Machine learning for liver biopsy outcome prediction

Parash Ghale
School of Electronic Electrical Engineering and Computer Science, Queen's University Belfast
Belfast, Northern Ireland, United Kingdom
parashghale96@gmail.com

Abstract— Non-Alcoholic Fatty Liver Disease (NAFLD) is a widespread liver condition having significant health implications. It is rapidly becoming one of the most common liver diseases worldwide, closely linked to obesity, diabetes and metabolic syndrome. Current diagnostic techniques are either invasive, costly or impractical for large-scale screening. Machine learning offers a powerful solution for detecting NAFLD early by analyzing datasets with demographic, clinical and laboratory data.

One of the key challenges was the class imbalance of the dataset. These imbalances skew the results, causing the model to favor majority class (healthy patients) and underperform in NAFLD detection. Another challenge was the missing data in the dataset. Proper imputation strategies were needed to ensure the critical information is not lost and the dataset remains comprehensive for accurate predictions. Also, the dataset comprised of wide range of clinical and demographic features making it difficult to balance simplicity and accuracy. Identifying the most relevant features while maintain model interpretability is another significant challenge.

A robust predictive model capable of accurately identifying NAFLD cases was resulted by the integration of various techniques. The model demonstrated significant potential in aiding early diagnosis and improving patient outcomes by providing a reliable, data-driven approach to NAFLD detection. The best machine learning algorithm was Random Forest. The use of K-Nearest Neighbor (KNN) was found out to be optimum solution for imputation. Generative Adversarial Networks (GANs) was crucial to overcome the class imbalance in the dataset.

This work highlights the efficacy of advanced machine learning methods in handling complex, real-world healthcare datasets.

Keywords: NAFLD detection, Machine learning, Imputation techniques, Early detection, Class imbalance handling

Section 1: Introduction

Non-Alcoholic Fatty Liver Disease (NAFLD) is increasingly prevalent as one of the most common liver disorders globally affecting a significant proportion of the population. It impacts up to a third of the European population and estimated prevalence of 25% worldwide [1,2,3]. A meta-analysis in China showed a national NAFLD prevalence rate of 29.2% [5,6].

NAFLD can advance to Non-Alcoholic Steatohepatitis (NASH), which may occur with or without fibrosis, increasing the risk of liver cirrhosis, end-stage liver disease and hepatocellular carcinoma (HCC) [3]. Around 33% patients with NASH progress to fibrosis and 15% go on to develop cirrhosis [3,4]. NASH is currently the second most frequent reason for liver transplantation in the United States and is expected to surpass hepatitis C infection as the primary cause for liver transplants in the future [7,8]. Although only 10-20% of NAFLD patients progress to NASH, those who do are at a heightened risk for developing fibrosis, cirrhosis, HCC and liver-related mortality [9]. NAFLD is projected to increase direct and indirect healthcare costs by approximately 26% over a five-year period [25,26]. NAFLD is now regarded as a manifestation of metabolic syndrome, with its connections to obesity, type 2 diabetes, and cardiovascular disease likely stemming from common underlying pathogenetic mechanisms which makes it a critical public health concern [26,27,28].

Early detection and accurate diagnosis of NAFLD are dominant to preventing the progression of the disease to more serious liver condition. However, its diagnosis is often challenging due to the asymptomatic nature of the condition in its early stages. NAFLD is evaluated using a 7-stage system, which involves a thorough assessment of perisinusoidal and periportal fibrosis. However, liver biopsy, the standard method for this evaluation, carries a limited but notable risk of complications such as bleeding and infection. Additionally, there is a considerable risk of observer and sampling errors, particularly when assessing fibrosis [3,23,24]. The gold standard for NAFLD diagnosis is liver biopsy. Nevertheless, the significant side effects and the potential for sampling errors with this technique highlights the urgent need to identify reliable diagnostic biomarkers for the disease [10,29,30]. Ultrasonography is a noninvasive, fairly accurate and commonly used method for diagnosing NAFLD; yet, it lacks sensitivity for detecting mild steatosis [10,11].
Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are commonly used as indicators of liver function. Previous research has shown a correlation between ALT levels and NAFLD, with even normal-range ALT values linked to an increased risk of NAFLD. Additionally, the ALT/AST ratio can be used to assess the extent of liver fat accumulation and hepatic steatosis [17,18,19]. The hepatic steatosis index (HIS), which includes factors such as ALT, AST, BMI, gender and diabetes history, has proven effective for screening NAFLD [10,12]. But the use of these surrogate markers in large-scale data has not been extensively

documented [10]. Traditional diagnostic methods, including liver biopsy, imaging techniques and serum biomarkers are either invasive, expensive or lack sufficient sensitivity and specificity. These drawbacks highlight the need for reliable, non-invasive and cost-effective methods for early NAFLD detection.

