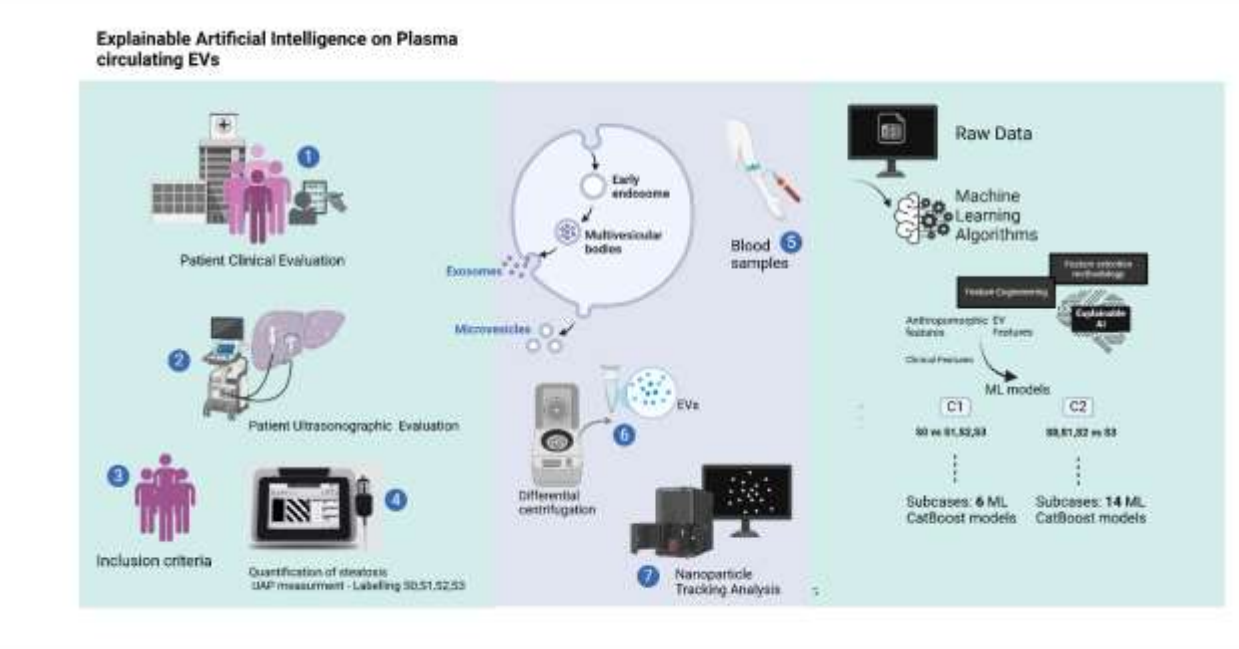
medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Extracellular Vesicles as Biomarkers for Steatosis Stages in MASLD Patients: an Algorithmic Approach Using Explainable Artificial Intelligence

Eleni Myrto Trifylli [1,2*], Athanasios Angelakis [3,4,5*], Anastasios G. Kriebardis[2], Nikolaos Papadopoulos[6], Sotirios P. Fortis[2], Vasiliki Pantazatou [2], Ioannis Koskinas[1], Hariklia Kranidioti[1], Evangelos Koustas[7,] Panagiotis Sarantis[8], Spilios Manolakopoulos[1] and Melanie Deutsch[1]

[1]GI-Liver Unit, 2nd Department of Internal Medicine National and Kapodistrian University of Athens, General Hospital of Athens "Hippocratio", 114 Vas Sofias, 11527 Athens, Greece
[2]Laboratory of Reliability and Quality Control in Laboratory Hematology (HemQcR), Department of Biomedical Sciences, Section of Medical Laboratories, School of Health & Caring Sciences, University of West Attica (UniWA), Ag. Spyridonos Str., 12243 Egaleo, Greece
[3]Amsterdam University Medical Center - Department of Epidemiology and Data Science, Amsterdam, Netherlands;
[4]Amsterdam Public Health Research Institute – Methodology, Digital Health, Amsterdam, Netherlands;
[5] University of Amsterdam - Data Science Center, Amsterdam, Netherlands
[6]Second Department of Internal Medicine, 401 General Military Hospital, 11527 Athens, Greece
[7]Oncology Department, General Hospital Evangelismos, 10676 Athens, Greece
[8]Department of Biological Chemistry, Medical School, National and Kapodistrian University of Athens, 11527 Athens, Greece

*Correspondence: trif.lena@gmail.com (Eleni Myrto Maria Trifylli); a.angelakis@amsterdamumc.nl (Athanasios Angelakis)
Both authors had equal contribution and are joint first authors



Graphical Abstract

Abstract

Background & Aims:

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Metabolic dysfunction-associated steatotic liver disease (MASLD), formerly known as NAFLD, is a leading cause of chronic liver disease worldwide. Current diagnostic methods, including liver biopsies, are invasive and have significant limitations, emphasizing the need for non-invasive alternatives. This study aimed to evaluate extracellular vesicles (EV) as biomarkers for diagnosing

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and staging steatosis in MASLD patients, utilizing machine learning (ML) and explainable artificial intelligence (XAI).

Methods:
This prospective, single-center cohort study was conducted at the GI-Liver Unit, Hippocration General Hospital, Athens. It included 76 MASLD patients with ultrasound-confirmed steatosis and at least one cardiometabolic risk factor. Patients underwent transient elastography for steatosis and fibrosis staging and blood sampling for EV analysis using nanoparticle tracking. Twenty machine learning models were developed. Six to distinguish non-steatosis (S0) from steatosis (S1-S3), and fourteen to identify severe steatosis (S3). Models incorporated EV measurements (size and concentration), anthropomorphic and clinical features, with performance evaluated using AUROC and SHAP-based interpretability methods.

Results:
The CB-C1a model achieved, on average on 10 random splits of 5-fold cross validation (5CV) of the train set, an AUROC of 0.71/0.86 (train/test) for distinguishing S0 from S1-S3 steatosis stages, relying on EV alone. The CB-C2h-21 model identified severe steatosis (S3), on average on 10 random splits of 3-fold cross validation (3CV) of the train set, with an AUROC of 0.81/1.00 (train/test), demonstrating superior performance when combining EV with anthropomorphic and clinical features such as diabetes and advanced fibrosis. Key EV features, including mean size and concentration, were identified as important predictors. SHAP analysis highlighted complex non-linear relationships between features and steatosis staging.

Conclusions:
EV are promising non-invasive biomarkers for diagnosing and staging MASLD. The integration of ML-enhanced EV analysis with clinical features offers a scalable, patient-friendly alternative to invasive liver biopsies, advancing precision in MASLD management. Further research should refine these methods for broader clinical application.

Keywords: Metabolic dysfunction-associated steatotic liver disease (MASLD), extracellular vesicles (EVs), non-invasive biomarkers, machine learning (ML), explainable artificial intelligence (XAI), steatosis staging, transient elastography, chronic liver disease, cardiometabolic risk factors, SHAP analysis, diagnostic precision, advanced fibrosis, severe steatosis (S3), liver biopsy alternatives, hepatology diagnostics.

1. Introduction

MASLD is one of the most frequently diagnosed and continuously increasing chronic liver diseases, particularly in the Western world [1]. It is characterized by significant morbidity and mortality, especially in advanced stages, and is currently the leading cause of liver transplantation in the US [2]. MASLD includes hepatic steatosis which may progress to metabolic dysfunction-associated steatohepatitis (MASH), fibrosis, cirrhosis, and hepatocellular carcinoma. The main challenges in everyday clinical practice include firstly the diagnosis of MASLD, secondly the diagnosis of MASH and the determination of the steatosis and fibrosis extent [3]. Although the extend of liver fibrosis is the strongest predictor of end stage liver disease, the presence and the degree of steatosis are also important factors that should be accessed since they are closely implicated in the development of hepatic and extrahepatic outcomes. More particularly the two most common causes of mortality in those patients are cardiovascular outcomes and extrahepatic malignancy, while the dissemination of different primary cancer in hepatic parenchyma is closely correlated with the degree of steatosis (HR of 1.34) based on EASL–EASD–EASO Clinical Practice Guidelines, whereas the least common cause is interestingly due liver-related outcomes [4]. Interestingly, MASLD patients have a higher risk of developing extrahepatic malignancies, than obese patients, implying its significant implication in obesity-mediated cancers [4, 5].

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Liver ultrasound is the most commonly used diagnostic method, especially for cirrhosis identification. However, it is an operator-dependent diagnostic method and it is considered insufficient for the evaluation of low-grade steatosis or fibrosis [6]. Liver biopsy remains the gold standard for the definitive diagnosis of MASH, and fibrosis evaluation, even at the early disease stages. However, its invasive nature, high cost, and variability in diagnostic accuracy limit its widespread use [7, 8]. Liver stiffness measurement (LSM) with transient elastography (TE) and attenuation parameters are reliable approaches for identifying liver fibrosis and steatosis respectively. However, there is an emerging need of developing new and easily-performed noninvasive tests (NITs) or combinations of tests in order to assess liver or non-liver prognosis in MASLD patients. [9].

In this context, extracellular vesicles (EVs) have emerged as probable diagnostic biomarkers and could be an interesting alternative [10, 11]. EV's are double-membraned nanoparticles, secreted by multiple cells, that carry several molecules such as DNA, RNA, proteins, lipids, and autophagosomes [10]. EVs significantly contribute to intercellular communication by altering the recipient cells' functional and biological status [11]. Lipotoxic hepatocytes are a major source of EVs in the systemic circulation. During lipotoxicity, increased lipid storage in liver parenchyma induces the release of EVs, which play a pivotal role in MASLD progression by promoting inflammatory reactions, recruiting macrophages and monocytes, and activating hepatic-stellate and liver sinusoidal endothelial cells, enhancing fibrosis and angiogenesis [10-13]. It is important to underline the crucial role of EVs within the framework of steatotic liver that can induce the development of extrahepatic malignancies and secondary hepatic lesions.

