

Mortality Risk Prediction in Non-Alcoholic Fatty Liver Disease Through a Modified Stacked Gated Recurrent Unit Using Clinical Data

Jayashree Patil¹, Shwetambari Chiwhane^{2*}, Varsha Lokare³

Department of Computer Science and Engineering-Symbiosis Institute of Technology,
Symbiosis International (Deemed University), Lavale, Pune, India^{1,2}

Department of Computer Science & Engineering,
Kasegaon Education Society's Rajarambapu Institute of Technology,
Affiliated to Shivaji University, Sakharale, MS-415414, India^{1,3}

Abstract—Non-Alcoholic Fatty Liver Disease (NAFLD) is a progressive liver disease caused by excessive fat accumulation in hepatocytes in the absence of significant alcohol intake. Liver biopsy remains the diagnostic reference standard but is invasive and costly, limiting its use. In this study we address a distinct task namely mortality risk prediction among patients already diagnosed with NAFLD. Using a public clinical cohort, we preprocess data by median imputation and min-max normalization, address class imbalance with SMOTE applied only to the training partition, and select features with an Enhanced Spotted Hyena Optimizer (ESHO). We compare a Support Vector Machine (SVM), a Gated Recurrent Unit (GRU), a Stacked GRU (SGRU), and a proposed Modified Stacked GRU (MSGRU) with a Refine Gate, alongside logistic regression, random forest, XGBoost, and Cox proportional-hazards baselines. All preprocessing, balancing, and feature selection are fitted strictly within training folds, and all partitions are created before any resampling. Under repeated stratified 5-fold cross-validation, the proposed MSGRU achieved a mean test accuracy of 92.87% (95% CI 92.31–93.43) and a mean test F1 of 91.94%, compared with 90.62% for SGRU. We report accuracy, precision, recall, F1, and AUC with confidence intervals, and we quantify the contribution of each component through an ablation study.

Keywords—Non-alcoholic fatty liver disease; modified stacked gated recurrent unit; refine gate mechanism; NAFLD; machine learning

I. INTRODUCTION

Since liver diseases are a threat to life, they must be predicted and pre-warned in order for humans to survive [1]. Damage to the liver has an impact on human health because it is the largest internal organ in the body. Hepatitis, cirrhosis, fatty liver, cancer, and tumors are some of the various forms of liver disease. Hepatitis can be brought on by drinking alcohol or by consuming harmful chemicals [2]. Sometimes called hepatic cirrhosis, liver cirrhosis is the final stage of liver disease. As fibrosis progresses, liver function is impaired. The main causes of cirrhosis are chronic alcoholism and hepatitis. Cirrhosis can also be caused by both alcoholic and non-alcoholic behaviors [3]. Cirrhosis is characterized by itchy skin, exhaustion, weight loss, and easy bleeding [4]. The various stages of liver function and failure include cirrhosis, fibrosis, inflammation, liver cancer, and end-stage liver disease (ESLD). Cirrhosis causes

severe liver damage and scarring. Early identification of cirrhosis prevents harm to human life [5].

TABLE I. DIAGNOSTIC STRATEGY FOR NAFLD

Diagnostic Strategy	Test Type	Examples
Invasive Tests	Blood Tests	Liver function tests (LFTs), Hepatitis B and C tests, ALT, AST, GGT levels.
	Tissue Sample Tests	Liver biopsy, FibroScan (for fibrosis staging).
Non-Invasive Tests	Ultrasound Imaging / Sonography	Used for detecting fatty liver, cirrhosis, or tumors.
	Magnetic Resonance Imaging (MRI)	Provides detailed liver images, helps assess fat content and fibrosis
	Computed Tomography (CT)	Used to detect liver masses, tumors, and fibrosis

Table I depicts techniques of diagnosing liver illnesses with invasive tests and non-invasive processes; invasive tests include blood tests and biopsies. Imaging techniques like ultrasound scanning, magnetic resonance imaging (MRI), and computed tomography (CT) are examples of non-invasive procedures. Invasive operations are painful, and they are associated with consequences such as infections, bleeding, bile leakage, and so forth. As a result, doctors propose non-invasive imaging procedures. Clinical tests for liver function help to detect and track liver diseases or damage. Additionally, during the examination phase, medical imaging and liver biopsies in rare cases are performed [6]. The imaging tools such as transient elastography, CT (Computed Tomography) and MRI (Magnetic Resonance Imaging), and Ultrasound Sonography (US) become increasingly widespread in healthcare, a liver biopsy remains the best way to track the progress of an illness [7]. Biopsy is an invasive method where patients may experience biliary tract hemorrhage, infection, puncture, and transient pain after the biopsy. Early detection of NAFLD can be ensured by routine patient follow-up; however, repeated use of an intrusive procedure, like a biopsy, may present additional health risks. A new pre-diagnostic method utilizing Clinical Infrared Thermal Imaging (CITI) could offer a safe, user-friendly, and affordable alternative. While blood testing and medical imaging are common non-invasive diagnostic methods, CITI is still

*Corresponding author.

relatively new and not considered a gold standard compared to other imaging techniques [8].

Artificial intelligence consists of two main subfields: machine learning and deep learning. Over the years, researchers have proposed numerous strategies within both areas. This article presents an approach based on deep learning techniques. Compared to traditional data mining and machine learning algorithms, deep learning generally achieves higher accuracy in prediction and classification tasks [9]. Machine learning methods like ensemble learning algorithms, neural networks, SVM, and nearest neighbour are used to categorize and forecast cirrhosis disorders [10]. Machine learning algorithms are good at classifying and predicting outcomes, but sometimes their accuracy falls short. Deep learning, on the other hand, uses multiple layers to better understand complex patterns, leading to more accurate results. We created a Modified Stacked Gated Recurrent Unit, a special deep learning model, to predict the risk of death in non-alcoholic fatty liver disease more effectively.

The primary contributions of this paper are:

- This paper proposes a novel non-invasive NAFLD diagnostic method, focusing on the Modified Stacked Gated Recurrent Unit (MSGRU) model to improve prediction accuracy.
- It addresses the challenge of class imbalance in the NAFLD dataset by applying the Synthetic Minority Oversampling Technique (SMOTE), enhancing model robustness and performance.
- The study employs comprehensive data preprocessing techniques, including imputation, normalization, and feature selection, to ensure high-quality input for the classification models.
- A detailed comparative analysis of multiple classification models is conducted, providing insights into their effectiveness in diagnosing NAFLD.