This research is situated within the broader area of medical diagnostics and machine learning, focusing on leveraging advanced computational techniques to address challenges in NAFLD detection. Specifically, the early detection of NAFLD is the problem being investigated utilizing the development of a robust machine learning framework. In the third National Health and Nutrition Examination Survey (NHANES) in the United States, it was found that liver steatosis could be identified by ultrasound in 7.4% of non-obese adults [17,31], and in Asia the prevalence was 8-19% [17,32]. The difference in NAFLD prevalence between Asian and Western populations indicates the presence of potentially significant metabolic risk factors in individuals with normal BMI. Identifying these factors could aid in the early prevention of NAFLD in people with normal BMI. Furthermore, early detection of non-obese NAFLD may lower the risk of developing cardiovascular disease and diabetes [17,36-38]. The frequent occurrence of incomplete clinical data and significant class imbalance between healthy individuals and those affected by the disease are the predominant factors for difficulty in diagnosing NAFLD. New tests incorporating cell death markers, developed for assessing fibrosis in NAFLD, are gaining attention but still need validation through large-cohort studies [3,15]. Validation studies have shown that individual biomarkers were unable to accurately predict fibrosis [3,33]. As a result, diagnostic multipanel tests, incorporating various individual parameters, have recently been proposed as non-invasive methods for accessing fibrosis [3,34,35]. This project utilizes advanced machine learning techniques, including Random Forest, Generative Adversarial Networks (GAN), and K-Nearest Neighbors (KNN), to enhance NAFLD detection. In essence machine learning follows four key steps: defining the problem, collecting and preparing data, building the model and making predictions with the model [10,20,21]. These steps are implemented in this research.

The primary aim of this MSc AI project is the application of an existing machine learning model which is capable of accurately detecting NAFLD using clinical and demographic data. The main scope of this project is to explore and implement advanced machine learning techniques to improve predictive accuracy for detecting NAFLD. It aims to evaluate the effectiveness of machine learning models, particularly Random Forest, by measuring performance metrics such as sensitivity, specificity, and overall accuracy. The ultimate goal is to enhance predictive modeling for early detection and better clinical decision-making in NAFLD detection and choose the optimum algorithm for our case. The study is confined to the data and techniques relevant to these approaches.

The specific research questions that this project seeks to answer are: Is the algorithm used show promise in early detection of NAFLD? Can the use of GAN-generated synthetic data effectively mitigate class imbalance and how does it impact on the overall performance of machine learning models? What is the comparative performance among Random Forest, Balanced Random Forest and LightGBM models in terms of diagnostic accuracy, sensitivity and specificity? Which of these machine learning models demonstrate the best trade-off between handling class imbalance and maintain high diagnostic accuracy in NAFLD detection? These research questions help to explore the effectiveness of the chosen techniques and models in addressing the challenges in context of NAFLD detection.

The methodology adopted in this project involves the preprocessing of a real-world clinical dataset, the implementation of KNN imputation for missing data handling, application of GANs to address class imbalance and development and evaluation of various machine learning algorithms for AFLD detection. The models are evaluated based on several metrics including accuracy, sensitivity and specificity. These metrics provide a comprehensive assessment of the model's ability to accurately detect NAFLD. The research employs a combination of data-driven techniques and experimental validation to address the research questions.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive literature review, highlighting relevant studies and methodologies in the context of NAFLD detection and machine learning. Section 3 comprises of the methodology, including data preprocessing, model development and application of various imputation and oversampling techniques along with machine learning models. Section 4 presents the experimental setup and results discussing the model's performance and the impact of the proposed models. Finally, Section 5 concludes the paper by summarizing the findings, discussing their implications and suggesting directions for future research.

SECTION 2: LITERATURE REVIEW

In this section, we review relevant studies that have explored machine learning approaches in the past for NAFLD detection. Unlike previous research, which primarily focuses on epidemiological predictions, this research specifically addresses real-world data challenges like missing data and class imbalance.

D.Y.Z. Lim et al. [1] explored machine learning models for prediction of NAFLD in large Asian cohorts for early detection and risk assessment. The study employed advanced machine learning techniques but failed to fully address the issue of class imbalance which is critical in medical datasets where occurrence of the disease is relatively low. The authors identified the need for more robust methods to handle imbalanced data, which is a key focus of this project through the use of GANs.

S. W. Wong et al. [2] and C.K. Argo & S. H. Caldwell [4] provided public health insights into NAFLD and its progression to more severe liver conditions, such as hepatocellular carcinoma. These studies underscored the importance of early and accurate detection of NAFLD to prevent progression. However, the traditional diagnostic tools provided are limited by their invasiveness and difficulty in capturing early stage of NAFLD. This highlights the need for non-invasive machine learning based diagnostic approaches.

