Supplementary material

1. C1, C2 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.
- 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. Algorithm

- Input: VE, anthropomorphic, clinical features, and the UAP 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 of 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.

8 Feature Selection Methodology:

- Apply FS

9 Identify the best model for each case:

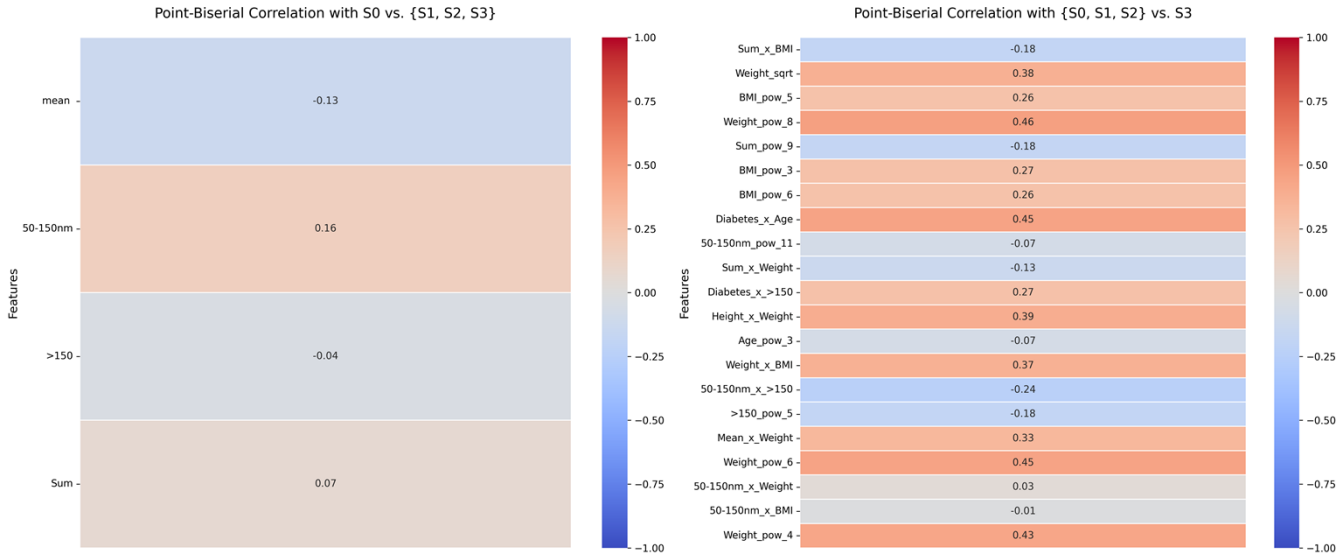
- From the statistics on the kCV and the performance on the test set identify the best model for each case.

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.



Supplementary Figure 1 Point-Biserial correlation of CB-C1a and CB-C2h-21 features with C1 and C2 cases accordingly.

Supplementary Table 1 Information regarding all the ML models of the C1. The name of the model, the number of features for each case, the dataset, 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 of C1 and the test set, appear at the table below1. With bold appear the best values for the corresponding metric

Model	Feats	Dataset	Spec. (+/-std) 95%CI	Sens. (+/-std) 95%CI	ROC-AUC (+/-std) 95%CI	F1 (+/-std) 95%CI
CB-C1a	4	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	Test	**0.83**	**0.89**	**0.86**	**0.89**
CB-C1b	54	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	Test	0.67	0.89	0.78	0.84
CB-C1c	9	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	Test	0.67	0.89	0.78	0.84
CB-C1d	125	5CV	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	Test	0.83	0.89	0.86	0.89
CB-C1e	6	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	Test	0.75	1.00	0.93	0.92
CB-C1f	11	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	Test	1.00	0.80	0.90	0.89

[1] For each model the values of its hyper-parameters appear in parenthesis (iterators, learning rate, depth, weight for class 0, weight for class 1): CB-C1a(21, 0.20, 13, 0.68, 0.32); CB-C1b(20, 0.60, 3, 0.68, 0.32); CB-C1c(10, 0.10, 8, 0.68, 0.32); CB-C1d(3, 0.10, 5, 0.68, 0.32); CB-C1e(6, 0.70, 8, 0.68, 0.32); CB-C1f(2, 0.60, 4, 0.65, 0.35).

