Computers in Biology and Medicine 174 (2024) 108321

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Predictive analytics for cardiovascular patient readmission and mortality: An explainable approach

Leo C.E. Huberts [a,*], Sihan Li [a], Victoria Blake [a,b], Louisa Jorm [a], Jennifer Yu [c,d], Sze-Yuan Ooi [c,d], Blanca Gallego [a,1]

[a] Centre for Big Data Research in Health, University of New South Wales, Sydney, NSW, Australia
[b] Eastern Heart Clinic, Prince of Wales Hospital, Sydney, NSW, Australia
[c] School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia
[d] Prince of Wales Hospital, South Eastern Sydney Local Health District, NSW, Australia

ARTICLE INFO

Keywords:
Cardiovascular disease (CVD)
Predicting readmission
Predicting mortality
Explainable machine learning
Risk factors

ABSTRACT

Background: Cardiovascular patients experience high rates of adverse outcomes following discharge from hospital, which may be preventable through early identification and targeted action. This study aimed to investigate the effectiveness and explainability of machine learning algorithms in predicting unplanned readmission and death in cardiovascular patients at 30 days and 180 days from discharge.

Methods: Gradient boosting machines were trained and evaluated using data from hospital electronic medical records linked to hospital administrative and mortality data for 39,255 patients admitted to four hospitals in New South Wales, Australia between 2017 and 2021. Sociodemographic variables, admission history, and clinical information were used as potential predictors. The performance was compared to LASSO regression, as well as the HOSPITAL and LACE risk score indices. Important risk factors identified by the gradient-boosting machine model were explored using Shapley values.

Results: The models performed well, especially for the mortality outcomes. Area under the receiver operating characteristic curve values were 0.70 for readmission and 0.87–0.90 for mortality using the full gradient boosting machine algorithms. Among the top predictors for 30-day and 180-day readmission were increased red cell distribution width, old age (especially above 80 years), high measured troponin and urea levels, not being married or in a relationship, and low albumin levels. For mortality, these included increased red cell distribution width, old age (especially older than 70 years), high measured troponin and urea levels, high neutrophil and monocyte counts, and low eosinophil and lymphocyte counts. The Shapley values gave clear insight into the dynamics of decision-tree-based models.

Conclusions: We demonstrated an explainable predictive algorithm to identify cardiovascular patients who are at high risk of readmission or death at discharge from the hospital and identified key risk factors.

1. Introduction

Cardiovascular disease (CVD) continues to be one of the leading causes of morbidity and mortality worldwide, causing an estimated 17.9 million deaths each year [1]. A significant proportion of patients hospitalised for CVD are readmitted after discharge due to recurrent events or complications, leading to increased healthcare costs and resource utilisation. Furthermore, these unplanned readmissions contribute to a decreased quality of life, an increased risk of mortality, and increased stress on patients and caregivers [2,3]. The rate of unplanned readmission is recognised worldwide as an indicator of the quality and safety of hospital care, primary care, and transitions between the two, including in countries such as Australia, Canada, the United Kingdom, and the United States (US) [4].

From a clinical perspective, readmissions have considerable negative impacts on both patients and the healthcare system. For example, a US study exploring Medicare claims data revealed that unplanned readmissions within 30 days of discharge impose a yearly cost of over $17.4 billion (17%) on the American healthcare system, from a total hospital spend of $102.6 billion [5]. An Australian study demonstrated

* Correspondence to: AGSM Building, G27 Botany St, Kensington NSW 2052, Australia.
E-mail address: l.huberts@unsw.edu.au (L.C.E. Huberts).

[1] Blanca Gallego and Sze-Yuan Ooi are co-senior authors of this paper. Blanca Gallego is the senior technical author and Sze-Yuan Ooi is the senior clinical author.

https://doi.org/10.1016/j.compbiomed.2024.108321
Received 11 November 2023; Received in revised form 6 February 2024; Accepted 13 March 2024
Available online 20 March 2024

that readmission costs constituted almost 30% of patient admission costs for atherothrombotic diseases [6].

In Australia, CVD remains a major health problem, with an adult prevalence rate of approximately 6.2%. It imposes a cost exceeding $11.8 billion on the Australian healthcare system, representing 8.7% of total expenditure [7]. Between July 2020 and June 2021, there were more than 600,000 acute CVD hospitalisations in Australia, and CVD accounted for 25% of all deaths. An estimated 646,000 healthy life years were lost in Australia in 2018 due to CVD [7]. The reported rate of readmission within 30 days for Australian patients hospitalised for CVD varies between 6.3% and 27% [8], noting significant heterogeneity in methodologies between studies.

Although challenges in data quality, data sharing, and standardisation of readmission definitions persist, harnessing big data in healthcare provides a significant opportunity to understand disease progression and improve patient outcomes [9]. The ability to identify CVD patients at high risk of readmission and mortality is critical to enable preventive measures, such as remote patient monitoring, especially given constrained health care resources and the increasing number of CVD-related hospitalisations per year [7].

Various parametric and semiparametric models have been developed to predict readmissions and mortality. For example, Mišić et al. [10] Mišić et al. [11] and Lee et al. [12] showed the potential of the use of machine learning in general postoperative readmission and mortality prediction within emergency departments. A systematic review by Smith et al. covered 18 traditional CVD predictive risk stratification models, such as the HOSPITAL score and the AMI registry model, which were designed specifically to predict the readmission risk associated with AMI in the US [13–15]. However, the performances of these models were modest, with an average area under the receiver operating characteristic curve (AUC) of 0.65 (0.53–0.79).

A common issue that many electronic medical record (EMR)-based machine learning studies share is the relative lack of clarity in explaining how their models make predictions. The models they use tend to function as 'black boxes', offering little insight into the reasoning behind specific predictions. Although this does not detract from the accuracy of the predictions, it can limit the utility of these models in clinical environments where the basis of a prediction is often as crucial as the prediction itself.

In the current study, we used machine learning models to identify cardiovascular patients at high risk of readmission and death within 30 days and 180 days from hospital discharge, which could allow for the implementation of preventive measures. We harnessed Australian EMR data to predict the risk of readmission and mortality using gradient boosting machines, compared the results to more traditional risk scoring methods and subsequently extensively analysed the model and individual risk factors using Shapley values.

