2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A New Machine Learning-Based Complementary Approach for Screening of NAFLD (Hepatic Steatosis)

Suranjan Panigrahi[1], Ridhi Deo[2], and Edward A. Liechty[3]

Abstract— **Non-Alcoholic Fatty Liver Disease (NAFLD) is the major reason for liver disease globally. Early warning of liver disease at the beginning of a progressive disease spectrum is critical for reduced mortality and increased longevity. Current clinical practices focus on disease management but can be improved in terms of screening & early detection. This paper focuses on machine learning-based intelligent model development using liver functionality and physiological parameters for Hepatic Steatosis (Non-alcoholic Fatty Liver) screening. Gender-specific models were developed separately. Customized data processing techniques were incorporated. Publicly available, population data (NHANES-III) was used. The maximum sensitivity provided by the models were approximately 72% and 71% for male and female, respectively. Maximum specificities obtained by the models were 74% and 75% for male and female, respectively. Performance comparison of different models has been discussed.**

I. Introduction

Non–Alcoholic Fatty Liver Disease (NAFLD) is a series of liver conditions in individuals without any history of significant alcohol consumption [1]–[3]. The range of NAFLD varies from simple fatty liver to steatohepatitis [1]–[3]. Hepatic Steatosis (HS) or Non-Alcoholic Fatty Liver (NAFL) is a condition in which an individual's liver accumulates fat but does not have any inflammation or injury [1], [3]. When left undiagnosed or untreated, HS can advance to Non-Alcoholic Steatohepatitis (NASH), potentially leading to liver fibrosis, cirrhosis, or carcinoma.

NAFLD is the largest cause of liver conditions around the world [1]. Because of the nature of the disease, and the lack of data related to it, current literature reflects the estimates of NAFLD prevalence in the population. A 2016 study estimated the global NAFLD prevalence as 25.24% and within the USA, 80 million (estimated) individuals are affected by NAFLD [4].

Literature indicates that multiple risk factors such as obesity, sedentary lifestyle, diabetes (type II), insulin resistance, and hyperlipidemia can contribute to NAFLD [1], [5]. However, the exact cause for the disease is not known. In addition, NAFLD does not show any symptoms in the early stages but can progress undetected over time [3], [6]. The progression of NAFLD to advanced stages is linked to a higher risk of other terminal liver disease and/or high mortality [7]. Further, there is no specific cure for NAFLD yet [3]. Current guidelines recommend disease management via lifestyle intervention [1]. Despite its chronic nature, if detected early, NAFLD progression can be slowed down or managed for cure (partial or complete reversal) [8], [9]. Studies show that reversal of the disease to a certain extent is also possible, if detected at or before a specific stage in the progression [10]. Early detection and disease reversal have many benefits - improved quality of life, reduced chance for a liver transplant, increased longevity, and reduced mortality [10]. Subsequently, this results in increased social and economic benefits to the society.

Although there are benefits to early detection, current clinical guidelines do not recommend screening for NAFLD, due to a lack of cost-effective screening methods [1], [11]. Further, there are no specific biomarkers for NAFLD detection [1], [2]. Typically, serum-based liver function tests (LFTs) are performed to monitor liver health. LFTs measure the levels of enzymes and hormones (i.e., ALT (alanine aminotransferase), AST (aspartate transaminase), ASP (alkaline phosphatase), Albumin, GGT (gamma-glutamyl-transpeptidase), and Bilirubin). Note that the variations of LFTs and relevant physiological parameters (HDL, triglycerides, etc.) are not specific to NAFLD [12].

Currently, there are two potential pathways of NAFLD detection in normal adults (with no symptoms related to liver disease). They are: 1) For adults residing in regions where annual checkups for liver functionality (Liver Function Tests or LFTs) are conducted, the health care practitioners could recommend an individual for further testing (i.e. Ultrasound or MRI) based on LFT results. 2) For adults residing in regions where current practices do not involve proactive, regular health screening, detection of NAFLD occurs randomly while being diagnosed/screened/tested for other health conditions [3], [6], [13]. In both the above pathways, the health care practitioner recommends the suspected NAFLD cases for further investigation via additional testing/screening (ultrasound, MRI, etc.) based on the liver functionality and related physiological parameters of an individual. However, this recommendation is subjective among healthcare practitioners. In this condition, the practitioner makes the decision solely based on available individual data (liver functionality, physiological data, etc.). We postulate that the relationship of HS with liver functionality and physiological parameters determined from population data could provide valuable insight about the disease etiology. This might also be useful in the decision-making process/screening by the healthcare practitioner.