J. -P. Sowa et al. [3] developed a novel algorithm for noninvasive assessment of fibrosis in NAFLD, using machine learning to enhance the accuracy of existing methods. While this study made significant strides in improving non-invasive diagnostics, it mainly focused on fibrosis rather than early NAFLD detection. Also, the approach did not incorporate advanced techniques to handle missing data, which is addressed in this project through KNN imputation.

Many studies have applied machine learning models to predict the progression and risk of NAFLD. G. Huang et al. [5] predicted the 5-year risk of NAFLD using various machine learning models for long-term risk assessment. However, the study's reliance on imbalanced datasets without sufficient correction methods limits its predictive accuracy. Similarly, A. Atsawarungruangkit et al. [7] employed machine learning models to predict NAFLD in U.S. population but models were not optimized to handle missing data or class imbalance which are addressed in this research.

H. Ma et al. [10] and J. Calvert et al. [9] explored the application of machine learning technique like Random Forest and other ensemble methods, for clinical predictive modeling. These studies showed the potential of machine learning in improving diagnostic accuracy. But they also highlighted challenges posed by imbalanced data, which can lead to biased predictions. The use of Random Forest proposed in this project aims to mitigate these issues.

The reviewed literature identifies several gaps in current approaches to NAFLD detection. Many studies do not adequately address the combination of missing data handling and class imbalance leading to suboptimal model performance. While machine learning models have showed promise, there is lack of studies that integrate advanced imputation and oversampling techniques to enhance model performance.

This project aims to bridge these gaps by implementing KNN imputation for missing data, utilizing GANs for generating synthetic NAFLD-positive cases and applying advanced machine learning models like Random Forest, Balanced Random Forest and LightGBM. This approach is expected to improve diagnostic accuracy and offers a more comprehensive solution for NAFLD detection.

SECTION 3: METHODS

This section outlines the specific steps and procedures used to conduct the research. The dataset details, AI model architectures, training procedures, hyperparameter tuning and ethical considerations are covered in this step.

Dataset details:

The dataset used in this research was sourced from a hospital's clinical database focusing mainly on liver conditions such as Cirrhosis, Fibrosis and the three stages of Steatosis. The dataset comprised of 1629 individual records each with 20 features including demographic data, liver function indicators, lipid profile and other clinical parameters. The dataset was highly imbalanced with many of the columns appearing to have null values of unnamed suggesting either missing data or placeholders for future data entry.

Preprocessing steps:

The data cleaning process involved removing unnamed columns, handling missing values with forward fill and KNN imputation, and addressing outliers using percentile capping and Isolation Forest. The dataset was normalized using MinMax scaling to ensure equal contribution of features. Final checks confirmed no duplicates or missing values, and nonnegativity was ensured for dimensionality reduction using Nonnegative Matrix Factorization (NMF). The data cleaning process involved removing unnamed columns and those with significant missing values, converting object data types to numeric, and addressing null values using forward fill method. KNN imputation was applied to fill remaining missing values by averaging similar data points. Residual null values were replaced with the column means, preserving data structure and correlations for more accurate imputations. Outliers were identified using the 0.5th and 99.5th percentiles as bounds, and extreme values were capped using the 'clip' function. Since outliers persisted, Isolation Forest with a 5% contamination rate was applied to reduce them. The dataset was then normalized using Min-Max scaling to ensure all features contributed equally during model training. After confirming no duplicates or missing values, the dataset's non-negativity was verified before applying Non-negative Matrix Factorization (NMF) for dimensionality reduction.

Data augmentation using Generative Adversarial Networks (GANs):

The dataset used in this research was highly imbalanced, with a higher prevalence of "no fibrosis" and "no cirrhosis" cases compared to positive cases. To address this, GANs were employed to generate synthetic data for the minority classes, balancing the dataset and improving the model's ability to generalize across all classes. The dataset consisted of 1500 samples with 29 features. The GAN architecture included a generator that created synthetic data from a noise vector using a three-layer neural network, and a discriminator that identified real versus synthetic data through its own three-layer network. This approach helped mitigate the imbalance and enhanced model performance.

The training procedure involved using a GAN to generate synthetic data for underrepresented classes, balancing the dataset for classification model training. After generating the balanced data, Non-negative Matrix Factorization (NMF) was applied to reduce the feature space to 12 components. Several machine learning models were employed for classification, including Balanced Random Forest (BRF), Random Forest (RF), Logistic Regression (LR), XGBoost, and LightGBM. These models were configured with class balancing techniques and custom decision thresholds to handle the dataset's imbalance, with hyperparameter tuning using methods like GridSearchCV and RandomizedSearchCV. The models aimed to improve classification accuracy across all classes, especially for the minority classes, by leveraging techniques such as resampling, boosting, and regularization.

