



OPEN Balancing accuracy and cost in machine learning models for detecting medial vascular calcification in chronic kidney disease: a pilot study

Urszula Bialonczyk^{1✉}, Malgorzata Debowska¹, Lu Dai², Abdul Rashid Qureshi³, Leon Bobrowski^{1,4}, Magnus Soderberg⁵, Bengt Lindholm³, Peter Stenvinkel³, Tomasz Lukaszuk⁴ & Jan Poleszczuk¹

Machine learning algorithms that integrate multiple biomarkers are increasingly used in disease detection, yet economic considerations are often overlooked. Medial vascular calcification (mVC), a pathology associated with elevated cardiovascular risk in chronic kidney disease (CKD), requires cost-effective diagnostic approaches. This pilot study evaluated the cost-effectiveness of machine learning models for mVC detection using traditional risk markers and circulating biomarkers in 152 CKD patients undergoing living donor kidney transplantation. Patients were classified as having no/minimal ($n = 93$) or moderate/extensive ($n = 59$) mVC. Five classification frameworks with automatic variable selection identified predictors of mVC. Age and copeptin were selected by all algorithms, while diabetes, male sex, choline, and osteoprotegerin were chosen by four methods. The number of features selected ranged from 5 to 21. Although accuracy differences among classifiers were limited to 3%, models using more features nearly tripled the procedure's cost. By incorporating the incremental cost-effectiveness ratio, the study highlighted significant disparities in performance versus cost among classifiers. The present findings suggest that machine learning has the potential to complement imaging techniques for mVC detection and uncover novel biomarkers. However, modest performance improvements may not justify higher costs, underscoring the importance of considering cost-effectiveness when selecting classification models.

Keywords Classification, Feature selection, Medial vascular calcification, Chronic kidney disease, Incremental cost-effectiveness ratio

Medial vascular calcification (mVC) is a pathological condition, with estimated prevalence rates ranging from 27 to 80% in the chronic kidney disease (CKD) population^{1–3}. The pathology contributes to the high cardiovascular morbidity and mortality in this group of patients^{4,5}. Moreover, a recent study has revealed that it is associated with the progression of CKD⁶. While the pathogenesis of mVC is not fully understood and a causal therapy is not available as of today, new therapeutic possibilities are currently being studied^{7–10}. Moreover, the feasibility of slowing down mVC progression in patients with CKD has been demonstrated^{11,12}. Therefore, improved methods for mVC detection, especially at early stages, is highly warranted.

At present, there is a lack of a dedicated and reliable method of mVC assessment in clinical practice^{2,13,14}. Invasive techniques such as artery biopsy¹⁵ or transcutaneous ultrasound¹⁶ are rarely performed and cannot be considered as screening procedures. Both direct semi-quantitative methods such as computed tomography, plain X-rays, or ultrasound¹³, and indirect methods such as measurement of pulse wave velocity that reflects

¹Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland.

²Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden. ³Renal Medicine and Baxter Novum, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden. ⁴Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland. ⁵Pathology, Clinical Pharmacology and Safety Sciences, AstraZeneca R&D, Gothenburg, Sweden. ✉email: ubialonczyk@ibib.waw.pl

increased arterial stiffness in calcified arteries^{17,18}, are not always available and not easy to perform; therefore, the presence of mVC is likely underestimated. Moreover, currently used tools struggle to differentiate between the two types of vascular calcification: medial and intimal^{19–21}; this is clinically significant, as these types have distinct implications and require different patient care strategies. Recently, a method enabling this differentiation, involving the identification of mVC patterns on PET-CT scans, has been introduced²². Nonetheless, the expense and limited availability of PET-CT scans highlight the need for an approach that can indicate the presence of mVC and readily determine which patients truly require this imaging technique.

Machine learning algorithms, which are designed to detect patterns in data, are thought to have the potential to radically improve our ability to diagnose and treat diseases. The large number of potential mVC markers complicates mVC diagnosis and statistical feature selection procedures may therefore play a crucial role in establishing future diagnostics. In previous studies, numerous biomarkers have been linked with vascular calcification including serum biomarkers²³, vitamin-K dependent proteins²⁴, various phenotypic features^{25,26} and risk factors such as high age, male sex, and diabetes mellitus²⁰. While models for mVC detection have demonstrated promising performance quality²⁵, the variability in their cost-effectiveness across different frameworks remains unexplored.

In a clinical setting, besides evaluating the statistical performance of the newly introduced methods, their overall applicability is a crucial consideration. This covers factors such as the procedure's availability, safety, and the overall expense of the diagnostic procedure. One of the indicators that can characterize the latter is the incremental cost-effectiveness ratio (ICER) which provides insights into the method's cost in relation to the potential benefits for patients²⁷.

The objective of this pilot study was to investigate the cost-effectiveness of various machine learning frameworks for mVC detection in the chronic kidney disease population. For each of the tested models, in addition to conventional classification correctness metrics, ICER was calculated to incorporate both performance and cost considerations into evaluation. The most favorable model in terms of ICER was further investigated to showcase its possible clinical utility. Finally, we discussed possible pathophysiological associations between mVC and the variables selected by the applied algorithms.

Methods

Data investigation, model building process, and performance evaluation were implemented in R version 4.0.5 and Python version 3.7.

Patients and study design

In this retrospective study, a cohort of patients with advanced CKD undergoing living donor kidney transplantation at Karolinska University Hospital was included. The study's eligibility criteria aligned with those established for patients eligible for kidney transplantation. Exclusion criteria were age under 18 years and unwillingness to participate in the study. The clinical procedures and protocol of measurements were described previously²⁸. The patients gave their informed consent for all performed procedures. The study was approved by the regional ethical review board in Stockholm and adhered to the Declaration of Helsinki.

The participants were classified into two groups according to the extent of medial calcification in inferior epigastric artery biopsies assessed by an experienced pathologist. 'Group 0' included patients with *no* and *minimal* signs of VC ($n=93$), whereas patients having *moderate* and *extensive* signs of VC were classified into 'Group 1' ($n=59$). The procedure of histological mVC examination was presented in detail in^{25,28}.

