BMJ Health & Care Informatics

Biomarker and clinical data–based predictor tool (MAUXI) for ultrafiltration failure and cardiovascular outcome in peritoneal dialysis patients: a retrospective and longitudinal study

Eva María Arriero-País ,[1] María Auxiliadora Bajo-Rubio,[2,3] Roberto Arrojo-García,[1] Pilar Sandoval,[1,3] Guadalupe Tirma González-Mateo,[3,4] Patricia Albar-Vizcaíno,[5] Gloria del Peso-Gilsanz,[3,5] Marta Ossorio-González,[3,5] Pedro Majano,[6] Manuel López-Cabrera,[1,3] Redes de Investigación Cooperativa Orientadas a Resultados en Salud (RICORS 2040-Renal)

To cite: Arriero-País EM, Bajo-Rubio MA, Arrojo-García R, *et al*. Biomarker and clinical data–based predictor tool (MAUXI) for ultrafiltration failure and cardiovascular outcome in peritoneal dialysis patients: a retrospective and longitudinal study. *BMJ Health Care Inform* 2025;**32**:e101138. doi:10.1136/bmjhci-2024-101138

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2024-101138).

Received 21 May 2024
Accepted 12 February 2025

For numbered affiliations see end of article.

Correspondence to
Prof. Dr. Manuel López-Cabrera; mlcabrera@cbm.csic.es

ABSTRACT

Objectives To develop a machine learning-based software as a medical device to predict the endurance and outcomes of peritoneal dialysis (PD) patients in real time using effluent-measured biomarkers of the mesothelial-to-mesenchymal transition (MMT).

Methods Retrospective, longitudinal, triple blind study in two independent hospitals (Spain), designed under information-theoretical approaches for feature selection and machine learning-based modelling techniques. A total of 151 (train set) and 32 (validation) PD patients in 1979–2022 were included. PD outcomes were analysed in four categories (endurance, exit from PD, cause of PD end, technical failure) by using MMT biomarkers in effluents and clinical databases.

Results MMT biomarkers and clinical data can predict PD with a mean absolute error of 16.99 months by using an Extra Tree (ET) regressor. Linear discriminant analysis (LDA) discerns among transfer to haemodialysis or death, predicts whether the cause of PD end is ultrafiltration failure (UFF) or cardiovascular disease (CVD) and anticipates the type of CVD (receiver operating characteristic curve under the area>0.71).

Discussion Our combination of longitudinal PD datasets, attribute shrinkage and gold-standard algorithms with overfitting testing and class imbalance ensures robust predictions in PD. Biomarkers displayed proper mutual information and SHapley values, indicating that MMT processes may have a causal relationship in the development of UFF and CVD.

Conclusions MMT biomarkers and clinical data may be associated in a causal manner with ultrafiltration failure (local effect) and cardiovascular events (systemic effect) in PD. The machine learning-based software MAUXI provides applicability of ET-LDA models with ≤38 variables to predict PD endurance and type of PD technique failure related to peritoneal membrane deterioration.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Prediction of peritoneal dialysis (PD) endurance and technique failure (ultrafiltration failure (UFF)-cardiovascular disease) are still unravelled. Previous machine learning (ML) techniques had been tested with moderate accuracy within a time span of 5–7 years.

⇒ Ultrafiltration and cardiovascular events take place within the first 29–60 months of PD treatment, giving importance to accurate predictions to implement prophylactic interventions.

WHAT THIS STUDY ADDS

⇒ Our study demonstrates that mesothelial-to-mesenchymal transition biomarkers and clinical data under ML models (MAUXI software) can predict endurance and different PD technique failures, opening new avenues to individual treatments.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ MAUXI software implies novel interpretability of complex models based on artificial intelligence in the cardiorenal field.

⇒ Paving the way to accurate predictions in PD technique will lead to an unprecedented development of prophylactic interventions related to the neurological, cardiovascular and UFF events. This will make possible the reduction of the cost burden of the European budget on PD withdrawal.

INTRODUCTION

Peritoneal dialysis (PD) is a home care, cost-effective kidney replacement therapy for removal of excess water, electrolytes and toxic metabolic products from the body. PD is based on infusing a sterile hyperosmotic solution into the peritoneal cavity. During PD, there is an ultrafiltration (UF) process

based on hydrostatic pressure (convection) and oncotic pressure (diffusion) between the blood and the PD solution (PDS) through the peritoneal membrane (PM).[1] Together with haemodialysis (HD), PD is a life-saving treatment for chronic kidney disease (CKD) and end-stage renal disease (ESRD). UF failure (UFF) occurs when patients experience long-term ultrafiltration rate (UFR) of less than 400 mL water removal in a 4-hour dwell (UFR4H) using a dextrose solution. The 6-year UFF incidence ranges from 30% to 60%, where around 54% will be transferring to HD or dying on cardiovascular disease (CVD) in concomitance with UFF.[2]

The structure of the PM is composed of a single layer of mesothelial cells (MCs) that lines a compact zone of connective tissue containing few fibroblasts, mast cells, macrophages and vessels. The PM is a semipermeable membrane, which is responsible for the UFR and UFF.[3] During PD, mesothelial genes are silenced to allow induction of mesenchymal signatures. This process is known as mesothelial-to-mesenchymal transition (MMT), with prototypical epithelioid and non-epithelioid morphologies representing early and advanced MMT. Advanced MMT is associated with deregulated secreted MMT biomarkers and mass transfer coefficient (MTC) of creatinine ≥11 mL/min.[4 5]

Machine learning (ML) is a branch of artificial intelligence (AI) with learning capability of data-driven experiences, avoiding theory-driven priors and factor-balanced hypotheses. Despite limitations, ML and deep learning algorithms are being used in CKD, ESRD and PD, among others.[6] Soluble surrogate effluent biomarkers are researched for the non-invasive validation of the PM function, but guidelines are still scarce—being this the rationale for this study.[7] Therefore, we propose a novel prediction software as a medical device, MAUXI, based on ML and MMT-associated biomarkers with robust accuracy to determine PD endurance and technique failure.

METHODS

Patients and data registry

Our study was built on the biobank of the Hospital Universitario La Paz (Madrid, Spain), comprising a total of 921 patients initiating PD from 1979 to the present times (table 1, online supplemental tables 1–3). Thus, we generated a platform with electronic medical records of each patient. This biobank benefits from semi-annual peritoneal equilibration tests (PETs), CVD and peritonitis events and a collection of longitudinal effluents, plasm and sera of every single patient. The PET study was performed following the method of Twardowski.[8]

Importantly, definitions of the four outcomes were set up by medical doctors of our team and stayed in line with the current body of research on major cardiac events.[9]

The first PD outcome to predict (primary endpoint) was endurance as the remaining time for a given patient until PD technique failure due to CVD or UFF.