The training procedure involved using a GAN to generate synthetic data for underrepresented classes in the dataset, balancing it for further training of classification models. The GAN's generator was trained to produce synthetic data resembling the real data, while the discriminator aimed to distinguish real from synthetic data, both optimized using binary cross-entropy loss over 5000 epochs. After generating balanced data, Non-negative Matrix Factorization (NMF) was applied for dimensionality reduction, reducing features to 12 components.

Several AI models were employed for classification:

1. Balanced Random Forest (BRF)**:** Designed for imbalanced datasets, BRF used 100 trees, resampling to balance the classes, with hyperparameter tuning via GridSearchCV.
2. Random Forest (RF)**:** A robust ensemble model of 100 decision trees, handling complex datasets with class weighting, using RandomizedSearchCV for hyperparameter optimization.
3. Logistic Regression (LR)**:** A basic model for binary classification, using L2 regularization and class balancing, with hyperparameter tuning and a custom threshold for classification.
4. XGBoost: A gradient boosting algorithm, handling class imbalance with customized weight adjustment and threshold tuning for accurate classification.
5. LightGBM (LGB)**:** Another gradient boosting model, utilizing class balancing and boosting rounds to enhance classification performance.

Ethical considerations:

All patient data used in this research was anonymized to protect patient privacy, Identifiable information such as names, addresses and social security numbers were not included in this study.

SECTION 4: EXPERIMENTATION AND RESULTS

This section provides a comprehensive overview of the hardware and software environment used, the evaluation metrics, limitations and detailed results of the experiments conducted. The results are presented with proper analysis, compared to baseline models and discussed in the context of existing literature.

The experiments were conducted using a hardware setup consisting of a 2.6 GHz 6-core Intel Core i7 processor, 16 GB RAM, Intel UHD Graphics 630 GPU, and 500 GB storage, running on the Sonoma 14.3 operating system. The software environment included Python 3.8 within Jupyter Notebook (Anaconda distribution), utilizing several key machine learning libraries. Scikit-learn (0.24.1) was used for implementing Logistic Regression, Random Forest, and Balanced Random Forest models, while XGBoost (1.3.3) and LightGBM (0.8.0) were employed for their respective classifiers. TensorFlow (2.4.1) was leveraged for generating synthetic data using GANs, and Numpy (1.19.5) and Pandas (1.2.3) facilitated numerical computations and data manipulation. This comprehensive setup supported the execution and analysis of the machine learning models and experiments.

The model struggled to capture sensitivity (recall) for positive cases across all classes before the imputation of missing data. While the overall accuracy appeared high, this was misleading because the model primarily achieved increased precision and recall for negative cases, which were more abundant in the dataset. In cases of imbalanced data, accuracy can be deceptive, as it may reflect the model's ability to correctly classify the dominant negative class while missing many positive cases. The missing data likely caused the model to overlook key features or misinterpret patterns necessary for detecting positive cases, leading to a significant drop in recall for those instances. Consequently, the high accuracy was primarily due to the model's bias towards predicting the negative class, masking its poor performance in identifying positive cases, which is critical in many real-world applications. After imputation there was a significant increase in sensitivity for the positive classes.

Optimum machine learning model:

Among various models evaluated for predicting liver biopsy outcomes, including Logistic Regression, Balanced Random Forest (BRF), XGBoost, LightGBM (LGB), and Random Forest, the Random Forest model demonstrated the best performance across all liver biopsy conditions. Logistic regression had lower sensitivity (recall) for the negative classes. XGBoost, although had high accuracy and recall for negative classes, it struggled to capture the sensitivity of positive classes. LightGBM also performed similar to XGBoost with a slight increase in sensitivity of positive classes.

The Random Forest model outperformed other models in predicting liver biopsy outcomes. It achieved high accuracy across all conditions: 90% for Cirrhosis LB**,** 91% for Fibrosis LB, 87% for Mild Steatosis LB, 88% for **Severe** Steatosis LB**,** and 89% for Moderate Steatosis LB. The model effectively balanced precision, recall, and F1-scores, but it predicted

negative classes more accurately than the positive ones. To increase the model's prediction for positive classes hyperparameter tuning is done.

After hyperparameter tuning using Randomized Search CV:

After tuning the models using Randomized Search CrossValidation, the performance in predicting liver biopsy outcomes showed significant improvement, achieving an overall weighted accuracy of 86%.