The dataset consisted of 60 features in 152 patients. All 60 features were available in 71% of the patients; in total, 8.3% of the data were missing. The full data flow is described in²⁵. The dataset included demographic and clinical data, circulating biomarkers, body composition and anthropometric measurements, and skin content of advanced glycation end products measured by autofluorescence. The investigated features are presented in Table 1.

Data preprocessing

First, we standardized the predictors proportionally within the range from 0 to 1. Missing values imputation was performed using the k-nearest neighbors algorithm with $k=3$ and Euclidean distance measure between the patients. mVC, as an outcome variable, was not involved in the imputation process. The distributions of the imputed and non-imputed variables did not exhibit statistical differences (Kolmogorov-Smirnov and Chi-square test for continuous and discrete distributions, respectively). Feature selection and patient classification were performed on the complete, standardized set of variables, while the univariable analysis was performed on the raw data.

Data investigation

To choose the appropriate feature selection and classification algorithms, a preliminary data investigation was conducted. Firstly, the Spearman rank correlation coefficient was used to reveal interdependencies between the analyzed features. Categorical variables (sex, smoking, and diabetes mellitus) were excluded from the analysis. Secondly, logistic regression was carried out to assess the interrelationship between a single feature and mVC. To account for multiple comparisons, p-values were adjusted using Benjamini-Hochberg correction²⁹.

Methods of feature selection and patient classification

In the process of feature selection and patient classification, the following methods were applied: logistic regression with forward Akaike feature elimination process (LR)³⁰, support vector machine with recursive feature elimination (SVM)³¹, random forest with permutation importance (RF)³², logistic regression with elastic net penalty (EN)^{33,34}, and, less explored, relaxed linear separability method (RLS)³⁵.

Feature	Median [IQR] or number (%)	Univariable logistic regression OR (95% CI), <i>p</i> -value ^a	<i>P</i> -value adjusted ^b
Clinical and demographic			
Age, years	46 [32–56]	4.14 (2.08–8.48), < 0.001	< 0.001
Diabetes mellitus	14 (9%)	NA ^c	
Mean arterial blood pressure (meanBP), mmHg	103 [93–111]	1.10 (0.57–2.12), 0.77	0.94
Sex, male	100 (66%)	3.44 (1.63–7.73), 0.002	0.03
Smoker	63 (41%)	1.59 (0.80–3.18), 0.19	0.54
Biomarkers			
Albumin, g/L	35 [33–38]	0.79 (0.41–1.52), 0.48	0.78
Alkaline phosphatase (ALP), U/L	66.2 [51.7–85.1]	1.48 (0.77–2.86), 0.25	0.57
Angiopietin 2, pg/mL	4368 [3353–5987]	2.17 (1.08–4.44), 0.03	0.17
Apolipoprotein A1 (ApoA1), mmol/L	1.38 [1.19–1.56]	0.68 (0.35–1.31), 0.25	0.57
Apolipoprotein B1 (ApoB1), mmol/L	0.87 [0.72–1.04]	0.91 (0.47–1.76), 0.79	0.94
Betaine, μ mol/L	43.0 [29.0–63.0]	1.36 (0.68–2.72), 0.39	0.70
Bone alkaline phosphatase (BALP), μ g/L	16.40 [11.3–25.5]	1.06 (0.55–2.03), 0.87	0.99
Calciprotein particles (CPPs), nm	57,170 [23221–134954]	1.00 (0.46–2.16), 1.00	1.00
Calcium, mmol/L	2.28 [2.18–2.41]	0.86 (0.45–1.66), 0.66	0.90
Carboxylated osteocalcin (GlaOC), ng/mL	35.9 [19.8–67.3]	0.69 (0.35–1.33), 0.27	0.57
Carboxy-terminal collagen crosslinks (CTX), pg/mL	1.88 [1.00–3.38]	0.79 (0.39–1.60), 0.52	0.78
Cholesterol, mmol/L	4.4 [3.8–5.1]	0.69 (0.36–1.33), 0.27	0.57
Choline, μ mol/L	94 [77–115]	2.03 (1.01–4.13), 0.047	0.22
Copeptin, pg/mL	386 [312–490]	2.96 (1.47–6.12), 0.003	0.04
Desphospho-uncarboxylated MGP (duMGP), pmol/L	1323 [948–1687]	2.20 (1.07–4.62), 0.03	0.17
Fibroblast growth factor 19 (FGF19), pg/mL	126 [67–194]	0.56 (0.28–1.09), 0.09	0.32
Fibroblast growth factor 23 (FGF23), pg/mL	4055 [1634–14998]	1.18 (0.62–2.28), 0.62	0.88
Folate, ng/mL	11.0 [8.8–16.0]	0.59 (0.29–1.20), 0.15	0.48
Free thyroxine (fT4), pg/mL	14.4 [13.3–16.7]	1.13 (0.59–2.19), 0.71	0.92
Free triiodothyronine (fT3), pg/mL	3.29 [2.57–3.94]	0.72 (0.37–1.39), 0.33	0.66
Hemoglobin, g/L	114 [105–121]	1.04 (0.50–2.14), 0.92	0.99
High-density lipoprotein (HDL), mmol/L	1.35 [1.10–1.60]	0.68 (0.35–1.30), 0.25	0.57
High sensitivity C-reactive protein (hsCRP), mg/L	0.72 [0.31–2.10]	2.07 (1.07–4.07), 0.03	0.17
Homocysteine, μ mol/L	36 [27–47]	1.91 (0.95–3.89), 0.07	0.27
Humanin, pg/mL	385 [358–421]	0.90 (0.43–1.85), 0.77	0.94
8-hydroxydeoxyguanosine (8-OHdG), ng/mL	0.21 [0.15–0.30]	1.36 (0.70–2.66), 0.36	0.68
IgM anti-MDA, U/mL	102 [67–139]	0.67 (0.33–1.37), 0.28	0.57
IgM anti-PC, U/mL	102 [69–130]	0.45 (0.21–0.92), 0.03	0.17
Insulin-like growth factor 1 (IGF1), ng/mL	236 [158–293]	0.35 (0.17–0.72), 0.005	0.05
Interleukin 6 (IL6), pg/mL	1.20 [0.52–2.01]	0.96 (0.48–1.94), 0.91	0.99
Interleukin 8 (IL8), pg/mL	5.41 [3.77–8.73]	1.28 (0.64–2.54), 0.49	0.78
Klotho, pg/mL	399 [296–537]	1.06 (0.55–2.04), 0.87	0.99
Lipoprotein (a), (LPA), mg/L	122 [50–319]	1.29 (0.60–2.76), 0.52	0.78
Mitochondrial open reading frame of the 12 S rRNA-c (MOTSc), ng/mL	178 [140–226]	0.71 (0.34–1.47), 0.35	0.68
Osteoprotegerin, pg/mL	6.26 [5.00–8.14]	4.32 (2.01–9.66), < 0.001	< 0.001
Parathyroid hormone, intact (PTH), ng/L	253 [153–377]	1.91 (0.97–3.81), 0.06	0.25
Pentraxin-related protein (PTX3), ng/mL	3.83 [2.00–6.39]	1.20 (0.62–2.32), 0.60	0.87
Phosphate, mmol/L	1.7 [1.4–2.0]	0.90 (0.46–1.75), 0.76	0.94
Sclerostin, pg/mL	436 [343–651]	2.22 (1.14–4.41), 0.02	0.16
Soluble receptor activator of nuclear factor- κ B ligand (RANKL), pmol/L	0.07 [0.03–0.12]	1.02 (0.47–2.22), 0.96	1.00
Tartrate resistant acid phosphatase 5a (TRAP5a), U/L	4.01 [2.93–5.38]	0.64 (0.30–1.31), 0.22	0.57
Thyroid-stimulating hormone (TSH), mIU/L	0.85 [0.52–1.36]	0.97 (0.50–1.86), 0.92	0.99
Triglycerides (fPTG), mmol/L	1.3 [1.0–1.8]	1.57 (0.82–3.04), 0.18	0.53
Trimethylamine N-oxide (TMAO), μ mol/L	62.0 [42.0–96.0]	1.01 (0.51–2.01), 0.98	1.00
Troponin T, μ g/L	23.0 [14.0–38.5]	2.70 (1.30–5.76), 0.009	0.08
Tumor necrosis factor (TNF), pg/mL	11.0 [8.96–14.12]	1.27 (0.63–2.58), 0.50	0.78
Undercarboxylated osteocalcin (GluOC), ng/mL	17.4 [5.3–56.15]	0.97 (0.50–1.87), 0.92	0.99
Uric acid (UricAcid), μ mol/L	363 [308–442]	0.88 (0.44–1.74), 0.70	0.92
Continued			