2. Methods

2.1. Objective

The primary aim of this study was to develop machine learning models capable of leveraging data from EMRs to predict the risk of readmission and death for cardiac patients at the time of hospital discharge. To this purpose, we processed and transformed EMR data and linked external data sources to build an extensive feature set of demographic and clinical information. We then predicted readmission and death after cardiac admissions for 30-day and 180-day time frames. A secondary objective of this study was to identify important risk factors, which we explored using Shapley values.

2.2. Study design, setting and study population

In this retrospective study, we used the Cardiac Analytics and Innovation (CardiacAI) Data Repository [16]. CardiacAI is a research-ready data resource that includes deidentified electronic medical record data from four hospitals in New South Wales (NSW), Australia[2] for patients admitted under a cardiologist, cardiothoracic surgeon or vascular surgeon ('cardiovascular patients'). These hospitals serve a diverse population, including patients from lower socioeconomic backgrounds and culturally and linguistically diverse communities. Data were extracted from the EMR systems, including admissions between January 1, 2017, and November 25, 2021. The EMR data set comprises demographic information, medical history, medications, allergies, vital signs, laboratory test results, diagnostics, treatments, and notes. The EMR data set was linked to state-wide administrative data for hospital admissions and death registrations.[3] A comprehensive overview of the data set is published [16].

The EMR data contained records for 61,720 cardiovascular index admissions for 44,201 patients between 01 January 2017 and 25 November 2021. The inclusion criteria to define an index admission were as follows.

1. The patient was admitted alive (not for organ procurement) on a date between 1 January 2017 and 25 November 2021.
2. The EMR had a valid link to the state-wide data set on hospital admissions and deaths.
3. The discharge date was K days before the maximum date of the linked data, where K refers to 30 or 180 days, depending on the outcome under consideration.
4. There was at least one pathology result and one vital sign result available in the EMR.
5. The patient was discharged alive.

2.3. Definition of target outcomes

We defined four target outcomes, all identified from the linked administrative data: unplanned (emergency) readmissions to a NSW hospital and all-cause deaths, within both 30 or 180 days from the discharge date of index admission.

2.4. Patient characteristics

Patient sociodemographic variables included gender, language, religion, marital status, and residential postcode. The residential postcodes of the patients were mapped according to the four Socio-Economic Indexes for Areas (SEIFA) 2016 developed by the Australian Bureau of Statistics [17]. For index admissions in which the postcodes did not link to a SEIFA score, we used the median score.

General administrative information about the inpatient visit included the patient's age, financial class, length of stay, admission facility, admission specialty, source of admission (planned or unplanned), and destination of discharge.

Clinical information derived from the EMR contained the last available result for each patient's pathology test and vital sign measurements, the procedures performed during admission, consult note types, allergies, discharged medications, and diagnoses. An overview of the pathology tests and vital signs included in the study is presented in Appendix 1. All categorical variables in the data were transformed into binary representations using one-hot encoding.

[2] Prince of Wales Public Hospital, St George Public Hospital, The Sutherland Hospital, and The Wollongong Hospital.

[3] The NSW Admitted Patient Data Collection (APDC) and the NSW mortality data recorded in the NSW Registry of Births Deaths & Marriages (RBDM).

2.5. Models

We used Chi-squared tests to compare the main characteristics of the two groups. To model readmission and mortality outcomes, for our main approach we used gradient-boosted decision trees (GBDT), which are designed to handle large datasets and high-dimensional feature spaces [18]. For comparison, we used Least Absolute Shrinkage and Selection Operator (LASSO) regression and calculated the HOSPITAL and LACE risk scores. Subsequently, we used the GBDT importance values and Shapley values for feature evaluation.

Data were divided into a training set (80% of the sample, used to train the machine learning models), a validation set (10%, used for probability calibration), and a test set (10%, used for model evaluation) using a grouped stratified random method to ensure that an even proportion of classes was obtained for each set. No patient was included in more than one set. To tune the GBDT hyperparameters, we ran 100 trials on the training data using an automatic hyperparameter optimisation framework to maximise the area under the precision recall curve (AUPRC) [19]. In each trial, the training set underwent 5-fold cross-validation.

LASSO (Least Absolute Shrinkage and Selection Operator) regression was applied to the same training, validation, and test sets as the GBDT model.[4] This method is commonly used for prediction, with feature selection and regularisation integrated into the regression estimation [20].

Probability calibration is needed to map predicted probabilities to the true likelihood of the event, as uncalibrated models may not have a nominal coverage probability [21,22]. Calibrated models for both the GBDT and LASSO models were developed by fitting the base models on the validation sets using isotonic regression. The calibrated model probability predictions were classified using thresholds that maximised Youden's J statistic within the validation set [23].

In addition to the GBDT and LASSO models, we calculated two established clinical risk score models: HOSPITAL and LACE. The HOSPITAL score, designed to predict the risk of hospital readmission, incorporates factors like hemoglobin level, discharge from an oncology service, sodium level, procedure during the hospital stay, index type of admission, number of hospital admissions, and length of stay [24]. The LACE index, another commonly used tool, predicts the risk of death or unplanned readmission within 30 days after hospital discharge. It includes length of stay, acuity of the admission, comorbidity of patients (measured by the Charlson comorbidity index), and emergency department visits in the preceding six months [25]. These two risk scores were calculated for the full set and classified using the threshold calculated by the maximising the Youden's J statistic.

As an alternative modelling approach, we discuss the use of multi-class labels for the two time frames in Appendix 2.

2.6. Model evaluation

For each combination of outcome and time period, we used the top 50 variables determined by the GBDT split-based importance to estimate Shapley values. The split-based importance of a predictor is based on the number of times the predictor was used in splits in all trees in training. This two-step approach uses the flexibility of the GBDT method to perform feature selection and the Shapley values to interpret the parameters.

Shapley values are calculated by considering all possible combinations of features and then assessing the change in the model prediction when a specific feature is added to a combination. See Appendix 3 for more details on these feature evaluation techniques.

[4] We used the default parameters of the sklearn LASSO module in Python.