Alterations in EV quantity and quality are closely implicated in MASLD pathogenesis and could facilitate diagnosis, patient stratification, and prognosis [14]. Moreover, it has been demonstrated in animal models (murine) that the EV levels are closely related to the EV circulating levels, especially in cases of MASH compared to healthy controls that have been strongly related to increased angiogenesis and advanced fibrosis [14, 15]. Additionally, it was demonstrated in another study that EVs were increased in pre- cirrhotic patients and were even more abundant in cirrhotic patients, in comparison to patients with advanced MASH and healthy individuals [16]. It has been also demonstrated in the same study by Povero et al. [16], that liver constitutes a significant source of EV in blood circulation. On the top of that, another study has reported a depletion of small-sized EVs after bariatric surgery and the subsequent weight loss [17]. Moreover, it was shown that plasma EVs especially the hepatocyte-derived ones, are notably increased in the case of MASLD patients, in comparison to post-weight loss patients, which implies the potential use of circulating EVs of hepatocyte origin as a diagnostic and prognostic biomarker in a point-of case manner [17]. The patient's response to surgical intervention (bariatric surgery) or other lifestyle modifications, hints their potential utilization also as monitoring tools for treatment response.

Additionally, several conditions can increase their levels and size such as in cirrhosis and YPE 2 Diabetes Mellitus (T2DM). More particularly, T2DM can alter the levels and the size of the detected plasma EVs [18]. More specifically, it was demonstrated in several studies the presence of higher levels of microvesicles in patients with T2DM, compared to euglycemic individuals [19]. A meta-analysis concluded that the levels of circulating large EVs (microvesicles) were indeed higher in T2DM patients, especially the ones that were derived from monocytes, endothelial cells, and platelets, in contrast to the levels that originated by leucocytes, which have been proven no statistically elevated and inversely correlated with the grade of fibrotic injury [20]. On top of that, there is a close relation between MASLD and LSM progression (≥20% increase in LSM values) in T2DM, which constitutes a statistically significant independent predictor for LSM increase based on the recent study by Huang, Daniel Q. et al. [21].

This study aimed on the assessment of the severity of steatosis, as well as the presence of steatosis,

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

in MASLD patients using a combination of two parameters: circulating levels of plasma EV and the ultrasound attenuation parameter (UAP) measured by TE [22, 23].

Additionally, we employed data science (DS), ML [24] and XAI [25] methods to understand the non-linear relationships between features and predict severe and presence of steatosis in MASLD patients.

2 Materials and Methods

2.1 Patients

This study is a prospective single-center study GI-Liver Unit of the Second Academic Department of Internal Medicine Hippocration General Hospital University of Athens), which is composed of 76 consecutive patients with ultrasound findings of steatosis and at least one cardiometabolic risk factor as it was suggested on multi-society Delphi consensus on new nomenclature and no other causes of steatosis. More particularly, the cardiometabolic risk factors included, a BMI≥ 25 kg/m2 or increased waist circumference for men (>80cm) and women (>94 cm), low HDL levels for males <40mg/dl and for females <50mg/dl, high triglycerides (>150 mg/dl) or the intake of treatment for dyslipidemia, hypertension (≥130/80 mmHg) or the intake of antihypertensive drug regime, as well as the presence high levels of fasting glucose, or high 2h-postprandial glucose levels or diagnosed diabetes mellitus or high levels of glycosylated hemoglobin or the intake of anti-diabetic treatment [26]. Patients with alcohol abuse (≥2 drinks for females and ≥ 3 drinks for males per day, with 1 drink being equal to 10g of alcohol), inflammatory or autoimmune diseases, other chronic viral or metabolic liver diseases, hepatocellular cancer, or non-hepatic malignancy have been excluded [26]. After acquiring the informed consent of all the participants in this study, we collected demographic data, patient's history and routine blood tests as well as we performed transient elastography (TE) by an experienced operator. The evaluation was made via the "iLivTouch" FT100 device by Wuxi Hisky Medical Technologies Co., Ltd. (Hisky Med), China. This device performs LSM using TE based on controlled low frequency shear wave synchronously with UAP using one universal probe applicable to patients of different body sizes [24, 25]. Steatosis and fibrosis were assessed with the evaluation of the UAP and the LSM, respectively, after the performance of ≥10 valid measurements, with a ratio between the successful to the total number of measurements ≥ 60%, as well as the ratio between the interquartile range (IQR) to the LSM being ≤ 30% based on the guidelines of World Federation for Ultrasound in Medicine and Biology (WFUMB) [27]. The thresholds for steatosis based on the estimation of UAP value were: ≥S3 296dB/dl, ≥S2 269dB/dl, and ≥S1 244dB/dl [10]. However, we further categorized the patients based on the presence of severe steatosis (≥S3 296dB/dl) , S0-S2 (< 296 dB/dl) and S0 < 244 dB/dl [23]. In Table 1 we provide patient characteristics and in Table 2 the percentages of T2DM and advanced fibrosis in different stages of steatosis.

2.2 Pre-analytical protocol

The protocol of pre-analysis was the same for each blood sample, including the time of collection, the blood processing interval (BPI), the handling, and the processing to minimize the variability of EV analysis. We chose the collection tubes with citrate for EV analysis, while the collection of the samples followed the Declaration of Helsinki (DoH) after the patients had given written informed consent. Blood drawing was performed at the same time of day, while the patients were fasting before the procedure. Blood samples were collected in a tube with Citrate (Vacuette sodium citrate 3.2%, volume 3,5 mL, Greiner Bio-One) [27], the time interval for blood drawing was approximately five minutes and the tube was turned vertically and back 10 times exactly after their collection. All the samples were transported to the lab after the collection in special conditions to avoid deterioration with a BPI of 1 hour. These samples were initially centrifuged for 20 min at 3000 x g at room temperature to deplete cell debris and the collected supernatant was immediately stored at −80°C, until the following processing, including differential centrifugation [28].

2.3 Nanoparticle Tracking Analysis (NTA)

Nanoparticle Tracking Analysis (NTA) was conducted using the NanoSight NS300 instrument

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

(Malvern Instruments, Amesbury, UK). Plasma samples were diluted in particle-free PBS (0.02 μm filtered, Cytiva Whatman, UK) for NTA by which we acquired data about their total concentration levels, size (between 50-1000nm), and distribution in the sample [29]. All the measurements were performed under constant flow and temperature, while autofocus was adjusted to ensure clarity by avoiding indistinct particles. The type of camera and laser were sCMOS and green, respectively. More particularly, five 30-second videos were recorded per measurement under the following conditions: cell temperature at 25°C and syringe speed at 100 μL/second. During the measurements, there was no evidence of vibration, which can alter the sizing of particles. Videos were analyzed using NanoSight NTA 3.4 build 3.4.4 software (Malvern, 2020) in script control mode, comprising 1,500 frames per sample. Each sample was measured five times, and the size distribution data were averaged. Additionally, we evaluated the possible inter-day variability of the method, by performing NTA for each sample on different days, which did not show any significant difference between the values. Lastly, we obtained data regarding the mean size of vesicles for each patient, the amount of vesicles sized between 50 to 150nm or >150nm and the sum of vesicles between 50 to 1000nm.

2.4 Data Science

Based on the bibliographic data regarding the alterations in EV quantity, size, and quality in MASLD pathogenesis, we utilized ML algorithms and feature engineering (FE) [30], techniques to understand non-linear relationships between features. FE boosts ML performance by creating more features, and feature selection methods identify the most important ones [30]. We applied the CatBoost algorithm [31-34], FE techniques, XAI methods, and feature selection (FS) methodologies [30, 35-38] to create a data science pipeline for predicting non-severe or severe steatosis in MASLD patients in two main cases: Case 1 (C1) and Case 2 (C2), respectively. More specifically, we primarily used UAP cut-off values to label patients between severe and non-severe steatosis and quantified their circulating plasma EV levels to show associations and potential diagnostic performance for C1: S0, S1, S2 vs. S3 and for C2: S0 vs. S1, S2, S3.

To further explore the relationships between features and steatosis stages, we conducted a comprehensive correlation analysis [39] using both parametric and non-parametric tests, including Pearson, Kendall, Spearman, and Point-Biserial correlations [40-43]. This analysis was performed separately for both C1 (S0 vs. S1, S2, S3) and C2 (S0, S1, S2 vs. S3) cases. Correlations were evaluated against a moderate threshold of 0.6, chosen to identify meaningful yet not overly stringent associations. This approach enabled us to assess both linear and rank-based relationships, providing a robust understanding of the interactions between features and steatosis stages.

We used the following EV features: ‘mean’, ‘50-150nm’, ‘>150’, and ‘Sum’, anthropomorphic ones: ‘Sex’, ‘Age’, ‘Height', 'Weight', ‘BMI’ and clinical: advanced fibrosis ‘Adv_Fibrosis’, and ‘Diabetes’. Concretely, the ‘mean’ represents the mean size of vesicles for each patient, ‘50-150nm’ the amount of vesicles sized between 50 to 150nm , “>150” the vesicles sized over 150nm and the “sum” that represents the sum of vesicles between 50 to 1000nm. We tuned twenty CatBoost ML models. Additionally, we used XAI and FS [35-38] to identify the most important features. For tuning and to understand the robustness of our models, we used non-stratified k-fold cross validation 5CV, 3CV for C1 and C2 respectively, and a test-set to measure the performance of our ML models. Moreover, since the dataset’s volume was relatively low we used ten times k-fold cross-validation, splitting with different random seeds, to eliminate possible randomness splitting effect [30].

2.4.1 Cases Definition: We investigated the role of EV as biomarkers in two cases. Namely, in C1 we defined the binary classification problem of distinguishing non-steatosis vs steatosis S0 vs. S1, S2, S3. In C2 we defined the binary classification problem of identifying an individual’s severe steatosis stage S0, S1, S2 vs. S3.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

C1: We created six sub-cases:

- C1a: Using EV features (mean, 50-150nm, >150, Sum).
- C1b: Applying FE on C1a.
- C1c: Combining EV with anthropomorphic features (Sex, Age, Height, Weight, BMI).
- C1d: Applying FE on C1c.
- C1e: Combining EV features with the presence of advanced fibrosis and diabetes.
- C1f: Applying FE on C1e.