Feature	Median [IQR] or number (%)	Univariable logistic regression OR (95% CI), <i>p</i> -value ^a	<i>P</i> -value adjusted ^b
Vitamin D 25 (D25Vitamin), nmol/L	34 [26–48]	1.29 (0.66–2.50), 0.46	0.78
Anthropometric measurements			
Body mass index (BMI), kg/m ²	24.3 [22.3–26.5]	3.33 (1.70–6.73), 0.001	0.02
Fat body mass index (FBMI), kg/m ²	6.29 [4.73–8.16]	2.13 (0.99–4.68), 0.06	0.24
Hand grip strength (HandGripStrength), % of control	97.7 [78.1–108.1]	0.80 (0.41–1.57), 0.52	0.78
Lean body mass index (LBMI), kg/m ²	18.1 [16.4–19.7]	1.57 (0.73–3.39), 0.25	0.57
Other measurements			
Advanced glycation end products (skin autofluorescence), (AGEAF)	3.0 [2.6–3.4]	1.81 (0.86–3.87), 0.12	0.41
Aortic augmentation index (AorticAI), %	20 [11–26]	1.17 (0.59–2.31), 0.66	0.90

Table 1. The association of individual features, median with interquartile (IQR) range or frequency, with medial vascular calcification (VC) as an outcome of univariable logistic regression measured by odds ratio (OR) with a 95% confidence interval (95% CI). ^aThe analysis was performed per median value of each quantitative feature (e.g., odds ratio of 4.14 for age means that the odds ratio of mVC in patients above the age of 46 years was 4.14 times higher than in a group younger than the median age) or per each dichotomized categorical variable (e.g., males were 3.44 times more likely to have mVC than females). ^bAfter adjusting for multiple comparisons, age, male sex, copeptin, IGF1, osteoprotegerin, and BMI remained statistically significant (marked with bold font). ^cNA, not applicable, because all patients with diabetes belong to one group, i.e., mVC.

Each of the methods was applied in its standard configuration, with algorithm-specific hyperparameter optimization conducted where appropriate. For feature selection, we opted for well-established algorithms commonly used within the applied classification frameworks. A brief description of the chosen methods can be found in the supplementary material. LR, EN, RF and SVM models were built using R *caret* package, for training RLS we used our own MATLAB implementation.

Performance evaluation

All methods were validated in the leave-one-out cross-validation (LOOCV) process. In the algorithms where hyperparameter tuning was required, a nested 5-fold cross-validation was incorporated aiming to maximize accuracy as the primary optimization criterion. The metrics used to evaluate the predictions were accuracy, area under the receiver operating characteristic curve (AUC), precision, recall, and F-score, which are discussed in the supplementary material. Additionally, confidence intervals for the LOOCV AUC values were estimated using the bootstrap method with 1,000 resamples.

Incremental cost-effectiveness ratio

The incremental cost-effectiveness ratio (ICER)^{27,36} represents the additional cost incurred for achieving an additional unit of health outcome, usually measured in quality-adjusted life years (QALYs). It allows decision-makers to ensure that limited healthcare resources are directed towards treatments that provide the most substantial health benefits relative to their associated costs. Thus, the evaluation of ICER facilitates informed decisions about the adoption and funding of medical interventions. In our study, ICER was calculated as:

$$ICER = \frac{measure_cost + (prevalence * TPR + (1 - prevalence) * FPR) * ct_price}{prevalence * TPR * years_gained}$$

Where:

- *measure_cost* – expense associated with evaluating the biomarkers. For certain biomarkers, their costs are considered hyperparameters (parameters with unknown true value) since they are not routinely measured - see Supplementary Table S2 for a list. For the biomarkers with unknown costs, where only the kit price is available, we introduce an additional factor called the *unavailability weight* which used to scale the kit prices accordingly.
- *prevalence* – a hyperparameter indicating true prevalence of mVC among the advanced CKD population.
- *TPR* – the rate of correctly identified true positive cases by the evaluated method.
- *FPR* – the rate at which the evaluated method incorrectly identifies cases as positive when they are actually negative.
- *ct_price* – the price of a PET-CT scan to confirm mVC presence; sourced from a polish laboratory in June 2023 and converted from PLN to USD at a rate of 0.23, was assumed to be 1127 USD.
- *years_gained* – quality of life years gained due to mVC detection. A hyperparameter.