Supplementary Table A2. Information regarding all the machine learning models of C2. The name of the model, the number of features for each case, the dataset, the specificity, sensitivity, ROC-AUC, F1 scores, their mean and standard deviation as well as the 95% confidence interval for the 3CV on the train set of C2, and the test set, appear at the table below2. With bold appear the best values for the corresponding metric.

Model	Feats	Dataset	Spec. (+/-std) 95%CI	Sens. (+/-std) 95%CI	ROC-AUC (+/-std) 95%CI	F1 (+/-std) 95%CI
CB-C2a	4	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	Test	0.73	1.00	0.86	0.73
CB-C2b	54	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	Test	1.00	1.00	1.00	1.00
CB-C2c	9	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	Test	1.00	1.00	1.00	1.00
CB-C2d	125	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	Test	1.00	1.00	1.00	1.00
CB-C2d-53	53	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	Test	1.00	0.75	0.88	0.86
CB-C2d-11	11	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	Test	0.70	1.00	0.85	0.73
CB-C2d-5	5	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	Test	1.00	1.00	1.00	1.00
CB-C2e-6	6	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	Test	0.83	1.00	0.92	0.86
CB-C2f-87	87	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	Test	0.83	1.00	0.92	0.86
CB-C2f-22	22	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	Test	0.92	1.00	0.96	0.92
CB-C2g-11	11	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	Test	1.00	0.83	0.92	0.91
CB-C2h-151	151	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	Test	1.00	1.00	1.00	1.00
CB-C2h-46	46	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, 0.76]

CB-C2h-46	46	Test	1.00	1.00	1.00	1.00
CB-C2h-21	21	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	Test	**1.00**	**1.00**	**1.00**	**1.00**

[2] For each model the values of its hyper-parameters appear in parenthesis (iterators, learning rate, depth, weight for class 0, weight for class 1): CB-C2a:(5, 0.60, 5, 0.25, 0.75); CB-C2b(50, 0.80, 11, 0.17, 0.83); CB-C2c(40, 0.40, 4, 0.23, 0.77); CB-C2d(16, 0.85, 4, 0.23, 0.77); CB-C2d-53(36, 0.60, 4, 0.17, 0.83); CB-C2d-11(4, 0.30, 8, 0.17, 0.83); CB-C2d-5(84, 0.82, 4, 0.17, 0.83); CB-C2e-6(25, 060, 7, 0.25, 0.75); CB-C2f-87(5, 0.40, 6, 0.26, 0.74); CB-C2f-22(4, 0.70, 13, 0.28, 0.72); CB-C2g-11(44, 0.30, 6, 0.18, 0.82); CB-C2h-151(151, 0.70, 5, 0.32, 0.68); CB-C2h-46(75, 0.90, 6, 0.16, 0.84); CB-C2h-21(15, 0.30, 5, 0.14, 0.86).

49 Spadina Ave. Suite 200
Toronto ON M5V 2J1 Canada
www.biorender.com

Confirmation of Publication and Licensing Rights - Open Access

April 22nd, 2025

Subscription Type:	*Postdoc Plan - Academic*
Agreement number:	*MC286HVIPV*
Publisher Name:	*World Journal of Gastroenterology*
Figure Title:	*Graphical Abstract*
Citation to Use:	*Created in BioRender. Fortis, S. (2025) https://BioRender.com/rv900sp*

To whom this may concern,

This document ("Confirmation") hereby confirms that Science Suite Inc. dba BioRender ("BioRender") has granted the following BioRender user: Sotirios Fortis ("User") a BioRender Academic Publication License in accordance with BioRender's Terms of Service and Academic License Terms ("License Terms") to permit such User to do the following on the condition that all requirements in this Confirmation are met:

1) publish their Completed Graphics created in the BioRender Services containing both User Content and BioRender Content (as both are defined in the License Terms) in publications (journals, textbooks, websites, etc.); and
2) sublicense such Completed Graphics under "open access" publication sublicensing models such as CC-BY 4.0 and more restrictive models, so long as the conditions set forth herein are fully met.