In recent years, a combination of data access, computational power, and advanced mathematical/analytical techniques has contributed to intelligent models for healthcare

[1] S. Panigrahi is a Professor at the School of Engineering Technology at Purdue University, West Lafayette, IN, 47907, USA (spanigr@ purdue.edu).
[2] R. Deo is a PhD Fellow at the School of Engineering Technology at Purdue University, West Lafayette, IN, 47907, USA (deor@purdue.edu)
[3] E. A. Liechty is a Professor Emeritus of Pediatrics at Indiana School of Medicine, Indianapolis, IN, 46202, USA (eliecht@iu.edu)

applications, including increased understanding of the associated cause and effect relationships [14]–[16]. The goal of this study is to investigate the association of HS (NAFL) with liver functionality and physiological parameters using a population dataset (NHANES). Based on our analysis, no such investigation has been conducted before.

II. PROCEDURE

In this paper, a publicly available dataset - National Health and Nutrition Exam Survey (NHANES) III (1988 – 1994), provided by the Centers for Disease Control and Prevention was used [17]. NHANES III data are available publicly for use by researchers. The institutional review board (IRB) at Purdue University determined the data used in this paper to be exempt (category-4: IRB-2020-1309). In the following paragraphs, data extraction, data analysis, and model training methods are explained.

A. Data extraction

Adult data (persons aged 18 and above) from the NHANES III dataset were used in this paper. The data extraction was conducted using a customized program ("Initial extraction module") using SAS 9.4 and further analysis was conducted using customized programs ("Analysis module") developed in MATLAB R2020b.

The initial extraction module retrieved twelve features from NHANES III. They are: gender, age, BMI, HDL, plasma glucose, AST, ALT, and ASP. Exclusion criteria were applied using three alcohol-related features to fit the criteria of NAFLD. "Hepatic Steatosis (HS)" was used as the output feature. The HS determination for individuals is further described in NHANES III documentation [17]. The parameter "Sequential numbers" from the NHANES III dataset was also extracted for internal cross-referencing. In this model, the observations with missing information in any of the features were eliminated. This logic has advantages over the conventional approach of estimating missing feature values in the dataset.

B. Data Analysis

The "Initial extraction module" was developed in SAS 9.4 to combine the data files from NHANES III. The features of interest were then exported into MATLAB R2020b for further analysis using the "Analysis module".

The observations with missing information in any of the selected features were deleted. This reduced the dataset size from 20,050 observations to 12,195. Next, the dataset was split into two sub-datasets based on HS (yes or no). The size of the HS sub-dataset was 2,956 and the size of the no-HS sub-dataset was 9,959. Subsequently, each of the two sub-datasets was split further by gender. This logic is supported by findings from the literature on the varied implication of NAFLD pathogenesis between the genders [18]–[20]. Then, the gender-specific alcohol exclusion criteria defined by the American Association of Study of Liver Diseases (AASLD) were applied. The criteria defines "significant" alcohol consumption as consuming >3 drinks/day and >2 drinks/day for men and women, respectively [1]. Consequently, 437 Male HS, 1,228 Male no-HS, 195 Female HS, and 716 female no-HS observations were excluded.

The flowchart for the "Analysis module" is shown in figure 1. Four derived features were created using the equations (1) – (4). An ALT value of > 33 units/liter (u/l) and > 25 u/l is considered above normal for male and female, respectively [21]. The upper limits of normal (ULN) for AST in male and female are 30 u/l and 20 u/l, respectively [22]. The normal level for BMI and diabetes were 25 kg/m^2 and 120 mg/dL, respectively for both the genders [23], [24]. In the context of HS, higher BMI (>25) is considered as a major risk factor [25]. Thus, we used 25 kg/m^2 as the ULN for BMI.

$$ALT_i\% = \frac{ALT_i - ALT\_ULN}{ALT\_ULN}\ x\ 100. \quad (1)$$

$$AST_i\% = \frac{AST_i - AST\_ULN}{AST\_ULN}\ x\ 100. \quad (2)$$

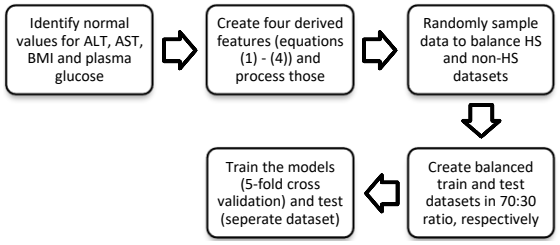
$$BMI_i\% = \frac{BMI_i - 25}{25}\ x\ 100. \quad (3)$$

$$Plasma\ glucose_i\% = \frac{Plasma\_glucose_i - 120}{120}\ x\ 100. \quad (4)$$

i = ith sample in the dataset; ALT_ULN: 33u/l - male, 25 u/l - female; AST_ULN: 30 u/l – male, 20 u/l – female

The negative values in the derived features were replaced with 0, to indicate "normal". The positive values were left unchanged.