The pricing details for the biomarkers, sourced from Polish laboratories in June 2023 are presented in Table 2. The prices were converted from PLN to USD for clarity using an exchange rate of 0.23. Biomarkers denoted with an asterisk (*) represent hyperparameters. In addition, we performed a sensitivity analysis to assess how the assumed prices influence the results; see supplementary material. We decided to incorporate the cost of a PET-

Feature	LR	SVM	RF	EN	RLS	Cost (USD)	Sum
Age	X	X	X	X	X	0	5
Copeptin	X	X	X	X	X	58	5
Diabetes mellitus	X	X		X	X	2	4
Choline		X	X	X	X	9	4
Osteoprotegerin		X	X	X	X	5*	4
Sex, male	X	X		X	X	0**	4
BMI		X	X	X		0**	3
FBMI	X	X			X	0**	3
Sclerostin		X	X	X		5*	3
CTX		X			X	29	2
duMGP		X		X		5*	2
Homocysteine		X		X		23	2
IgMantiPC		X		X		7*	2
AGEAF		X				0**	1
Angiopoietin 2		X				5*	1
ApoB1					X	21	1
fT3					X	8	1
fPTG					X	5	1
GlaOC					X	23	1
GluOC		X				5*	1
hsCRP		X				7	1
IGF1		X				21	1
IgMantiMDA		X				23	1
LBMI		X				0**	1
PTX3					X	5*	1
TMAO					X	44	1
TroponinT		X				14	1
TSH					X	9	1
Uric Acid					X	9	1
Total number of features	5	21	6	11	16		

Table 2. Features selected using (1) logistic regression (LR), (2) support vector machine (SVM), (3) random forest (RF), (4) elastic net (EN), and (5) relaxed linear separability method (RLS). For full feature names, see Table 1. *Not measured in clinical practice. The cost is calculated based on the price of the kit per measurement. **Cost disregarded because of the relatively low machine expenses; the price per measurement is negligible when assuming testing of numerous individuals.

CT scan in the equation as we presume that, irrespective of how well the classifiers perform, cases with a certain likelihood of being positive would be additionally verified using a more direct method.

Results

Data investigation

Spearman correlation analysis revealed the presence of collinearity among certain feature pairs. Associations are presented as a heat map in Fig. S1.

Using a univariable logistic regression model, we identified age, male sex, angiopoietin 2, choline, copeptin, duMGP, hsCRP, IgM anti-PC, insulin-like growth factor 1, osteoprotegerin, sclerostin, troponin T, and body mass index as factors associated with mVC (Table 1). However, after adjusting for multiple comparisons, only age, male sex, copeptin, IGF1, osteoprotegerin, and BMI remained statistically significant (Table 1).

Classification frameworks

In a multivariable analysis, we applied five classification frameworks with appropriate variable selection methods. To fine-tune SVM, RF, and EN, we conducted hyperparameter optimization. Table S1 in the supplementary material presents the calculated optimal values and short parameter descriptions.

The algorithms applied to the data differed regarding features identified as being potentially associated with mVC (Table 2). Only age and copeptin were chosen by all five methods (Table 2). The number of selected features varied between the methods with 21 features being selected by SVM, 16 by RLS, 11 by EN, 6 by RF, and 5 features chosen by LR.

The classification ability of the applied methods was measured, among others, by the area under the receiver operating characteristic curve (AUC). In the cross-validation evaluation process, the highest AUC was achieved

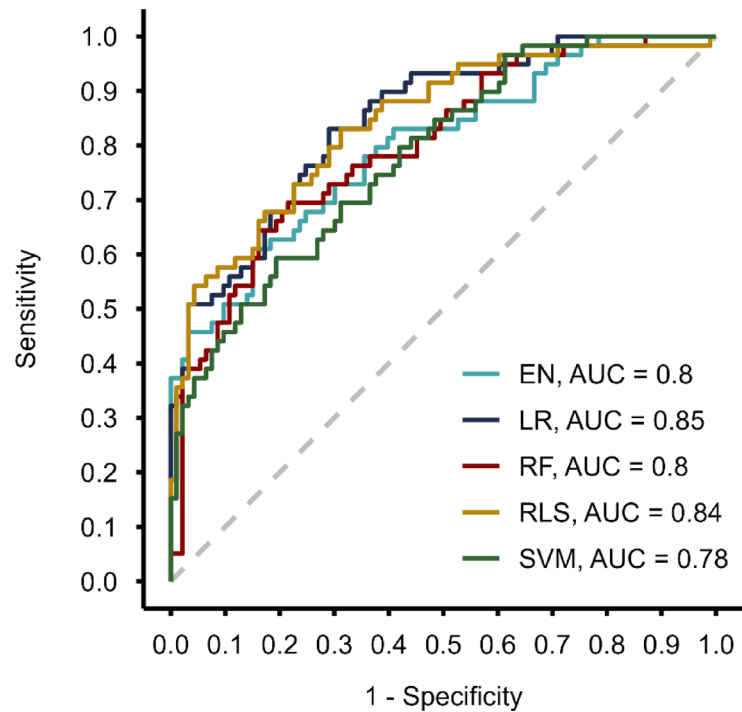


Fig. 1. Receiver operating characteristic (ROC) curves with area under the curve (AUC) for elastic net (EN), logistic regression (LR), random forest (RF), relaxed linear separability method (RLS) and support vector machine (SVM).