C2: We created eight sub-cases:

- C2a: Using only EV features.
- C2b: Applying FE on C2a.
- C2c: Combining EV features with anthropomorphic features.
- C2d: Applying FE on C2c.
- C2e: Combining EV features with the presence of advanced fibrosis and diabetes.
- C2f: Applying FE on C2e.
- C2g: Combining EV features with anthropomorphic features and the presence of advanced fibrosis and diabetes

C2h: Applying FE on C2g.

2.4.2 Data Preprocessing: For the C2 we used the cut-off UAP value of >= 296 to label the 'target' feature, which was severe steatosis (S3) and not severe steatosis (S0, S1, S2). For the C1 we used the cut-off UAP value of >= 244a. We dropped data instances with null values and no rnomalization/scaling applied**.**

2.4.3 Datasets Creation: For both cases, we split the initial dataset of 74 or 72 data instances, depending on the subcase, to 80% as train and 20% as a test set with a fixed random seed for reproducibility. The percentage of the 'target' was similar in the train and test sets in both cases so that the ML models we tuned would be fair. A patient who belonged to the train datasets did not belong to the test sets so there would be no data/information leakage [39]. Furthermore, our approach to random split to train and test set is aligned with the guidelines from the DELFI TRIPOD-AI statement [40] so that our ML models would be applied to the examination center where the data have been retrieved from. The feature selection methodology [6a-9a] uses both datasets but the inductive learning [41] approach of it makes sure there is no information/data leakage [39]. We had two approaches for the examination of the role of EV regarding their predictability and associativity, due to the following reasons: (i) lack of bibliographic data about the assessment of EV role and their predictability on MASLD using XAI and ML models and (ii) the relatively low volume of data instances of the dataset (number of patients). In FE sub-cases we divided the EVs with 10^9, we raised the numeric features to 11^{th} power and we calculated their square root, and finally we calculated the pair-wise products for all features.

2.4.4 Feature selection methodology: Applying the FE methodology [35-38] on the sub-cases C2d, C2f and C2h, lead us to five, eleven and twenty-one respectively, most important features. Analyzing these features using SHAP [42] and CatBoost's feature importance, provided us with

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

insights regarding the features in C2 cases.

2.4.5 Models Evaluation: All ML models tuned using k-fold cross validation (kCV) on the train dataset for each case, and have been applied to the corresponding per case test set for validation. We used 5 and 3 folds for the C1 and C2 respectively. Each model could correspond to potentially a new method for classifying in an automatic way individuals that belong to each of the classes for each case. Since the data volume was relatively small, after the tuning we applied ten splits with different random seeds for the kCV to eliminate the impact of randomness. We tuned the iterators, the learning rate, the depth and the weights from CatBoost. For all the other hyperparameters we used the default values.

2.4.6 Algorithm:

A deterministic algorithm to create two predictive ML models for the cases C1 and C2, and to identify the most important features and their associations.

- Input: VE, anthropomorphic, clinical features and the UPA values of the 76 patients described in 2.1.
- Output: a. The best predictive model for the case C1 and the best one for the case C2.
  b. The most important features for each model and their associations to the outcome

1. Initialize Data and Libraries:
   - Import necessary libraries (CatBoost, SHAP, etc.).
   - Load the dataset.
2. Data Preprocessing:
   - Drop data instances with null values.
   - Label the target feature for C1 (S0 vs. S1, S2, S3) and C2 (S0, S1, S2 vs. S3) using the cut-off UAP values (≥ 244a for C1 and ≥ 296 for C2).
3. Dataset Splitting:
   - Split the dataset into train (80%) and test (20%) sets using a fixed random seed to ensure reproducibility.
   - Ensure similar percentages of ‘target’ in both train and test sets.
4. Features’ Combinations:

   For each case (C1 and C2), create at least four sub-cases that correspond to :
   - Use only EV features.
   - Combine EV with anthropomorphic features.
   - Combine EV with clinical features.
   - Combine EV with anthropomorphic and clinical features.
5. Feature Engineering
   - For numeric features:
     - Divide by 10^9.
     - Raise to the 11th power.
     - Calculate square roots.
   - For all features compute the pair-wise products
6. Model Training:
   - Use CatBoost model for training.
   - Perform k-fold cross-validation (kCV) for tuning hyperparameters (iterators, learning rate, depth, weights).
7. Model Evaluation:
   - Evaluate the ML models validating them on the test set.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

8. Feature Selection Methodology:
    - Apply FS
9. Identify the best model for each case:
    - From the statistics on the kCV and the performance on the test set identify the best model for each case.
10. Identification of Important Features:
    - After feature selection, identify the most important features using SHAP.
11. Identify the most important combination of family of features:

Conclude about the contribution of each one of the families of features (EV, anthropomorphic, clinical) for C1 and C2

3 Results

3.1 C1

In C1, the training set consisted of 59 individuals, and the test set included 15. Six CatBoost models were developed, tuning hyperparameters such as iterations, learning rate, depth, and class weights. Performance metrics, including sensitivity, specificity, ROC-AUC, and F1 scores, are summarized in Table 3. The CB-C1a model emerged as the best-performing configuration, achieving an AUROC of 0.70 (±0.15) and an F1 score of 0.73 (±0.14). The model demonstrated balanced performance with a sensitivity of 0.66 (±0.17) and specificity of 0.74 (±0.14).

We conducted correlation analysis on C1. Across all correlation tests (Pearson, Kendall, Point-Biserial, and Spearman), Figure certain features, such as transformed BMI and diabetes markers, exhibit moderate associations with steatosis stages. However, none of the correlations surpass the 0.6 threshold, indicating that no feature demonstrates a moderately strong or strong relationship with steatosis severity in this comparison. This suggests that the relationships between individual features and the steatosis stages are relatively weak and potentially complex.

The violin plots in the top row of Figure 3 depict the distributions of selected features (>150, >150 nm particle count, mean, and sum). These plots do not reveal any distinguishable patterns or separation between the groups, indicating that these features, when analyzed individually and linearly, fail to provide insight into the differences between steatosis stages. This lack of separation highlights the challenge of identifying meaningful patterns using traditional visualization techniques, particularly for multifactorial diseases like MASLD. Conversely, the SHAP scatter plots in the bottom row, derived from XAI, reveal non-linear patterns in the associations between these features and the target outcome. The scatter plots show that even in cases where traditional statistical methods and visualizations fall short, the XAI approach uncovers nuanced relationships. For example, SHAP values illustrate the varying impact of features like >150, >150 nm particle count, mean, and sum on the classification of steatosis stages, demonstrating their combined importance in the predictive model.

The addition of anthropomorphic and clinical features, such as advanced fibrosis and diabetes, or the application of FE, did not improve the model's performance. In the S1- S3 group, advanced fibrosis was present in only 6/53 cases, diabetes in 13/53, and both conditions in 5/53 (Table 2), supporting the sufficiency of EV for distinguishing S0 from steatosis stages. Feature importance and SHAP analyses, presented in Figure 1, respectively, identified "Mean," "50-150nm," ">150nm," and "Sum" as the most influential predictors, with "Mean" ranking highest. The SHAP

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

scatter plots (Figure 3). highlighted the non-linear relationships between these features and model predictions, showcasing the necessity of advanced ML techniques to capture such complexity.

3.2 Case 2: S0, S1, S2 vs. S3 Steatosis

In Case 2, the training set included 58 individuals, and the test set consisted of 13. Fourteen CatBoost models were tuned, and their results are presented in Table 4. The CB-C2h-21 model demonstrated superior performance, with an AUROC of 0.89 (±0.03) on the training set (3CV) and a perfect AUROC of 1.00 on the test set. Sensitivity and specificity values of 0.92 (±0.12) and 0.87 (±0.10), respectively, further underscored its reliability in distinguishing mild-to- moderate steatosis (S0-S2) from severe steatosis (S3).

Correlation analysis showed that features like >150, >150 nm particle counts, BMI transformations, and weight-related interactions show some associations, but, similar to C1, none exceed the 0.6 threshold. This highlights that while certain features are associated with advanced steatosis, the strength of these associations are not even moderate, emphasizing the multifactorial nature of MASLD.

The plots in Figure 4 correspond to the C2 case, where S0, S1, and S2 are compared against S3. The violin plots (top panel) illustrate the distribution of various features across the two groups. Similar to the C1 case, the violin plots do not show a clear separation between the groups for any of the features, indicating that traditional visualization methods fail to provide significant insights or distinguishable patterns. In contrast, the SHAP scatter plots (bottom panel) reveal complex, non-linear relationships between features and the target variable. These plots highlight how features such as >150, >150 nm particle count, transformed weights, and BMI interactions contribute to the predictive model. The nuanced and dispersed patterns observed in the SHAP scatter plots underscore the importance of these features in classifying advanced steatosis (S3), despite the lack of clarity from traditional statistical methods.

In contrast to C1, the inclusion of anthropomorphic and clinical features, such as advanced fibrosis and diabetes, significantly enhanced model performance, increasing the AUROC from 0.81 to 0.89. The S3 group had a higher prevalence of advanced fibrosis (3/19), diabetes (9/19), and their co-occurrence (3/19) (Table 2), highlighting their relevance in identifying severe steatosis. Feature importance and SHAP analyses, shown in Figure 2 and Figure 4, revealed that EV features remained imporant predictors. Additionally, engineered features such as "Mean × Weight" and "Sum × Age" contributed significantly, indicating non-linear relationships between predictors and outcomes.

The consistent significance of EV across both cases underscores their diagnostic value. In C1, features like "Mean" and "50-150nm" were the strongest predictors, while in C2, engineered features such as "Mean × Weight" and "Sum_sqrt" played crucial roles. SHAP scatter plots in both cases (Figures 3 and 4) illustrated the absence of linear correlations, emphasizing the necessity of non-linear ML models to capture intricate interactions.