- Endurance (in remaining months in PD).

As secondary endpoints, other specific PD outcomes 'to predict' (categorical variables) in the models of this study were defined as:

- Exit: transfer to HD, exitus (known as exitus letalis or death).
- Cause of PD end: UFF, CVD.
- Technical failure: ictus (brain haemorrhage), ischaemic cardiac congestion (ICC), vascular events (amputations and peripheral artery disease, among others), MMT, peritonitis-induced MMT (MMT-peritonitis).

Thus, we provide the description of the outcomes to predict in each ML model in the online supplemental methods, as well as the details of the external positive and negative datasets to validate algorithms.

Patient involvement

Patients could not participate in the study (see online supplemental methods).

Sample preservation and cell culture

For all the duration of PD, patients were extracted regularly every 6 months effluent, serum and plasm, which were frozen at −80°C until their analysis. Six dwells of effluents (0–240 min) were measured routinely for 188 variables, including anthropometrics and peritoneal transport as seen in PETs. Phenotypic information of MCs was recycled from a database from other previous studies, in which isolation and culture of MCs from human effluents was performed, as previously described.[4]

Biomarker selection and multiplex ELISA

To predict PD failure and endurance in our PD population, we selected 13 biomarker proteins related to the MMT that were specifically produced and secreted by MCs, as provided in our previously reported microarrays.[4] Selected biomarkers were: matrix metalloproteinase-2 (MMP-2), tenascin-C (TN-C), interleukin-11 (IL-11), plasminogen activator inhibitor-1 (PAI-1), periostin (PSTN), vascular endothelial growth factor A (VEGF-A), collagen-13 (COL-13), cadherin-13 (CDH-13), thrombospondin-1 (TSP-1), bone morphogenetic protein-7 (BMP-7), IL-6, fibroblast activation protein (FAP) and IL-33.

All biomarkers have been associated with MMT or epithelial-mesenchymal transition (EMT) processes, including in malignancies and on contact with PDS, though different pathways—being BMP-7 the prototypical counteractor.[10–13] Proteins found in diabetes and glucose imbalances are VEGF-A, TSP-1, PSTN and CDH-13. In cardiotoxic status related to hypertrophy, cardiomyopathy, infarction, coronary diseases or neurological disorders, all proteins have been found to be involved.[14–20] Merely TSP-1, CDH-13, PAI-1, BMP-7 and TN-C in certain levels or isoforms can act as cardioprotective.[21–23]

Multiplex ELISA Quantibody (RayBio Human QuantiBody, RayBiotech Inc., Peterborough, UK) arrays were suitable for the simultaneous quantification of soluble proteins in effluents, sera and plasm with the guarantee

Table 1 Patient characteristics, PETs and biomarkers of the PD treatment

Variable (abbreviation or acronym)	Definition of variable abbreviation or acronym	C0 n=187	C1 n=113	C2 n=55	C3 n=14	P value	N
Age	Years; chronological age of patient	61.8 (52.0; 72.6)	61.4 (46.4; 71.0)	52.6 (42.5; 66.0)	54.7 (45.0; 61.0)	<0.001	369
Months	Time in current PD treatment in our hospitals when sample is measured	24.9 (10.7; 49.2)	11.5 (1.97;27.0)	7.93 (1.63; 24.6)	5.10 (1.93; 25.3)	<0.001	369
Endurance	Months; remaining lifespan of treatment (how long a doctor can treat the patient with PD)	6.62 (2.16; 13.4)	31.2 (24.9; 39.7)	63.1 (56.9; 75.0)	104 (98.9; 123)	<0.0001	369
PRIORPD	Months; prior PD treatment in other hospital, transferred patient to our hospital	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)	<0.001	369
PERITONITISFREC ACCUM	Accumulated sum of amount of peritonitis until the date of this sample	0.00 (0.00; 1.00)	0.00 (0.00; 1.00)	0.00 (0.00; 1.00)	0.00 (0.00; 0.00)	<0.001	369
CVDSUM ACCUM	Accumulated sum of number of all CVD and neurological events until the date of this sample	1.00 (0.00; 2.00)	1.00 (0.00; 2.00)	1.00 (0.00; 1.00)	0.00 (0.00; 0.00)	<0.001	369
Ethnicity Latin-merican	Given ethnicity (if not white and Latin, then black)					<0.001	369
	No	176 (94.1%)	(97.3%)	50 (90.9%)	14 (100%)		
	Yes	11 (5.88%)	3 (2.65%)	5 (9.09%)	0 (0.00%)		
Ethnicity white	Ethnicity					<0.001	369
	No	12 (6.42%)	5 (4.42%)	5 (9.09%)	0 (0.00%)		
	Yes	175 (93.6%)	108 (95.6%)	50 (90.9%)	14 (100%)		
Sex	Gender					<0.001	369
	Female	83 (44.4%)	37 (32.7%)	23 (41.8%)	9 (64.3%)		
	Male	104 (55.6%)	76 (67.3%)	32 (58.2%)	5 (35.7%)		
DB	Diabetes (diagnosed)					<0.001	369
	No	114 (61.0%)	67 (59.3%)	30 (54.5%)	13 (92.9%)		
	Yes	73 (39.0%)	46 (40.7%)	25 (45.5%)	1 (7.14%)		
HTA	Hypertension (arterial)					<0.001	369
	No	17 (9.09%)	4 (3.54%)	1 (1.82%)	0 (0.00%)		
	Yes	170 (90.9%)	109 (96.5%)	54 (98.2%)	14 (100%)		
DL	Dyslipidaemia (simple, mixed, hyper, cholesterolaemia)					<0.001	369
	No	72 (38.5%)	44 (38.9%)	15 (27.3%)	3 (21.4%)		
	Yes	115 (61.5%)	69 (61.1%)	40 (72.7%)	11 (78.6%)		
Exit						<0.001	369