The model performed well across various liver biopsy conditions, with strong recall values indicating a good ability to detect true positive cases. For Mild Steatosis LB and Severe Steatosis LB, the recall was also solid at 76% and 77%, respectively. For Cirrhosis LB and Fibrosis LB it was 77% and 74%, with Moderate Steatosis LB showing the highest recall of 89%, highlighting the model's effectiveness in identifying moderate cases. Despite the high recall, precision was lower, especially for Cirrhosis LB (60%) and Severe Steatosis LB (59%), reflecting a higher rate of false positives. Overall, the model demonstrates strong sensitivity, particularly in detecting positive cases, while maintaining a balanced accuracy across all conditions.

The tuning process particularly enhanced recall, although there was a decrease in precision, across various NAFLD-related conditions, reflecting the models' effectiveness in identifying true cases.

In the context of NAFLD detection, maintaining high recall is crucial to ensure that no potential cases are missed, particularly since early intervention can prevent the progression of the disease to more severe liver conditions. Although lower precision results in more false positives, this is considered an acceptable trade-off, as the risk of missing true cases and allowing the disease to progress is far greater than the inconvenience of follow-up tests to rule out false positives.

Section 5: discussions

This study examined the effectiveness of KNN imputation, the impact of GAN-generated synthetic data on class imbalance, and the comparative performance of Random Forest, Balanced Random Forest, and LightGBM models in detecting Non-Alcoholic Fatty Liver Disease (NAFLD).

KNN Imputation and Dataset Quality

Prior to imputation, the model showed comparatively lower performance, with reduced precision, recall, and F1-scores, particularly for the minority classes. The application of KNN imputation resulted in better overall accuracy, precision, recall, and F1-scores across all classes, indicating that the imputation process helped the model better capture the underlying patterns in the data. The confusion matrices also reflect fewer misclassifications after imputation, especially for the minority class, where performance was previously suboptimal. This demonstrates that KNN imputation effectively addressed missing data, enhancing model performance and providing more balanced predictions.

KNN imputation significantly improved dataset completeness, which enhanced model performance in detecting NAFLD. Given the importance of early detection in NAFLD, improving the completeness of the dataset directly impacted the models' sensitivity to early disease stages, making KNN a reliable tool for handling incomplete medical data. The resulting increase in accuracy and recall demonstrates that KNN imputation is an effective method for handling incomplete medical data, ensuring more reliable machine learning models.

Synthetic Data and Class Imbalance

NAFLD datasets often suffer from class imbalance, where more severe disease states are underrepresented. The use of Generative Adversarial Networks (GANs) to generate synthetic data helped overcome this issue by creating additional samples for the minority classes. This approach significantly improved recall for underrepresented classes, ensuring that the models were more sensitive to detecting the more critical cases, such as advanced fibrosis or cirrhosis. High recall in these cases is crucial, as failing to detect early signs of progression can have serious implications for patient outcomes. By balancing the data, GAN-generated synthetic samples helped models to generalize better, increasing their robustness in real-world applications where class imbalance is common. This also aligns with the clinical need to prioritize sensitivity over specificity in the early detection of NAFLD to ensure no cases are missed.

Comparative Performance of Machine Learning Models

When comparing the different machine learning models, Random Forest emerged as the most effective overall, delivering a balanced combination of accuracy, sensitivity (recall), and specificity (precision). Random Forest's ability to handle complex datasets with many features makes it highly suitable for NAFLD detection, which involves a range of clinical indicators like ALT, AST, and their ratios. Its ensemble nature helps reduce overfitting while capturing the nuanced patterns that could indicate the early stages of liver disease. Furthermore, its performance in balancing both sensitivity and specificity ensure it can not only detect the disease but also minimize false positives, which is crucial in clinical diagnosis to avoid unnecessary treatments. Although there was a slight decrease in precision during tuning, the overall sensitivity was increased. The Random Forest model emerged as the best option for balancing class imbalance with high diagnostic accuracy, making it the most suitable model for NAFLD detection in this study.

Implications and Significance

These findings highlight the value of KNN imputation for data quality, the effectiveness of synthetic data in addressing class imbalance, and the superiority of Random Forest in balancing accuracy and sensitivity. These insights are crucial for

improving machine learning applications in medical diagnostics, particularly for NAFLD.

Limitations and Challenges

This study highlights several limitations and challenges in using machine learning models for detecting NAFLD. Data quality is crucial, as KNN imputation relies on the assumption of predictable missing data, which may introduce bias if the data is incomplete or non-random. GAN-generated synthetic data helped address class imbalance but may not fully capture realworld complexities, leading to potential overfitting. Handling class imbalance remains difficult, as improving recall for minority classes often lowers precision, which can lead to false positives in clinical settings. Model interpretability is another issue, especially with "black box" models like Random Forest, which are hard to explain in medical contexts. Additionally, the models may not generalize well to different datasets and require significant computational resources, particularly for GANs and ensemble models. Lastly, **computational complexity** and the limited testing of other advanced models (e.g., deep learning or XGBoost) suggest room for further exploration and improvement. These factors present significant challenges for real-world clinical application. Future research should address these issues and explore the scalability of these models in clinical settings.