The XAI analysis provided interpretable insights into feature contributions. Figure 1 and Figure 3 highlighted "Mean" and "50-150nm" as the top predictors in C1, while Figure 2 and Figure 4 identified "Mean × Weight" and "Weight^7" as the most critical features in C2. These findings reinforce the physiological relevance of EV as biomarkers and their potential in MASLD diagnosis.

This study demonstrated the crucial role of EV in diagnosing and staging steatosis in MASLD. The CB-C1a model, which relied solely on EV, was sufficient for distinguishing S0 from S1, S2, and S3

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

stages in C1. Conversely, the CB-C2h-21 model showed that combining EV with anthropomorphic and clinical features improved the prediction of severe steatosis in C2. These results highlight the diagnostic potential of EV-related features and the utility of XAI in interpreting complex ML models for clinical applications in Hepatology.

4. Discussion

There is a continuous effort to identify new non-invasive biomarkers and tests for MASLD patients, due to the various limitations of liver biopsy. The development of blood biomarkers that can be reproducible, more easily applicable, and potentially available for utilization in clinical practice and treatment response monitoring, is in the spotlight of scientific research. Steatosis has a key role in the progression of the disease from MASL to MASH and fibrosis. Even though fibrosis strongly predicts long-term prognosis for MASLD patients and end-stage liver outcomes, early identification of steatosis can significantly alter the management of MASLD and its outcomes at hepatic and extrahepatic levels. In addition, there are notably fewer tests and biomarkers in the context of steatosis evaluation than in fibrosis, a phenomenon that implies an open horizon for new diagnostic perspectives.

Some of the limitation that have been reported regarding current biomarkers is the inclusion of age as a parameter for their calculation. More particularly, age as a parameter can importantly alter the interpretation of the results, a phenomenon that is mainly correlated with the significantly higher frequency of several diseases such as hyperlipidemia (increased triglyceride and cholesterol levels), insulin resistance, and diabetes (altered fasting glucose levels) due to aging [43].

Aging can also be implicated in the distribution of visceral fat, as well as in the noteworthy reduction of muscle mass, resulting in a normal BMI, altering the interpretation of several tests that include BMI [44]. Similarly, several enzymes can be influenced by aging and medication for age-related comorbidities, such as transaminases and GGT that may lead to inaccurate conclusions regarding steatosis [45]. Interestingly, a normal BMI can be also presented in lean patients with MASLD, which can lead to misinterpretation of the test, despite the presence of steatosis [46]. Similarly, some other anthropometric measurements such as waist-to-hip ratio or waist circumference also be normal in lean MASLD patients, leading to underestimated scores [47]. Meanwhile, it is reported that lean MASLD patients have a lower prevalence of hypertension, dyslipidemia, and T2DM, as well as lower levels of transaminases, higher HDL, and lower levels of TG and total cholesterol [47].

The absence of correlations exceeding 0.6 in both C1 and C2 suggests that no single feature serves as a strong predictor of steatosis stages in MASLD. This is consistent with the complex and multifactorial nature of the disease, where individual features may only have a limited explanatory or predictive role. The results emphasized the need for integrative approaches that combine multiple moderately correlated features to capture the heterogeneity of MASLD. The Pearson test identifies linear associations, while the Kendall and Spearman tests assess rank-based relationships, making them robust to non-linearity. The Point-Biserial test, tailored for binary classification, complements these analyses by quantifying the ability of features to separate distinct steatosis stages. Across all these methods, the moderate correlations (none exceeding 0.6) reinforce the idea that MASLD severity cannot be effectively characterized by isolated markers. This finding underlines the importance of ML models that leverage the combined contributions of several moderately associated features. Additionally, the moderate correlations align with clinical experience, where MASLD progression is driven by the interplay of metabolic, genetic, and environmental factors rather than a single dominant feature. These insights, derived from a robust and multi-faceted correlation analysis, provided a strong foundation for predictive modeling and biomarker discovery using ML and XAI.

The present study demonstrates the significant potential of EV as biomarkers for distinguishing

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

steatosis stages in MASLD. By employing XAI and ML we elucidated the diagnostic importance of EV, particularly their size distributions and concentrations. The integration of advanced ML models like CB-C1a and CB-C2h-21 provides a robust framework for non-invasive, accurate diagnostics.

Figure 3 highlights the limitations of traditional statistical analysis and visualization techniques, such as violin plots, in capturing complex relationships within biological data, as demonstrated by the indistinguishable patterns that fail to show strong, separable linear relationships with steatosis stages in MASLD. In contrast, the SHAP scatter plots illustrate the power of machine learning and explainable AI (XAI) to detect and interpret non-linear associations, uncovering intricate patterns and feature interactions that are critical for understanding and differentiating steatosis stages, particularly in the C2 case where advanced steatosis (S3) is compared to earlier stages. The non-linear relationships revealed by SHAP emphasize the need for advanced analytical approaches to disentangle the multifactorial nature of MASLD and highlight the importance of features that might otherwise be overlooked due to their lack of linear separability. This ability to extract and interpret complex patterns not only enhances our understanding of MASLD but also compensates for the limitations of conventional methods, setting the stage for developing robust, data-driven predictive models. These findings demonstrate the transformative potential of XAI in hepatology research, paving the way for improved disease stratification and targeted interventions.

Our results underscore the central role of EV-related features in differentiating between various stages of steatosis. For C1 EV independently achieved high diagnostic performance, with the CB-C1a model achieving an AUROC of 0.70 in 5CV and 0.86 on the test set. Anthropomorphic and clinical features such as diabetes and advanced fibrosis did not improve diagnostic accuracy in this scenario. These findings highlight the potential of EV as standalone non-invasive biomarkers for early-stage steatosis detection. Conversely, for C2 the inclusion of anthropomorphic and clinical features significantly enhanced the model's diagnostic performance, as evidenced by the CB-C2h-21 model's AUROC improvement from 0.81 to 0.89 in 3CV. The higher prevalence of advanced fibrosis (3/19), diabetes (9/19), and their co-occurrence (3/19) in the S3 group suggests that these features are more relevant for identifying advanced stages of steatosis. Importantly, EV remained pivotal predictors in this context, underscoring their physiological relevance in MASLD progression. The results using 10 times iterative cross validation of C1 and C2 cases with AUROC +/- std of 0.71+/-0.03 and 0.81+/-0.04, respectfully, as well as the performance regarding the specificity, sensitivity and F1 score in Table 5, shows that even though the dataset volume was relatively low, the performance of the two best models was very good and promising. From the explainability and the insights on the associations of the most important features it is clear that VEs are potentially very strongly non-linear correlated with steatosis stages.

The observed diagnostic performance aligns with the established role of EV in MASLD pathogenesis. EV, particularly those with specific size distributions, are known mediators of intercellular communication, inflammation, and angiogenesis. These mechanisms are crucial in the progression of MASLD from simple steatosis to advanced fibrosis and hepatocellular carcinoma. Our findings suggest that EV size distributions ("Mean," "50-150nm," ">150nm") and concentrations ("Sum") could serve as reliable, non-invasive biomarkers for disease staging and risk stratification.

By leveraging XAI tools such as SHAP analysis, we provided interpretable insights into the non-linear relationships between EV, anthropomorphic, clinical features, and steatosis stages. This transparency is critical for clinical adoption, as it enables healthcare professionals to understand the factors driving model predictions. The high sensitivity, specificity, and AUROC values achieved by CB-C1a and CB-C2h-21 models demonstrate their robustness and potentially clinical applicability.

Additionally, the use of a standardized protocol for EV analysis and rigorous validation of ML models ensures the reliability and reproducibility of our results. While liver biopsy remains the gold standard for steatosis assessment, our findings support the potential of ML-enhanced EV as

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

non-invasive alternatives. Our approach could reduce the reliance on invasive procedures, making steatosis staging more accessible and patient-friendly.

Despite the promising findings, the study has some limitations such as the variability in blood sample processing and EV analysis that could possibly introduce bias [48-50]. More particularly, the count of particles could be influenced and overestimated due to similar-sized contaminants or lipoproteins [49,50], while serial centrifugations and freeze- thawing cycles can cause EV breakage [51]. Finally, another limitation of this study is the lack of biopsy validation for many of the patients, due to the ethical reasons, incompliance, and high cost of the procedure. However, we followed strict selection- criteria after excluding other causes of steatosis, other chronic liver diseases, and several conditions that can significantly alter the total circulating levels and the size of EVs.

Future research should explore the in-depth characterization of EV cargoes and their specific roles in MASLD progression, as well as the Apoptotic EVs (ApoEVs) that exceed 1000nm and their role in inflammation-related condition, such as lipotoxicity, their implication in tissue regeneration and their possible utilization as drug vectors [52]. Combining EV with advanced imaging techniques and other non- invasive biomarkers could further improve diagnostic accuracy.

5. Conclusion

This study demonstrates the potential of extracellular vesicles (EVs) as non-invasive biomarkers for diagnosing and staging steatosis in MASLD patients. Using advanced machine learning (ML) techniques combined with explainable artificial intelligence (XAI), we successfully identified EV, particularly their size distributions and concentrations, as critical features for differentiating between steatosis stages. The CB-C1a model achieved strong performance for distinguishing S0 from S1-S3 steatosis stages using EV alone, while the CB-C2h-21 model demonstrated the added diagnostic value of incorporating anthropomorphic and clinical features, such as diabetes and advanced fibrosis, for identifying severe steatosis (S3).
Correlation analysis revealed that no single feature exceeded a moderate threshold (0.6), emphasizing the multifactorial and non-linear nature of MASLD. Traditional statistical and visualization techniques, such as violin plots, failed to provide meaningful separations between groups, further highlighting the need for advanced analytical approaches. SHAP-based XAI enabled the detection of complex non-linear patterns, offering interpretable insights into feature contributions and their interactions, thus addressing the limitations of conventional methods.
These findings underscore the diagnostic utility of EVs in MASLD and the transformative potential of ML-enhanced EV analysis to provide scalable, patient-friendly alternatives to invasive liver biopsies. Future work should focus on further characterizing EV cargoes and integrating these biomarkers with other non-invasive diagnostic tools to refine disease staging and improve clinical outcomes in MASLD management.