Method	Accuracy	AUC	Precision	Recall	F-score
LR	0.74	0.85	0.71	0.58	0.64
SVM	0.71	0.78	0.65	0.54	0.59
RF	0.74	0.80	0.74	0.49	0.59
EN	0.76	0.80	0.87	0.46	0.60
RLS	0.77	0.84	0.71	0.68	0.69

Table 3. Performance evaluation metrics, including area under the curve (AUC), calculated for the applied classification frameworks using leave-one-out cross-validation: (1) logistic regression (LR), (2) support vector machine (SVM), (3) random forest (RF), (4) elastic net (EN) and (5) relaxed linear separability method (RLS).

by LR (0.85 [0.78–0.90]), followed by RLS (0.84 [0.77–0.90]), EN (0.80 [0.72–0.87]), RF (0.80 [0.73–0.86]), and SVM (0.78 [0.70–0.85]) (Fig. 1). The values in square brackets represent bootstrapped 95% confidence intervals. All computed performance evaluation metrics are summarized in Table 3. None of the applied methods outperformed the others across all the assessed measures.

Incremental cost-effectiveness ratio

Figure 2 illustrates the Incremental Cost-Effectiveness ratio for the built models across three unknown parameters: unavailability weights (1, 10, 20, 30), reflecting the possible increase in procedure costs caused by the biomarkers with the unknown prices; true mVC prevalence in CKD population (0.4, 0.6, 0.8); and Quality Adjusted Life Years gained. In general, ICER decreases as QALYs gained increase, indicating better cost-effectiveness with more QALYs. Higher unavailability weights lead to higher ICER values for models relying on features with unknown costs (all except LR). Additionally, higher disease prevalence tends to result in lower ICER values suggesting better cost-effectiveness of the models. Moreover, the examination of the ICER indicate that irrespective of the model employed, the procedural costs remain notably low³⁷ when compared with the potential gain in quality-adjusted life years (Fig. 2). Logistic Regression (LR), a model requiring only five input features, of which only one incurs a substantial cost, remains the cheapest procedure, while SVM, which takes 21 features as an input, remains the most expensive (Fig. 2). When sticking to the current state of knowledge about the prices, i.e., taking into account kit price for the features unavailable to examine in a laboratory, the order of ICER follows the order of the number of features. However, when considering the scenario where features not currently routinely measured are presumed to be significantly more expensive than the kit price, which is much more plausible, there is a shift in ICER outcomes among the methods evaluated. Averaged over

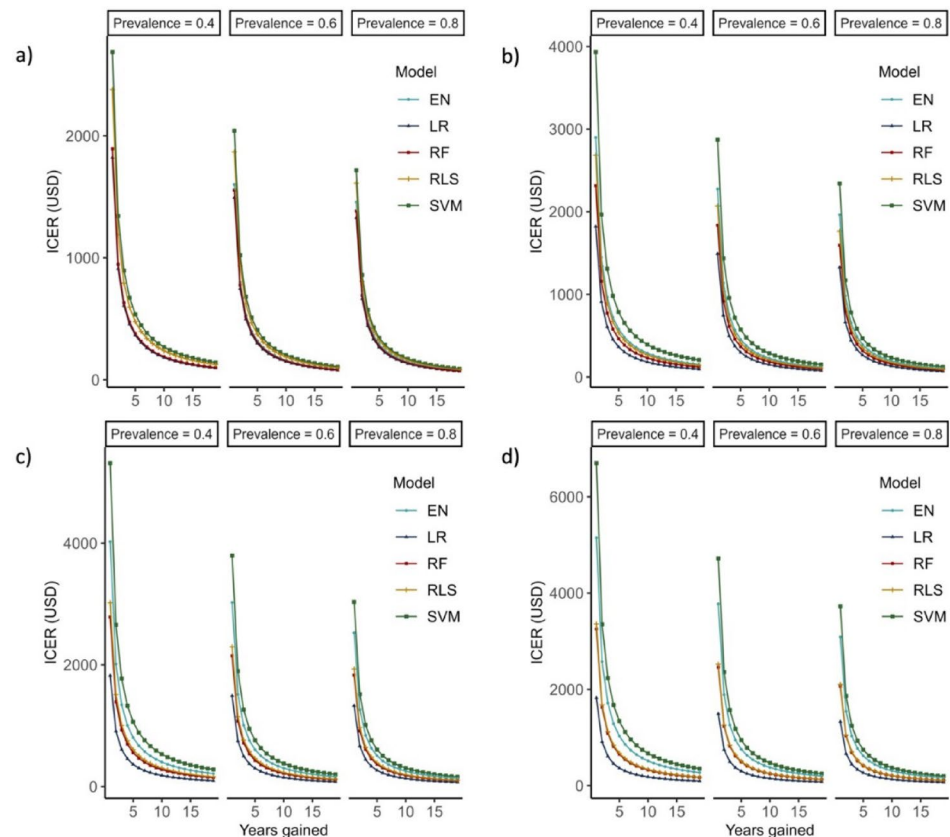


Fig. 2. ICER plots with respect to mVC prevalence, quality of life years gained, and unavailability weight. (a) unavailability weight = 1, (b) unavailability weight = 10, (c) unavailability weight = 20, (d) unavailability weight = 30. The prices were converted from PLN to USD for clarity.

prevalence, QALYs, and unavailability weights, LR emerges as the most cost-effective option with mean ICER equal to \$278, followed by RF (\$412), RLS (\$445), EN (\$608), and SVM (\$769). Sensitivity analysis revealed that the presented results are consistent regardless of the established feature prices. The only exception is sclerostin; assuming a 50% increase in its cost, RLS is favored over RF.

In the supplementary material we explored the LR model's coefficients and showcase its possible clinical utility by calculating ICER for various probability thresholds.

Discussion

In our research, based on the data from 152 participants, we demonstrated the cost-effectiveness of five machine learning frameworks for detecting medial vascular calcification in CKD patients, a group susceptible to mVC. The algorithms were assessed in terms of statistical performance (Table 3) and cost-effectiveness assessed by the incremental cost-effectiveness ratio, ICER (Fig. 2).