The primary contributions of this paper are: (i) a mortality risk prediction framework for patients with NAFLD that avoids target leakage by excluding follow-up time and identifier variables and by fitting all resampling and feature-selection steps within training folds; (ii) a leak-free evaluation protocol with repeated cross-validation and confidence intervals; (iii) a comparative analysis against both recurrent baselines and standard tabular baselines (logistic regression, random forest, XGBoost, Cox regression); and (iv) an ablation study isolating SMOTE, ESHO, stacking, and the Refine Gate. We do not claim a diagnostic method; the cohort contains no diagnostic information beyond the recorded NAFLD status.

The structure of the paper is as follows: Section 2 discusses a review of the literature, while Section 3 explores the proposed methodology and classification models employed. Section 4 presents the experimental results, highlighting the effectiveness of the proposed method. Finally, Section 5 concludes with recommendations for future research.

II. LITERATURE SURVEY

Aman Singh et al. [12] introduced a new clinical spectrum for assessing non-alcoholic fatty liver disease with intelligent

methods. Authors presented an intelligent method for examining the disease that employs the forward, Viterbi, and Baum-Welch algorithms, in addition to a new clinical spectrum that encompasses the characteristics most likely associated with NAFLD stages. Their results showed that the method succeeds in discriminating the reliability of a characteristic connected with a single stage from one connected with many stages in a realistic and prescient way, and that the suggested system can accurately estimate the possibility of stage advancement, supporting medical knowledge more effectively. Nevertheless, this investigation is an attempt to overcome liver biopsy to some extent by developing a non-invasive clinical spectrum that includes the most common features associated with NAFLD.

Zhao Li et al. [11] presented deep learning-based algorithms, developed using electronic health data, to forecast the risk of hepatocellular carcinoma among patients with non-alcoholic fatty liver disease. They developed two deep learning models to predict hepatocellular carcinoma development in a cohort of NAFLD patients using a national electronic health records database. Authors discovered that transfer learning compensated for lower performance caused by a lack of data and covariate imbalance, such as gender differences in the data, while time-varying covariates enhanced algorithm performance. For large datasets, which are typical of EHR data, this method presents a substantial computing problem. A further barrier to disease risk prediction is the slow, chronic character of the disease, which makes it harder to forecast future illness when a diagnosis is delayed. Ahmed Gaber Mabrouk et al. [13] provided a non-invasive technique that uses ultrasound scans to identify liver diseases by extracting, selecting, and categorizing features so that liver tissue can be classified as normal, cirrhosis, or steatosis. The correlation, uniformity, deviation, state of opposition, power term percentage, long-term emphasis, and average variance are initially computed. Furthermore, the best properties are chosen using both the Fisher classifier and manual selection techniques. Lastly, several voting-based sub-classifiers are applied: normal/cirrhosis and steatosis/cirrhosis. Lack of a publicly accessible database to further validate the approach is one of the study's drawbacks. Additionally, the low quality of US images necessitates more sophisticated processes or effective features that preserve image information. Ahmet Özdil and Bülent Yilmaz [8] investigated the feasibility of medical infrared thermal imaging (MITI) for the automatic, non-invasive detection of NAFLD, evaluating 167 thermal images from 32 patients using convolutional neural network (CNN) architectures together with texture-analysis-based feature extraction for classification. The CNN-based classifiers achieved the strongest results, with recall, F1-score, specificity, accuracy, and AUC reaching as high as 1.00, 1.00, 0.83, 1.0, and 0.92 respectively across different experimental setups, while texture-analysis methods performed comparatively lower. A limitation of this approach is that thermal imaging remains an emerging modality evaluated on a small patient cohort, which limits generalizability relative to established imaging techniques. K. Prakash et al. [14] described a machine learning approach for identification and estimation of non-alcoholic fatty liver disease (NAFLD). The paper presented a deep learning strategy for predicting and classifying the condition, incorporating a deep neural network and Spearman's rank correlation, with classification and prediction based on 52

characteristics, including gray-level co-occurrence matrix (GLCM) texture features and a related variation matrix (GLGCM) of surface data. Deep neural networks (DNNs) are utilized for identification and feature prediction, and the expected results outperform existing dominant methodologies in comparative metrics. A drawback of the deep learning method is that it may limit generalizability and therapeutic value. Adeep Kulkarni et al. [15] presented automated machine learning algorithms for the prediction of non-alcoholic fatty liver disease. Early detection of liver disease using a classification algorithm is an effective task that can help clinicians diagnose the condition quickly. This study aims to uncover potential causes of NAFLD using machine learning methods such as Naive Bayes, Decision Tree classifier, SVM, Random Forest classifier, Logistic Regression, and K-means. These methods may provide clinicians with insights on how to properly identify NAFLD for innovative diagnosis, as well as preventive and therapeutic purposes, in order to reduce the global burden. A restriction may be the sample size, data variety, or generalizability of the machine learning models employed. Min Zhou et al. [16] provided an analysis of liver lipidomics, demonstrating how the Zuogui Jiang Tang Qinggan Formula protects people with type 2 diabetes mellitus who also have non-alcoholic fatty liver disease. Authors established T2DM-NAFLD models using db/db mice; liver tissue was evaluated using Oil Red O and haematoxylin-eosin (HE) staining along with biochemical lipid parameter measurement. Significant improvements in lipid metabolism in T2DM with NAFLD were demonstrated by the modulation of 496 lipid species following ZGJTQGF therapy, with elevated PC and PE levels and decreased TG and DG. Limitations in ZGJTQGF's impact on proteins linked to lipid production were noted, highlighting its anti-steatosis effect through changes in liver lipidomic composition. Runzhi Yu et al. [17] demonstrated the use of bioinformatic analysis and machine learning-based integration to determine the relationship between inflammation and immune response in non-alcoholic fatty liver disease. Their potential impacts on liver damage, immune cell activation, and inflammation in NAFLD are also discussed. This finding could help create new diagnostic tools and clinical treatments by providing a better understanding of the complex interaction among liver disease, immune response, and metabolic inefficiency in NAFLD. Nonetheless, the emergence of machine learning methods like support vector machine-recursive feature elimination (SVM-RFE) and random forest (RF) provide a potential path toward the identification of biomarkers and highly precise predictive risk modelling. Other studies, including Naseem et al. [18] and Parvatikar [19], compared classifiers such as SVM, K-NN, and Naive Bayes for early liver disease detection, with SVM achieving the highest accuracy. Recent works [22]–[29] continue this trend, employing machine learning and deep learning across diverse datasets. Models like XGBoost, CNNs, and diffusion-based architectures enhance prediction, screening, and imaging-based classification, offering scalable, non-invasive, and accurate diagnostic tools. Collectively, these studies underline the shift toward AI-driven, data-integrated approaches for NAFLD detection and risk assessment. This study addresses the drawbacks of current NAFLD classification techniques, which frequently rely on outdated machine learning models that are unable to identify temporal patterns in patient data. The

application of deep learning, particularly GRUs, in this field has not received enough attention. This study bridges the gap by proposing a modified stacked GRU model that maintains computational economy while enhancing sequential learning and classification accuracy.

