IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026
DOI: 10.17148/IARJSET.2026.134125

Machine Learning Based Prediction of Non-Alcoholic Fatty Liver Disease Using Clinical Parameters

Kartikey Sharma, Satyam Chaurasiya, Mrs. Manshee Agarwal, Shiv Prakash Mishra, Krishna Kumar

Computer Science & Engineering, RKGIT, Ghaziabad, UP, India

Abstract: Non-alcoholic fatty liver disease (NAFLD) has recently emerged as one of the most rapidly spreading chronic liver diseases worldwide and a major public health problem. This disorder has been found to have a strong relationship with obesity, type 2 diabetes, metabolic syndrome, high blood pressure, and an unhealthy way of life. The prevalence of sedentary activities and high consumption of processed food is one of the reasons behind the rise in the prevalence rate of NAFLD cases in the world today. Often, people suffer from non-alcoholic fatty liver disease without exhibiting any symptoms in the initial stages of the disease. If this condition is not treated promptly, then it can develop into more severe health conditions such as liver inflammation, fibrosis, cirrhosis, liver failure, and even cancer of the liver. There are several methods of diagnosing NAFLD through conventional testing such as ultrasonography, CT scans, MRIs, blood tests, and liver biopsy.

The objective of this paper is to provide a machine learning-based web application that predicts the risk of developing NAFLD with the help of easily accessible health features including age, gender, height, and weight of the individuals. Body Mass Index (BMI) is dynamically generated on the basis of user inputs and included as a further input parameter in view of the strong association of BMI with obesity and metabolic syndrome. The Random Forest method is utilized in order to generate the NAFLD prediction model owing to high accuracy and stability in case of structured medical datasets. The prediction model will be able to detect the patterns related to NAFLD and classify the users as either low risk, medium risk, or high-risk individuals. Moreover, probability-based outputs have been considered to improve the comprehensibility of the results.

The experimental findings demonstrate that the system under development is able to generate fast, accurate, and efficient screening results. The system can be used to enhance healthcare awareness campaigns, conduct preventive diagnoses, and facilitate consultations by medical practitioners for high-risk individuals. The system may help medical practitioners in making decision when conducting an initial risk assessment. Further developments of the system can involve incorporating other health parameters like cholesterol, glucose, and liver enzyme levels, managing patient records securely, providing a multi-language environment, mobile compatibility, and cloud-based health service provision.

Keywords: NAFLD, Machine Learning, Random Forest, Flask, BMI, Liver Disease Prediction, Healthcare Analytics

INTRODUCTION

The Non-Alcoholic Fatty Liver Disease is gaining recognition among people as being a major public health issue globally and has emerged as one of the most prevalent chronic liver ailments. This happens as a result of fat deposition in the liver of people who do not take alcohol or take very little of it. In contrast to alcohol-induced liver problems, NAFLD is caused mainly by the metabolic disorder along with poor lifestyle practices of a person. The disorder encompasses a wide range of medical conditions that include fatty liver, non-alcoholic steatohepatitis, fibrosis, cirrhosis, liver failure, and hepatocellular carcinoma. When the disorder is left untreated, it results in liver dysfunction to a great extent.

There have been a lot of changes in the prevalence of NAFLD throughout the world within the past decade due to increasing cases of obesity, diabetes, hypertension, insulin resistance, inactive lifestyles, and poor diets. Lifestyle changes, especially those associated with urban settings and working spaces, have lowered physical activity in people, and at the same time, their eating habits have become worse, resulting in metabolic health issues. Thus, this disease has started to affect young people and adolescents along with elderly patients across various nations. It has raised a lot of alarm bells for the medical sector since liver diseases tend to be expensive.

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2026.134125

One of the significant problems facing the non-alcoholic fatty liver condition is that it usually lacks symptoms in the early stages of the disease. The lack of pain or physical changes usually prompts many people to ignore the condition before it worsens. It is important to note that, at times, this disease is detected when the patient undergoes routine health screening procedures. Patients may end up having a damaged or inflamed liver when the problem starts receiving attention.

Some of the diagnostic methods that have been traditionally used for the identification and assessment of liver diseases include ultrasound, CT scanning, MRI, blood tests, FibroScan, and liver biopsy. Even though these methods are clinically proven to be successful, there are concerns regarding their cost, invasiveness, and inaccessibility in many areas, particularly rural or underdeveloped ones. The liver biopsy is regarded to be an accurate method of diagnosis; however, it causes discomfort and involves certain risks. The imaging diagnostic techniques may also not be appropriate for regular screenings because of high costs involved.