Figure 1. Analysis Module - Flowchart



Class balance is essential prior to training machine learning models to avoid poor sensitivities, specifically from a healthcare perspective [26]. Data from the majority class (no-HS) were randomly sampled to match the size of the minority class (HS) in both genders, separately. The class balanced sub-datasets for male and female had 1,038 and 1,212 samples each, respectively, in the HS and no-HS categories. The male class-balanced training and testing sub-dataset sizes were 1,454 and 622, respectively. The female class-balanced training and testing sub-dataset sizes were 1,696 and 728 respectively. In this paper, we also assessed an extension of the above model that used imbalanced data (male HS:1,038; male no-HS:3,177; female HS:1,212; female no-HS:4,591).

C. Model training

Machine Learning (ML) models (class-balanced) from multiple families (total 17 models): tree-based, ensemble-based (random forest, boosted trees), K-nearest neighbors, support vector machines (SVM) and logistic regression, were analyzed.

The best performing models (SVMs, in this case) are reported. SVM models are suitable for binary classification (e.g. HS: yes/no). They classify data by using an optimal hyperplane and can be implemented using a linear or non-linear kernel [27].

TABLE I. PERFORMANCE METRICS– SUMMARY OF 10 RUNS

SVM Models	Performance Metrics of the class-balanced models							
	Training Accuracy (%) ± SD		*Test Accuracy (%) ± SD*		*Test Sensitivity (%) ± SD*		*Test Specificity (%) ± SD*	
Gender	M[a]	F[b]	M	F	M	F	M	F
Linear	69.360 ± 0.007	71.385 ± 0.011	68.553 ± 0.011	70.865 ± 0.016	65.852 ± 0.022	68.104 ± 0.026	71.254 ± 0.026	73.626 ± 0.023
Quadratic	68.954 ± 0.009	71.362 ± 0.010	68.778 ± 0.014	71.002 ± 0.014	65.562 ±0.02	66.620 ± 0.027	71.993 ± 0.021	75.384 ± 0.027
Gaussian scale 1	66.334 ± 0.010	69.504 ± 0.013	66.559 ± 0.011	69.148 ± 0.011	72.122 ± 0.023	70.659 ± 0.02	60.996 ± 0.028	67.637 ± 0.021
Gaussian scale 2	69.133 ± 0.009	71.462 ± 0.011	68.987 ± 0.018	70.659 ± 0.012	66.463 ± 0.015	67.032 ± 0.02	71.511 ± 0.027	74.285 ± 0.015
Gaussian scale 3	68.954 ± 0.007	71.409 ± 0.009	68.794 ± 0.011	70.178 ± 0.009	63.826 ± 0.021	67.225 ± 0.015	73.762 ± 0.026	73.131 ± 0.02

[a]Male, [b]Female

In this study, SVMs with linear, quadratic, and Gaussian kernels were implemented. Within the Gaussian kernel-based models, three separate kernel scales (0.83, 3.3, and 13) were incorporated. All SVM models used a box constraint level of 1 to avoid overfitting. The models were trained using separate training data with 5-fold cross-validation to enhance generalization. The trained models were then tested using a separate test dataset. The training and testing processes were repeated 10-times to compute the average performance metrics. The 10-run averages of the class-balanced models are summarized in table I. Moreover, the results of the extension models using imbalanced data were analyzed.

III. RESULTS AND DISCUSSION

Performance of the SVM (Linear, Quadratic and Gaussian-Scales 1,2,3) models are shown in table I. For male-specific models, the average testing accuracies ranged from approximately 67 - 69 % with a maximum standard deviation (SD) of 0.018. The average sensitivity ranged from 64 - 72% with a maximum SD of 0.023. The average specificity ranged from 61 - 74% with a maximum SD of 0.028. The best performing model (male) was the Gaussian Scale-II with 69% accuracy. The model with Gaussian Scale I provided 72% sensitivity (male). For this specific screening application, it is preferable to emphasize sensitivity over specificity. Hence, the Gaussian scale I model would be preferred.

For female-specific models, the average testing accuracies ranged from approximately 69% - 71% with a maximum SD of 0.016. The sensitivity ranged from 67% - 71% with a maximum SD of 0.027. Although the maximum specificity of 75% was obtained with the quadratic SVM model, the Gaussian Scale I model with 71% sensitivity is preferred, for the above-explained reasons.