6. Abbreviations

- Blood Processing Interval (BPI)
- Cross-Validation (CV)
- Data Science (DS)
- Explainable Artificial Intelligence (XAI)
- Extracellular Vesicles (EVs)
- Feature Engineering (FE)

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

- Feature Selection (FS)
- Interquartile Range (IQR)
- Liver Stiffness Measurement (LSM)
- Machine Learning (ML)
- Metabolic dysfunction-associated steatohepatitis (MASH)
- Metabolic dysfunction-associated steatotic liver disease (MASLD)
- Nanoparticle Tracking Analysis (NTA)
- Non-alcoholic Fatty liver disease (NAFLD)
- Non-invasive tests (NITs)
- Machine Learning Enhanced Non-invasive Tests (MLE-NITs)
- Steatotic Liver Disease (SLD).
- Type 2 Diabetes Mellitus (T2DM)
- Ultrasound Attenuation Parameter (UAP)
- Vibration-Controlled Transient Elastography (VCTE)

7 . References

1. Hsu CL, Loomba R. From NAFLD to MASLD: implications of the new nomenclature for preclinical and clinical research. Nat Metab . 2024;6(4):600–2. Available from: http://dx.doi.org/10.1038/s42255-024-00985-1

2. Miao L, Targher G, Byrne CD, Cao Y-Y, Zheng M-H. Current status and future trends of the global burden of MASLD. Trends Endocrinol Metab . 2024;35(8):697–707. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1043276024000365

3. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology . 2023;77(5):1797–835. Available from: http://dx.doi.org/10.1097/hep.0000000000000323

4. Tacke F, Horn P, Wai-Sun Wong V, Ratziu V, Bugianesi E, Francque S, et al. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol . 2024;81(3):492–542. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168827824003295

5. Allen AM, Hicks SB, Mara KC, Larson JJ, Therneau TM. The risk of incident extrahepatic cancers is higher in non-alcoholic fatty liver disease than obesity – A longitudinal cohort study. J Hepatol . 2019;71(6):1229–36. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168827819304854

6. Miele L, Zocco MA, Pizzolante F, De Matthaeis N, Ainora ME, Liguori A, et al. Use of imaging techniques for non-invasive assessment in the diagnosis and staging of non-alcoholic fatty liver disease. Metabolism . 2020;112:154355. Available from: http://dx.doi.org/10.1016/j.metabol.2020.154355

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

7. Sanyal AJ, Jha P, Kleiner DE. Digital pathology for nonalcoholic steatohepatitis assessment. Nat Rev Gastroenterol Hepatol . 2023;21(1):57–69. Available from: http://dx.doi.org/10.1038/s41575-023-00843-7

8. Zoncapè M, Liguori A, Tsochatzis EA. Non-invasive testing and risk-stratification in patients with MASLD. Eur J Intern Med . 2024;122:11–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0953620524000220

9.Wattacheril JJ, Abdelmalek MF, Lim JK, Sanyal AJ. AGA clinical practice update on the role of noninvasive biomarkers in the evaluation and management of nonalcoholic fatty liver disease: Expert Review. Gastroenterology . 2023;165(4):1080–8. Available from: http://dx.doi.org/10.1053/j.gastro.2023.06.013

10. Trifylli E-M, Kriebardis AG, Koustas E, Papadopoulos N, Deutsch M, Aloizos G, et al. The emerging role of extracellular vesicles and autophagy machinery in NASH-future horizons in NASH management. Int J Mol Sci . 2022;23(20). Available from: http://dx.doi.org/10.3390/ijms232012185

11. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. Nat Rev Mol Cell Biol . 2018 [cited 2024 Nov 28];19(4):213–28. Available from: https://www.nature.com/articles/nrm.2017.125

12. Li W, Yu L. Role and therapeutic perspectives of extracellular vesicles derived from liver and adipose tissue in metabolic dysfunction-associated steatotic liver disease. Artif Cells Nanomed Biotechnol . 2024;52(1):355–69. Available from: http://dx.doi.org/10.1080/21691401.2024.2360008

13. Yu J, Sane S, Kim J-E, Yun S, Kim H-J, Jo KB, et al. Biogenesis and delivery of extracellular vesicles: harnessing the power of EVs for diagnostics and therapeutics. Front Mol Biosci . 2024;10. Available from: http://dx.doi.org/10.3389/fmolb.2023.1330400

14. Hammerich L, Tacke F. Hepatic inflammatory responses in liver fibrosis. Nat Rev Gastroenterol Hepatol . 2023 [cited 2024 Nov 28];20(10):633–46. Available from: https://www.nature.com/articles/s41575-023-00807-x

15. Keingeski MB, Longo L, Brum da Silva Nunes V, Figueiró F, Dallemole DR, Pohlmann AR, et al. Extracellular vesicles and their correlation with inflammatory factors in an experimental model of steatotic liver disease associated with metabolic dysfunction. Metab Syndr Relat Disord . 2024;22(5):394–401. Available from: http://dx.doi.org/10.1089/met.2023.0284

16. Povero D, Yamashita H, Ren W, Subramanian MG, Myers RP, Eguchi A, et al. Characterization and proteome of circulating extracellular vesicles as potential biomarkers for NASH. Hepatol Commun . 2020;4(9):1263–78. Available from: http://dx.doi.org/10.1002/hep4.1556

17. Nakao Y, Amrollahi P, Parthasarathy G, Mauer AS, Sehrawat TS, Vanderboom P, et al. Circulating extracellular vesicles are a biomarker for NAFLD resolution and response to weight loss surgery. Nanomedicine . 2021;36(102430):102430. Available from: http://dx.doi.org/10.1016/j.nano.2021.102430

18. Noren Hooten N, Evans MK. Extracellular vesicles as signaling mediators in type 2 diabetes mellitus. Am J Physiol Cell Physiol . 2020;318(6):C1189–99. Available from: http://dx.doi.org/10.1152/ajpcell.00536.2019

19. Freeman DW, Noren Hooten N, Eitan E, Green J, Mode NA, Bodogai M, et al. Altered extracellular vesicle concentration, cargo, and function in diabetes. Diabetes . 2018;67(11):2377–88. Available from: http://dx.doi.org/10.2337/db17-1308

20. Li S, Wei J, Zhang C, Li X, Meng W, Mo X, et al. Cell-derived microparticles in patients with type 2 diabetes mellitus: A systematic review and meta-analysis. Cell Physiol Biochem . 2016;39(6):2439–50. Available from: http://dx.doi.org/10.1159/000452512

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

21. Huang DQ, Wilson LA, Behling C, Amangurbanova M, Kleiner DE, Kowdley KV, et al. Liver stiffness progression in biopsy-proven metabolic dysfunction-associated steatotic disease among people with diabetes versus people without diabetes: A prospective multicenter study. Hepatology . 2024; Available from: http://dx.doi.org/10.1097/HEP.0000000000001015

22. Liao Y, Liu L, Yang J, Zhou X, Teng X, Li Y, et al. Analysis of clinical features and identification of risk factors in patients with non-alcoholic fatty liver disease based on FibroTouch. Sci Rep . 2023 [cited 2024 Nov 28];13(1):1–11. Available from: https://www.nature.com/articles/s41598-023-41596-2?fbclid=IwY2xjawG1yTZleH-RuA2FlbQIxMAABHTBkdtMpMul0QAop43V639jPE5n2N-C1Nfu0nr-EDJL-PITDYu_a7Zytz6g_aem_tvCV3sNCohPU6GM8ltq70Q

23. Qu Y, Song YY, Chen CW, Fu QC, Shi JP, Xu Y, Xie Q, Yang YF, Zhou YJ, Li LP, Xu MY, Cai XB, Zhang QD, Yu H, Fan JG, Lu LG. Diagnostic performance of FibroTouch ultrasound attenuation parameter and liver stiffness measurement in assessing hepatic steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Clin Transl Gastroenterol. 2021 Apr;12(4):e00323. doi:10.14309/ctg.0000000000000323.

24. Bishop CM. Pattern Recognition and Machine Learning. Springer; 2006. ISBN: 978-0387310732.

25. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd ed. Independently Published; 2022. ISBN: 979-8409805166.

26. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol . 2023;79(6):1542–56. Available from: http://dx.doi.org/10.1016/j.jhep.2023.06.003

27. Buntsma NC, Gąsecka A, Roos YBWEM, van Leeuwen TG, van der Pol E, Nieuwland R. EDTA stabilizes the concentration of platelet-derived extracellular vesicles during blood collection and handling. Platelets . 2022 [cited 2024 Sep 12];33(5):764–71. Available from: https://pubmed.ncbi.nlm.nih.gov/34697987/

28. ScienceDirect. Differential centrifugation . Available from: https://www.sciencedirect.com/topics/chemistry/differential-centrifugation

29. Bachurski D, Schuldner M, Nguyen P-H, Malz A, Reiners KS, Grenzi PC, et al. Extracellular vesicle measurements with nanoparticle tracking analysis – An accuracy and repeatability comparison between NanoSight NS300 and ZetaView. J Extracell Vesicles . 2019 [cited 2024 Sep 12];8(1). Available from: https://pubmed.ncbi.nlm.nih.gov/30988894/

30. Kuhn M, Johnson K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press; 2019. ISBN: 978-1138079229.

31. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems 31 (NeurIPS 2018); 2018. p. 6638–6648.

32. Angelakis A. Cats on the classification of benign and malignant breast lesions using ultrasound shear wave elastography features and BI-RADS score. J. Ultrasound. Med.. 2021.

33. Angelakis A, Chen T. Using FIB-4's parameters an explainable black-box machine learning model outperforms FIB-4 index on the diagnosis of advanced fibrosis of non alcohol related fatty liver disease patients in three cohorts from China, Malaysia and India. Journal of Hepatology. 2023 Jun 1;78:S100-1.