Continued

Table 1 Continued

Variable (abbreviation or acronym)	Definition of variable abbreviation or acronym	C0 n=187	C1 n=113	C2 n=55	C3 n=14	P value	N
EXITUS	Death of patient as main cause of peritoneal dialysis failure	75 (40.1%)	55 (48.7%)	25 (45.5%)	8 (57.1%)		
HD	Transfer to haemodialysis (HD) due to ultrafiltration failure (UFF) of patient as main cause of peritoneal dialysis failure	112 (59.9%)	58 (51.3%)	30 (54.5%)	6 (42.9%)		
PD cause end						<0.001	369
CVD	Diagnosed cardiovascular disease of patient as main cause of peritoneal dialysis failure	49 (26.2%)	27 (23.9%)	9 (16.4%)	4 (28.6%)		
UFF	Diagnosed UFF of patient as main cause of peritoneal dialysis failure	138 (73.8%)	86 (76.1%)	46 (83.6%)	10 (71.4%)		
Failure						<0.001	369
ICC	Diagnosed cardiac arrest of patient as main cause of peritoneal dialysis failure	24 (12.8%)	16 (14.2%)	2 (3.64%)	3 (21.4%)		
ICTUS	Diagnosed fatal cerebrovascular infarct of patient as main cause of peritoneal dialysis failure	7 (3.74%)	3 (2.65%)	2 (3.64%)	1 (7.14%)		
MMT	Presumed MMT in patient leading to diagnosed UFF as main cause of peritoneal dialysis failure	102 (54.5%)	58 (51.3%)	29 (52.7%)	6 (42.9%)		
MMT PERITONITIS	Diagnosed peritonitis–induced MMT in patient leading to diagnosed UFF as main cause of peritoneal dialysis failure	40 (21.4%)	30 (26.5%)	19 (34.5%)	4 (28.6%)		
VASCULAR	Diagnosed vascular events –such as amputations or peripheral vascular disease (PVD)– of patients as main cause of peritoneal dialysis failure	14 (7.49%)	6 (5.31%)	3 (5.45%)	0 (0.00%)		

Protein concentration (pg/mL) in effluents and blood is provided with Rayplex QuantiBody, together with the ratio dialysate/plasma. Statistical analysis was performed in imputations as in Methods. No=0, and Yes=1. p values**** <0.0001, p values*** < 0.001, p values** < 0.01, p values* < 0.05.
CVD, cardiovascular disease; ICC, ischaemic cardiac congestion; MMT, mesothelial-to-mesenchymal transition; PD, peritoneal dialysis.

of no cross-reaction and exclusion of homologues and orthologues (online supplemental figure 1). The assay was performed following the protocol given in the manufacturer's instructions with the longest recommended incubation times and a dilution factor of 1:2 for serum and 1:1 for effluents. Additionally, we compared protein detection ranges of effluents and sera of former studies and other protein kits.[4]

Data preprocessing

We digitalised and used <70 000 clinical observations, with available effluents and 166 variables, to train models. We removed variables with >45% of missingness and implemented one-hot encoding to binarise categorical variables to proceed with imputation (online supplemental methods). Further, we added biomarker values of the sera and the dialysate-serum ratio or dialysate-plasm ratio (D/P) and performed a second miceRanger run without compromising important variables. We concluded independently with 16 subdatasets for each the train (Hospital Universitario La Paz, Madrid, Spain) and for the validation datasets (Hospital Universitario La Princesa, Madrid, Spain).

Statistical analysis

In the present study, we used RStudio 2022.02.2+485 'Prairie Trillium' Release (2022-04-19) and Python 3.9 for Windows. The modules used for each software are showcased in online supplemental tables 4,5.

For the canonical statistical analysis, retrieve online supplemental methods.

Clinical feature selection

Linear correlation

We calculated the Pearson correlation of all variables with RStudio to demonstrate linear dependencies among the collected variables and PD outcomes.

Mutual information

Correlation measures often miss dependency between features when dealing with non-linear relationships.[24] Normalised mutual information feature selection (NMIFS) is powered to demonstrate relationships between features and response targets, as it is more sensitive to linear and non-linear relationships.[25] To accomplish the explanation of most of the variability of the outcomes to predict (PD cause end, exit, failure, endurance), the NMIFS approach was conducted among the 162 remaining preprocessed variables.

Mutual information (MI) or Shannon's entropy is defined as the decrease in uncertainty (or informational gain).[26] MI quantifies the amount of information shared by two random variables and is a non-linear dimensionality reduction to spare computational cost (online supplemental methods).[27] The normalised MI (NMI) takes real values within the range [0, 1] like the correlation coefficient. Each target to predict required the following number of variables: 30 (endurance), 38 (PD cause end), 36 (exit), 34 (failure).

Advanced machine learning (ML) analysis for model selection

A representative dataset is important to ensure generalisability. However, AI algorithms do not necessarily require data that mirrors specific distributions of other countries or studies. Instead, AI models need diverse and high-quality data to learn effectively (retrieve online supplemental methods).

Ranking of machine learning (ML) regressors for numerical outcomes

To allocate models without overfitting and memory learning (see online supplemental methods), regressors were ranked by similarity of distribution to the original values, performance of mean absolute error (MAE), mean squared error (MSE) and root MSE (RMSE). The outcome variable was deleted as input for model training and validation to avoid data leakage.

Ranking of machine learning classifiers for categorical outcomes

In statistical inference, the limitations of the information-theoretic performance are commonly expressed in reference to statistical divergence between the underlying statistical models. We avoided indomitable data dimension growth, by which computation of the decision-making statistics and attendant performance limits (divergence metrics) underpin complexity and instability. So, we used the following metrics of robustness to rank classifiers: Kullback–Leibler (KL) divergence, precision, recall, F1 score, MI and similarity of classifications to the original values. Additionally, we deleted outcome variables as model inputs in the training and validation phase to avoid data leakage.[28]

Sequentially, we selected models with the best recall, precision and receiver operating characteristic curve under the area (ROC-AUC). We computed for the majority of the metrics the macro-, micro- and weighted averages since all subclasses were imbalanced. Nonetheless, medical publications—even with class imbalances—normally consider direct metrics or the macro-average.[29] Further, we scored classifiers by precision, recall and the confusion matrix (see online supplemental methods).

Training and validation of machine learning algorithms

Model development was split into four common steps and classified on the abovementioned metrics: a transformation and centering of the datasets; a prior training with the 'La Paz' dataset and a selection of the best raw algorithms by metrics of robustness and a final optimisation of (hyper-)parameters; and a last verification of correct predictions with 'unseen' datasets by using the external positive and negative validation datasets (see online supplemental methods).