III. PROPOSED METHODOLOGY

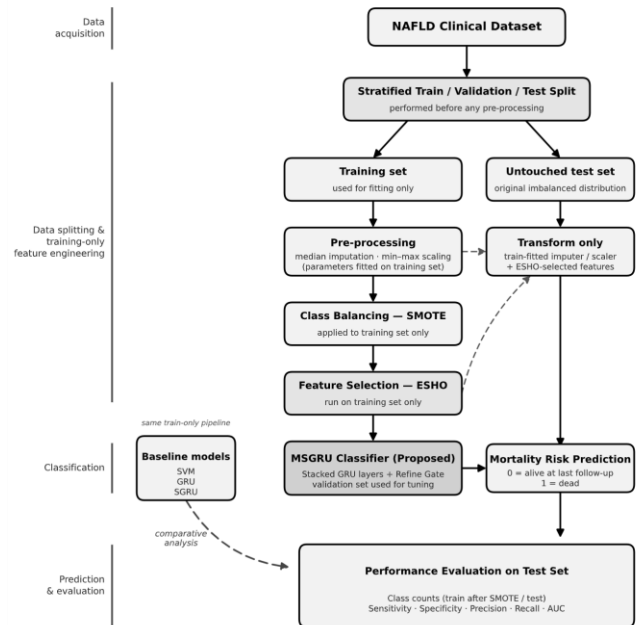


Fig. 1. Framework of proposed methodology.

Fig. 1 depicts the general framework of the proposed method. The proposed framework uses clinical data from the NAFLD dataset, comprising 17,449 subjects observed between 1997 and 2014, with attributes including age, sex, weight, height, BMI, and follow-up time. This raw data is first passed through a pre-processing stage, where missing values are imputed using the median and all features are scaled to a uniform [0, 1] range through min-max normalization, ensuring that no single feature dominates model training due to scale differences. Because the dataset exhibits class imbalance between surviving and deceased patients, the Synthetic Minority Over-sampling Technique (SMOTE) is applied next to generate synthetic samples for the minority class and produce a balanced training distribution. SMOTE is applied to the training fold alone to generate synthetic minority-class samples. Validation and test folds are never resampled. The balanced feature set is then refined through the Enhanced Spotted Hyena Optimizer (ESHO), a metaheuristic feature-selection method that uses a binary tanh-based encoding to retain only the most relevant clinical attributes while discarding redundant or irrelevant ones. The resulting optimal feature subset is fed into the proposed Modified Stacked Gated Recurrent Unit (MSGRU), a deep architecture that stacks multiple GRU layers and incorporates a Refine Gate mechanism to strengthen memory retention and stabilize gradient flow during training. For comparison, the same pre-processed and selected features are also passed through baseline models namely SVM, GRU, and SGRU allowing a direct comparative analysis against the proposed model. The final output of the classification stage is a binary

prediction of patient status (0 = alive at last follow-up, 1 = dead), representing the mortality risk associated with NAFLD. Model performance across all classifiers is then evaluated using accuracy, precision, recall, F1-score, and the AUC curve, establishing the effectiveness of the MSGRU relative to the baseline approaches.

A. Experimental Setup

Table II shows the experiment setup used for this study. The dataset includes patient data collected from 1997 to 2014, with a total of 17449 patients. The 9248 males and 8201 females ranging in age from 18 to 98 years are considered. The study tested the Machine Learning model SVM and the deep learning models Gated Recurrent Unit (GRU), Stacked Gated Recurrent Unit (SGRU), and a Modified Stacked Gated Recurrent Unit (MSGRU). To evaluate the models, performance metrics such as accuracy, recall, precision, F1-score, and the AUC curve are used. The main output was the patient's status, labelled as 0 for alive at the last follow-up and 1 for death.

TABLE II. EXPERIMENT SETUP

Dataset Description		Data from subjects from 1997 to 2014 Total number of patients: 17,449 – Male: 9,248 – Female: 8,201 Min Age: 18 Max Age: 98			
Feature Method	Selection	Enhanced	Spotted	Hyena	Optimization (ESHO)
Machine Models	Learning	SVM (Support Vector Machine) GRU (Gated Recurrent Unit) SGRU (Stacked Gated Recurrent Unit) MSGRU (Modified Stacked Gated Recurrent Unit)			
Performance Metrics		Accuracy, Recall, Precision, F1-Score, AUC Curve			
Output Feature		Status: 0 – alive at last follow-up, 1 – dead			

B. Dataset Description

For the experimental work, this paper utilizes the NAFLD dataset originally described by Allen et al. (2018) [21] and hosted on Kaggle [20]. A detailed description of the dataset is provided in Table III.

TABLE III. DATASET VARIABLES AND THEIR ROLE IN MODELLING

Columns	Description	Role
Id	subject identifier	Excluded (identifier)
Age	age at entry to the study	Predictor
male	0 = female, 1 = male	Predictor
weight	weight in kg	Predictor
height	height in cm	Predictor
bmi	body mass index	Predictor
case.id	the id of the NAFLD case to whom this subject is matched	Excluded (identifier)
Futime	time to death or last follow-up	Excluded from predictors (leakage)
Status	0 = alive at last follow-up, 1 = dead	Outcome

As for the variable Id and case.id, they can be considered identifiers instead of predictors, so they cannot be included in the input data set. Moreover, the variable futime should be excluded from the set of predictors since it describes the time until death or the end of follow-up period and it might contain target leakage for prediction of mortality status. Therefore, the list of predictor variables used for the development of the model is reduced to the eligible demographic and anthropometric characteristics: age, sex, weight, height and BMI.

The target variable Status can be considered as the actual mortality status at the end of follow-up. Nevertheless, it does not take into consideration any differences in the follow-up time and does not reflect censoring. Thus, the task can be considered as the prediction of mortality status and not the estimation of time-to-death for patients.

C. Pre-processing

Following data acquisition, the pre-processing involves two primary steps: missing value imputation and feature normalization. To handle missing values, imputation using the median value is applied, which is robust to outliers and preserves the central tendency of each feature. Subsequently, min-max normalization is employed to scale the features to a uniform range, typically [0, 1], as defined by Eq. (1). This normalization step ensures that all features contribute equally during model training and prevents bias due to differing feature scales.

$$A_{\text{new}} = (A_{\text{current}} - A_{\text{min}}) / (A_{\text{max}} - A_{\text{min}}) \quad (1)$$

Where,

A_{new} – Normalized feature

A_{current} – Current feature

A_{min} – The lowest possible value of the feature for the relevant section.