34. Angelakis A. WED-347 Diagnosis of advanced liver fibrosis: the synergy of open data, synthetic data generation, CatBoost, and feature engineering. Journal of Hepatology. 2024 Jun 1;80:S561.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

35. Angelakis A, Soulioti I, Filippakis M. Diagnosis of acute myeloid leukaemia on microarray gene expression data using categorical gradient boosted trees. Heliyon. 2023;9(10).

36. Angelakis A, Nathoe R, Filippakis M. Towards the gene profile of acute myeloid leukaemia using machine learning and blood transcriptomics. Preprints. 2024;2024020593. doi: 10.20944/preprints202402.0593.v1.

37. Angelakis A, Nathoz R, Filippakis M, et al. Uncovering the gene profile of acute myeloid leukaemia using gradient boosted trees. Research Square. Preprint. 2022 Dec 20. doi: 10.21203/rs.3.rs-2374064/v1.

38. Angelakis A, Soulioti I. Diagnosis of acute myeloid leukaemia using machine learning. arXiv preprint. 2021;arXiv:2108.07396.

39. Sheskin DJ. Handbook of Parametric and Nonparametric Statistical Procedures. 5th ed. CRC Press; 2020.

40. Zar JH. Biostatistical Analysis. 5th ed. Pearson Prentice Hall; 2010.

41. Kendall MG. Rank Correlation Methods. 4th ed. Charles Griffin; 1975.

42. Conover WJ. Practical Nonparametric Statistics. 3rd ed. Wiley; 1999.

43. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. Houghton Mifflin; 2003.

44. Kaufman S, Rosset S, Perlich C. Leakage in data mining: Formulation, detection, and avoidance. ACM Transactions on Knowledge Discovery from Data (TKDD). 2012;6(4):1-21. doi: 10.1145/2382577.2382579.

45. Collins GS, Moons KGM, Riley RD, et al. TRIPOD-AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi: 10.1136/bmj.q902.

46. Mitchell TM. Machine Learning. 1st ed. New York: McGraw-Hill; 1997. Inductive learning; p. 52-79.

47. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017;30:4765-4774.

48.Sebastiani P, Thyagarajan B, Sun F, Honig LS, Schupf N, Cosentino S, et al. Age and sex distributions of age-related biomarker values in healthy older adults from the long Life Family Study. J Am Geriatr Soc [Internet]. 2016;64(11):e189–94. Available from: http://dx.doi.org/10.1111/jgs.14522

49. Gortan Cappellari G, Zanetti M, Donini LM, Barazzoni R. Detecting sarcopenia in obesity: emerging new approaches. Curr Opin Clin Nutr Metab Care. 2024;27(5):402-409. doi:10.1097/MCO.0000000000001062.

50. Petroff D, Bätz O, Jedrysiak K, Kramer J, Berg T, Wiegand J. Age dependence of liver enzymes: An analysis of over 1,300,000 consecutive blood samples. Clin Gastroenterol Hepatol [Internet]. 2022;20(3):641–50. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1542356521000896

51.Rastogi A, Rath I, Varadarajan A, Ramakrishna G, Bihari C, Maiwall R. Non-alcoholic fatty liver disease (NAFLD) in lean individuals – Single centre large cohort clinicopathologic and immunophenotypic study. Pathol Res Pract [Internet]. 2022;238(154112):154112. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0344033822003569

52. Sato-Espinoza K, Chotiprasidhi P, Huaman MR, Díaz-Ferrer J. Update in lean metabolic dysfunction-associated steatotic liver disease. World J Hepatol. 2024;16(3):452-464. doi:10.4254/wjh.v16.i3.452. PMID: 38577539; PMCID: PMC10989317.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

53. Gelibter S, Marostica G, Mandelli A, Siciliani S, Podini P, Finardi A, et al. The impact of storage on extracellular vesicles: A systematic study. J Extracell Vesicles [Internet]. 2022;11(2):e12162. Available from: http://dx.doi.org/10.1002/jev2.12162

54. Nelson BC, Maragh S, Ghiran IC, Jones JC, DeRose PC, Elsheikh E, et al. Measurement and standardization challenges for extracellular vesicle therapeutic delivery vectors. Nanomedicine (Lond) [Internet]. 2020;15(22):2149–70. Available from: http://dx.doi.org/10.2217/nnm-2020-0206

55. Kashkanova AD, Blessing M, Reischke M, Baur J-O, Baur AS, Sandoghdar V, et al. Label-free discrimination of extracellular vesicles from large lipoproteins. J Extracell Vesicles [Internet]. 2023;12(8):e12348. Available from: http://dx.doi.org/10.1002/jev2.12348

56. Sivanantham A, Jin Y. Impact of storage conditions on EV integrity/surface markers and cargos. Life (Basel) [Internet]. 2022;12(5). Available from: http://dx.doi.org/10.3390/life12050697

57. Zou X, Lei Q, Luo X, Yin J, Chen S, Hao C, et al. Advances in biological functions and applications of apoptotic vesicles. Cell Commun Signal [Internet]. 2023;21(1). Available from: http://dx.doi.org/10.1186/s12964-023-01251-9\

8. Tables

Feature	Count	Missing (%)	Mean	Std
Age	74	2.63	55.34	13.20
Height	75	1.32	1.68	0.09
Weight	75	1.32	80.81	14.30
Sex	75		0.59	0.50
Diabetes	76	0	0.20	0.40
S2, S2, S3	76	0	0.71	0.46
S3	76	0	0.26	0.44
Adv_Fibrosis	76	0	0.11	0.31
Mean size	74	2.63	138.16	19.76
50-150nm	74	2.63	8.27x10^10	4.37x10^10
>150 nm	74	2.63	5.84x10^10	4.85x10^10
Sum (50-1000nm)	74	2.63	1.41x10^11	7.04x10^10

Table 1. Patient Characteristics.

Group	Total Patients	Advanced Fibrosis (n/%)	Diabetes (n/%)	Advanced Fibrosis & Diabetes (n/%)

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

S0	21	0/21 (0.0%)	0/21 (0.0%)	0/21 (0.0%)
S1-S3	53	6/53 (11.3%)	13/53 (24.5%)	5/53 (9.4%)
S3	19	3/19 (15.8%)	9/19 (47.4%)	3/19 (15.8%)
S0-S2	55	3/55 (5.5%)	4/55 (7.3%)	2/55 (3.6%)
S1-S2	34	3/34 (8.8%)	4/34 (11.8%)	2/34 (5.9 %)

Table 2. Percentages of advanced fibrosis and diabetes in the groups of S0, S1-S3, S3, S0-S2 and S1-S2.

Model	Feats	Iters	LR	Depth	W0	W1	Dataset	Spec. (+/-std) 95%CI	Sens. (+/-std) 95%CI	AUROC (+/-std) 95%CI	F1 (+/-std) 95%CI
CB-C1a	4	21	0.20	13	0.68	0.32	5CV	0.74 (+/-0.14) [0.60, 1.00]	0.66 (+/-0.17) [0.43, 0.89]	0.70 (+/-0.10) [0.51, 0.78]	0.73 (+/-0.14) [0.50, 0.89]
CB-C1a	4	21	0.20	13	0.68	0.32	Test	0.83	0.89	0.86	0.89
CB-C1b	54	20	0.60	3	0.68	0.32	5CV	0.60 (+/-0.33) [0.00, 1.00]	0.80 (+/-0.07) [0.71, 0.89]	0.70 (+/-0.15) [0.44, 0.88]	0.81 (+/-0.06) [0.71, 0.88]
CB-C1b	54	20	0.60	3	0.68	0.32	Test	0.67	0.89	0.78	0.84
CB-C1c	9	10	0.10	8	0.68	0.32	5CV	0.60 (+/-0.23) [0.29, 0.80]	0.67 (+/-0.24) [0.47, 1.00]	0.63 (+/-0.01) [0.63, 0.63]	0.69 (+/-0.10) [0.61, 0.83]
CB-C1c	9	10	0.10	8	0.68	0.32	Test	0.67	0.89	0.78	0.84
CB-C1d	125	3	0.1	5	0.68	0.32	10CV	0.56 (+/-0.26) [0.25, 1.00]	0.71 (+/-0.19) [0.50, 1.00]	0.64 (+/-0.18) [0.44, 0.94]	0.72 (+/-0.13) [0.57, 0.93]
CB-C1d	125	3	0.1	5	0.68	0.32	Test	0.83	0.89	0.86	0.89
CB-C1e	6	6	0.70	8	0.68	0.32	5CV	0.69 (+/-0.21) [0.50, 1.00]	0.69 (+/-0.29) [0.20, 1.00]	0.69 (+/-0.19) [0.35, 0.93]	0.63 (+/-0.22) [0.31, 0.92]
CB-C1e	6	6	0.70	8	0.68	0.32	Test	0.75	1.00	0.93	0.92
CB-C1f	11	2	0.60	4	0.65	0.35	5CV	0.63 (+/-0.34) [0.00, 1.00]	0.75 (+/-0.16) [0.50, 1.00]	0.69 (+/-0.20) [0.39, 0.99]	0.78 (+/-0.12) [0.62, 0.90]
CB-C1f	11	2	0.60	4	0.65	0.35	Test	1.00	0.80	0.90	0.89