SHapley Additive exPlanations (SHAP)

The goal with SHapley Additive exPlanations (SHAP) is to explain the individual feature attribution of a model by computing the contribution of each to the prediction when the feature is present or absent. The methodology is based on SHapley values from coalitional game

theory.[30] SHAP values determine how the members of a group should receive individual payoffs according to their marginal contributions (see online supplemental methods).

MAUXI: building the automatised calculator with embedded machine learning algorithms

Creation of a local server and a web-host was made using the Bulma interface template. To that extent, we created a Python server, in which a dashboard ('get') was included. Further, we included four different 'post' scripts to receive the patient data, inputted by any medical professional. The post inbox was connected indirectly with the pretrained algorithm for each target (see online supplemental methods). The output referred to the numerical and categorical variables.

RESULTS

Baseline characteristics of clinical cohorts

Patient characteristics, biomarkers and clinical parameters are described longitudinally (n=369) in table 1 for a total of 151 independent patients, stratified by gender in online supplemental figure 1.

Patients were further stratified into groups with similar features (CCuts) based on the Jenks natural breaks classification method, to determine descriptive statistics and logical assumptions. Exploration of the CCuts (table 1) revealed that C3 patients with increased endurance (median 104 months) displayed a decreased time undergoing PD (median 5.10 months), whereas C0 patients enduring shorter already have spent a median of 24.9 months in PD. Long PD endurers were younger and had undergone less time in PD in other external centres (PRIORPD). The female population accounted for 33–64%, being overall 6–9% and <1% Latin-American and Black population, respectively. The diabetic status was reported to be as high as 46% for the mid-endurers C2, appearing in higher frequency the type 2. Hypertension, dyslipidaemia, smoking habits, any type of cancer and co-infection were annotated.

Patients being permanently transferred to HD were >43%, whereas patients undergoing fatal events were >40%. Drop-off (UFF or PM failure) was reported at a rate of >71%. CVD events leading to death entailed 4–21%. High PD endurers displayed higher levels of biomarkers in effluents, except calcium and glucose in blood (online supplemental table 1). The interslide variation in the biomaker array was [−2.20, 4.65] %, indicating assay replicability.

Feature selection

Missingness of datasets

The missingness of the total train dataset (17.7%) and each variable (0–43%) are displayed in online supplemental figures 2, 3, as well as the missingness of the positive and the negative validation cohorts (online supplemental figure 4) with >50% missingness.

Linear correlation

We observed linearly high MTCs in UFF patients (online supplemental figure 5). Clinical variables related to renal function, PM functionality and ultrafiltration efficiency correlated to each other positively: residual renal function (RRF), UFR4H, generation of urea (GENERUREA) or creatinine (GENER) in effluent and urine, DWELL, ratio dialysate plasma of creatinine (D/P CREATININE) and dialysate ratio of glucose in the dwell time 240 min and 0 min (D5/D0 glucose ratio).

Table 2 Final models to predict endurance were trained, optimised with hyperparameters and ranked on RMSE and MAE.

Algorithm	Category	MAE	MSE	RMSE	MI
DT	Train	5.21	92.01	8.89	0.53
ET	Train	6.55	132.46	11.51	0.52
kNN	Train	0.00	0.00	0.00	1.00
DT	Test	25.11	1234.44	35.40	0.09
ET	Test	20.27	891.46	29.86	0.08
kNN	Test	25.64	1352.55	36.78	0.08
DT	Validation	19.99	796.11	28.31	0.16
ET	Validation	16.99	633.30	25.17	0.21
kNN	Validation	21.78	932.25	30.53	0.13
DT	Negative validation	31.87	2377.57	52.06	0.18
ET	Negative validation	32.14	2389.35	48.88	0.25
kNN	Negative validation	38.04	3143.27	56.06	0.21

The trained algorithms were submitted to additional validation with the train, test and validation datasets. Merely the three best algorithms, with robust metrics and distributions resembling the original, are displayed.
DT, decision tree; ET, Extra Tree; kNN, k-nearest neighbours; MI, mutual information; MSE, mean squared error; RMSE, root MSE.

Table 3 Final models to predict PD cause end, PD exit and PD failure were trained, optimised with hyperparameters and ranked on classifier metrics

PD outcome (to predict)	Algorithm	Tested dataset	F1 macro	F1 micro	F1 weighted	MI	Precision macro	Precision micro	Precision weighted	Recall macro	Recall micro	Recall weighted
PD cause end	**DT**	Train	0.96	0.97	0.97	0.78	0.96	0.97	0.97	0.96	0.97	0.97
	LDA	Train	0.66	0.79	0.77	0.11	0.72	0.79	0.77	0.64	0.79	0.79
	RF	Train	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	DT	Test	0.53	0.66	0.66	0.00	0.53	0.66	0.66	0.53	0.66	0.66
	LDA	Test	0.54	0.72	0.69	0.01	0.57	0.72	0.68	0.54	0.72	0.72
	RF	Test	0.52	0.66	0.66	0.00	0.53	0.66	0.65	0.52	0.66	0.66
	DT	Validation	0.49	0.63	0.64	0.00	0.49	0.63	0.65	0.49	0.63	0.63
	LDA	Validation	0.51	0.66	0.66	0.00	0.51	0.66	0.66	0.51	0.66	0.66
	RF	Validation	0.50	0.62	0.64	0.00	0.51	0.62	0.66	0.51	0.62	0.62
	DT	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LDA	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RF	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PD exit	**DT**	Train	0.98	0.98	0.98	0.88	0.98	0.98	0.98	0.98	0.98	0.98
	LDA	Train	0.68	0.69	0.69	0.10	0.69	0.69	0.69	0.68	0.69	0.69
	RF	Train	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	DT	Test	0.52	0.53	0.53	0.00	0.52	0.53	0.53	0.52	0.53	0.53
	LDA	Test	0.51	0.52	0.52	0.00	0.51	0.52	0.52	0.51	0.52	0.52
	RF	Test	0.52	0.53	0.53	0.00	0.52	0.53	0.53	0.52	0.53	0.53
	DT	Validation	0.35	0.36	0.32	0.00	0.51	0.36	0.60	0.51	0.36	0.36
	LDA	Validation	0.50	0.53	0.55	0.00	0.51	0.53	0.60	0.51	0.53	0.53
	RF	Validation	0.40	0.40	0.39	0.00	0.51	0.40	0.59	0.51	0.40	0.40
	DT	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LDA	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RF	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Continued