The result from the imbalanced-models (test-data) shows very poor sensitivities (i.e., poor detection of HS cases). The average accuracy of the imbalanced models increased slightly – with a maximum of 78% and 81% for male and female, respectively. Interestingly, the specificities were very high with a maximum of 98% and 99% for male and female, respectively. However, the sensitivities were in the range of 6-23% and 6 - 18% for male and female, respectively. This information indicated the lack of usefulness of these imbalanced models - thus they were not pursued further. Detailed results of the imbalanced models are not shown in this paper.

The findings from the class-balanced models demonstrate the potential of an intelligent model relating liver functionality and physiological parameters with HS. The performance of the models could be improved by using other modeling techniques and additional relevant data. Once developed, tested, and validated, the proposed intelligent model(s) could serve as a new, complementary approach for NAFLD screening. We postulate that such a validated model(s) could be used as a decision support system by the healthcare practitioners at the point-of-care settings.

The performance of this model is similar to the model(s) in our previous work [15]. The previous work involved different predictor features and utilized a synthetic data generation approach for balancing the output classes (HS). The novelty of the model(s) (current paper) is in the use of original data without any synthetic generation. Therefore, we are pursuing this approach.

IV. CONCLUSION AND FUTURE WORK

In this paper, machine learning-based (SVM), intelligent models (separate for male and female) for NAFLD screening were evaluated. The male-specific model demonstrated approximately 72% sensitivity and 74% specificity. Similarly, the female-specific model provided 71% and 75% sensitivity and specificity, respectively.

NAFLD screening and diagnosis is a complex issue with no specific biomarkers but with multiple metabolic co-morbidities. Our current and future work involves developing and identifying advanced models for NAFLD screening. In parallel, we are testing the developed models on different population datasets. We might also consider including the underweight BMI category (<18.5 kg/m^2) in the algorithm of our model. We envision that such a complementary screening approach could contribute to detect NAFLD at an early stage, in a proactive manner. Subsequently, this effort could contribute to achieve better health outcomes and enhanced quality of life for people.

REFERENCES

[1] N. Chalasani *et al.*, "The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases: Hepatology," *Hepatology*, vol. 67, no. 1, pp. 328–357, Jan. 2018, doi: 10.1002/hep.29367.

[2] T. G. Cotter and M. Rinella, "Nonalcoholic Fatty Liver Disease 2020: The State of the Disease," *Gastroenterology*, vol. 158, no. 7, pp. 1851–1864, May 2020, doi: 10.1053/j.gastro.2020.01.052.

[3] "Nonalcoholic Fatty Liver Disease & NASH | NIDDK." https://www.niddk.nih.gov/health-information/liver-disease/nafld-nash (accessed Apr. 26, 2021).

[4] Z. M. Younossi, A. B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, and M. Wymer, "Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes," *Hepatology*, vol. 64, no. 1, pp. 73–84, Jul. 2016, doi: 10.1002/hep.28431.

[5] V. Ratziu, M. Ghabril, M. Romero-Gomez, and G. Svegliati-Baroni, "Recommendations for Management

and Treatment of Nonalcoholic Steatohepatitis," *Transplantation*, vol. 103, no. 1, pp. 28–38, Jan. 2019, doi: 10.1097/TP.0000000000002483.

[6] "Nonalcoholic fatty liver disease - Diagnosis and treatment - Mayo Clinic." https://www.mayoclinic.org/diseases-conditions/nonalcoholic-fatty-liver-disease/diagnosis-treatment/drc-20354573 (accessed Mar. 31, 2021).

[7] P. S. Dulai *et al.*, "Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis," *Hepatol. Baltim. Md*, vol. 65, no. 5, pp. 1557–1565, May 2017, doi: 10.1002/hep.29085.

[8] S. H. Caldwell and C. K. Argo, "Reversing Advanced Hepatic Fibrosis in NASH: Clearly Possible, but Widely at Hand?," *Dig. Dis. Sci.*, vol. 60, no. 4, pp. 810–812, Apr. 2015, doi: 10.1007/s10620-015-3540-0.

[9] J. Serpaggi *et al.*, "Direct and indirect evidence for the reversibility of cirrhosis," *Hum. Pathol.*, vol. 37, no. 12, pp. 1519–1526, Dec. 2006, doi: 10.1016/j.humpath.2006.07.007.

[10] P. Stål, "Liver fibrosis in non-alcoholic fatty liver disease - diagnostic challenge with prognostic significance," *World J. Gastroenterol. WJG*, vol. 21, no. 39, p. 11077, Oct. 2015, doi: 10.3748/wjg.v21.i39.11077.