Whereas the tested methods had similar predictive power with AUC values between 0.80 and 0.84 and most of them identified traditional risk factors including age, diabetes, male sex, and body mass index (BMI) as important predictors of mVC in patients with CKD, they yielded different results regarding mVC-related features (Table 2). However, the cost differs significantly between the frameworks with LR working on 5 features appearing as the most efficient option.

The accuracies of the models were not perfect, underscoring that there is still much to uncover regarding the biomarkers associated with mVC and that machine-learning-based algorithms cannot serve as a standalone method for assessing mVC presence in CKD patients. However, they can help reduce the frequency of performing unnecessary CT scans for individuals who are found to be less likely to have the pathology, based on the initial assessment of the biomarkers. This reduction can lead to significant savings in healthcare costs, limit radiation exposure, and decrease the time required for diagnostic procedures. In the supplementary material, we provide a detailed example using logistic regression to illustrate how model outputs can be translated into clinical decision-making. Lowering the cut-off threshold for recommending scans increases diagnostic accuracy but reduces potential savings from avoiding unnecessary imaging. The final choice of threshold should be guided by clinical context and resource availability, allowing practitioners to balance diagnostic performance with operational constraints.

In this pilot study, logistic regression emerged as the most effective method. Besides favorable cost-effectiveness, as well as simplicity, and interpretability of the coefficients, it offered another advantage over

the other built classifiers: it required only 5 easily obtainable features (Table 2). This minimizes the likelihood of encountering missing values, a situation more common in complex models. However, this interpretation is possible only after looking at the models' cost-effectiveness and the sets of their required features. Solely examining performance evaluation metrics (Table 3) makes determining the best of the built models much more complex.

Furthermore, examining a panel of different outcomes of the applied feature selection frameworks may provide valuable insights into biomarkers potentially related to mVC. A predictor that emerged as particularly important in our analysis is copeptin that was chosen by all utilized algorithms (Table 2). This confirms findings from a previous study on this topic³⁸. Osteoprotegerin and sclerostin, chosen by 4 and 3 models, respectively, have also been demonstrated to be associated with mVC presence^{15,39}. Hence, it would be worthwhile to perform a longitudinal study to assess whether it is justified to incorporate one or more of these three biomarkers into regular clinical practice.

Finally, we highlight some of the well-established or plausible underlying pathophysiological links between the selected variables and mVC (Table S4). This may reinforce the rationale for including some of the identified predictors when designing studies aiming at detecting mVC in future investigations.

In the context of applied biomedicine, it is increasingly recognized that the criteria for assessing a successful statistical model should extend beyond the predictive power of the classifiers; they ought to also be tailored to align with the medical facilities' condition and capabilities. Thus, the cost of the procedure, the availability, and interpretability of the utilized features, should be also considered. Our findings demonstrated that, given certain conditions, a framework employing less expensive variables can outperform another that relies on fewer but costly ones. This was exemplified by RLS, which produced better results in terms of ICER when compared to EN despite utilizing 5 additional features (Fig. 2b–d) and obtaining far worse precision. Moreover, it produced equivalent results when compared to RF which employed 10 additional features (Fig. 2c, d). Although effective therapies specifically targeting mVC are currently lacking, there are interventions available that can slow its progression^{11,12}. This supports the inclusion of *years_gained* in the ICER calculation, as early detection of mVC followed by appropriate clinical management may lead to gains in quality-adjusted life years. In the future, the development of therapies capable of reversing mVC would likely increase the expected *years_gained*, thereby reducing the relative cost of using biomarkers as a pre-screening tool, as illustrated in Fig. 2.

A major strength of our study is the comprehensiveness of the performed analysis and that it is based on a unique clinical material with histological identification of mVC in artery specimens. To the best of our knowledge, this represents one of the most extensive clinical datasets of arterial biopsies gathered from chronic kidney disease patients. The collected database includes, inter alia, an evaluation of several factors with documented involvement in the disturbed mineral metabolism in CKD and plausible involvement in the etiology of mVC such as sclerostin³⁸, osteoprotegerin³⁹, calciprotein particles⁴⁰, FGF23⁴¹, klotho⁴¹, and parathyroid hormone⁴². We showed the interdependencies between features (Spearman rho, Fig. S1), univariable associations between mVC and each one of the 60 investigated features (Table 1) and performed a multivariable analysis that allowed us to select subsets of features associated with mVC, which entered classification models (Table 2). To the best of our knowledge, no previous studies on mVC detection analyzed ICER or any other price-related metrics of the evaluated procedures.

Our study has several limitations which should be considered when interpreting the results. First, the database includes missing values. Whereas their imputation can change the original dataset, including only complete cases may result in a considerable reduction of the number of included patients and features and therefore, a loss of statistical power. Additionally, many statistical tools and algorithms require a complete dataset; for this reason, and considering the relatively small sample size, we decided to fill in the missing data and ensured that the variable distribution did not alter significantly post-imputation. It should also be noted that imputation may interfere with the stability of feature selection. Furthermore, the lack of external validation is a key limitation, as it prevents us from fully assessing the generalizability and robustness of the developed models.

Moreover, due to the retrospective nature of this long-lasting study, some potentially relevant features were not analysed which may limit the comprehensiveness of our findings. Missing features include, for example, N-terminal pro b-type natriuretic peptide (NT-proBNP) and Gla-rich protein, a vitamin K dependent calcification inhibitor^{43,44}.

Another issue is that the costs related to the measurements needed for ICER analysis can vary significantly between countries, laboratories, and over time. While the sensitivity analysis revealed the consistency of the presented results, it is important to emphasize that the conducted investigation is only a rough estimation of the potential costs associated with each procedure. Before implementation of such a detection method, medical facilities should estimate the costs based on their resources and capabilities.

Lastly, it is important to note that mVC distribution varies across different vascular beds^{45,46}. In the past, mVC presence assessed in the inferior epigastric artery was linked with higher values of coronary artery calcification (CAC) score¹⁵, which altogether demonstrates the complexity and variability of the condition. However, further studies are needed to assess the impact of the selected features on calcification in different vascular beds, as the current findings may not be universally applicable.