Table 3 Continued

PD outcome (to predict)	Algorithm	Tested dataset	F1 macro	F1 micro	F1 weighted	MI	Precision macro	Precision micro	Precision weighted	Recall macro	Recall micro	Recall weighted
PD failure	DT	Train	0.79	0.85	0.85	0.55	0.83	0.85	0.85	0.76	0.85	0.85
	LDA	Train	0.48	0.65	0.63	0.20	0.55	0.65	0.64	0.45	0.65	0.65
	RF	Train	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00
	DT	Validation	0.12	0.36	0.42	0.00	0.15	0.36	0.52	0.10	0.36	0.36
	LDA	Validation	0.15	0.52	0.52	0.00	0.15	0.52	0.53	0.14	0.52	0.52
	RF	Validation	0.11	0.33	0.40	0.00	0.14	0.33	0.52	0.09	0.33	0.33
	DT	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LDA	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RF	Negative validation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DT	Test	0.24	0.41	0.40	0.02	0.24	0.41	0.40	0.24	0.41	0.41

The trained algorithms were submitted to additional validation with the train, test and validation datasets. Analysis guidelines were reported in Methods.
DT, decision tree ; LDA, linear discriminant analysis; PD, peritoneal dialysis; RF, Random Forest.

Mutual information

Overall, 30–38 variables were found to be sufficient to train and test robustly ML algorithms, by selecting the top first 20 of 162 variables, in addition to the biomarkers (online supplemental figures 6, 7).

Advanced ML analysis for model selection

Training and validation of machine learning algorithms

Conceptually, we designed a brute force regressor pipeline (online supplemental table 6 and online supplemental figure 8) of 28 algorithms, which was used to predict endurance. We discovered, based on the MAE, that Extra Tree (ET), random forest (RF) and k-nearest neighbours regressors were the most robust, including distribution similarities and overfitting (online supplemental figure 8). Several models were skewed to the mean of the distribution, acquiring great metrics but poor distributions (Bernoulli Naïve Bayes, etc).

Further, the pipeline for categorical targets to predict (PD cause end, exit, failure) comprised 16 algorithms (online supplemental tables 6–9 and online supplemental figures 9, 10). The most optimal models were linear discriminant analysis (LDA), decision tree (DT) and RF.

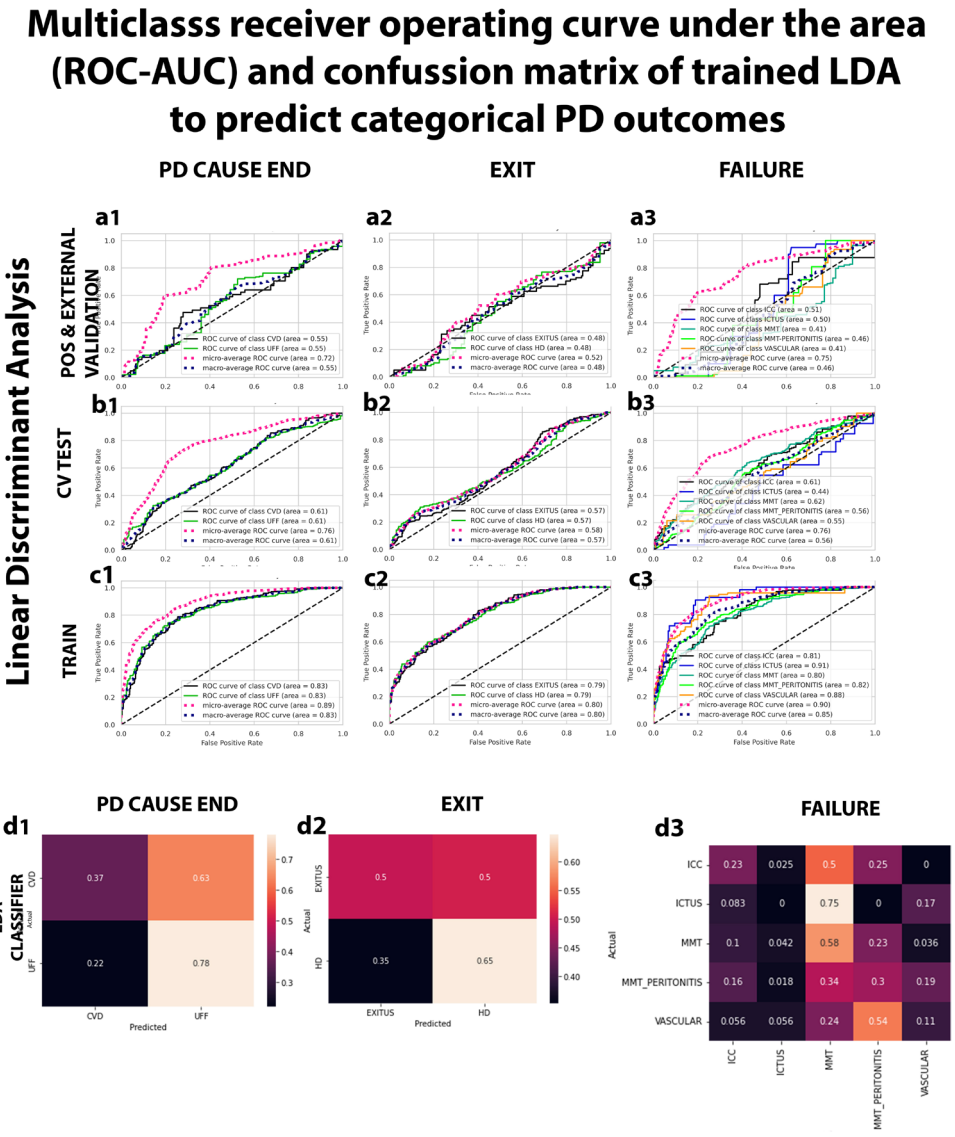

Figure 1 ROC-AUC of LDA (a–c) and confusion matrix of LDA (d1–d3) for categorical targets: PD cause end, exit and failure. Each class to be predicted is highlighted in the rainbow colours and indicated as appropriate in the legend. (a1), (b1) and (c1) display values for the prediction of PD cause end, by using the three datasets: train, cross-validation (CV) test, and positive (POS) external validation. (a2), (b2) and (c2) display values for the prediction of exit, by using the three datasets, whereas (a3), (b3) and (c3) reflect values on predicting failure. Overfitting is implied in a perfect fit of 1:1 true-positive and false-positive cases by using the train dataset by which the algorithm was trained. CVD, cardiovascular disease; ICC, ischaemic cardiac congestion; MMT, mesothelial-to-mesenchymal transition; LDA, linear discriminant analysis; PD, peritoneal dialysis; UFF, ultrafiltration failure; HD, hemodialysis. Additional abbreviations are found table 1 and online supplemental file 1.