Table 3. ML models of the C1: the name of the model, the number of features for each case, the values of their tuned hyper-parameters, the classes' weights, the specificity, sensitivity, ROC-AUC, F1 scores, their mean and standard deviation as well as the 95% confidence interval for the 5CV on the train set and on the test set.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Model	Feats	iters	lr	depth	a	b	Dataset	spec (+/-std) 95%CI	sens (+/-std) 95%CI	roc (+/-std) 95%CI	f1 (+/-std) 95%CI
CB-C2a	4	5	0.6	5	0.25	0.75	3CV	0.62 (+/-0.12) [0.50, 0.79]	0.79 (+/-0.03) [0.75, 0.83]	0.71 (+/-0.06) [0.66, 0.79]	0.55 (+/-0.10) [0.43, 0.67]
CB-C2a	4	5	0.6	5	0.25	0.75	Test	0.73	1.00	0.86	0.73
CB-C2b	54	50	0.8	11	0.17	0.83	3CV	0.89 (+/-0.08) [0.81, 1.00]	0.62 (+/-0.10) [0.50, 0.75]	0.75 (+/-0.02) [0.73, 0.78]	0.62 (+/-0.03) [0.60, 0.67]
CB-C2b	54	50	0.8	11	0.17	0.83	Test	1.00	1.00	1.00	1.00
CB-C2c	9	40	0.4	4	0.23	0.77	3CV	0.71 (+/-0.07) [0.81, 1.00]	0.72 (+/-0.11) [0.50, 0.75]	0.71 (+/-0.09) [0.73, 0.78]	0.59 (+/-0.09) [0.60, 0.67]
CB-C2c	9	40	0.4	4	0.23	0.77	Test	1.00	1.00	1.00	1.00
CB-C2d	125	16	0.85	4	0.23	0.77	3CV	0.80 (+/-0.17) [0.50, 1.00]	0.75 (+/-0.21) [0.50, 1.00]	0.78 (+/-0.11) [0.68, 1.00]	0.62 (+/-0.23) [0.29, 1.00]
CB-C2d	125	16	0.85	4	0.23	0.77	Test	1.00	1.00	1.00	1.00
CB-C2d-53	53	36	0.6	4	0.17	0.83	3CV	0.76 (+/-0.02) [0.73, 0.77]	0.81 (+/-0.05) [0.75, 0.86]	0.79 (+/-0.03) [0.74, 0.81]	0.67 (+/-0.09) [0.55, 0.75]
CB-C2d-53	53	36	0.6	4	0.17	0.83	Test	1.00	0.75	0.88	0.86
CB-C2d-11	11	4	0.3	8	0.17	0.83	3CV	0.71 (+/-0.04) [0.67, 0.77]	0.86 (+/-0.10) [0.75, 1.00]	0.79 (+/-0.07) [0.71, 0.88]	0.73 (+/-0.13) [0.50, 0.82]
CB-C2d-11	11	4	0.3	8	0.17	0.83	Test	0.70	1.00	0.85	0.73
CB-C2d-5	5	84	0.82	4	0.17	0.83	3CV	0.86 (+/-0.05) [0.80, 0.92]	0.77 (+/-0.05) [0.71, 0.83]	0.81 (+/-0.03) [0.78, 0.84]	0.71 (+/-0.08) [0.60, 0.77]
CB-C2d-5	5	84	0.82	4	0.17	0.83	Test	1.00	1.00	1.00	1.00
CB-C2e-6	6	25	0.6	7	0.25	0.75	3CV	0.77 (+/-0.13) [0.60, 0.91]	0.77 (+/-0.18) [0.57, 1.00]	0.77 (+/-0.02) [0.74, 0.80]	0.64 (+/-0.10) [0.50, 0.75]
CB-C2e-6	6	25	0.6	7	0.25	0.75	Test	0.83	1.00	0.92	0.86
CB-C2f-87	87	5	0.4	6	0.26	0.74	3CV	0.75 (+/-0.16) [0.53, 0.90]	0.82 (+/-0.13) [0.71, 1.00]	0.79 (+/-0.03) [0.77, 0.83]	0.66 (+/-0.14) [0.46, 0.80]
CB-C2f-87	87	5	0.4	6	0.26	0.74	Test	0.83	1.00	0.92	0.86
CB-C2f-22	22	4	0.7	13	0.28	0.72	3CV	0.90 (+/-0.08) [0.80, 1.00]	0.86 (+/-0.20) [0.57, 1.00]	0.88 (+/-0.07) [0.79, 0.95]	0.78 (+/-0.12) [0.67, 0.94]
CB-C2f-22	22	4	0.7	13	0.28	0.72	Test	0.92	1.00	0.96	0.92
CB-C2g-11	11	44	0.3	6	0.18	0.82	3CV	0.82 (+/-0.14) [0.67, 1.00]	0.77 (+/-0.21) [0.50, 1.00]	0.79 (+/-0.07) [0.73, 0.89]	0.67 (+/-0.16) [0.50, 0.73]
CB-C2g-11	11	44	0.3	6	0.18	0.82	Test	1.00	0.83	0.92	0.91
CB-C2h-151	151	151	0.7	5	0.32	0.68	3CV	0.85 (+/-0.12) [0.71, 1.00]	0.79 (+/-0.29) [0.38, 1.00]	0.82 (+/-0.10) [0.69, 0.92]	0.68 (+/-0.12) [0.55, 0.83]
CB-C2h-151	151	151	0.7	5	0.32	0.68	Test	1.00	1.00	1.00	1.00
CB-C2h-46	46	75	0.9	6	0.16	0.84	3CV	0.85 (+/-0.11) [0.75, 1.00]	0.83 (+/-0.12) [0.75, 1.00]	0.84 (+/-0.05) [0.77, 0.88]	0.74 (+/-0.11) [0.60, 0.86]
CB-C2h-46	46	75	0.9	6	0.16	0.84	Test	1.00	1.00	1.00	1.00
CB-C2h-21	21	15	0.3	5	0.14	0.86	3CV	0.87 (+/-0.10) [0.75, 1.00]	0.92 (+/-0.12) [0.75, 1.00]	0.89 (+/-0.03) [0.88, 0.93]	0.81 (+/-0.04) [0.77, 0.86]
CB-C2h-21	21	15	0.3	5	0.14	0.86	Test	1.00	1.00	1.00	1.00

Table 4. ML models of the C2: the name of the model, the number of features for each case, the values of their tuned hyper-parameters, the classes' weights, the specificity, sensitivity, ROC-AUC, F1 scores, their mean and standard deviation as well as the 95% confidence interval for the 5CV on the train set and on the test set.

	spec_cv_m	sens_cv_m	roc_cv_m	f1_cv_m
mean	0.76	0.66	0.71	0.72
std	0.05	0.06	0.03	0.05
min	0.65	0.53	0.67	0.64
25%	0.73	0.63	0.68	0.68
50%	0.76	0.66	0.71	0.73
75%	0.79	0.70	0.73	0.75
max	0.83	0.74	0.77	0.78

	spec_cv_m	sens_cv_m	roc_cv_m	f1_cv_m
mean	0.80	0.81	0.81	0.71
std	0.06	0.06	0.04	0.05
min	0.68	0.74	0.77	0.64
25%	0.78	0.76	0.78	0.67
50%	0.81	0.80	0.79	0.69
75%	0.85	0.85	0.82	0.73
max	0.89	0.92	0.89	0.81

Table 5. Performance on the the CB-C1a (left) and CB-C2h-21 (right) using 10 times iterative 5CV and 3CV, for C1 and C2 accordingly. Here x_cv_m for x in {spec, sens, roc, f1} corresponds to {specificity, sensitivity, AUROC, F1}.

9. Figure Legends

Figure 1. Plots of feature importance on the predictability of the ML model CB-C1a and on SHAP on the train set (ROC-AUC train: 0.70, test: 0.86).

Figure 2. Plots of feature importance on the predictability of the ML model CB- C2h-21 and on

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

SHAP on the train set (ROC-AUC train: 0.89, test: 1.00)

Figure 3. Plots of EVs distribution above and SHAP scatter plots for the train set of the ML CB-C1a model.

Figure 4. Plots of features distribution above and SHAP scatter plots for the train set of the ML CB-C2h-21 model.

Figure 5. Plots of correlations of parametric and non-parametric tests for C1 and C2

10. Figures

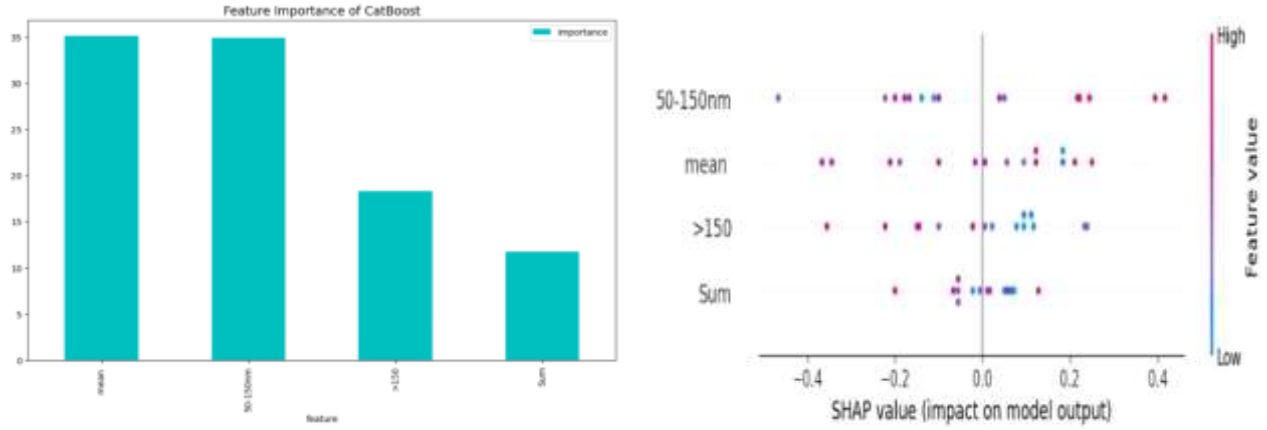


Figure 1. Plots of feature importance on the predictability of the ML model CB-C1a and on SHAP on the train set (ROC-AUC train: 0.70, test: 0.86).