3. Results

3.1. Patient characteristics

There were 61,721 admissions in the full set, 58,350 were successfully linked to the APDC, 57,524 were discharged alive, 55,790 (53,299) had at least 30 (180) days of follow-up in the data, and 39,255 (36,462) had at least one pathology and vital sign result available in the EMR.[5]

For the 30-day follow-up, the 39,255 index admissions (for just over 31,000 patients) included 5039 (12.8%) unplanned readmissions and 338 (0.86%) deaths. The 36,462 index admissions (for just over 29,000 patients) for the 180-day follow-up included 11,241 (30.1%) unplanned readmissions and 1499 (4.1%) deaths.

Table 1 compares the baseline demographic characteristics of the index admissions, split into groups based on the observed results at 30 days and 180 days. Given the 30-day time frame, the sex, age, length of stay, marital status, language, emergency/non-emergency type of admission, and the SEIFA indexes all showed a statistically significant association with unplanned readmission. In contrast, sex and the SEIFA indexes did not have a statistically significant association with mortality, while the discharge destination was significant. The readmission rates for women (14.0%) were higher than those for men (12.1%). Older age and length of stay showed a clear correlation with readmission rates and death. Readmission rates increased with age, from 9.5% for ages below 50 to 17.8% for ages 90 and up. The increase in death rates was even more pronounced, from only 0.2% for patients under 50 years of age to 3.6% for patients older than 90 years. Mortality rates for patients over 90 years of age were double those for patients aged between 80 and 90 years.

The patterns of relationships among baseline variables and outcomes for the 180-day time frame 2, were broadly similar to those for the 30-day time frame.

3.2. Models' performances

Table 3 shows the performance of the uncalibrated GBDT, LASSO, HOSPITAL and LACE models in terms of AUC, accuracy, sensitivity, and specificity for the 30-day and 180-day time frames. Note that the scores for the GBDT and LASSO models were based on a test set, whereas the HOSPITAL and LACE indices were calculated for the full set. In both time frames for both outcomes, the GBDT model achieved the highest AUC values.

The calibrated results for the GBDT and LASSO models are presented in Table 4, the GBDT performance remained superior in terms of AUC and sensitivity. The GBDT readmission models achieved an AUC of 0.70 for both time frames. The calibrated sensitivities for the readmission models were 0.703 and 0.722 with specificities of 0.611 and 0.587 for the 30-day and 180-day time frames, respectively. The calibrated mortality models achieved superior performance, with AUC scores of 0.89 and 0.87, sensitivities of 0.855 and 0.891 and specificities of 0.809 and 0.705 for the 30-day and 180-day time frames, respectively. Fig. 1 shows the calibration plots for the GBDT models.

3.3. Risk factor evaluation

We calculated Shapley values for the 50 variables with the highest importance in the separated-outcome GBDT models to evaluate the effects of individual variables on readmission and death for cardiac patients.

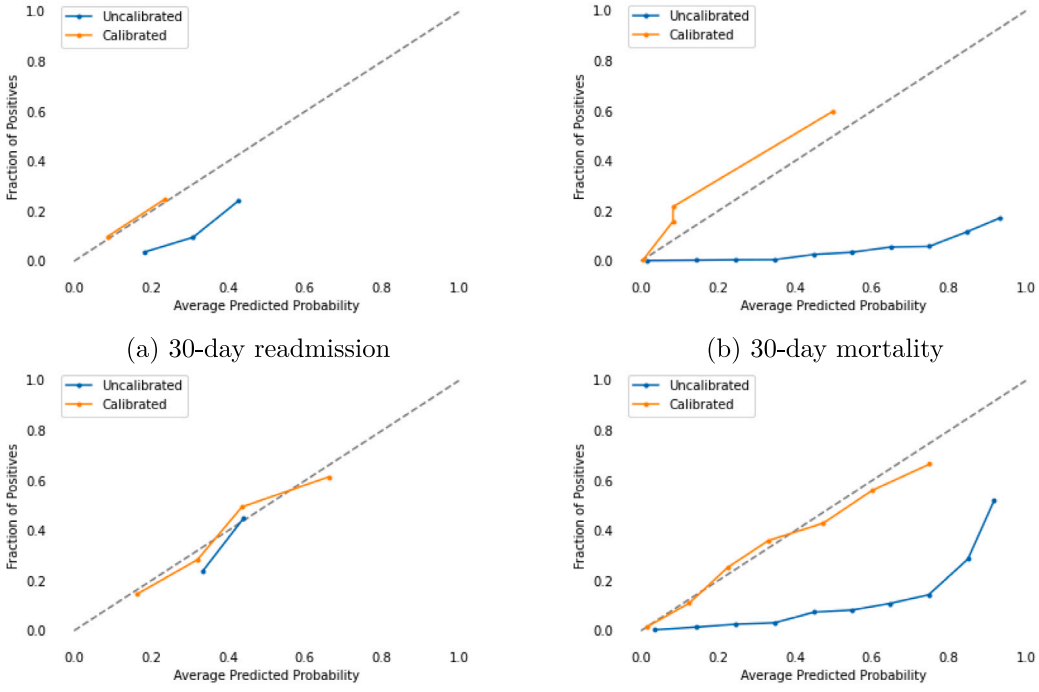
Six variables were among the 20 most predictive for each outcome and time frame: the red cell distribution width (RDW), age, troponin,

[5] Note that patients admitted under a vascular surgeon were excluded, as there were no pathologies and vital signs available at the time.

Table 1
Baseline characteristic table for 30-day events.