Requirements of User:

1) All Completed Graphics to be published in any publication (journals, textbooks, websites, etc.) must be accompanied by the following citation either as a caption, footnote or reference for each figure that includes a Completed Graphic:
"Created in BioRender. Fortis, S. (2025) https://BioRender.com/rv900sp ".
2) All terms of the License Terms including all Prohibited Uses are fully complied with. E.g. For Academic License Users, no commercial uses (beyond publication in journals, textbooks or websites) are permitted without obtaining or switching to a BioRender Industry Plan.
3) A Reader (defined below) may request that the User allow their figure to be a public template for Readers to view, copy, and modify the figure. It is up to the User to determine what level of access to grant.

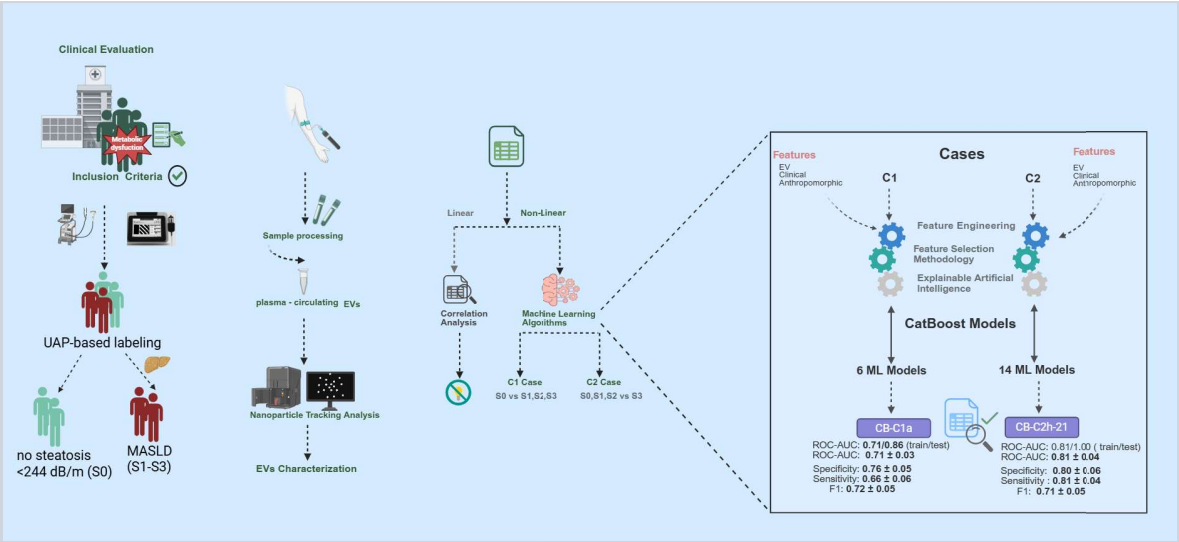
Open-Access Journal Readers:

Open-Access journal readers ("Reader") who wish to view and/or re-use a particular Completed Graphic in an Open-Access journal subject to CC-BY sublicensing may do so by clicking on the URL link in the applicable citation for the subject Completed Graphic.

The re-use/modification options below are available after the Reader requests the User to adapt their

figure as a BioRender template and the User has granted such access.

1) View-Only/Free Plan Use: A Reader who wishes to only view the Completed Graphic may do so in the BioRender Services as either a BioRender Free Plan user or simply as a viewer. By becoming a BioRender Free Plan user, the Reader may view, modify and re-use the Completed Graphic as permitted under BioRender's Basic License Terms (e.g. personal use only, no publishing or commercial use permitted).
2) Re-Use/Publish with No Modifications: For any re-use and re-publication of a Completed Graphic with no modification(s) to the Completed Graphic made by the Reader, a Reader may do so by citing the original author using the citation noted above with the Completed Graphic. The Reader must also comply with the underlying License Terms which apply to the Completed Graphic as noted above (e.g. no commercial use for Academic License).
3) Re-Use/Publish with Modifications: For any re-use and re-publication of a Completed Graphic with a modification(s) made by the Reader, the Reader may do so by becoming a BioRender user themselves under either an Academic or Industry Plan, citing the original author using the citation noted above with the Completed Graphic and complying with the applicable License Terms.



For any questions regarding this document, or other questions about publishing with BioRender, please refer to our BioRender Publication Guide, or contact BioRender Support at support@biorender.com.