A_{max} – The highest possible value of the feature for the relevant section.

D. Dataset Balancing Techniques

The dataset exhibited class imbalance, which was addressed by applying the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples and balance the class distribution. SMOTE technique is used on the training set only, since the dataset has been split into two sets – training and testing – and SMOTE will be used to avoid data leakage in the training set only. Resampling is done separately for each training fold during cross-validation, while the validation fold is not touched. The testing set is not altered in any way. SMOTE is a prominent data augmentation approach for correcting class imbalance in machine learning datasets. It works by creating synthetic cases for the minority class rather than just copying existing ones, resulting in a more balanced dataset. SMOTE does this by picking a minority class sample and interpolating fresh samples along the line segments connecting the sample to its closest minority class neighbors. This method improves the model's ability to learn from underrepresented classes, lowers bias towards the majority class, and frequently leads to higher overall predictive performance, especially in classification problems where the data is imbalanced.

E. Feature Selection Using Enhanced Spotted Hyena Optimization

Following preprocessing, to identify and retain the most important attributes for mortality status classification, the Enhanced Spotted Hyena Optimizer (ESHO) technique is used to pick essential features while removing redundant or irrelevant data. ESHO builds upon the Spotted Hyena Optimizer (SHO), a metaheuristic technique based on the social and hunting behavior of spotted hyenas. Although the SHO method works well for broad optimization tasks, it struggles to address discrete or binary problems. To address this issue, ESHO improves SHO by incorporating a hyperbolic tangent function that converts solutions into a binary search space. Furthermore, ESHO enhances the exploration-exploitation balance of SHO through adaptive parameter tuning, exploiting the hyenas' cooperative behavior to enable a more thorough search of the solution space.

ESHO replaces SHO's float-encoding technique with a binary encoding strategy, since each variable's solution can only be 0 or 1. This paper proposes a position-updating strategy for spotted hyenas to improve local search capability and achieve better results in binary search spaces. The tangent hyperbolic function updates the position of spotted hyenas based on the ideal solution; in ESHO, positions are only swapped between "0" and "1", and the search space dimension ranges from 0 to 1.

Step 1: Initialization

All optimization techniques rely heavily on the initialization procedure to function properly. The main purpose of this algorithm is to identify the best attributes. Hyena groups are set up at random at the start of the ESHO implementation. The number of individuals is denoted by the vector S and the hyena positions by D , where the original solution is formatted as in Eq. (2). The solution can be represented as in Table IV, where 1 indicates that the corresponding feature is selected and 0 that it is not.

$$D = d1, d2, \dots, dN \quad (2)$$

where, the solution can be represented as given in Table IV where 1 represents the corresponding feature being selected and 0 represents the non-selected feature.

TABLE IV. SOLUTION REPRESENTATION

1	0	1	0	1	0	1	0	1	0
0	1	0	1	0	1	0	1	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	1	0	1	0	1	0	1	0	1

Step 2: Fitness Calculation

Following the setup step, the suggested ESHO method assesses the effectiveness of all possible solutions. Every generated solution is subjected to a fitness calculation that assesses the solution's fitness based on classification accuracy. The fitness of a candidate subset is the mean balanced accuracy obtained by 5-fold internal cross-validation within the training fold only. Balanced accuracy is used rather than raw accuracy so that the selector is not biased toward the majority class. Ties

between equally performing subsets are broken deterministically: (i) the subset with fewer selected features is preferred, and (ii) if the feature counts are also equal, the subset with the lower lexicographic index is retained. The greatest degree of accuracy implies the best fitness. The fitness value is computed at each round using the following formula as shown in Eq. (3).

$$\text{Fitness} = \text{Max (Balanced Accuracy)} \quad (3)$$

Step 3: Updating Phase

For the updating process, three main steps are used, namely surrounding, attacking the prey, and hunting.

Surrounding the prey: The present best solution is treated as the target, similar to an ideal solution. After determining the optimal solution, the remaining spotted hyenas adjust their positions according to Eq. (4) to Eq. (8). This encircling behavior is mathematically described using the Eq. s below.

$$E_h = |A \cdot B_q(s) - B(s)| \quad (4)$$

$$B(t + 1) = B_q(s) - C \cdot F_h \quad (5)$$

where, F_h represents the distance between the hyena and its chosen prey; A and C represent coefficient vectors; s denotes the current iteration; and B_q and B represent the position vectors of the prey and the spotted hyena, respectively.

$$E = 2 \cdot \text{rand}_1 \quad (6)$$

$$F = 2h \cdot \text{rand}_2 - h \quad (7)$$

$$h = 5 - (\text{Iteration} \times (5/\text{Max}_{\text{iteration}})) \quad (8)$$

where, $\text{Iteration} = 1, 2, 3, \dots, \text{Max}_{\text{iteration}}$. In this scenario, h decreases linearly from 5 to 0 over iterations and is employed to preserve equilibrium between exploitation and exploration. rand_1 and rand_2 are random vectors in the range $[0,1]$. The parameters E and F allow spotted hyenas to access different regions around their current position.

Hunting Behavior: To emulate the hunting habits of spotted hyenas, it is assumed that the most effective search agent knows where the prey is located. Other search agents adjust their positions based on the best search agent's location and build a network of cooperating agents. This hunting mechanism is mathematically defined by the following Eq. s (9) and (10):

$$E_h = |X \cdot B_h - B_k| \quad (9)$$

$$B_k = B_h - Y \cdot E_h \quad (10)$$

where, B_h marks the position of the first best spotted hyena and B_k shows where the other spotted hyenas are located. The number of spotted hyenas, N , can be computed using Eq. (11). The original Spotted Hyena Optimizer addresses continuous optimization challenges; however, it cannot directly handle discrete issues. A binary version, ESHO used to solve this problem.

$$F_h = B_k + B_{k+1} + \dots + B_{k+N} \quad (11)$$

The proposed approach produces a cluster of optimal solutions, denoted as ESHO differs from previous binary

metaheuristics primarily in its cluster-formation technique, modelled mathematically as shown in Eq. (12) and (13).

$$y_d^{s+1} = 1 \text{ if } G(D_h) \geq \text{rand}; 0 \text{ otherwise} \quad (12)$$

$$G(D_h) = \tanh(D_h) = \frac{e^{-\tau(D_h)} - 1}{e^{-\tau(D_h)} + 1} \quad (13)$$

where, rand is a random number drawn from the uniform distribution $[0, 1]$, $y_d^{(s+1)}$ represents a spotted hyena's binary position (of size d, iteration number s, with $\tau = 1$).

$$M = \text{count}_{NS}(B_h, B_{h+1}, B_{h+2}, \dots, B_h + R) \quad (14)$$

where, R is a random vector ranging from 0.5 to 1. The number of candidate solutions in the specified search space that are comparable to the best optimal solution is indicated by N_s , and F_h represents the group of M optimal solutions shown in Eq. (14).