Demographics	Subgroup	Readmission		No readmission		p-values	Mortality		Alive		p-values
		#	%	#	%		#	%	#	%	
Sex	Female	2140	42.6%	13 172	38.5%	<0.01	126	37.3%	15 186	39.0%	0.55
	Male	2899	57.4%	21 044	61.5%		212	62.7%	23 731	61.0%	
Age group	<50	428	6.0%	4076	11.9%	<0.01	9	2.7%	4495	11.6%	<0.01
	50–60	565	10.7%	4969	14.5%		9	2.7%	5525	14.2%	
	60–70	833	16.4%	7170	21.0%		26	7.7%	7977	20.5%	
	70–80	1390	28.8%	9079	26.5%		68	20.1%	10 401	26.7%	
	80–90	1494	30.6%	7404	21.6%		160	47.3%	8738	22.5%	
	>90	329	7.5%	1518	4.4%		66	19.5%	1781	4.6%	
Length of stay	0–1	911	18.4%	10 817	31.6%	<0.01	30	8.9%	11 698	30.1%	<0.01
	2–3	1389	29.3%	9976	29.2%		63	18.6%	11 302	29.0%	
	4–7	1103	22.7%	6301	18.4%		70	20.7%	7334	18.8%	
	>7	1636	29.6%	7119	20.8%		175	51.8%	8580	22.0%	
Specialty	Cardiology	4627	95.4%	31 755	92.8%	0.134	322	95.3%	36 060	92.7%	0.084
	Cardiothoracic surgery	412	4.6%	2461	7.2%		16	4.7%	2857	7.3%	
Marital status	Married or Defacto	2644	52.7%	20 455	59.8%	<0.01	184	54.4%	22 915	58.9%	0.11
	Others	2395	47.3%	13 761	40.2%		154	45.6%	16 002	41.1%	
Language	English	4205	81.2%	29 444	86.1%	<0.01	274	81.1%	33 375	85.8%	0.0174
	Non_English	834	18.8%	4772	13.9%		64	18.9%	5542	14.2%	
Discharged to	Others	947	17.5%	6441	18.8%	0.973	111	32.8%	7277	18.7%	<0.01
	Priv Med Prac other than Psychiatric	4092	82.5%	27 775	81.2%		227	67.2%	31 640	81.3%	
Admission type	Emergency	4313	88.2%	28 300	82.7%	<0.01	299	88.5%	32 314	83.0%	<0.01
	Non_Emergency	726	11.8%	5916	17.3%		39	11.5%	6603	17.0%	
Index of relative socio-economic disadvantage decile group	≤5	1869	37.5%	12 321	36.0%	0.14	114	33.7%	14 076	36.2%	0.382
	>5	3170	62.5%	21 895	64.0%		224	66.3%	24 841	63.8%	
Index of relative socio-economic advantage and disadvantage decile group	≤5	1074	20.8%	7532	22.0%	0.27	60	17.8%	8546	22.0%	0.0725
	>5	3965	79.2%	26 684	78.0%		278	82.2%	30 371	78.0%	
Index of economic resources decile group	≤5	2639	52.8%	16 776	49.0%	<0.01	177	52.4%	19 238	49.4%	0.308
	>5	2400	47.2%	17 440	51.0%		161	47.6%	19 679	50.6%	
Index of education and occupation decile group	≤5	1093	22.0%	7912	23.1%	0.025	63	18.6%	8942	23.0%	0.0682
	>5	3946	78.0%	26 304	76.9%		275	81.4%	29 975	77.0%	

The event rates and p-values calculated from chi-square tests are reported. Index of Relative Socio-economic Disadvantage Decile, Index of Relative Socio-economic Advantage and Disadvantage Decile, Index of Economic Resources Decile, and Index of Education and Occupation Decile refer to the scores allocated in the SEIFA based on postcodes.



(a) 30-day readmission (b) 30-day mortality

(c) 180-day readmission (d) 180-day mortality

Fig. 1. Calibration plots for the LightGBM models.

Table 2
Baseline characteristic table for 180-day events.

Demographics	Subgroup	Readmission		No readmission		p-values	Mortality		Alive		p-values
		#	%	#	%		#	%	#	%	
Sex	Female	4735	42.1%	9519	37.7%	<0.01	596	39.8%	13 658	39.1%	0.608
	Male	6506	57.9%	15 702	62.3%		903	60.2%	21 305	60.9%	
Age group	<50	832	7.4%	3315	13.1%	<0.01	26	1.7%	4121	11.8%	<0.01
	50–60	1171	10.4%	3914	15.5%		49	3.3%	5036	20.0%	
	60–70	1721	15.3%	5757	22.8%		140	9.3%	7338	29.1%	
	70–80	3209	28.5%	6562	26.0%		342	22.8%	9429	37.4%	
	80–90	3496	31.1%	4787	19.0%		664	44.3%	7619	30.2%	
	>90	812	7.2%	886	3.5%		278	18.5%	1420	5.6%	
Length of stay	0–1	2316	20.6%	8397	33.3%	<0.01	148	9.9%	10 565	30.2%	<0.01
	2–3	3189	28.4%	7349	29.1%		322	21.5%	10 216	40.5%	
	4–7	2414	21.5%	4471	17.7%		340	22.7%	6545	26.0%	
	>7	3322	29.6%	5002	19.8%		689	46.0%	7635	30.3%	
Specialty	Cardiology	10 478	93.2%	23 287	92.3%	<0.01	1437	95.9%	32 328	92.5%	<0.01
	Cardiothoracic surgery	763	6.8%	1934	7.7%		62	4.1%	2635	7.5%	
Marital status	Married or defacto	5874	52.3%	15 576	61.8%	<0.01	752	50.2%	20 698	59.2%	<0.01
	Others	5367	47.7%	9645	38.2%		747	49.8%	14 265	40.8%	
Language	English	9365	83.3%	21 895	86.8%	<0.01	1220	81.4%	30 040	85.9%	<0.01
	Non_English	1876	16.7%	3326	13.2%		279	18.6%	4923	14.1%	
Discharged to	Others	2117	18.8%	4767	18.9%	0.890	365	24.3%	6519	18.6%	<0.01
	Priv Med Prac other than psychiatric	9124	81.2%	20 454	81.1%		1134	75.7%	28 444	81.4%	
Admission type	Emergency	9638	85.7%	20 642	81.8%	<0.01	1300	86.7%	28 980	82.9%	<0.01
	Non_Emergency	1603	14.3%	4579	18.2%		199	13.3%	5983	17.1%	
Index of relative socio-economic disadvantage decile group	≤5	4030	35.9%	9195	36.5%	0.271	510	34.0%	12 715	36.4%	0.0686
	>5	7211	64.1%	16 026	63.5%		989	66.0%	22 248	63.6%	
Index of relative socio-economic advantage and disadvantage decile group	≤5	2288	20.4%	5762	22.8%	<0.01	280	18.7%	7770	22.2%	<0.01
	>5	8953	79.6%	19 459	77.2%		1219	81.3%	27 193	77.8%	
Index of economic resources decile group	≤5	5764	51.3%	12 288	48.7%	<0.01	748	49.9%	17 304	49.5%	0.777
	>5	5477	48.7%	12 933	51.3%		751	50.1%	17 659	50.5%	
Index of education and occupation decile group	≤5	2343	20.8%	6068	24.1%	<0.01	289	19.3%	8122	23.2%	<0.01
	>5	8898	79.2%	19 153	75.9%		1210	80.7%	26 841	76.8%	

The event rates and p-values calculated from chi-squared tests are reported. Index of Relative Socio-economic Disadvantage Decile, Index of Relative Socio-economic Advantage and Disadvantage Decile, Index of Economic Resources Decile, and Index of Education and Occupation Decile refer to the scores allocated in the SEIFA based on postcodes.

