Appendix I

Figure A1: Clinical frailty scale

Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ. 2005 Aug 30;173(5):489–95.

Appendix II. The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1 - 1.3 are all reported in abstract (page 1).
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Pages 2 - 3
Objectives	3	State specific objectives, including any prespecified hypotheses			Page 3
Methods					
Study Design	4	Present key elements of study design early in the paper			Page 4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Page 4
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Pages 4 - 5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Page 5
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Page 5

Appendix II. Continued

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Bias	9	Describe any efforts to address potential sources of bias			Page 6
Study size	10	Explain how the study size was arrived at			Page 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Page 5
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Pages 5 - 7
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Pages 4 - 5
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	N/A
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Pages 4 - 6
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Pages 7 - 9

Appendix II. Continued

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			Page 8
fMain results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			Page 9 - 10
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			N/A
Discussion					
Key results	18	Summarise key results with reference to study objectives			Pages 10 - 11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Pages 10 - 13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			Page 13
Generalisability	21	Discuss the generalisability (external validity) of the study results			Page 13
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Page 14
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 14

Appendix III

Figure 1: ROC Curves of models trained on original dataset (cut-off of 4)

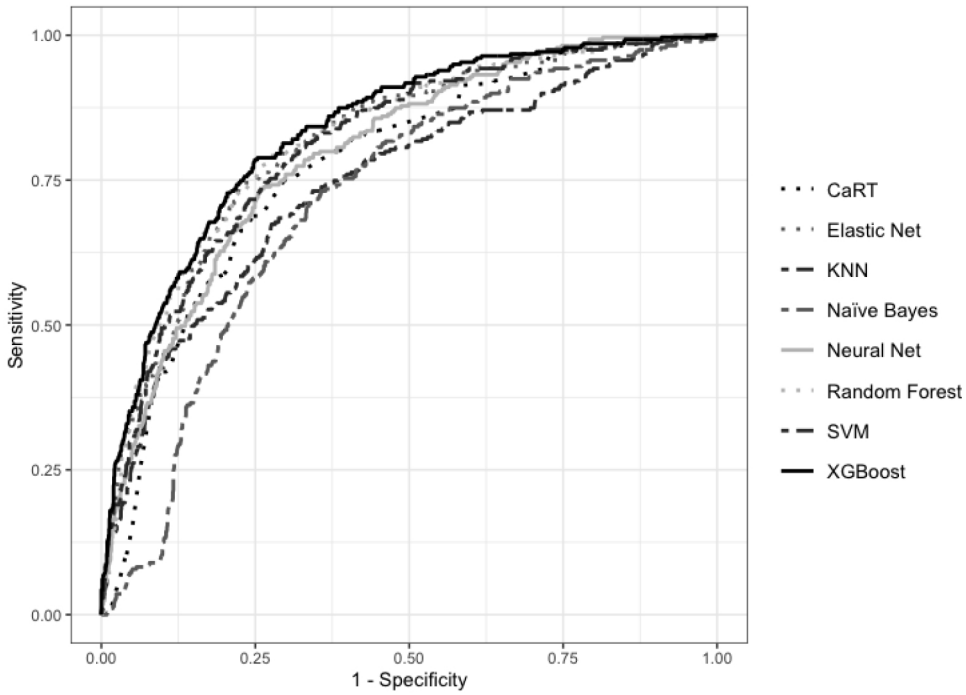

Table 1: Performance metrics of models trained on original data using default threshold (cut-off of 4)

Model	AUC	Accuracy	F1	Sensitivity	Specificity	PPV	NPV
Elastic Net Logistic Regression	81.43%	77.30%	51.56%	70.97%	78.60%	40.49%	92.96%
SVM	79.01%	73.52%	60.62%	57.19%	82.56%	64.48%	77.70%
KNN	73.46%	72.18%	45.71%	32.88%	93.93%	75.00%	71.66%
Naïve Bayes	68.48%	66.20%	42.65%	73.84%	64.63%	29.99%	92.33%
CaRT	75.67%	82.98%	46.66%	68.82%	74.12%	35.29%	92.05%
Random Forest	79.36%	75.41%	63.50%	58.39%	85.88%	69.59%	78.85%
XGBoost	81.91%	76.08%	53.06%	73.12%	78.97%	41.63%	93.47%
Feedforward NN	79.56%	81.03%	47.02%	49.46%	87.50%	44.81%	89.41%