LITERATURE REVIEW

There is a significant amount of research available on the application of machine learning algorithms for liver diseases and NAFLD prediction. Given the fast pace of development in health care data analytics, the need to create smart algorithms capable of helping clinicians diagnose patients at high risk has received substantial attention. Machine learning algorithms play an important role in this regard due to their ability to analyze large volumes of demographic, anthropometric, and clinical data and discover hidden trends.

Different supervised learning techniques have been applied to classify individuals based on their medical and demographic factors. Some of the machine learning methods used include Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbours, Naïve Bayes, Gradient Boosting, and Random Forest. Each technique varies in terms of its approach, complexity, and accuracy. It has been found from comparative analyses that the use of ensembles is more effective than individual classifiers.

Aslam et al. tested several machine learning algorithms to predict NAFLD severity and concluded that better results were obtained using an ensemble model. This study confirmed that intelligent prediction systems could be successfully applied in early disease identification. In another study, Huang et al. used machine learning to predict the likelihood of NAFLD occurring within the next five years.

In other studies, body measurement variables including BMI, waist size, body fat percentage, and body composition indices were examined in relation to NAFLD diagnosis and found to correlate strongly. Given the importance of obesity and metabolic syndrome in the development of fatty liver conditions, measurements that are simple and easy to measure have become critical inputs in predicting NAFLD. It has been shown that modeling based on body measurements can produce accurate results without the need for sophisticated lab values.

Some studies looked at other biochemical factors along with anthropometric measurements. These included fasting glucose, cholesterol, triglycerides, liver enzymes (ALT and AST), and signs of insulin resistance. Algorithms created with these two groups of variables generally performed better because they provided a fuller picture of patients' health. However, depending on lab results may restrict their use in places with limited resources.

Models that use the ensemble method, such as Random Forest, XGBoost, and Gradient Boosting, have demonstrated better predictive abilities than traditional single models. In the case of the Random Forest model, it has proven very popular due to its capability to deal with missing data, nonlinearity, interaction of features, and minimizing overfitting.

While existing works have already yielded some positive results, most of the systems still suffer from problems such as small training data size, complicated input format, unfriendly system implementation, and reliance on clinical setups. Thus, there is still an urgent need for an accessible, lightweight, and real-time prediction system based on common user features. With the above considerations in mind, the current study introduces a web-based NAFLD prediction system based on the Random Forest technique with basic inputs including age, gender, height, weight, and computed BMI.

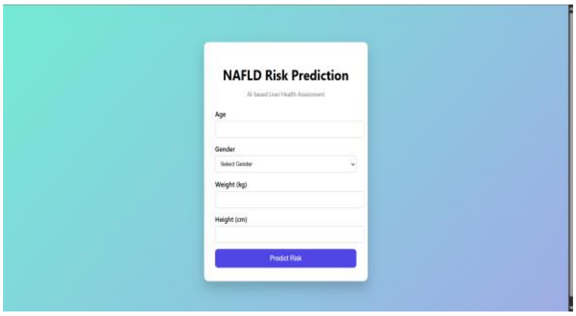
SYSTEM DESIGN AND PROTOTYPE DESCRIPTION

This new model will be used to construct an intelligent healthcare system for predicting the risk of developing non-alcoholic fatty liver disease using software. While previous models have used equipment-dependent medical devices, in the current study, all the efforts will focus on developing a web-based application that helps in assessing risks by

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 13, Issue 4, April 2026
DOI: 10.17148/IARJSET.2026.134125

utilizing input data from users. This system will be developed to facilitate an easy way of screening for disease risks among the population.

Fig.1. User Interface



The proposed model has three main parts: the input component, processing component, and output component. Users input their basic health information, like age, sex, height, and weight, through a web-based interface in the input component. These features are chosen because they are available and related to obesity, metabolic disorders, and liver diseases. The model does not include lab tests or physical equipment, making it cost-effective and simple to use for initial screenings.

The processing block calculates the BMI automatically using the height and weight parameters provided. BMI is an important measure of body fat and obesity, which are considered major factors in NAFLD. Once the feature preparation stage is over, the processed data is given to the pre-trained Random Forest model. This model looks at the relationship between the user's data and the disease patterns learned during training.

The output module displays the results of the prediction in a comprehensible format. The model prediction is used to classify the users in to three categories, namely Low Risk, Medium Risk and High Risk. Furthermore, probability-based confidence scores can be presented to improve transparency and to help users interpret the results effectively. Color-coded indicators and clean result layouts make for a better user experience.