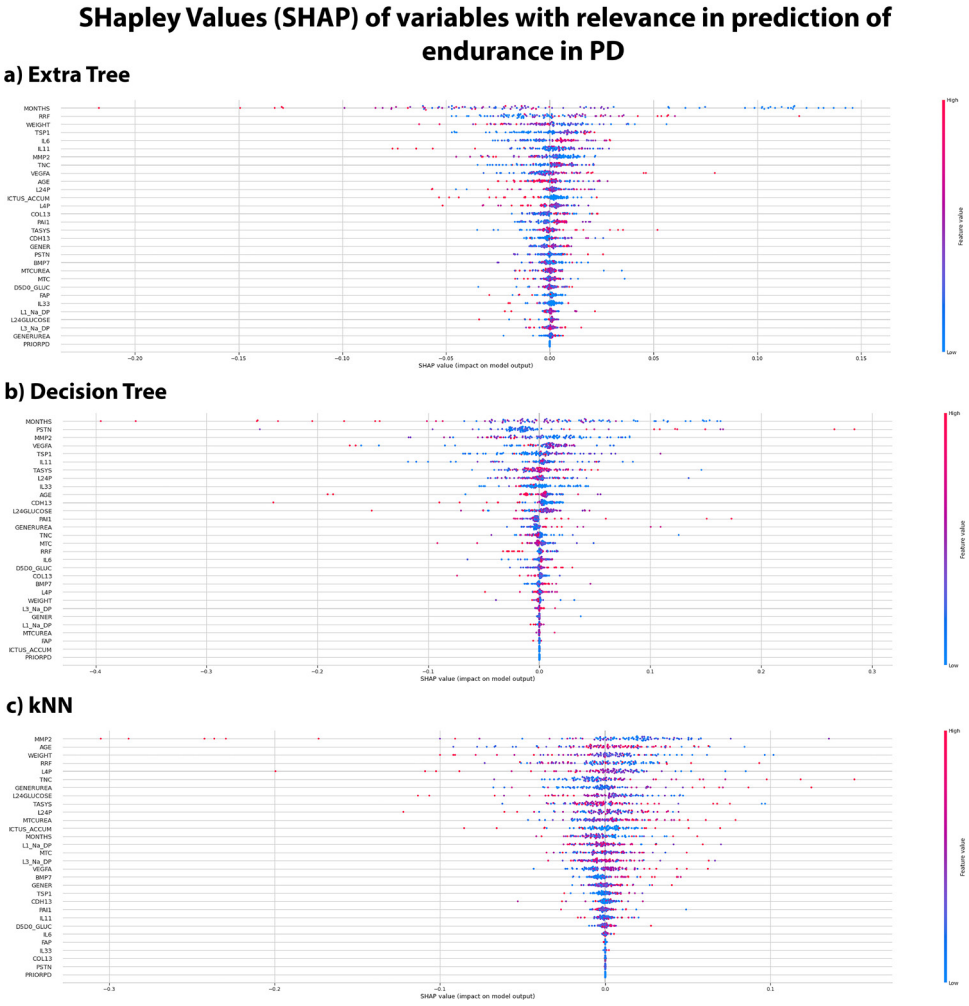

Figure 2 SHapley values of clinical attributes with importance in the algorithms for prediction of PD endurance. Data is shown for the best performing algorithms in the prediction of endurance. Importantly, the prediction accuracy in discriminating groups of patients among the estimated remaining time in PD (endurance) is determined by the shown variables. Positive feature values (pink) can possess a negative (left, negative SHAP value) or a positive (right, positive value) SHAP value. Thus, endurance is decreased (left, negative SHAP value) by incrementing the measure (pink) or by diminishing (blue) the measure and vice versa. CDH13, cadherin-13, COL13, collagen-13; GENER, generation of creatinine; GENERUREA, generation of urea; IL, interleukin; kNN, k-nearest neighbours; MMP2, matrix metalloproteinase-2; PD, peritoneal dialysis; PSTN, periostin; RRF, residual renal function; TSP1, thrombospondin-1; VEGFA, vascular endothelial growth factor A. Additional abbreviations are found in table 1 and online supplemental file 1.

Optimisation and validation of selected algorithms

Algorithms generally performed properly when comparing all robustness indices on hyperparameter optimisation (tables 2 and 3). ET regressor seemed more suitable for prediction of PD endurance (table 3), since validating with the CV test, positive external and negative validation datasets entailed the lowest MAE and highest MI.

For binary and multiclass variables, the LDA classifier (figure 1) resembled the original distribution without overfitting and displayed the highest metrics, excluding the KL divergence, for all the categorical outcomes (online supplemental tables 7–9). Generally, it could be said that predictions were robustly drawn for PD cause end with metrics >0.72 (figure 1, online supplemental figures 11, 12). The other predictions were scoring >0.72 in the micro-averaged ROC-AUC (for almost all the datasets). Infinitesimal-branching classifiers (DT, RF) displayed again perfect metrics biased to the most frequent class (online supplemental tables 7–12). Negative validation showed sensitivity. Consistently, we obtained robust classifications of MMT, peritonitis-induced MMT, ICC, HD and CVD with up to 78% of correct classifications.

Overfitted ML algorithms worsened metrics of robustness (online supplemental figures 11, 12). This performance analytics further proved that LDA (figure 1, online supplemental figures 13–16) was the most robust model for PD cause end, PD exit and PD failure. So, ET and LDA models were embedded in the MAUXI software (online supplemental methods). Metrics worsened by predicting patients with other causes of leave.

All algorithms were trained with the proposed dynamic ranges in protein detection of the selected kit. Other kits detecting different levels of protein concentrations, as

shown in online supplemental table 13, may interfere with correct predictions. Thus, harmonised measurements are imperative for proper AI and ML use and performance.

SHapley values for model selection

We chose SHAP values as our major provider for causality insights about predicting PD endurance (figure 2).

Variables impacting negatively on the remaining time until failure were weight, IL-11, MMP-2, age, accumulated ictus and MTCs. Opposingly, increasing values of RRF, TSP-1, IL-6, TN-C, VEGF-A, COL-13, systolic arterial tension (TASYS), PAI-1 and FAP provided patients with longer survival in PD.