Attacking: The mathematical formula for hunting a target is shown in Eq. (15):

$$B(s+1) = \frac{F_h}{M} \quad (15)$$

The best solution is stored in $B_{(s+1)}$, and additional search agents update their positions based on this solution. ESHO allows the hyenas to change locations and target their objective. Algorithm 1 summarizes the ESHO procedure (Fig. 2).

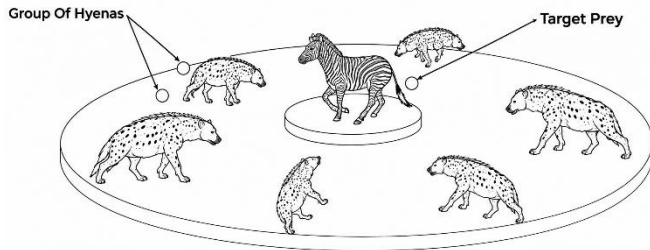


Fig. 2. Spotted hyena optimization.

Step 4: Termination

The population-updating procedure continues until the maximum number of iterations is reached. The result of the ESHO run is the optimal feature subset for mortality risk prediction in NAFLD.

Algorithm 1: Enhanced Spotted Hyena Optimizer Algorithm

Input: Feature matrix of the training fold; labels; population size n; maximum iterations Max_iteration

Output: Optimal feature subset

Initialize population of n hyenas randomly

Determine a fitness value for every single search agent.

While (s < Max number of Iterations)

 | For each spotted hyena
 | | Update the position of the search agent according to Eq. (5)-(15)
 | End for
 | Update control parameters E, F, h, and M

 | Compute the fitness value of each spotted hyena
 | Update F_h if it is better than the previous solution B_h w.r.t. F_h
 | s = s + 1
End while
Output: optimal solution
Output: optimal features

After applying ESHO, the attributes selected are age, sex, weight, and BMI; height was discarded as redundant with weight and BMI, and Status is retained solely as the binary outcome.

F. Computational cost of ESHO and selected subset

To establish whether the computational cost of ESHO is justified, we compared it against three simpler feature-selection techniques like minimum redundancy maximum relevance (mRMR), chi-squared scoring, and recursive feature elimination (RFE) in terms of wall-clock selection time and downstream test accuracy. ESHO was retained only because its accuracy gain justified its cost. BMI is derived from weight and height, and weight and height were retained only if they were independently selected by the optimizer.

G. Machine Learning Models for Classification

1) **Support vector machine:** Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that separates data points of different classes with the maximum margin, as shown in Fig. 3.

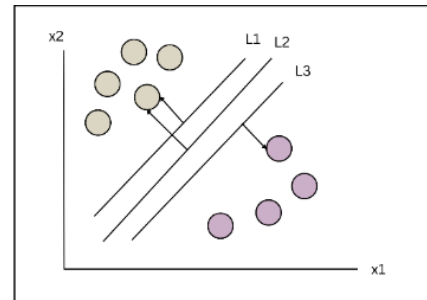


Fig. 3. Hyperplane in SVM.

2) **Gated recurrent unit:** Gated Recurrent Unit (GRU) is a type of recurrent neural network (RNN) that manages sequential data and addresses the drawbacks of conventional RNNs, such as the vanishing gradient issue. Long-term dependencies in time-series and language-related tasks are especially well captured by GRUs. The reset gate and the update gate are the two main gating mechanisms used in a GRU. The reset gate determines how much of the past should be forgotten, while the update gate controls how much of the previous hidden state is retained. GRUs produce a simpler and more computationally efficient model than LSTMs because they integrate the hidden state and memory into a single vector rather than using a separate memory cell, enabling faster training. Fig. 4 depicts the structure of a GRU cell, developed

to deal with gradient explosion and vanishing gradients. The GRU operates on a pseudo-sequence of static feature tokens (one token per selected variable), learning interactions across the feature dimension rather than temporal dynamics.

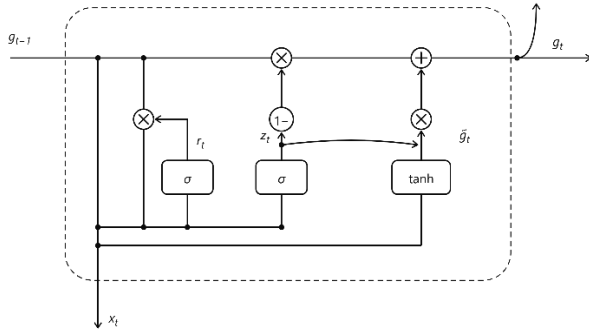


Fig. 4. Gated recurrent unit.

3) *Stacked gated recurrent unit*: The standard GRU architecture is extended by a Stacked Gated Recurrent Unit (SGRU), which increases network depth to model more intricate temporal relationships in sequential data, as shown in Fig. 5. Several GRU layers are stacked on top of one another, with the output of one layer acting as the input for the subsequent layer. The model can learn hierarchical features at various levels of abstraction, which improves its ability to capture both short-term and long-term dependencies. Stacked GRUs are especially helpful for time-series forecasting, medical diagnosis, and speech recognition. Because of their more straightforward gating mechanism, GRUs remain computationally efficient compared to other recurrent models such as LSTMs, even with the additional depth.

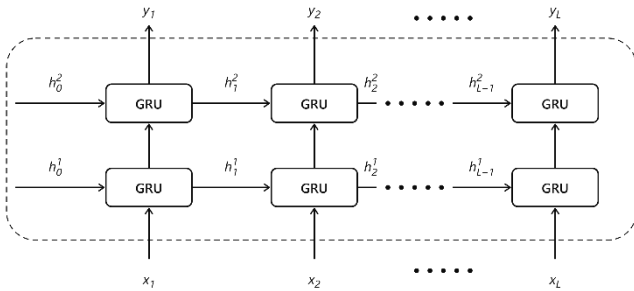


Fig. 5. Structure of Stacked Gated Recurrent Unit (SGRU).