Section 6:Conclusion and Future Work Considerations

This project aimed to enhance NAFLD detection through machine learning, focusing on improving data quality with KNN imputation, addressing class imbalance with GANgenerated synthetic data, and comparing the performance of Random Forest, Balanced Random Forest and LightGBM models.

a) Summary of Findings

The Random Forest model emerged as the best performer, balancing accuracy, sensitivity, and specificity. KNN imputation effectively improved dataset quality, leading to better model performance. GAN-generated synthetic data successfully mitigated class imbalance, particularly in detecting severe NAFLD cases.

b) Limitations and Recommendations

Key limitations include the potential for overfitting with synthetic data, the computational demands of Random Forest, and the variable effectiveness of KNN imputation depending on the data type. Future research should focus on optimizing model scalability, refining synthetic data techniques, and exploring alternative imputation methods. Validation in clinical settings is also recommended to ensure practical applicability. In conclusion, while significant progress has been made in using machine learning for NAFLD detection, further research is needed to address these limitations and enhance the models' clinical utility.

Future Recommendations

Future research should focus on optimizing the scalability of the Random Forest model, ensuring it can be applied effectively in clinical settings without being hindered by high computational demands. This could involve developing more efficient versions of the model or exploring hybrid approaches that combine the strengths of different algorithms. Additionally, refining synthetic data generation techniques is crucial thereby improving GAN models or exploring alternative methods could help generate more representative data minimizing the risk of overfitting. Further research should also explore alternative imputation methods that may be better suited to different types of missing data, ensuring that data quality improvements are robust across various scenarios. Finally, validating these models in real-world clinical settings is essential to confirm their practical applicability and effectiveness in routine medical practice, ensuring that they can be integrated into healthcare workflows to improve NAFLD detection and patient outcomes.

References

1. D. Y. Z. Lim, G. E. Chung, P. H. Cher, R. Chockalingam, W. Kim, and C. K. Tan, "Use of machine learning to predict onset of NAFLD in an all-comers cohort: Development and validation in two large Asian cohorts," *Gastro Hep Advances*, vol. 495, 2024. [Online]. Available: https://doi.org/10.1016/j.gastha.2024.06.007
2. S. W. Wong, Y. W. Ting, and W. K. Chan, "Epidemiology of nonalcoholic fatty liver disease-related hepatocellular carcinoma and its implications," *JGH Open: An Open Access Journal of Gastroenterology and Hepatology*, 2018. [Online]. Available: https://doi.org/10.1002/jgh3.12070
3. J.-P. Sowa, D. Heider, L. P. Bechmann, G. Gerken, D. Hoffmann, and A. Canbay, "Novel algorithm for non-invasive assessment of fibrosis in NAFLD," *PLoS ONE*, vol. 8, no. 5, 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0062439
4. C. K. Argo and S. H. Caldwell, "Epidemiology and natural history of non-alcoholic steatohepatitis," *Clinical Liver Disease*, vol. 13, no. 4, pp. 511–531, 2009. [Online]. Available: https://doi.org/10.1016/j.cld.2009.07.005
5. 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 Name*, vol. 2023. [Online]. Available: https://doi.org/10.2196/46891
6. F. Zhou, J. Zhou, W. Wang, X. Zhang, Y. Ji, P. Zhang, *et al.*, "Unexpected rapid increase in the burden of NAFLD in China from 2008 to 2018: A systematic review and meta-analysis," *Hepatology*, vol. 70, no. 1, pp. 111–123, 2019. [Online]. Available: https://doi.org/10.1002/hep.30702
7. A. Atsawarungruangkit, P. Laoveeravat, and K. Promrat, "Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population," *World Journal of Hepatology*, vol. 13, no. 10, pp. 1417–1428, 2021. [Online]. Available: https://doi.org/10.4254/wjh.v13.i10.1417
8. Z. Younossi, M. Stepanova, J. P. Ong, I. M. Jacobson, E. Bugianesi, A. Duseja, Y. Eguchi, V. W. Wong, F. Negro, Y. Yilmaz, M. Romero-Gomez, J. George, A. Ahmed, R. Wong, I. Younossi, M. Ziayee, and A. Afendy, "Nonalcoholic steatohepatitis is the fastest growing cause of hepatocellular carcinoma in liver transplant candidates," *Clinical Gastroenterology and Hepatology*, vol. 17, no. 4, pp. 748–755, 2019. [Online]. Available: https://doi.org/10.1016/j.cgh.2018.05.057
9. J. Calvert, S. Ghandian, R. Thapa, A. Garikipati, G. Barnes, A. Green-Saxena, Q. Mao, and R. Das, "Machine learning to predict progression of non-alcoholic fatty liver to non-alcoholic steatohepatitis or fibrosis," *Journal of Gastroenterology and Hepatology*, vol. 37, no. 7, pp. 1040–1047, 2022. [Online]. Available: https://doi.org/10.1002/jgh3.12716

