, *Hepatology: A Textbook of Liver Disease*, 4th ed, B. Zakim, Ed. Philadelphia, PA, USA: Saunders, 2003, p. 1765.

[53] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *J. Hepatology*, vol. 70, no. 1, pp. 151–171, Jan. 2019, doi: 10.1016/j.jhep.2018.09.014.

[54] A. Lonardo, S. Ballestri, G. Marchesini, P. Angulo, and P. Loria, "Nonalcoholic fatty liver disease: A precursor of the metabolic syndrome," *Digestive Liver Disease*, vol. 47, no. 3, pp. 181–190, Mar. 2015, doi: 10.1016/j.dld.2014.09.020.

[55] A. Lonardo, F. Nascimbeni, A. Mantovani, and G. Targher, "Hypertension, diabetes, atherosclerosis and Nash: Cause or consequence?" *J. Hepatology*, vol. 68, no. 2, pp. 335–352, Feb. 2018, doi: 10.1016/j.jhep.2017.09.021.

[56] A. Lonardo, S. Lugari, S. Ballestri, F. Nascimbeni, E. Baldelli, and M. Maurantonio, "A round trip from nonalcoholic fatty liver disease to diabetes: Molecular targets to the rescue?" *Acta Diabetologica*, vol. 56, no. 4, pp. 385–396, Apr. 2019, doi: 10.1007/s00592-018-1266-0.

[57] T. Kanda, T. Goto, Y. Hirotsu, R. Masuzaki, M. Moriyama, and M. Omata, "Molecular mechanisms: Connections between nonalcoholic fatty liver disease, steatohepatitis and hepatocellular carcinoma," *Int. J. Mol. Sci.*, vol. 21, no. 4, p. 1525, Feb. 2020.
[58] M. Giulio, P. Christopher, D. Jean-François, C. Ali, N. Valerio, R. Vlad, T. Herbert, R. Michael, G. Amalia, J. HanneleYki, S. Fritz, V. Roberto, F. Gema, V. Lisbeth, 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), "EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease," *J. Hepatol.*, vol. 64, no. 6, pp. 1388–1402, 2016, doi: 10.1016/j.jhep.2015.11.004.
[59] M. E. Rinella, "Nonalcoholic fatty liver disease: A systematic review," *Jama*, vol. 313, no. 22, pp. 2263–2273, 2015.

AMAN SINGH is working as an Assistant Professor in computer science and engineering with Lovely Professional University, Phagwara, India. He has about seven years of teaching experience, and his areas of interest are biomedical engineering, medical expert systems, artificial intelligence in medicine, and health information systems. He has published scientific research publications in reputed international/national journals and conferences, which are indexed in various international databases. He was invited as a Faculty Resource Person/Session Chair/Reviewer/TPC Member of different FDP, conferences, and journals.

PINKU NATH received the Ph.D. degree in material sciences from the IIT Kanpur, India. He is currently an Assistant Professor with Lovely Professional University, Phagwara. He has over six years of teaching and research experience and has published around 20 research papers and articles. His research interests include intelligent computing techniques, high-performance computing, machine learning, and artificial intelligence.

VIVEK SINGHAL received the Ph.D. degree from the IIT Roorkee, India. He is currently an Associate Professor with the Department of Information Technology, JSS Academy of Technical Education, Noida, India. He has over eight years of teaching and research experience, and has published around several research papers and articles in refereed journals. His main research interests are artificial intelligence, machine learning, intelligent transportation systems, and multi-agent systems and its applications to medicine.

DIVYA ANAND is an Assistant Professor with the School of Computer Science and Engineering, Lovely Professional University, Phagwara, India. She has over two years of research experience. Her areas of interest include expert systems and artificial intelligence.

KAVITA (Member, IEEE) received the B.Tech. and M.Tech. degrees from Maharishi Markandeshwar University, Mullana, India, in 2012, and the Ph.D. degree from Jaipur National University, Jaipur, India, in 2018, all in computer science and engineering. She is guiding six Ph.D. degree students. She has visited many universities out of these two are international universities in Italy and Czech Republic. She is currently working as an Associate Professor with Lovely Professional University, Phagwara, India. She has many research contributions in the areas of cloud computing, the Internet of Things, vehicular ad-hoc networks, WSNs, and MANETs. Some of her research findings are published in top-cited journals such as the IEEE, Wiley, and various international reputed conferences. She is a member of the ACM and the IAENG, and an editorial board member of many international journals, and she has chaired many sessions in reputed international conferences abroad and in India.

SAHIL VERMA (Member, IEEE) received the B.Tech. and M.Tech. degrees from Maharishi Markandeshwar University, Mullana, India, in 2012, and the Ph.D. degree from Jaipur National University, Jaipur, India, in 2017, all in computer science and engineering. He has guided 14 master's degree students and three Ph.D. degree students and six are ongoing. He has visited many universities abroad and in India. He is currently working as an Associate Professor with Lovely Professional University, Phagwara, India. He has many research contributions in the areas of cloud computing, the Internet of Things, vehicular ad-hoc networks, WSNs, and MANETs. Some of his research findings are published in top-cited journals such as the IEEE, Wiley, and various reputed international conferences. He is a member of the ACM and the IAENG, and an editorial board member of many international journals, and he has chaired many sessions in reputed international conferences abroad and in India.

TZUNG-PEI HONG (Senior Member, IEEE) received the B.S. degree in chemical engineering from National Taiwan University, in 1985, and the Ph.D. degree in computer science and information engineering from National Chiao Tung University, in 1992. He served at the Department of Computer Science, Chung Hua Polytechnic Institute, from 1992 to 1994, and the Department of Information Management, I-Shou University, from 1994 to 2001. He was in charge of the whole computerization and library planning and preparation with the National University of Kaohsiung, from 1997 to 2000, and served as the first Director of the Library and Computer Center, National University of Kaohsiung, from 2000 to 2001, the Dean of Academic Affairs, from 2003 to 2006, the Administrative Vice President, from 2007 to 2008, and the Academic Vice President, in 2010. He is currently a Distinguished and Chair Professor with the Department of Computer Science and Information Engineering and the Department of Electrical Engineering, and the Director of the AI Research Center, National University of Kaohsiung, Taiwan. He is also a joint Professor with the Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan. He got the first national flexible wage award from the Ministry of Education, Taiwan. He has published more than 600 research papers in international/national journals and conferences, and has planned more than fifty information systems. His current research interests include knowledge engineering, data mining, soft computing, management information systems, and www applications. He is also a board member of more than forty journals and the program committee member of more than nine hundred conferences.