Table 2: Sensitivity and specificity of models trained on original data using best threshold – cut-off of 4

Model	Sensitivity	Specificity	Threshold
Elastic Net Logistic Regression	74.55%	77.06%	0.4787
SVM	71.75%	72.51%	0.3754
KNN	65.92%	70.05%	0.2914
Naïve Bayes	61.30%	70.33%	0.0000
CaRT	64.90%	76.02%*	0.2540
Random Forest	72.26%	73.36%	0.4070
XGBoost	76.37%*	71.75%	0.3385
Feedforward NN	73.48%	74.19%	0.5534

Figure 2: ROC Curves of models trained on original dataset (cut-off of 6)

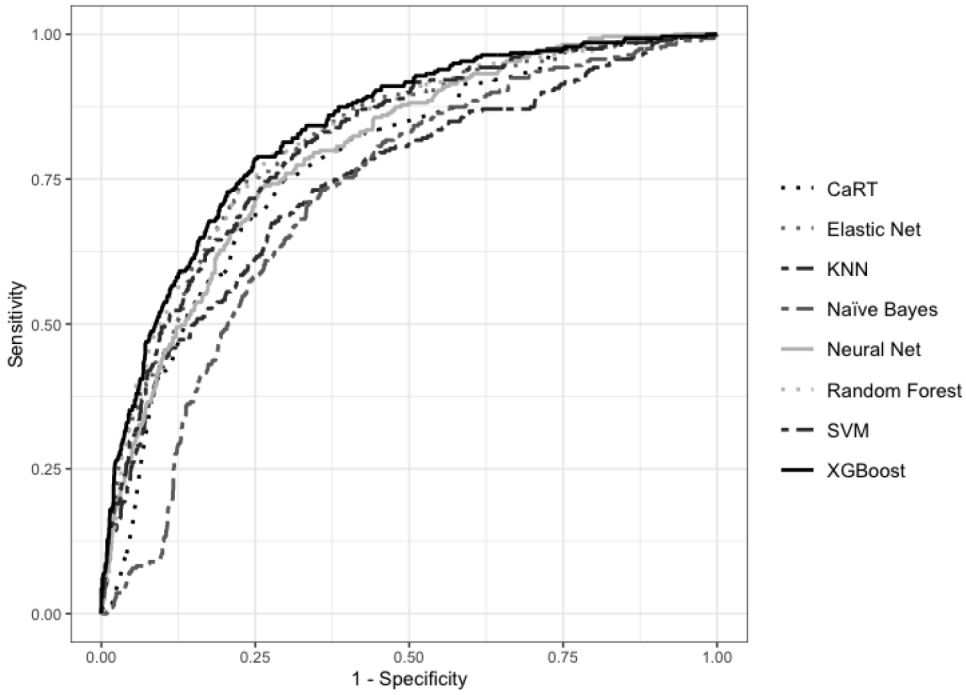

Table 3: Performance metrics of models trained on original data using default threshold (cut-off of 6)

Model	AUC	Accuracy	F1	Sensitivity	Specificity	PPV	NPV
Elastic Net Logistic Regression	80.83%	84.32%	23.28%	13.98%	98.75%	69.64%	84.84%
SVM	71.49%	78.95%	39.79%	40.86%	86.76%	38.78%	87.73%
KNN	73.06%	64.92%	4.64%	2.40%	99.53%	73.68%	64.81%
Naïve Bayes	72.21%	71.63%	43.64%	64.52%	73.09%	32.97%	90.94%
CaRT	78.81%	90.85%	32.13%	32.67%	92.88%	31.61%	93.19%
Random Forest	78.86%	91.28%	11.18%	6.00%	99.87%	81.82%	91.34%
XGBoost	83.70%	91.64%	27.91%	17.20%	98.75%	73.85%	92.95%
Feedforward NN	75.04%	90.97%	11.90%	6.67%	99.46%	55.56%	91.36%