10. H. Ma, C.-f. Xu, Z. Shen, C.-h. Yu, and Y. Li, “Application of machine learning techniques for clinical predictive modeling: A cross-sectional study on nonalcoholic fatty liver disease in China,” *Hindawi BioMed Research International*, vol. 2018, Article 6958381, 2018. [Online]. Available: **https://doi.org/10.1155/2018/4304376**
11. G. Bedogni, S. Bellentani, and L. Miglioli, “The fatty liver index: A simple and accurate predictor of hepatic steatosis in the general population,” *BMC Gastroenterology*, vol. 6, Article 33, 2006. [Online]. Available: https://doi.org/10.1186/1471-230X-6-33
12. H.-J. Lee, D. Kim, H. J. Kim, *et al.*, “Hepatic steatosis index: A simple screening tool reflecting nonalcoholic fatty liver disease,” *Digestive and Liver Disease*, vol. 42, no. 7, pp. 503–508, 2010. [Online]. Available: https://doi.org/10.1016/j.dld.2009.11.008
13. R. Z. Ning, C. S. Kirkness, C. V. Asche, and H. Wang, "Risk of using logistic regression to illustrate exposure-response relationship of infectious diseases," *BMC Infectious Diseases*, vol. 14, no. 1, 2014. Available: https://doi.org/10.1186/1471-2334-14-540
14. S. C. Bagley, H. White, and B. A. Golomb, "Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain," *Journal of Clinical Epidemiology*, vol. 54, no. 10, pp. 979–985, 2001.
15. **P**. Calès, F. Lainé, J. Boursier, Y. Deugnier, V. Moal, et al., "Comparison of blood tests for liver fibrosis specific or not to NAFLD," *Journal of Hepatology*, vol. 50, pp. 165–173, 2009. Available: https://doi.org/10.1016/j.jhep.2008.07.038
16. F. Vizzutti, U. Arena, V. Nobili, R. Tarquini, M. Trappoliere, et al., "Non-invasive assessment of fibrosis in non-alcoholic fatty liver disease," *Annals of Hepatology*, vol. 8, pp. 89–94, 2009. Available: https://doi.org/10.1016/S1665-2681(19)31884-0
17. Y. Zou, L. Zhong, C. Hu, and G. Sheng, "Association between the alanine aminotransferase/aspartate aminotransferase ratio and newonset non-alcoholic fatty liver disease in a nonobese Chinese population: a population-based longitudinal study," *Lipids in Health and Disease*, vol. 19, no. 1, p. 245, 2020. Available: https://doi.org/10.1186/s12944-020-01419-z
18. **A**. A. Nanji, S. W. French, and J. B. Freeman, "Serum alanine aminotransferase to aspartate aminotransferase ratio and degree of fatty liver in morbidly obese patients," *Enzyme*, vol. 36, pp. 266–269, 1986.
19. M. T. Long, A. Pedley, L. D. Colantonio, J. M. Massaro, U. Hoffmann, P. Muntner, et al., "Development and validation of the Framingham Steatosis Index to identify persons with hepatic steatosis," *Clinical Gastroenterology and Hepatology*, vol. 14, pp. 1172–1180.e2, 2016. Available: https://doi.org/10.1016/j.cgh.2016.03.034\
20. **X**. Wu, V. Kumar, J. R. Quinlan, et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008. Available: https://doi.org/10.1007/s10115-007-0114-2
21. L. Jiang, H. Zhang, and Z. Cai**,** "A novel bayes model: hidden naive bayes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 10, pp. 1361–1371, 2009. Available: https://doi.org/10.1109/TKDE.2008.234
22. C. K. Argo and S. H. Caldwell, "Epidemiology and natural history of non-alcoholic steatohepatitis," *Clinics in Liver Disease*, vol. 13, pp. 511–531, 2009. Available: https://doi.org/10.1016/j.cld.2009.07.005
23. K. Grønbaek, P. B. Christensen, S. Hamilton-Dutoit, B. H. Federspiel, E. Hage, et al., "Interobserver variation in interpretation of serial liver biopsies from patients with chronic hepatitis C," *Journal of Viral Hepatitis*, vol. 9, pp. 443–449, 2002. Available: https://doi.org/10.1046/j.1365-2893.2002.00394.x
24. V. Ratziu, F. Charlotte, A. Heurtier, S. Gombert, P. Giral, et al., "Sampling variability of liver biopsy in nonalcoholic fatty liver disease," *Gastroenterology*, vol. 128, pp. 1898–1906, 2005. Available: https://doi.org/10.1053/j.gastro.2005.03.084