Table 3
Uncalibrated model performances (AUC, Accuracy, Threshold, Sensitivity, Specificity) using the test set for the separate binary models.

Days	Outcome	Model	AUC	Accuracy	Threshold	Sensitivity	Specificity
30	Readmission	GBDT	**0.701**	0.666	0.5	0.647	0.668
		LASSO	0.597	0.85	0.5	0.048	0.974
		HOSPITAL	0.590	0.604	2	0.545	0.613
		LACE	0.606	0.594	7	0.560	0.599
30	Mortality	GBDT	**0.898**	0.927	0.5	0.522	0.931
		LASSO	0.815	0.988	0.5	0.000	1.000
		HOSPITAL	0.669	0.596	2	0.702	0.595
		LACE	0.696	0.668	8	0.637	0.668
180	Readmission	GBDT	**0.7**	0.684	0.5	0.0	1.0
		LASSO	0.677	0.694	0.5	0.323	0.864
		HOSPITAL	0.587	0.599	2	0.515	0.638
		LACE	0.604	0.596	7	0.537	0.624
180	Mortality	GBDT	**0.876**	0.838	0.5	0.703	0.844
		LASSO	0.779	0.951	0.5	0.116	0.988
		HOSPITAL	0.658	0.606	2	0.684	0.602
		LACE	0.680	0.596	7	0.537	0.624

urea levels, albumin levels, and the number of medical progress notes registered by cardiology. Appendix 4 gives visualisations of the 20 variables with the highest absolute Shapley values for each of the four models.

The RDW was an important predictor for all four outcomes. RDW measurements were split into the coefficient of variation (RDW-CV) and the red cell distribution standard deviation (RDW-SD). The RDW-CV is the red cell distribution width relative to the average size; the RDW-SD is a direct measure of the RDW distribution. The second of the 20 most predictive variables for each outcome and time frame was the patient's age, with mortality and readmission risks increasing with age. 30-day readmission was the least affected by age. The partial dependence plots for the RDW and age variables are shown in Fig. 2.

The third of the 20 most predictive variables for each outcome and time frame was the level of blood troponin, an indicator of heart muscle injury (See Fig. 4 for the partial dependence plots). There was a largely constant increased risk for troponin values above the normal range.

The fourth common predictive variable was related to kidney failure: patient urea levels, where above normal urea levels elevated the probability of readmission and death in all timeframes. Figs. 3(a)–3(d)

Table 4
Calibrated model performances (AUC, Accuracy, Threshold, Sensitivity, Specificity) using the test set for the separate binary models. The calibrated threshold was determined by the maximum Youden's J statistic.

Days	Outcome	Model	AUC	Accuracy	Threshold	Sensitivity	Specificity
30	Readmission	GBDT	**0.698**	0.622	0.139	0.703	0.611
		LASSO	0.598	0.671	0.157	0.490	0.698
30	Mortality	GBDT	**0.887**	0.809	0.047	0.855	0.809
		LASSO	0.785	0.826	0.017	0.633	0.828
180	Readmission	GBDT	**0.700**	0.630	0.291	0.722	0.587
		LASSO	0.674	0.659	0.336	0.566	0.701
180	Mortality	GBDT	**0.873**	0.713	0.369	0.891	0.705
		LASSO	0.777	0.652	0.034	0.801	0.645

show the partial dependence plots for the urea level measurements, including the normal range between 1.8 and 7.1 mmol urea per litre. The estimated glomerular filter rate (eGFR) is another measurement used to assess kidney function. This measurement was predictive of readmission at 30 days and 180 days, as well as of mortality at 30 days. Figs. 3(f) and 3(g) show the dependence plots for the eGFR measurements for 180-day readmission and 30-day mortality.

The fifth variable among the 20 most predictive variables for each outcome was the albumin level, where lower values were related to increased risks. The sixth was the number of cardiology medical progress notes, where a higher number of notes was generally related to the risk of mortality and readmission (see Appendix 4).

Variables related to the immune response were generally important for mortality, less so for readmission. These included total white blood cell count, neutrophils, eosinophils, monocytes, and lymphocytes, see Figs. 4(g)–4(h) for the partial dependence plots.

Other common variables with high absolute Shapley values included the oxygen flow rate, mean platelet volume, and the patient's weight.

The SEIFA index of economic resources decile was mainly predictive of readmission within 30 days. Whether a patient was married or in a relationship was highly important for the 180-day readmission predictions.

Important predictors for both 180-day outcomes, but not for the 30-day outcomes, were the chloride level and mean corpuscular hemoglobin concentration. Low body temperatures were predictive of a higher risk of mortality. The red cell count and hematocrit measurements were only predictive of readmission.

See Appendix 4 for the full visualisations of the 20 variables with the highest absolute Shapley values for each of the four models. As an alternative to the use of Shapley values for risk factor evaluation, Appendix 5 presents the results of logistic regressions using the 50 most predictive variables in terms of the split-based importance of the GBDT models.

4. Discussion

This study developed machine learning models capable of leveraging EMR data to predict the risk of readmission and death at discharge for cardiovascular patients. Identifying high-risk patients can aid in prioritising resources and providing symptomatic, palliative and preventive treatment to those who would benefit the most.

The gradient-boosted trees models with Shapley values helped decode the complex interactions between patient characteristics and clinical variables, providing predictive insights into readmission and mortality. The models highlight significant predictors, including the red cell distribution width (RDW), age, troponin level, urea level, albumin level, and the number of recorded clinical progress notes.

The red blood cell distribution width (RDW) was found to have a large impact on the risk of readmission and mortality, consistent with previous findings [26]. Elevated variation in RDW can serve as a non-specific indicator of chronic inflammation, nutritional deficiencies, bleeding, or impaired bone marrow function, which can have implications for cardiovascular health. The Shapley values confirmed this, with higher coefficients leading to an increased risk of readmission and death for both the 30-day and 180-day time intervals.