Conclusion

Our findings showcase the importance of employing analysis that considers not only statistical accuracy but also economic implications of proposed machine learning frameworks. In the present study, the incremental cost-effectiveness ratio (ICER), was found to provide a suitable criterion for model selection, as analysis using ICER is where the difference between the models becomes evident. This highlights the importance of considering cost-effectiveness when selecting the final classifier, as a minor increase in model performance might not balance the costs related to measuring model-required inputs. While the findings from this pilot study warrant validation

on a larger dataset, we believe that it may encourage other researchers using machine learning algorithms for detection of medial vascular calcification to seek optimal solutions that consider not only predictive capabilities but also the applicability of the implemented methods.

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at Karolinska Institutet.

Received: 7 January 2025; Accepted: 13 May 2025

Published online: 20 May 2025

References

- London, G. M. et al. Arterial media calcification in end-stage renal disease: impact on all-cause and cardiovascular mortality. *Nephrol. Dialysis Transplantation*. <https://doi.org/10.1093/ndt/gfg414> (2003).
- Nelson, A. J. et al. Targeting vascular calcification in chronic kidney disease. *JACC Basic. Transl. Sci.* <https://doi.org/10.1016/j.jacbt.2020.02.002> (2020).
- Duhn, V. et al. Breast arterial calcification: A marker of medial vascular calcification in chronic kidney disease. *Clin. J. Am. Soc. Nephrol.* <https://doi.org/10.2215/CJN.07190810> (2011).
- London, G. M. et al. Arterial media calcification in end-stage renal disease: impact on all-cause and cardiovascular mortality. *Nephrol. Dialysis Transplantation*. **18**, 1731–1740. <https://doi.org/10.1093/ndt/gfg414> (2003).
- Erlundsson, H. et al. Scoring of medial arterial calcification predicts cardiovascular events and mortality after kidney transplantation. *J. Intern. Med.* **291**, 813–823. <https://doi.org/10.1111/joim.13459> (2022).
- Park, S. et al. Vascular calcification as a novel risk factor for kidney function deterioration in the nonelderly. *J. Am. Heart Assoc.* <https://doi.org/10.1161/JAHA.120.019300> (2021).
- Lin, Y.-L. & Hsu, B.-G. Vitamin K and vascular calcification in chronic kidney disease: an update of current evidence. *Tzu Chi Med. J.* **35**, 44. https://doi.org/10.4103/tcmj.tcmj_100_22 (2023).
- Bao, W. H. et al. Relationship between gut microbiota and vascular calcification in Hemodialysis patients. *Ren. Fail.* <https://doi.org/10.1080/0886022X.2022.2148538> (2023).
- Düsing, P. et al. Vascular pathologies in chronic kidney disease: pathophysiological mechanisms and novel therapeutic approaches. *J. Mol. Med.* <https://doi.org/10.1007/s00109-021-02037-7> (2021).
- Xu, C., Smith, E. R., Tiong, M. K., Ruderman, I. & Toussaint, N. D. Interventions to attenuate vascular calcification progression in chronic kidney disease: A systematic review of clinical trials. *J. Am. Soc. Nephrol.* <https://doi.org/10.1681/ASN.2021101327> (2022).
- Raggi, P. et al. The ADVANCE study: A randomized study to evaluate the effects of Cinacalcet plus low-dose vitamin D on vascular calcification in patients on Hemodialysis. *Nephrol. Dialysis Transplantation*. <https://doi.org/10.1093/ndt/gfq725> (2011).
- Chen, N. C., Hsu, C. Y. & Chen, C. L. The strategy to prevent and regress the vascular calcification in dialysis patients. *Biomed. Res. Int.* <https://doi.org/10.1155/2017/9035193> (2017).
- Marreiros, C., Viegas, C. & Simes, D. Targeting a silent disease: vascular calcification in chronic kidney disease. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms232416114> (2022).
- Raggi, P. & O'Neill, W. C. Imaging for vascular calcification. *Semin Dial.* <https://doi.org/10.1111/sdi.12596> (2017).
- Qureshi, A. R. et al. Increased circulating sclerostin levels in end-stage renal disease predict biopsy-verified vascular medial calcification and coronary artery calcification. *Kidney Int.* <https://doi.org/10.1038/ki.2015.194> (2015).
- Lanzer, P. et al. Medial arterial calcification: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* <https://doi.org/10.1016/j.jacc.2021.06.049> (2021).
- Ren, S. C. et al. Vascular calcification in chronic kidney disease: an update and perspective. *Aging Dis.* **13**, 673–697. <https://doi.org/10.14336/AD.2021.1024> (2022).
- Lanzer, P. et al. Medial vascular calcification revisited: review and perspectives. *Eur. Heart J.* **35**, 1515–1525. <https://doi.org/10.1093/eurheartj/ehu163> (2014).
- Hjortnaes, J., New, S. E. P. & Aikawa, E. Visualizing novel concepts of cardiovascular calcification. *Trends Cardiovasc. Med.* <https://doi.org/10.1016/j.tcm.2012.09.003> (2013).
- Lanzer, P. et al. Medial vascular calcification revisited: review and perspectives. *Eur. Heart J.* <https://doi.org/10.1093/eurheartj/ehu163> (2014).
- Jinnouchi, H. et al. Intravascular imaging and histological correlates of medial and intimal calcification in peripheral artery disease. *EuroIntervention*. <https://doi.org/10.4244/EIJ-D-20-01336> (2021).
- Konijn, L. C. D. et al. CT calcification patterns of peripheral arteries in patients without known peripheral arterial disease. *Eur. J. Radiol.* <https://doi.org/10.1016/j.ejrad.2020.108973> (2020).
- Gölüke, N. M. S. et al. Serum biomarkers for arterial calcification in humans: A systematic review. *Bone Rep.* <https://doi.org/10.1016/j.bonr.2022.101599> (2022).
- Wen, L., Chen, J., Duan, L. & Li, S. Vitamin K-dependent proteins involved in bone and cardiovascular health (Review). *Mol. Med. Rep.* <https://doi.org/10.3892/mmr.2018.8940> (2018).
- Dai, L. et al. Phenotypic features of vascular calcification in chronic kidney disease. *J. Intern. Med.* <https://doi.org/10.1111/joim.13012> (2020).
- Lyu, B. et al. Vascular calcification markers and hemodialysis vascular access complications. *Am. J. Nephrol.* <https://doi.org/10.1159/000493549> (2018).
- Bambha, K. & Kim, W. R. Cost-effectiveness analysis and incremental cost-effectiveness ratios: uses and pitfalls. *Eur. J. Gastroenterol. Hepatol.* **16**, 519–526. <https://doi.org/10.1097/00042737-200406000-00003> (2004).
- Qureshi, A. R. et al. Increased circulating sclerostin levels in end-stage renal disease predict biopsy-verified vascular medial calcification and coronary artery calcification. *Kidney Int.* **88**, 1356–1364. <https://doi.org/10.1038/ki.2015.194> (2015).
- Benjamini, Y. & Hochberg, Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educational Behav. Stat.* **25**, 60–83. <https://doi.org/10.3102/10769986025001060> (2000).
- Hastie, T., Tibshirani, R. & Friedman, J. The Elements of Statistical Learning Data Mining, Inference, and Prediction (12th printing). (2017).
- Ben-Hur, A. & Weston, J. A user's guide to support vector machines. *Methods Mol. Biol.* https://doi.org/10.1007/978-1-60327-241-4_13 (2010).
- James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction To Statistical Learning with Applications in R* 8th edn <https://doi.org/10.1201/9781315120256> (Springer, 2017).
- Freijeiro-González, L., Febrero-Bande, M. & González-Manteiga, W. A critical review of LASSO and its derivatives for variable selection under dependence among covariates. *Int. Stat. Rev.* **90**, 118–145. <https://doi.org/10.1111/insr.12469> (2022).
- Hastie, T., Tibshirani, R., James, G. & Witten, D. An introduction to statistical learning (2nd ed.). *Springer Texts* **102**. (2021).