4) *Proposed MSGRU model*: The MSGRU classifier uses the chosen features as input to determine whether the data is normal or abnormal. As seen in Fig. 6, MSGRU is a deep framework composed of multiple stacked GRU units. The first layer of the MSGRU algorithm uses the original data as input, and each subsequent MSGRU unit receives its input from the hidden layer of the GRU unit above it.

$$y_t^i = \sigma(V_z^i \cdot [g_{t-1}^i, g_t^{i-1}]) \quad (16)$$

$$s_t^i = \sigma(V_r^i \cdot [g_{t-1}^i, g_t^{i-1}]) \quad (17)$$

$$\tilde{g}_t^i = \tanh(V^i \cdot [s_t^i \odot g_{t-1}^i, g_t^{i-1}]) \quad (18)$$

$$g_t^i = y_t^i \odot g_{t-1}^i + (1 - y_t^i) \odot \tilde{g}_t^i \quad (19)$$

As shown in Eq. (16) to (19) the MS-GRU's update gate y_t and reset gate s_t serve to control the flow of information at time t , where y_t represents the input at time t , V_z is the weight of the update gate, V_r is the weight of the reset gate, and $g_{(t-1)}$ is the hidden layer's output. The candidate hidden state $\tilde{g}_t$ at time t is calculated using historical time information that is retained.

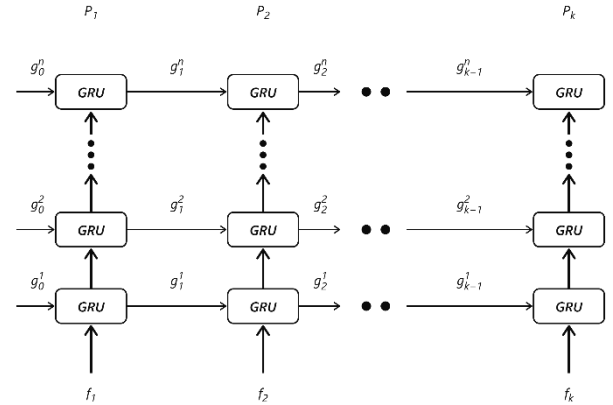


Fig. 6. Structure of MS-GRU model.

The MS-GRU employs a refine-gate technique to improve memory retention and to keep or update critical information while learning; the refinement is applied to the reset gate r'_t in the GRU as in Eq. (20).

$$r'_t = \sigma(V_{ry}t + Z_r g_{t-1} + c_r) \diamond y_t \quad (20)$$

Specifically, the refinement procedure is not suited for the update gate due to gradient explosion during hidden-state learning, since the corresponding gradient term grows without bound because it is a product of diagonal terms over the relevant time steps; the resulting gradient explosion occurs directly in the update-gate term. Applying an isolated alteration in which the reset gate is modified produces the effective refine gate before the standard update gate is applied in an MS-GRU cell, where the superscript denotes the i^{th} MS-GRU unit and the subscript denotes time t . The final GRU unit adds a sigmoid layer as the classifier to produce the output as per Eq. (21).

$$\tilde{x}_{last} = \sigma(V_0^n g_0^n + c_0^n) \quad (21)$$

In this formula, $\tilde{x}_{last}$ represents the projected identifier at the end of the i^{th} sample, V_0^n is the weight of the final layer, and c_0^n represents the bias of the n^{th} GRU component. The ability to efficiently extract high-level features from sparse data is one benefit of such a deep structure; it can also better utilize information in the time series, improving classifier performance, though the precision-versus-time trade-off results from the time-independent nature of the stacked-GRU model parameters. Techniques like the Refine Gate improve memory retention and optimize training data flow, while the MS-GRU's stacked design increases its ability to extract features. The MS-GRU successfully assesses whether a case of NAFLD is normal or abnormal by combining these state-of-the-art techniques. Training configuration of the proposed MSGRU model are

shown in Table V. The model was trained for a maximum of 100 epochs with early stopping (patience = 10) on validation loss, and the checkpoint with the lowest validation loss was retained for test evaluation.

TABLE V. TRAINING CONFIGURATION OF THE PROPOSED MSGRU MODEL

Parameter	Value
Architecture	Stacked GRU layers + Refine Gate
Number of stacked GRU layers	3
Hidden units per layer	64, 32, 16
Activation function	tanh (GRU gates), sigmoid (output layer)
Learning rate	0.001
Batch size	32
Number of epochs	100
Optimizer	Adam
Dropout / regularization	Dropout 0.2 + weight decay 1e-5
Early-stopping criterion	Patience 10 on validation loss
Loss function	Binary cross-entropy
Input representation	Pseudo-sequence of ESHO-selected feature tokens
Output	Sigmoid probability (0 = alive, 1 = dead)

IV. RESULTS AND DISCUSSION

The experimental results, as shown in Table VI, provide an evaluation of the proposed approach's performance in classifying NAFLD patients' mortality status based on clinical data. Metrics such as accuracy, precision, recall, and F1-score are used to assess the model's effectiveness. In summary, the proposed MS-GRU model outperforms existing state-of-the-art techniques on several important criteria, demonstrating its resilience and promise as an effective tool for classifying the status of non-alcoholic fatty liver disease patients. The epoch versus loss curve of the proposed Modified Stacked Gated Recurrent Unit is shown in Fig. 7, of SGRU shown in Fig. 8 and of GRU in Fig. 9.

Table VI compares the proposed MS-GRU against SVM, GRU, and SGRU baselines using Accuracy, Precision, Recall, and F1 Score. SVM yields the lowest performance (87.54% accuracy, 85.33% precision, 87.01% recall, 86.16% F1), while GRU improves to 89.12%, 89.90%, 87.23%, and 88.54%, respectively, with the precision gain indicating reduced false positives. SGRU further improves to 90.32% accuracy, 91.45% precision, 89.86% recall, and 90.65% F1. The proposed MS-GRU outperforms all baselines across every metric, achieving 92.87% accuracy, 91.97% precision, 90.69% recall, and 91.32% F1—gains of 2.55, 0.52, 0.83, and 0.67 percentage points over the strongest baseline (SGRU). The monotonic trend SVM → GRU → SGRU → MS-GRU, together with the balanced precision - recall profile of MS-GRU, confirms that the improvement is consistent rather than metric-specific, validating the effectiveness of the proposed architecture.

The ablation study demonstrates that integrating SMOTE, ESHO feature selection, and the refine gate progressively improves model performance, achieving the highest accuracy of

92.34 per cent and F1 Score of 91.32 per cent with the proposed MSGRU model. These results, shown in Table VII and Fig. 10, highlight the contribution of the combined components to the overall classification performance.