The Shapley values for age showed a nonlinear pattern with a consistently lower risk of mortality for ages below 60, followed by a small jump in 180-day mortality risk and a steeply increasing mortality risk from 80 years upward (see [27] for a discussion of age as a CVD risk factor). The risk of readmission in the short term was reduced for ages below 30, flat for ages 30–70, increased strongly for ages 70 and up and was especially high for 80+ year-olds. The 180-day predicted readmission risk was similarly reduced for ages below 75, with a large jump in risk around 75 years of age.

Elevated troponin levels in blood tests indicate heart muscle injury and infarction and this information is used to diagnose and monitor heart conditions. Consistent with the literature on troponin levels and myocardial infarction and heart failure, higher troponin levels increased the risk of readmission and death [28].

The urea level and eGFR levels, which contributed significantly to the predictions, are both indicators of renal function. Heart and kidney failure often coexist in patients with cardiovascular disease [29]. Reduced cardiac function leads to renal hypoperfusion and subsequent renal impairment. Furthermore, many cardiac treatments cause renal impairment.

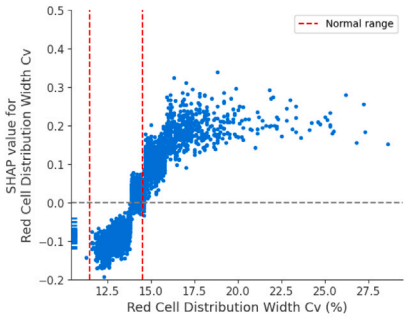
Decreased albumin levels resulted in increased risks of mortality and readmission (consistent with e.g. [30]). Albumin is synthesised by the liver and low levels may be indicative of failure of the hepatic synthetic function.

Variables related to the immune system, such as neutrophil, lymphocyte, eosinophil, and monocyte counts, were especially predictive of mortality. Lower lymphocyte counts or a higher neutrophil-to-lymphocyte ratio can be indicative of increased inflammation and potentially worse CVD outcomes [31,32]. Our results similarly showed that increased neutrophil counts and reduced lymphocyte counts resulted in increased mortality risk.

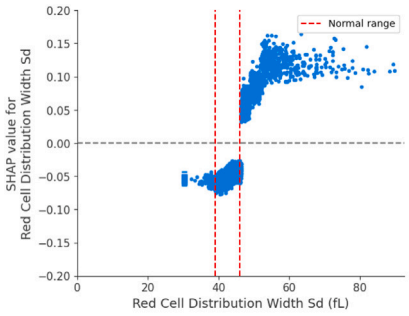
Unlike neutrophils, the relationship between eosinophils and CVD outcomes is less clear. Our results showed that higher levels of eosinophils were associated with a lower risk of mortality at 30 and 180 days from discharge.

Monocytes are a type of white blood cell that can differentiate into macrophages and contribute to the inflammatory response in atherosclerosis. Higher monocyte counts have been associated with a higher risk of acute coronary events [33]. Our results were in line with this.

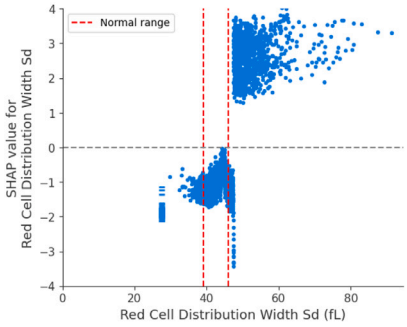
In addition, the number of cardiac progress notes was significant in predicting readmission and death. A potential explanation for this is that the variable serves as an instrument for the severity of a patient's condition. More notes, regardless of content, could be the result of more complications and a longer hospital stay. The patient's marital status, specifically whether a patient was not married or in a de facto relationship, was an important predictor for 180-day readmission, in terms of its Shapley values. A low SEIFA index of economic resources decile lead to increased 30-day readmission risk, in line with previous findings [34].



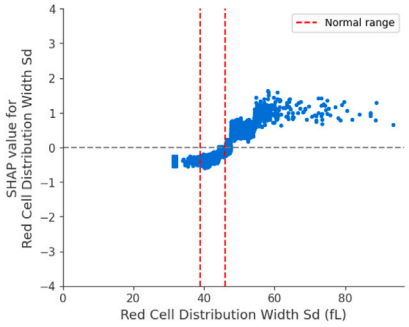
(a) 30-day readmission, Shapley values for RDW-CV, normal range is between 11.6% and 14.6%.



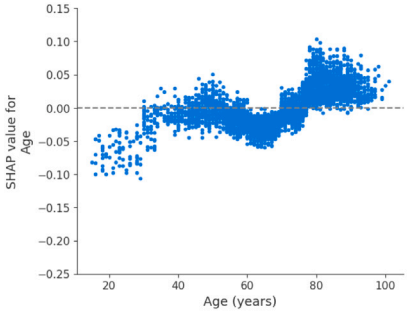
(b) 180-day readmission, Shapley values for RDW-SD, normal range is between 39 and 46 fL.



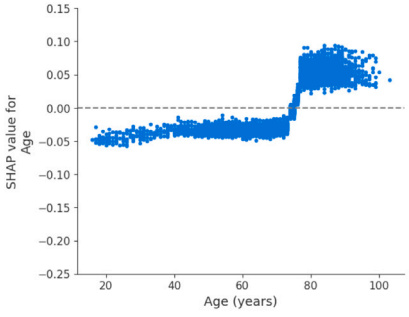
(c) 30-day mortality, Shapley values for RDW-SD, normal range is between 39 and 46 fL.



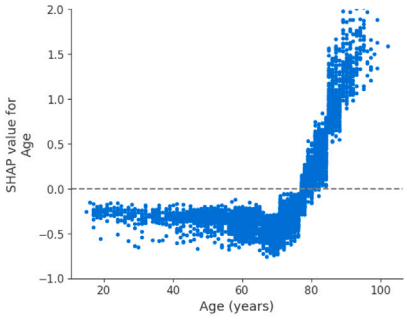
(d) 180-day mortality, Shapley values for RDW-SD, normal range is between 39 and 46 fL.