35. Bobrowski, L. et al. Selection of genetic and phenotypic features associated with inflammatory status of patients on dialysis using relaxed linear separability method. *PLoS ONE*. **9**, e86630. <https://doi.org/10.1371/journal.pone.0086630> (2014).
36. Weinstein, M. C. & Stason, W. B. Foundations of cost-effectiveness analysis for health and medical practices. *N. Engl. J. Med.* **296**, 716–721. <https://doi.org/10.1056/nejm197703312961304> (1977).
37. Appleby, J., Devlin, N. & Parkin, D. NICE'S cost effectiveness threshold. *Br. Med. J.* <https://doi.org/10.1136/bmj.39308.560069.BE> (2007).
38. Golembiewska, E. et al. Copeptin is independently associated with vascular calcification in chronic kidney disease stage 5. *BMC Nephrol.* <https://doi.org/10.1186/s12882-020-1710-6> (2020).
39. Makarović, S., Makarović, Z., Steiner, R., Mihaljević, I. & Milas-Ahić, J. Osteoprotegerin and vascular calcification: clinical and prognostic relevance. *Coll. Antropol.* **39**, (2015).
40. ter Braake, A. D. et al. Calciprotein particle Inhibition explains magnesium-mediated protection against vascular calcification. *Nephrol. Dialysis Transplantation.* <https://doi.org/10.1093/ndt/gfz190> (2020).
41. Yamada, S. & Giachelli, C. M. Vascular calcification in CKD-MBD: roles for phosphate, FGF23, and Klotho. *Bone* **100**, 87–93. <https://doi.org/10.1016/j.bone.2016.11.012> (2017).
42. Fujii, H. Association between parathyroid hormone and cardiovascular disease. *Therapeutic Apheresis Dialysis.* <https://doi.org/10.1111/1744-9987.12679> (2018).
43. Silva, A. P. et al. Gla-rich protein (GRP) as an early and novel marker of vascular calcification and kidney dysfunction in diabetic patients with CKD: A pilot cross-sectional study. *J. Clin. Med.* <https://doi.org/10.3390/jcm9030635> (2020).
44. Jouni, H., Rodeheffer, R. J. & Kullo, I. J. Increased serum N-terminal pro-B-type natriuretic peptide levels in patients with medial arterial calcification and poorly compressible leg arteries. *Arterioscler. Thromb. Vasc. Biol.* <https://doi.org/10.1161/ATVBAHA.110.216770> (2011).
45. Sinha, S. & Santoro, M. M. New models to study vascular mural cell embryonic origin: implications in vascular diseases. *Cardiovasc. Res.* <https://doi.org/10.1093/cvr/cvy005> (2018).
46. Muyor, K. et al. Vascular calcification in different arterial beds in ex vivo ring culture and in vivo rat model. *Sci. Rep.* <https://doi.org/10.1038/s41598-022-15739-w> (2022).

Author contributions

UB – conceptualization, methodology, software, visualization, writing – original draft; MD – conceptualization, writing – review & editing; LD – data curation, writing – review & editing; AQ – data curation; LB – software; MS – supervision; BL – supervision, writing – review & editing; PS – supervision; TL – software; JP – conceptualization, supervision, writing – review & editing.

Funding

Baxter Novum is the result of a grant from Baxter Healthcare Corporation to Karolinska Institutet.

Declarations

Competing interests

Peter Stenvinkel has a support from Bayer for conducting a randomized trial on testosterone supplementation in dialysis patients, participates in scientific advisory boards: Astra Zeneca, Glaxo, Vifor, Baxter, Fresenius, Invizius, and has payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Reata, Astra Zeneca, Baxter, Fresenius, Novo Nordisk, Astellas, Pfizer, Bayer; Magnus Soderberg is a full-time employee of AstraZeneca; Malgorzata Debowska has grant no 2018/31/D/ST7/03472 from National Science Center (Poland); Jan Poleszczuk has a grant grant No. 2018/31/D/ST7/03472 from National Science Center (Poland); Bengt Lindholm has a grant to Karolinska Institutet from Baxter Healthcare Corporation and was previously employed by Baxter Healthcare Corporation. He also has received stock or stock options from Baxter Healthcare Corporation; Urszula Bialonczyk – none; Lu Dai – none; Abdul Rashid Qureshi – none; Leon Bobrowski – none; Tomasz Lukaszuk – none.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02457-2>.

Correspondence and requests for materials should be addressed to U.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025