[11] S. Leoni, F. Tovoli, L. Napoli, I. Serio, S. Ferri, and L. Bolondi, "Current guidelines for the management of non-alcoholic fatty liver disease: A systematic review with comparative analysis," *World J. Gastroenterol.*, vol. 24, no. 30, pp. 3361–3373, Aug. 2018, doi: 10.3748/wjg.v24.i30.3361.

[12] A. R. Murali and W. D. Carey, "Liver Test Interpretation - Approach to the Patient with Liver Disease: A Guide to Commonly Used Liver Tests," *Cleveland Clinic - Center for Continuing Education*, Apr. 2014. https://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/hepatology/guide-to-common-liver-tests/ (accessed Oct. 27, 2020).

[13] "NASH Diagnosis — American Liver Foundation." https://liverfoundation.org/for-patients/about-the-liver/diseases-of-the-liver/nonalcoholic-steatohepatitis-information-center/nash-diagnosis/ (accessed Mar. 16, 2021).

[14] C.-C. Wu *et al.*, "Prediction of fatty liver disease using machine learning algorithms," *Comput. Methods Programs Biomed.*, vol. 170, pp. 23–29, Mar. 2019, doi: 10.1016/j.cmpb.2018.12.032.

[15] R. Deo and S. Panigrahi, "Prediction of Hepatic Steatosis (Fatty Liver) using Machine Learning," in *Proceedings of the 2019 3rd International Conference on Computational Biology and Bioinformatics - ICCBB '19*, Nagoya, Japan, 2019, pp. 8–12. doi: 10.1145/3365966.3365968.

[16] H. Ma, C. Xu, Z. Shen, C. Yu, and Y. Li, "Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China," *BioMed Res. Int.*, vol. 2018, pp. 1–9, Oct. 2018, doi: 10.1155/2018/4304376.

[17] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data, "NHANES III (1988-1994) - Data Files." https://wwwn.cdc.gov/nchs/nhanes/nhanes3/datafiles.aspx#core (accessed Apr. 23, 2021).

[18] A. Lonardo *et al.*, "Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps," *Hepatology*, vol. 70, no. 4, pp. 1457–1469, Oct. 2019, doi: 10.1002/hep.30626.

[19] S. Ballestri, F. Nascimbeni, E. Baldelli, A. Marrazzo, D. Romagnoli, and A. Lonardo, "NAFLD as a Sexual Dimorphic Disease: Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk," *Adv. Ther.*, vol. 34, no. 6, pp. 1291–1326, 2017, doi: 10.1007/s12325-017-0556-1.

[20] M. Balakrishnan *et al.*, "Women Have a Lower Risk of Nonalcoholic Fatty Liver Disease but a Higher Risk of Progression Vs Men: A Systematic Review and Meta-analysis," *Clin. Gastroenterol. Hepatol.*, p. S1542356520306121, Apr. 2020, doi: 10.1016/j.cgh.2020.04.067.

[21] P. Y. Kwo, S. M. Cohen, and J. K. Lim, "ACG Clinical Guideline: Evaluation of Abnormal Liver Chemistries," *Am. J. Gastroenterol.*, vol. 112, no. 1, pp. 18–35, Jan. 2017, doi: 10.1038/ajg.2016.517.

[22] "Understand Liver Enzyme Test Results — American Liver Foundation." https://liverfoundation.org/understand-liver-enzyme-test-results-2/ (accessed Mar. 30, 2021).

[23] "Risk Factors for Diabetes | NIDDK," *National Institute of Diabetes and Digestive and Kidney Diseases*. https://www.niddk.nih.gov/health-information/professionals/clinical-tools-patient-management/diabetes/game-plan-preventing-type-2-diabetes/prediabetes-screening-how-why/risk-factors-diabetes (accessed Mar. 30, 2021).

[24] "Hyperglycemia in diabetes - Diagnosis and treatment - Mayo Clinic." https://www.mayoclinic.org/diseases-conditions/hyperglycemia/diagnosis-treatment/drc-20373635 (accessed Mar. 30, 2021).

[25] "Association between body mass index and fatty liver risk: A dose-response analysis | Scientific Reports." https://www.nature.com/articles/s41598-018-33419-6 (accessed Jul. 27, 2021).

[26] V. Jain and J. M. Chatterjee, Eds., *Machine learning with health care perspective: machine learning and healthcare*. Cham: Springer, 2020.

[27] "Support Vector Machine Classification - MATLAB & Simulink." https://www.mathworks.com/help/stats/support-vector-machine-classification.html (accessed May 03, 2021).