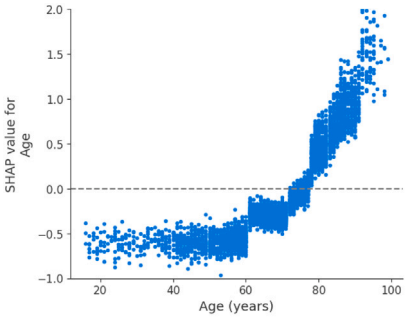
(e) 30-day readmission, Shapley values for age



(f) 180-day readmission, Shapley values for age



(g) 30-day mortality, Shapley values for age



(h) 180-day mortality, Shapley values for age

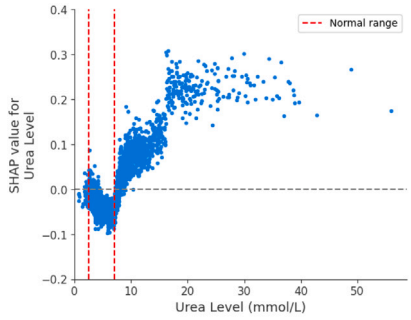
Fig. 2. Partial dependence plots showing the Shapley values for selected RDW levels (a–d) and the patient's age (e–h).

Patients who were on oxygen but had low oxygen flow rates had an increased risk of readmission and 30-day mortality. A high mean platelet volume predicted increased mortality and 30-day readmission, which is consistent with previous findings [35].

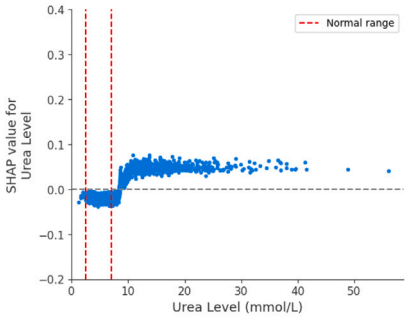
The mean corpuscular hemoglobin concentrations (MCHC) and chloride levels were predictive of the two 180-day outcomes, where lower values increased the risk of 180-day readmission and mortality. Related to the MCHC values and anemia are the hematocrit and hemoglobin values, which were also found to be important for the risk of readmission at 30 and 180 days (see [36] for a study on hematocrit).

Lower patient weights and body mass index values were predictive of increased risk of mortality and 30-day readmission. This phenomenon is often referred to as part of the 'obesity paradox' [37], which relates to findings suggesting that obesity may be protective in people with heart conditions.

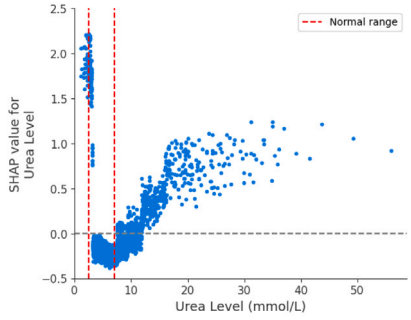
Interestingly, lower body temperatures at discharge was a risk factor for mortality at 30 and 180 days. There is limited research on



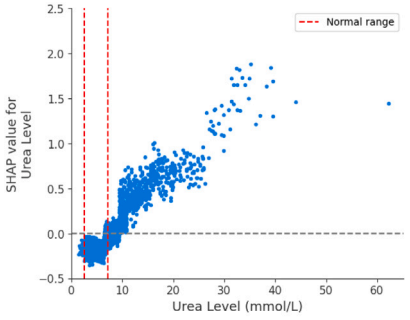
(a) 30-day readmission, Shapley values for urea level, normal range is between 1.8 and 7.1 mmol/L.



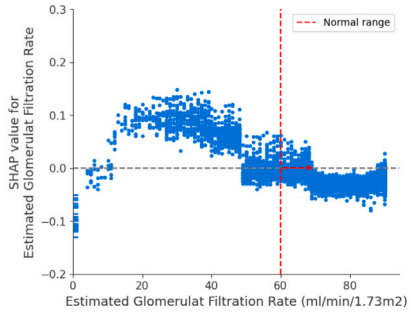
(b) 180-day readmission, Shapley values for urea level, normal range is between 1.8 and 7.1 mmol/L.



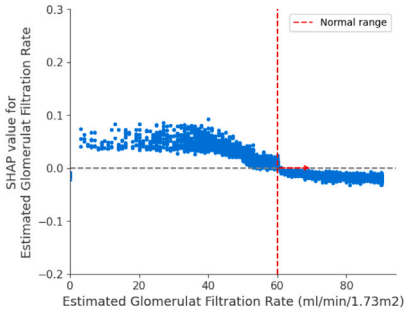
(c) 30-day mortality, Shapley values for urea level, normal range is between 1.8 and 7.1 mmol/L.



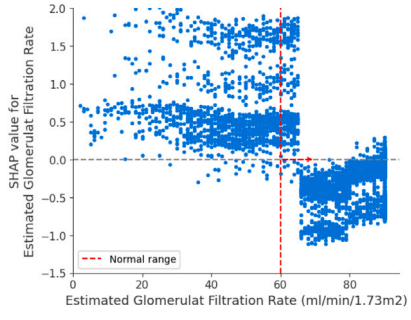
(d) 180-day mortality, Shapley values for urea level, normal range is between 1.8 and 7.1 mmol/L.



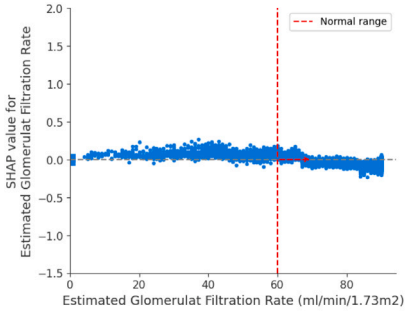
(e) 30-day readmission, Shapley values for eGFR, normal range is above 60 ml/min/1.73m^2.



(f) 180-day readmission, Shapley values for eGFR, normal range is above 60 ml/min/1.73m^2.



(g) 30-day mortality, Shapley values for eGFR, normal range is above 60 ml/min/1.73m^2.



(h) 180-day mortality, Shapley values for eGFR, normal range is above 60 ml/min/1.73m^2.

Fig. 3. Partial dependence plots showing the Shapley values for the urea levels (a–d) and the eGFR levels (e–h).