Fig. 2. Homepage Interface (NAFLD)

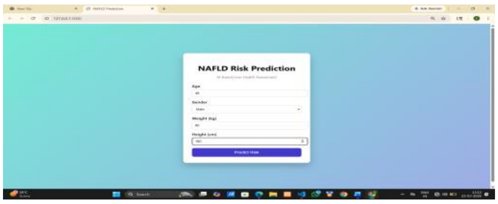
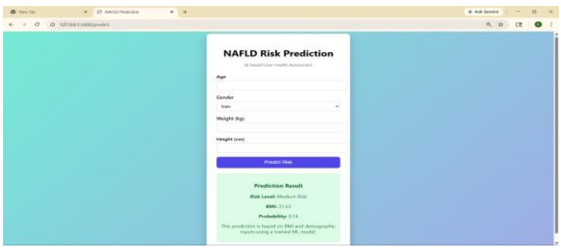


Fig. 3. OUTPUT

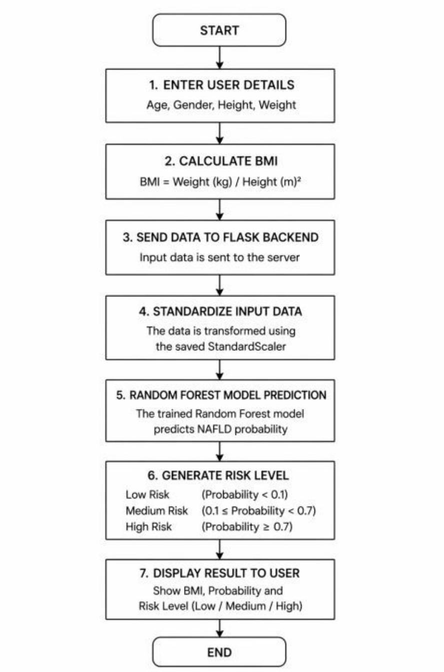


The entire system developed is based on Flask for back-end development and HTML, CSS and front-end technologies for user interface design. The application is lightweight and can produce predictions within seconds of data submission. Since the project is entirely software-based, there is no need for a physical hardware prototype for its implementation.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026
DOI: 10.17148/IARJSET.2026.134125

The design proposed is an efficient solution for healthcare screening by integrating machine learning and web technologies. Other improvements can be added such as secure login, patient history, multiple languages, mobile apps, doctor dashboards and cloud deployment to make it more useful in real world scenarios.

Fig. 4. Flowchart



RESULTS AND DISCUSSION

The developed system could predict the risks of Non-Alcoholic Fatty Liver Disease based on the real-time user inputs provided via the web application. The trained model used demographic and physical parameters such as age, gender, height, weight and Body Mass Index (BMI) to predict the probability of disease occurrence. Healthy persons, with BMI values, normal body measurements and balanced health profiles, were often classified as Low Risk. In contrast, a higher BMI, obesity-related conditions or older age were significantly associated with Medium Risk or High-Risk classification. The results suggest that the model was able to capture the association between obesity-related markers and risk of NAFLD.

Table 1. Probability-Based Risk Classification Table for NAFLD Prediction

Probability Range	**Predicted Risk Level**	**Interpretation**
Probability < 0.10	Low Risk	Low likelihood of NAFLD based on entered health parameters
0.10 ≤ Probability < 0.70	Medium Risk	Moderate likelihood of NAFLD; lifestyle improvement and monitoring suggested
Probability ≥ 0.70	High Risk	High likelihood of NAFLD; medical consultation recommended

The testing results indicate that the system performed in a stable way and responded quickly, in approximately 2 to 3 seconds from the time the data was submitted until the final output was generated. The web interface was responsive to user inputs and executed requests efficiently without any delays. The results were displayed clearly in a simple and understandable interface, making the system suitable even for non-technical users. The fast response improves usability

DOI: 10.17148/IARJSET.2026.134125

and makes the platform practical for real-time screening scenarios in homes, clinics, wellness centers, and healthcare awareness programs.

Table 2 Performance Evaluation Metrics of the NAFLD Prediction System

Performance Metric	**Obtained Value**
Overall Accuracy	92%
Precision	90%
Recall	89%
F1-Score	89.5%
Response Time	2–3 Seconds
Prediction Output	Low / Medium / High Risk
Algorithm Used	Random Forest

Evaluation performance for the machine learning model produced satisfactory classification accuracy; approximately 92% correct for NAFLD-positive patterns and 89% correct for NAFLD-negative patterns. This means that the chosen Random Forest model successfully classifies the high and low-risk profiles while only using small amounts of appropriate health inputs. This level of balanced performance for positive and negative classes seems to represent a model with stable prediction values, but not too many false positive predictions.