25. V. Ratziu, S. Bellentani, H. Cortez-Pinto, C. Day, and G. Marchesini, "A position statement on NAFLD/NASH based on the EASL 2009 special conference," *Journal of Hepatology*, vol. 53, pp. 372–384, 2010. Available: https://doi.org/10.1016/j.jhep.2010.04.008.
26. S. E. Baumeister, H. Volzke, and P. Marschall**,** "Impact of fatty liver disease on health care utilization and costs in a general population: a 5-year observation," *Gastroenterology*, vol. 134, pp. 85–94, 2008. Available: https://doi.org/10.1053/j.gastro.2007.10.016
27. G. Targher, "Non-alcoholic fatty liver disease, the metabolic syndrome and the risk of cardiovascular disease: the plot thickens," *Diabetic Medicine*, vol. 24, pp. 1–6, 2007. Available: https://doi.org/10.1111/j.1464-5491.2007.02168.x
28. A. Lonardo, C. Carani, N. Carulli, et al., "Endocrine NAFLD: a hormonocentric perspective of nonalcoholic fatty liver disease pathogenesis," *Journal of Hepatology*, vol. 44, pp. 1196–1207, 2006. Available: https://doi.org/10.1016/j.jhep.2006.03.005
29. N. A. Shackel and G. W. McCaughan, "Liver biopsy: Is it still relevant?," *Internal Medicine Journal*, vol. 36, no. 11, pp. 689–691, 2006. Available: https://doi.org/10.1111/j.1445-5994.2006.01210.x
30. E. Emanuele, "Is biopsy always necessary? Toward a clinicolaboratory approach for diagnosing nonalcoholic steatohepatitis in obesity," *Hepatology*, vol. 48, no. 6, pp. 2086–2087, 2008. Available: https://doi.org/10.1002/hep.22622
31. Z. M. Younossi, M. Stepanova, F. Negro, S. Hallaji, Y. Younossi, B. Lam, et al., "Nonalcoholic fatty liver disease in lean individuals in the United States," *Medicine (Baltimore)*, vol. 91, pp. 319–327, 2012. Available: https://doi.org/10.1097/MD.0b013e3182779d49
32. J. G. Fan, S. U. Kim, and V. W. Wong, "New trends on obesity and NAFLD in Asia," *Journal of Hepatology*, vol. 67, pp. 862–873, 2017. Available: https://doi.org/10.1016/j.jhep.2017.06.003
33. R. Malik, M. Chang, K. Bhaskar, I. Nasser, M. Curry, et al., "The clinical utility of biomarkers and the nonalcoholic steatohepatitis CRN liver biopsy scoring system in patients with nonalcoholic fatty liver disease," *Journal of Gastroenterology and Hepatology*, vol. 24, pp. 564–568, 2009. Available: https://doi.org/10.1111/j.1440-1746.2009.05793.x
34. V. Ratziu, J. Massard, F. Charlotte, D. Messous, F. Imbert-Bismut, et al., "Diagnostic value of biochemical markers (FibroTestFibroSURE) for the prediction of liver fibrosis in patients with nonalcoholic fatty liver disease," *BMC Gastroenterology*, vol. 6, p. 6, 2006. Available: https://doi.org/10.1186/1471-230X-6-6
35. I. N. Guha, J. Parkes, P. Roderick, D. Chattopadhyay, R. Cross, et al., "Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers," *Hepatology*, vol. 47, pp. 455–460, 2008. Available: https://doi.org/10.1002/hep.21984
36. T. Fukuda, M. Hamaguchi, T. Kojima, Y. Hashimoto, A. Ohbora, T. Kato, et al., "The impact of non-alcoholic fatty liver disease on incident type 2 diabetes mellitus in non-overweight individuals," *Liver International*, vol. 36, pp. 275–283, 2016. Available: https://doi.org/10.1111/liv.12906
37. S. S. Kim, H. J. Cho, H. J. Kim, D. R. Kang, J. R. Berry, J. H. Kim, et al., "Nonalcoholic fatty liver disease as a sentinel marker for the development of diabetes mellitus in non-obese subjects," *Digestive and Liver Disease*, vol. 50, pp. 370–377, 2018. Available: https://doi.org/10.1016/j.dld.2018.01.126
38. H. Yoshitaka, M. Hamaguchi, T. Kojima, T. Fukuda, A. Ohbora, and M. Fukui, "Nonoverweight nonalcoholic fatty liver disease and incident cardiovascular disease: A post hoc analysis of a cohort study," *Medicine (Baltimore)*, vol. 96, p. e6712, 2017. Available: https://doi.org/10.1097/MD.0000000000006712