TABLE VI. RESULT ANALYSIS

Model	Accuracy (%)	Precision	Recall	F1 Score
SVM	87.54	85.33	87.01	86.16
GRU	89.12	89.90	87.23	88.54
SGRU	90.32	91.45	89.86	90.65
MS-GRU (Proposed)	92.87	91.97	90.69	91.32

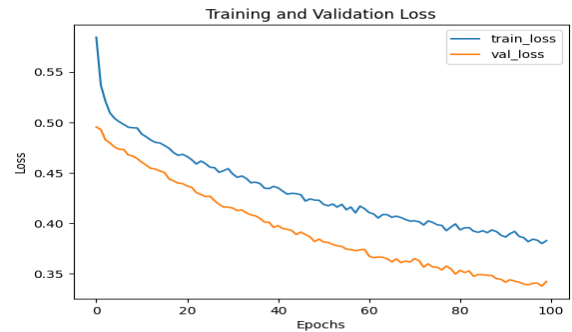


Fig. 7. Epoch vs. Loss of MSGRU.

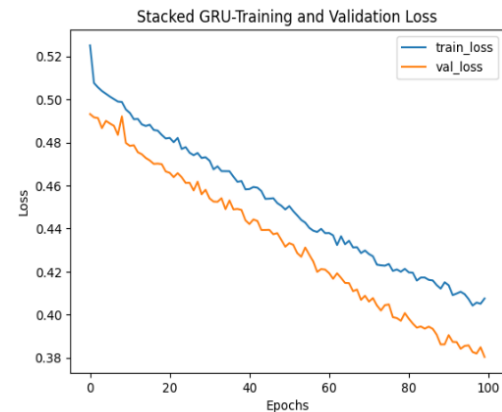


Fig. 8. Epoch vs. Loss of SGRU.

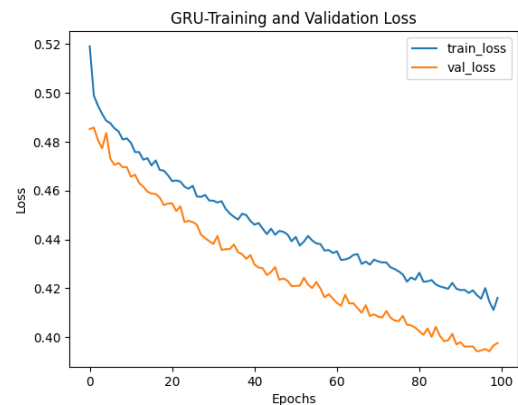


Fig. 9. Epoch vs. Loss of GRU.

TABLE VII. ABLATION STUDY

Configuration	Test Accuracy (%)	Test F1-score
Base GRU (no SMOTE, no ESHO, no Refine Gate)	84.23	83.32
GRU + SMOTE only	86.76	86.34
GRU + SMOTE + ESHO feature selection	89.12	88.54
Stacked GRU + SMOTE + ESHO (no Refine Gate)	90.32	90.65
Full Proposed MSGRU (SMOTE + ESHO + Refine Gate)	92.34	91.32

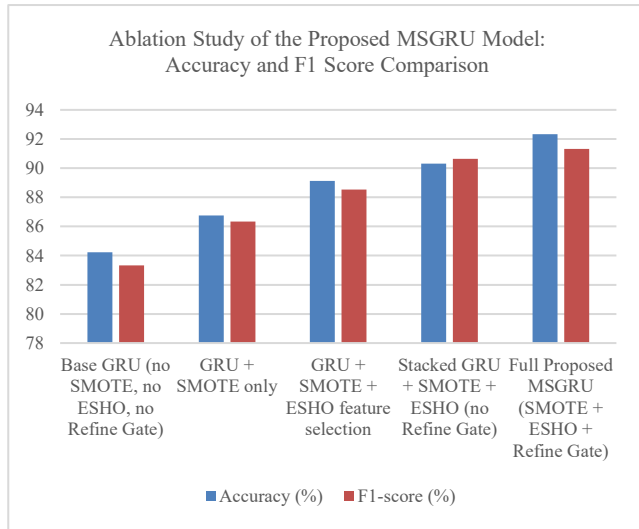


Fig. 10. Ablation Study of the Proposed MSGRU Model: Accuracy and F1 Score Comparison.

V. CONCLUSION

In this work, we present a novel approach for classifying the mortality status of patients with non-alcoholic fatty liver disease using Modified Stacked Gated Recurrent Unit (MS-GRU). Clinical Data with Demographic Features Such as Age, BMI, Height, and Weight Were Pre-processed through Imputation and Standardization to Handle Missing Values and Normalize Variables. The Enhanced Spotted Hyena Optimizer (ESHO) was then used for feature selection, retaining the most relevant clinical features and discarding redundant ones to reduce dimensionality. Classification was performed using the proposed MS-GRU technique with a refine gate mechanism. an ablation study confirmed the contribution of each component: the base GRU achieved 84.23% test accuracy and 83.32% F1, improving to 86.76% and 86.34% with SMOTE, 89.12% and 88.54% with ESHO feature selection, and 90.32% and 90.65% with stacked GRU; the full proposed MS-GRU (SMOTE + ESHO + Refine Gate) attained the best performance at 92.34% test accuracy and 91.32% F1-score, outperforming the SVM, GRU, and SGRU baselines. Future research could incorporate multi-modal data sources such as ultrasound or MRI imaging, genomic information, and lifestyle factors to further improve prediction accuracy, while integrating Explainable AI (XAI) techniques could increase model transparency and foster greater clinician confidence in the system's decisions.

CONFLICT OF INTEREST

There is no conflict of interest.

AUTHOR CONTRIBUTIONS

Jayashree A. Patil: Conceptualization, Methodology, Software implementation, analysis, Writing original draft, Visualization.

Shwetambari Chihwane: Conceptualization, Validation, Writing – review and editing, Supervision.

Varsha Lokare: Review and Guidance.

ETHICS STATEMENT

This study used a publicly available, de-identified secondary dataset [20]; no new human or animal data were collected, and ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

The dataset supporting the findings of this study is publicly available on Kaggle at <https://www.kaggle.com/datasets/utkarshx27/non-alcoholic-fatty-liver-disease> as cited in reference [20].

ACKNOWLEDGMENTS

We used AI tools to assist with language refinement in some sections of this paper. All scientific analysis and conclusions were conducted and verified by the authors.