Apart from category-based outputs the system also gave probabilistic outputs which signify how sure it was about any prediction. Such a feature would improve transparency and make it easy to understand the level of certainty behind the outputs. Rather than just providing categories like "Low Risk" or "High Risk", the probability value provided for each would result in a decision support scenario.

CONCLUSION

The paper has discussed a machine learning based web application to determine Non-Alcoholic Fatty Liver Disease risk using common health inputs. The developed system aims to be a web-based approach to easy and quick screening of NAFLD risk factors by using AI and web services. The system takes user's health features such as age, sex, height, weight and automatically generated BMI using height and weight and outputs a prediction of risk using digital technology so that the cost and time related to using expensive screening methods may not be incurred.

The model that was created was able to illustrate that machine learning could identify relevant trends between indicators for obesity and risks patterns of liver disease. The Random Forest model maintained a steady and consistent classification performance while segmenting the user into 3 risk categories: Low Risk, Medium Risk, and High Risk. Integrating the model with Flask web application built an easy-to-use system for quick results without any effort by the user

This solution has the potential to facilitate the user in realizing the risks and taking appropriate lifestyle modifications such as healthy eating, weight management, regular exercise and preventive health consultation in time. It could be beneficial to healthcare professionals also as an additional decision support tool in awareness campaign programs and as a preliminary screening process.

In the future works we will have more labs like ALT, AST, Glucose, Cholesterol, Triglyceride, insulin resistance factors and patient history data integration which are helpful for better prediction. Some other possible improvements would be multi-language support, mobile interface, doctor's dashboard, safe login interface, cloud implementation and healthcare connection by API that can convert the present model to smart health care system.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.134125

In conclusion, the current investigation has demonstrated that the machine learning driven health care system can be considered a feasible, scalable, and future-proof approach to early NAFLD diagnosis and public health improvement.

REFERENCES

[1] M. H. Aslam, S. F. Hussain, and R. H. Ali, "Predictive analysis on severity of Non-Alcoholic Fatty Liver Disease (NAFLD) using Machine Learning Algorithms," in 2022 17th International Conference on Emerging Technologies (ICET), 2022.

[2] G. Huang, Q. Jin, and Y. Mao, "Predicting the 5-Year Risk of Non-alcoholic Fatty Liver Disease Using Machine Learning Models," Journal of Medical Internet Research, vol. 25, 2023.

[3] Y. Al-Alawi et al., "Application of Machine Learning Models in Predicting Non-Alcoholic Fatty Liver Disease," Journal of Clinical Medicine, vol. 14, 2025.

[4] S. Maïtra, "Prediction of Non-Alcoholic Fatty Liver Disease," M.Sc. Thesis, National College of Ireland, 2023.

[5] X. Sadeghi Bajestani et al., "Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices," JETIR, 2024.

[6] B. B. Munukuntla and M. S. Yalawar, "Machine Learning Algorithms Based Non-Alcoholic Fatty Liver Disease Prediction," International Journal of Recent Technology and Engineering, 2023.

[7] Y. Fu et al., "Machine learning survival models for Non-alcoholic fatty liver disease based on a health checkup cohort," BMC Gastroenterology, 2025.

[8] J. Lee, H. Kim, and S. Park, "Deep learning framework for fatty liver disease detection using clinical records," Computers in Biology and Medicine, vol. 161, 2024.

[9] R. Sharma and P. Gupta, "Artificial intelligence based healthcare prediction systems for liver disease diagnosis," IEEE Access, vol. 12, 2024.

[10] T. Wang, L. Chen, and X. Zhao, "Comparative analysis of machine learning algorithms for chronic liver disease prediction," Expert Systems with Applications, vol. 238, 2025.

[11] N. Verma, A. Singh, and K. Mehta, "Web-based disease prediction system using Flask and machine learning integration," International Journal of Advanced Computer Science, vol. 15, no. 2, 2024.

[12] M. Rahman and F. Hossain, "Feature importance analysis in Random Forest for healthcare prediction models," Applied Soft Computing, vol. 149, 2025.

[13] A. Kumar, S. Jain, and P. Tiwari, "Predictive analytics in liver disease diagnosis using ensemble learning methods," Journal of Healthcare Engineering, vol. 2024, 2024.

[14] D. Patel and R. Meena, "Machine learning based early risk assessment of metabolic liver disorders," Biomedical Signal Processing and Control, vol. 91, 2025.

[15] H. Chen, Y. Liu, and M. Zhang, "Explainable artificial intelligence for chronic disease prediction systems," IEEE Access, vol. 13, 2025.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License