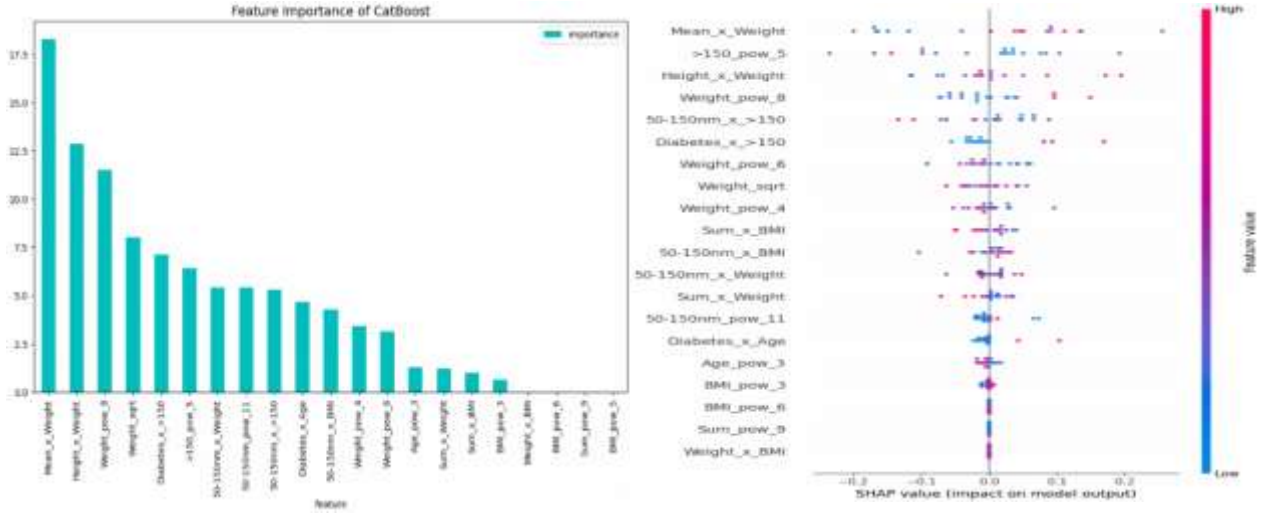


Figure 2 Plots of feature importance on the predictability of the ML model CB- C2h-21 and on SHAP on the train set (ROC-AUC train: 0.89, test: 1.00)

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

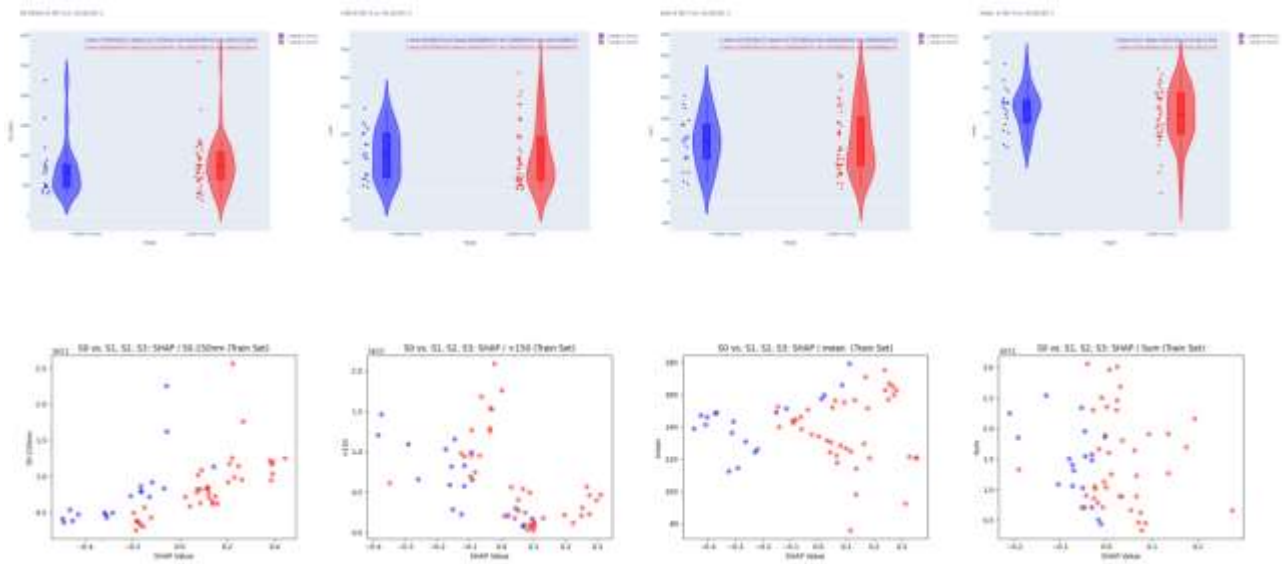


Figure 3. Plots of EVs distribution above and SHAP scatter plots for the train set of the ML CB-C1a model.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

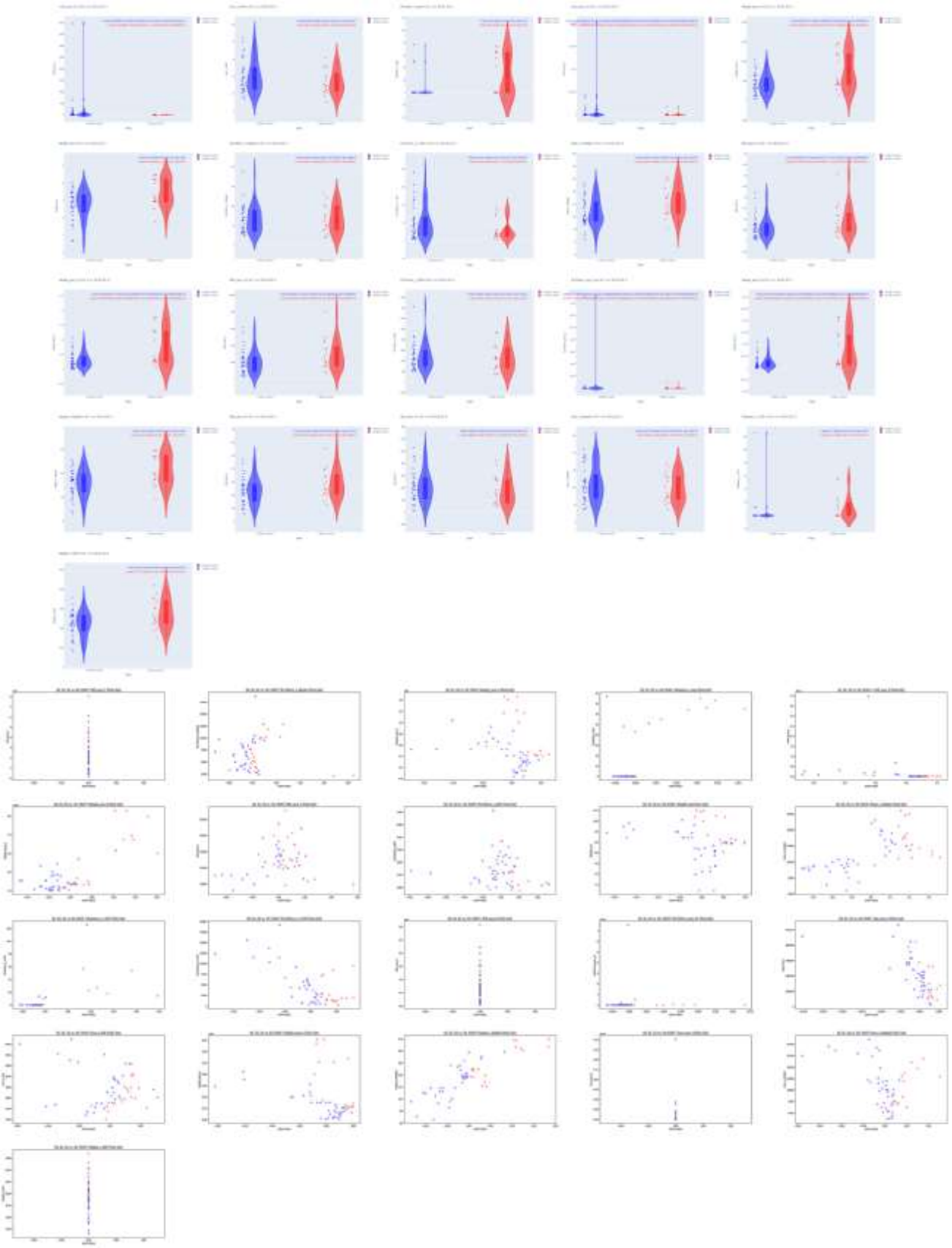


Figure 4. Plots of features distribution above and SHAP scatter plots for the train set of the ML CB-C2h-21 model.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.07.24318644; this version posted December 8, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

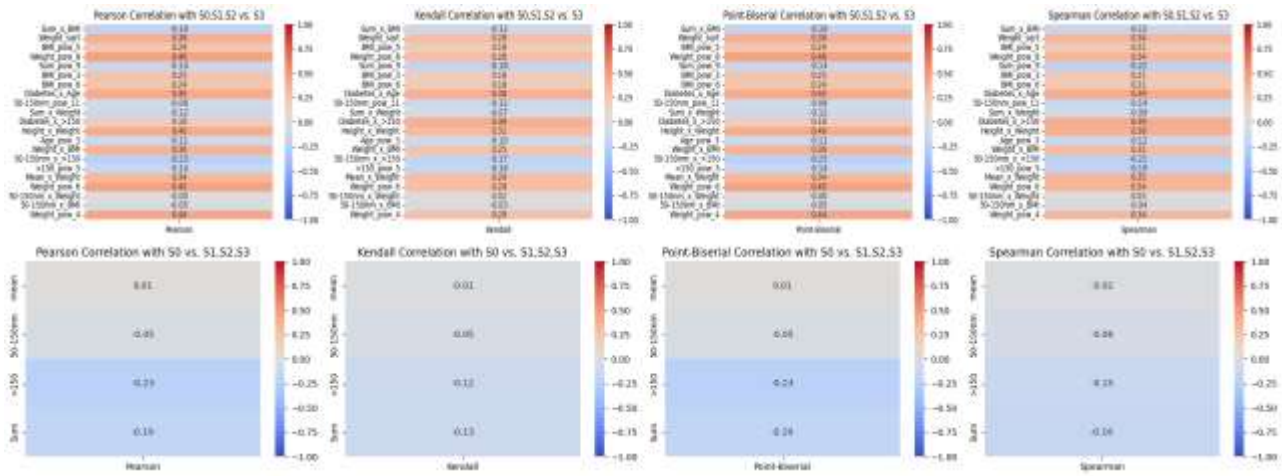


Figure 5. Plots of correlations of parametric and non-parametric tests for C1 and C2.