DISCUSSION

For the first time, we show that MMT-associated biomarkers are relevant for prediction of PD drop-out: MMP-2, TN-C, IL-11, PAI-1, PSTN, VEGF-A, COL-13, CDH-13, TSP-1, BMP-7, IL-6, FAP and IL-33. Interestingly, immunoassays have different detection ranges, and robust predictions are tied to those.[8] All biomarkers had proper NMIFS and SHAP values, indicating that MMT processes may have a causal relationship with the development of CVD-UFF in PD. We confirmed that small longitudinal PD datasets, attribute shrinkage and gold-standard algorithms (ET, LDA) with overfitting testing and class imbalances predict PD endurance and technique failure.[27]

We acknowledge that our study has several constraints. Our train and validation datasets possessed a preponderance of white males with class imbalances of PD technique failure. Most patients with PD dropoff due to CVD were only registered with missing samples and/or reports. We note that this study comprised merely a small cohort, limiting the prediction power.

Nevertheless, we expanded common knowledge that MTC, UFR4H, RRF and electrolyte sieving, among others, are indeed predictors of PD endurance and can even be cardioprotective.[4] External validations verified predictions only for MMT-UFF-CVD patients, preventing vague risk scores with a time resolution of 5–7 years.[6 20] MAUXI predictions, requiring only effluents, avoid patient invasiveness and time burden.[4]

We divided PD outcomes into four categories (cohort selection) based on the observed historical registry of the entire dataset of patients since the first registry in the training dataset (1979). There were no other outcomes in the hospitals with whom we collaborated in the European Union. In our study, only real-world patients were integrated. We included as well patients who left PD for a transplant, failing shortly afterwards and returning to PD until failure due to MMT or CVD. Further, we did not include any outcome related to final transplant, or any failure not related to MMT (catheter, abdominal perforations, surgery, etc.).

ET—as an ensemble method building multiple DTs with random splits—effectively handles repeated measures (time series data) by modelling complex relationships and robustness against noise. By correctly formatting temporal data (natural logarithm, identification of longitudinal samples), ET captured patterns across time points.

LDA was adapted to include temporal features, indirectly accounting for repeated measures, using feature engineering. Despite LDA's limitations compared with ET, it was effective for this dataset due to the prevalence of linear dependencies. The study highlights the importance of aligning models with specific data structures to enhance medical predictions.

MAUXI software intends to provide more predictability of the PD technique failure, avoiding unprecedented neurological, cardiovascular and UFF consequences. Further studies should unravel molecular mechanisms of these MMT biomarkers in PD.

CONCLUSION

In conclusion, the MAUXI medical device can help healthcare professionals to predict robustly PD patient fate, increasing the current knowledge in the field of prophylactic interventions.

Author affiliations
[1]Tissue and Organ Homeostasis Program, Cell-Cell Communication and Inflammation Unit, Centro de Biologia Molecular Severo Ochoa (CBM), CSIC-UAM, Fundacion General CSIC, Madrid, Spain
[2]Servicio de Nefrología, Hospital Universitario de la Princesa & Instituto de Investigación la Princesa (IP), Madrid, Spain
[3]Redes de Investigación Cooperativa Orientadas a Resultados en Salud (RICORS 2040-Renal), Madrid, Spain
[4]Premium Research, S.L, Guadalajara, Spain
[5]Servicio de Nefrología, Hospital Universitario La Paz & Instituto de Investigación Sanitaria la Paz (IdiPAZ), Madrid, Spain
[6]Unidad de Biología Molecular, Hospital Universitario de la Princesa & Instituto de Investigación la Princesa (IP), Madrid, Spain

Acknowledgements We thank Sara Alvarez Lopez de Rodas (Complutense University, Genomic Analysis Unit, Madrid, Spain) for her help with the Quantibody microarrays reading. We are kindly grateful for the support by the nurses of Hospital Universitario La Paz, Madrid (Spain). Finally, we thank Daniel Hernandez Lobato (Computer Science Department, Universidad Autónoma de Madrid, Cantoblanco, Spain) for his advisor tasks in machine learning. We thank Javier Arriero Pais for his invaluable insights in software development.

Collaborators Redes de Investigación Cooperativa Orientadas a Resultados en Salud (RICORS): María Auxiliadora Bajo-Rubio, Pilar Sandoval, Guadalupe Tirma González-Mateo, Gloria del Peso-Gilsanz, Marta Ossorio-González, Manuel López-Cabrera.

Contributors Project guarantor: MLC. project administration: MLC. investigation: MLC, PS. conceptualisation: EMAP, RAG, MLC, PS. data curation: EMAP. formal analysis: EMAP, RAG. funding acquisition: MLC, MABR. resources: MLC, MARB, GPG, MOG, RAG, EMAP. Methodology: EMAP, MLC, RAG, PS. Visualisation: EMAP, RAG. Software: EMAP, RAG. Supervision: MLC, PS, MABR, GGTM, GPG. Validation: MLC, PS, MOG, PLMR, PAV. Writing-original draft: EMAP. Writing-review and editing: MLC, PS, EMAP. We have used a pipeline of machine learning algorithms (artificial intelligence (AI)) to perform predictions of cardiorenal patients in 2 independent hospitals in Spain and determine their endurance and outcome in peritoneal dialysis. For this purpose, we generated own python/rstudio scripts. AI (ChatGPT, or any other) was not used to design, analyse and write the manuscript.

Funding This project has received funding from the grant IMPROVEPD from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 812699. This work was also supported by grants (PID2019-110132RB-I00/AEI/https://doi. org/ 10. 13039/ 50110 00110 33 and PID2022-142796OB-I00/AEI/https://doi. org/ 10. 13039/ 50110 00110 33) from the Spanish Ministry of Science and Innovation/Fondo Europeo de Desarrollo Regional (MICIN/FEDER) to ML-C. Instituto de Salud Carlos III provided the additional grant PI18/00882.

Competing interests None declared.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval The studies involving human participants were reviewed and approved by the Ethics Committees for Investigation with medicinal products (CEIm) of Hospital Universitario La Paz and Hospital Universitario La Princesa. Details of each CEIm are the following: (1) CEIm IdiPaz (ceic.hulp@salud.madrid.org), code 12028768952665218393 18.2) CEIm Hospital Universitario La Princesa (ceic.hlpr@salud.madrid.org). The patients provided their written informed consent to participate in this study. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. Data, as the excel files with the preprocessed data (clinical variables and outcomes) can be requested. Otherwise, all given data in the paper is the relevant one.