Table 4: Sensitivity and specificity of models trained on original data using best threshold – cut-off of 6

Model	Sensitivity	Specificity	Threshold
Elastic Net Logistic Regression	73.12%	76.47%	0.0806
SVM	67.33%	72.87%	0.1277
KNN	67.33%	66.62%	0.0255
Naïve Bayes	59.33%	80.05%	0.9973
CaRT	78.85%*	71.99%	0.0568
Random Forest	70.67%	70.92%	0.2086
XGBoost	76.00%	77.77%*	0.0875
Feedforward NN	74.19%	64.34%	0.0000

Table 5: Hyperparameters used for final models

Model	Original imbalanced data using cut-off of 5	SMOTE balanced data using cut-off of 5	Original imbalanced data using cut-off of 4	Original imbalanced data using cut-off of 6
Elastic Net Logistic Regression	alpha = 0.5318833, lambda = a0.005369339	alpha = 0.1764004, lambda = 0.002016792	alpha = 0.5600862, lambda = 7.090597	alpha = 0.1, lambda = 0.01925033
SVM	polynomial kernel, degree = 3, scale = 0.004422882, C = 0.1504941	radial kernel, sigma = 0.02996594, C = 170.478	degree = 2, scale = 0.0005473211, C = 267.0139	linear kernel, C = 181.4091
KNN	kmax = 55, distance = 0.2262503, kernel = triweight	kernel = rank, distance = 1, kmax = 500	kmax = 105, distance = 1.644928, kernel = cos	kmax = 1043, distance = 0.9733469, kernel = triweight
Naïve Bayes	fL = 0, usekernel = True, adjust = 1	fL = 0.1, no kernel usage, adjust = 0.5	fL = 0, usekernel= T, adjust = 1	fL = 0, usekernel = F, adjust = 1
CaRT	cp = 0.0002762431	cp = 0.009829198	cp = 0.00201909	cp = 0
Random Forest	mtry = 11, splitrule = gini, min.node.size = 9	mtry= 3, splitrule = gini, min.node.size = 2	mtry = 11, splitrule = gini, min.node.size = 9	mtry = 12
XGBoost	nrounds = 971, max_adepth = 2, eta = 0.2322766, gamma = 5.086296, colsample_bytree = 0.5705734, min_child_weight = 18, subsample = 0.9047023	nrounds = 365, max_depth = 2, eta = 0.2394084, gamma = 9.56787, colsample_bytree = 0.3579414, min_child_weight = 5, subsample = 0.6451248	nrounds = 707, max_depth = 6, eta = 0.06909712, gamma = 6.766357, colsample_bytree = 0.3710754, min_child_wight = 1, subsample = 0.7310282	nrounds = 714, max_depth = 7, eta = 0.06228869, gamma = 7.277172, colsample_bytree = 0.3480463, min_child_weight = 15, subsample = 0.5177022
Feedforward NN	epochs = 500, hidden = c(100, 100, 100, 100, 100), activation = 'MaxoutWithDropOut', dropout = 50%, loss =CrossEntropy	epochs = 500, hidden = c(100, 100, 100, 100, 100), activation = 'MaxoutWithDropOut', dropout = 50%, loss =CrossEntropy	epochs = 500, hidden = c(100, 100, 100, 100, 100), activation = 'MaxoutWithDropOut', dropout = 50%, loss =CrossEntropy	epochs = 500, hidden = c(100, 100, 100, 100, 100), activation = 'MaxoutWithDropOut', dropout = 50%, loss =CrossEntropy