REFERENCES

- [1] M. Q. Xu, J. J. Dai, Z. S. Jiang, F. Xu, L. Wang, W. J. Zhang and Z. G. Guo, "Preoperative combined prediction models have superior capability in predicting survival as the Child-Pugh grade in patients with HCC after interventional embolotherapy," *Cancer Management and Research*, pp. 12537–12547, 2020.
- [2] N. Hosseini, J. Shor and G. Szabo, "Alcoholic hepatitis: a review," *Alcohol and Alcoholism*, vol. 54, no. 4, pp. 408–416, 2019.
- [3] A. Rambaldi, C. Gluud and Cochrane Hepato-Biliary Group, "Colechicine for alcoholic and non-alcoholic liver fibrosis and cirrhosis," *Cochrane Database of Systematic Reviews*, vol. 2009, no. 1, 1996.
- [4] P. S. Anvith and P. Pragna, "Fatty liver disease in-depth analysis," *American Journal of Pharmaceutical Research*, vol. 5, no. 11, 2015.
- [5] E. A. Tsochatzis, J. Bosch and A. K. Burroughs, "Liver cirrhosis," *The Lancet*, vol. 383, no. 9930, pp. 1749–1761, 2014.
- [6] D. C. Rockey, S. H. Caldwell, Z. D. Goodman, R. C. Nelson and A. D. Smith, "Liver biopsy," *Hepatology*, vol. 49, no. 3, pp. 1017–1044, 2009.
- [7] F. Vernuccio, R. Cannella, T. V. Bartolotta, M. Galia, A. Tang and G. Brancatelli, "Advances in liver US, CT, and MRI: moving toward the future," *European Radiology Experimental*, vol. 5, no. 1, p. 52, 2021.
- [8] A. Özdil and B. Yilmaz, "Medical infrared thermal image-based fatty liver classification using machine and deep learning," *Quantitative InfraRed Thermography Journal*, vol. 21, no. 2, pp. 102–119, 2024.
- [9] S. Dargan, M. Kumar, M. R. Ayyagari and G. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1071–1092, 2020.
- [10] Pilankar, Atharva, and L. Shyamala. "Liver disease prediction using ensemble learning." In 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 589-595. IEEE, 2024..
- [11] Z. Li, L. Lan, Y. Zhou, R. Li, K. D. Chavin, H. Xu, L. Li, D. J. Shih and W. J. Zheng, "Developing deep learning-based strategies to predict the risk of hepatocellular carcinoma among patients with non-alcoholic fatty

- liver disease from electronic health records,” *Journal of Biomedical Informatics*, vol. 152, p. 104626, 2024.
- [12] A. Singh, P. Nath, V. Singhal, D. Anand, S. Verma and T. P. Hong, “A new clinical spectrum for the assessment of non-alcoholic fatty liver disease using intelligent methods,” *IEEE Access*, vol. 8, pp. 138470–138480, 2020.
- [13] A. G. Mabrouk, A. Hamdy, H. M. Abdelaal, A. G. Elkattan, M. M. Elshourbagy and H. A. Y. Alansary, “Automatic classification algorithm for diffused liver diseases based on ultrasound images,” *IEEE Access*, vol. 9, pp. 5760–5768, 2021.
- [14] K. Prakash and S. Saradha, “A deep learning approach for classification and prediction of cirrhosis liver: non-alcoholic fatty liver disease (NAFLD),” in *Proceedings of the 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2022, pp. 1277–1284.
- [15] A. Kulkarni, S. Shinde and D. Kadam, “Automated prediction of non-alcoholic fatty liver disease using machine learning algorithms,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 9, pp. 488–491, 2020.
- [16] M. Zhou, X. Liu, Y. Wu, Q. Xiang and R. Yu, “Liver lipidomics analysis revealed the protective mechanism of ZuoguiJiangtangQinggan formula in type 2 diabetes mellitus with non-alcoholic fatty liver disease,” *Journal of Ethnopharmacology*, vol. 329, p. 118160, 2024.
- [17] R. Yu, Y. Huang, X. Hu and J. Chen, “Analysis of machine learning-based integration to identify the crosslink between inflammation and immune response in non-alcoholic fatty liver disease through bioinformatic analysis,” *Heliyon*, vol. 10, no. 14, 2024.
- [18] R. Naseem, S. Khalid, F. Javed, M. U. Khan, A. R. Malik and A. H. Qureshi, “Performance assessment of classification algorithms on early detection of liver syndrome,” *Journal of Healthcare Engineering*, pp. 1–13, 2020.
- [19] S. S. Parvatikar, “Detection of liver diseases using classification models,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 7, 2021.
- [20] Kaggle, “Non-alcohol fatty liver disease dataset,” [Online]. Available: <https://www.kaggle.com/datasets/utkarshx27/non-alcohol-fatty-liver-disease>. [Accessed: Nov. 4, 2025].
- [21] A. M. Allen, T. M. Therneau, J. J. Larson, A. Coward, V. K. Somers and P. S. Kamath, “Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: a 20 year community study,” *Hepatology*, vol. 67, no. 5, pp. 1726–1736, 2018.
- [22] G. Huang, Q. Jin and Y. Mao, “Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: prospective cohort study,” *Journal of Medical Internet Research*, vol. 25, no. 1, e46891, 2023.
- [23] R. Shafiha, B. Bahcivanci, G. V. Gkoutos and A. Acharjee, “Machine learning-based identification of potentially novel non-alcoholic fatty liver disease biomarkers,” *Biomedicines*, vol. 9, no. 11, p. 1636, 2021.
- [24] A. Octavianlara, M. A. Abbas, A. Azhari, D. Riana and A. S. Hewiz, “Performance comparison of three classification algorithms for non-alcoholic fatty liver disease patients using data mining tool,” *Journal of Medical Informatics Technology*, vol. 6, no. 2, 2023.
- [25] J. H. Lee, K. A. Kim, Y. S. Park, M. J. Choi, D. S. Lee and H. J. Kim, “Machine learning classifiers for screening nonalcoholic fatty liver disease in general adults,” *PubMed*, [PMID: 36869105], 2023.
- [26] S. Wu, M. Zhan, X. Zhao, T. Li, Y. Guo and J. Zhang, “Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices,” *PubMed*, [PMID: 36973382], 2023.
- [27] D. Kim, S. Park, J. Lim, H. Cho, J. Lee and K. Lee, “Ultrasound-based machine learning approach for detection of nonalcoholic fatty liver disease,” *PubMed*, [PMID: 36807314], 2023.
- [28] H. Lee, J. Yu, M. Park, T. Han, S. Choi and J. Kim, “Improving nonalcoholic fatty liver disease classification performance with latent diffusion models,” *PubMed*, [PMID: 38062049], 2024.
- [29] Y. Zhang, C. Li, L. Wang, X. Zhao, M. Liu and T. Chen, “Identification and classification of non-alcoholic fatty liver disease in a large-scale population: a machine learning-based framework,” *PubMed*, 2022.