ORCID iD

Eva María Arriero-País http://orcid.org/0000-0001-7719-7022

REFERENCES

1 Khanna R. Solute and Water Transport in Peritoneal Dialysis: A Case-Based Primer. *Am J Kidney Dis* 2017;69:461–72.
2 Aguirre AR, Abensur H. Protective measures against ultrafiltration failure in peritoneal dialysis patients. *Clinics (Sao Paulo)* 2011;66:2151–7.
3 López-Cabrera M. Mesenchymal Conversion of Mesothelial Cells Is a Key Event in the Pathophysiology of the Peritoneum during Peritoneal Dialysis. *Adv Med* 2014;2014:473134.
4 Ruiz-Carpio V, Sandoval P, Aguilera A, *et al*. Genomic reprograming analysis of the Mesothelial to Mesenchymal Transition identifies biomarkers in peritoneal dialysis patients. *Sci Rep* 2017;7:44941.
5 Yáñez-Mó M, Lara-Pezzi E, Selgas R, *et al*. Peritoneal dialysis and epithelial-to-mesenchymal transition of mesothelial cells. *N Engl J Med* 2003;348:403–13.
6 Noh J, Yoo KD, Bae W, *et al*. Prediction of the Mortality Risk in Peritoneal Dialysis Patients using Machine Learning Models: A Nation-wide Prospective Cohort in Korea. *Sci Rep* 2020;10:7470.
7 Lopes Barreto D, Krediet RT. Current status and practical use of effluent biomarkers in peritoneal dialysis patients. *Am J Kidney Dis* 2013;62:823–33.
8 Szeto C-C, Chow K-M, Kwan BC-H, *et al*. The relationship between bone morphogenic protein-7 and peritoneal transport characteristics. *Nephrol Dial Transplant* 2008;23:2989–94.
9 Schrempf M, Kramer D, Jauk S, *et al*. Machine Learning Based Risk Prediction for Major Adverse Cardiovascular Events. *Stud Health Technol Inform* 2021;279:136–43.
10 Hao N, Chiou TT-Y, Wu C-H, *et al*. Longitudinal Changes of PAI-1, MMP-2, and VEGF in Peritoneal Effluents and Their Associations with Peritoneal Small-Solute Transfer Rate in New Peritoneal Dialysis Patients. *Biomed Res Int* 2019;2019:2152584.
11 Peng W, Zhou X, Xu T, *et al*. BMP-7 ameliorates partial epithelial-mesenchymal transition by restoring SnoN protein level via Smad1/5 pathway in diabetic kidney disease. *Cell Death Dis* 2022;13:1–12.
12 Pecoits-Filho R, Araújo MRT, Lindholm B, *et al*. Plasma and dialysate IL-6 and VEGF concentrations are associated with high peritoneal solute transport rate. *Nephrol Dial Transplant* 2002;17:1480–6.
13 Xiao J, Gong Y, Chen Y, *et al*. IL-6 promotes epithelial-to-mesenchymal transition of human peritoneal mesothelial cells possibly through the JAK2/STAT3 signaling pathway. *Am J Physiol Renal Physiol* 2017;313:F310–8.
14 Zhou Y, Ng DYE, Richards AM, *et al*. microRNA-221 Inhibits Latent TGF-β1 Activation through Targeting Thrombospondin-1 to Attenuate Kidney Failure-Induced Cardiac Fibrosis. *Mol Ther Nucleic Acids* 2020;22:803–14.
15 Song C, Burgess S, Eicher JD, *et al*. Causal Effect of Plasminogen Activator Inhibitor Type 1 on Coronary Heart Disease. *J Am Heart Assoc* 2017;6:e004918.
16 Gungor O, Unal HU, Guclu A, *et al*. IL-33 and ST2 levels in chronic kidney disease: Associations with inflammation, vascular abnormalities, cardiovascular events, and survival. *PLoS One* 2017;12:e0178939.
17 Nagaraju CK, Dries E, Popovic N, *et al*. Global fibroblast activation throughout the left ventricle but localized fibrosis after myocardial infarction. *Sci Rep* 2017;7:10801.
18 Dixon IMC, Landry NM, Rattan SG. Periostin Reexpression in Heart Disease Contributes to Cardiac Interstitial Remodeling by Supporting the Cardiac Myofibroblast Phenotype. *Adv Exp Med Biol* 2019;1132:35–41.
19 Philippova M, Suter Y, Toggweiler S, *et al*. T-cadherin is present on endothelial microparticles and is elevated in plasma in early atherosclerosis. *Eur Heart J* 2011;32:760–71.
20 Cho Y, Johnson DW, Vesey DA, *et al*. Baseline serum interleukin-6 predicts cardiovascular events in incident peritoneal dialysis patients. *Perit Dial Int* 2015;35:35–42.
21 Merino D, Villar AV, García R, *et al*. BMP-7 attenuates left ventricular remodelling under pressure overload and facilitates reverse remodelling and functional recovery. *Cardiovasc Res* 2016;110:331–45.
22 Imanaka-Yoshida K, Tawara I, Yoshida T. Tenascin-C in cardiac disease: a sophisticated controller of inflammation, repair, and fibrosis. *Am J Physiol Cell Physiol* 2020;319:C781–96.
23 Landry NM, Cohen S, Dixon IMC. Periostin in cardiovascular disease and development: a tale of two distinct roles. *Basic Res Cardiol* 2018;113:1.
24 Lopez de Prado M. Statistical Association (Presentation Slides). *SSRN Electronic Journal* 2020.
25 Hlaváčková-Schindler K, Paluš M, Vejmelka M, *et al*. Causality detection based on information-theoretic approaches in time series analysis. *Phys Rep* 2007;441:1–46.
26 Estévez PA, Tesmer M, Perez CA, *et al*. Normalized mutual information feature selection. *IEEE Trans Neural Netw* 2009;20:189–201.
27 Vollmer S, Mateen BA, Bohner G, *et al*. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* 2020;368:l6927.
28 Ji S, Zhang Z, Ying S, *et al*. Kullback-Leibler Divergence Metric Learning. *IEEE Trans Cybern* 2022;52:2047–58.
29 Kader A, Sharif S, Bhowmick P, *et al*. Effective Workflow for High-Performance Recognition of Fruits using Machine Learning Approaches. *International Research Journal of Engineering and Technology* 2020.
30 Shapley LS. Notes on the n-person game. US airforce; 1951.