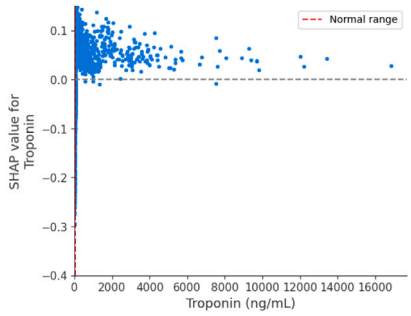
the link between body temperature and the risk of readmission and mortality [38,39].

The superior performance of the models in predicting mortality over readmission suggests that there are unmeasured factors that affect the risk of unplanned readmission more than mortality within the specified time frames. These could relate, for example, to the availability of social support, and access to primary care services.

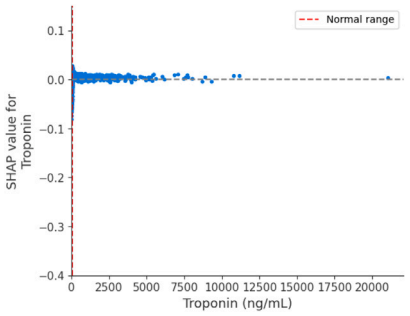
5. Limitations

Generalisability presents a challenge when applying machine learning algorithms in practice [40]. This study did not perform external validation, although multicentre data were used.

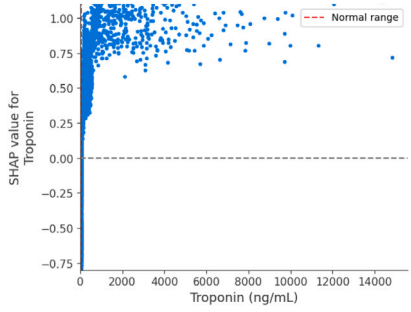
The sparsity of parts of the EMR data (in particular vital signs, pathologies, diagnoses, and medications) hindered the algorithms'



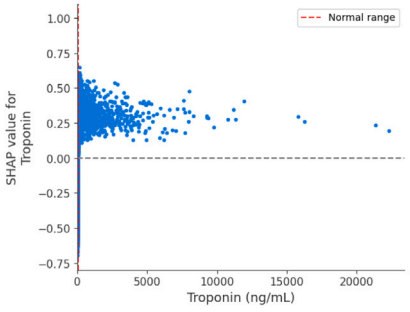
(a) 30-day readmission, Shapley values for troponin levels, normal range is below 0.04 ng/mL



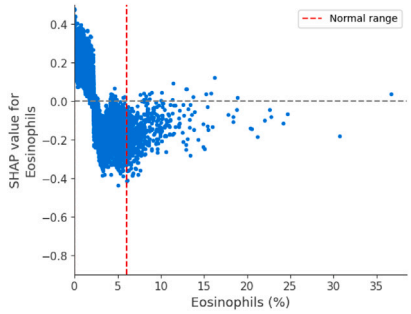
(b) 180-day readmission, Shapley values for troponin levels, normal range is below 0.04 ng/mL



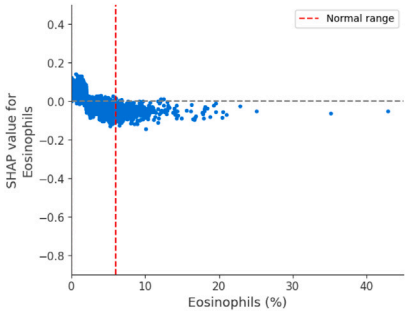
(c) 30-day mortality, Shapley values for troponin levels, normal range is below 0.04 ng/mL



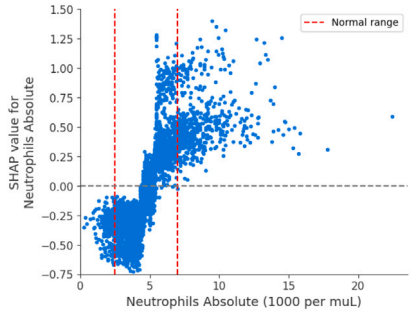
(d) 180-day mortality, Shapley values for troponin levels, normal range is below 0.04 ng/mL



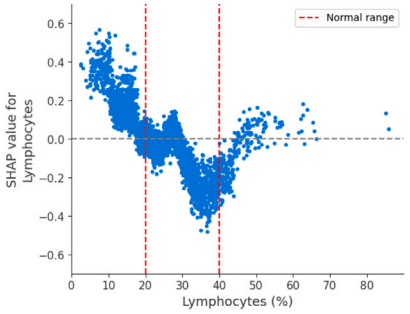
(e) 30-day mortality, Shapley values for eosinophils, normal range is below 6%.



(f) 180-day mortality, Shapley values for eosinophils, normal range is below 6%.



(g) 30-day mortality, Shapley values for neutrophils, normal range is between 2500 and 7000 per μL.



(h) 180-day mortality, Shapley values for lymphocytes, normal range is between 20% and 40%.

Fig. 4. Selected partial dependence plots showing the Shapley values for the troponin levels (a–d), eosinophils (e–f), neutrophils (g), and lymphocytes(h).

training process and predictive performance. Furthermore, some index admissions recorded in the EMR were not linked to the APDC data set, especially toward the end of the data set. This resulted in a reduced sample size and potentially unobserved events.

We used all-cause mortality as the cause of death data from the Australian Coordinating Registry (ACR) Cause of Death Unit Record File was unavailable at the time of conducting this study. Therefore, the mortality target outcomes may have been unrelated to hospital admission in some cases, affecting the estimations.

Many patients were readmitted to medical facilities other than their index hospitals [41]. Linking to administrative data (the APDC and RBDM NSW population data) from the local health registry provided administrative data for events outside of the index hospitals. However, events outside of NSW remained unobserved.

Lastly, there is no standard baseline control to evaluate performance. Many studies used different baselines such as traditional risk prediction models, logistic regression, or other machine learning algorithms to compare with individual algorithms or an ensemble of algorithms [42,43]. This makes comparisons challenging.

6. Conclusions and future directions

Our study sheds light on the prediction of mortality and readmission for cardiac patients at the time of discharge. Using a large data set of patient characteristics and clinical variables, our gradient-boosted trees models were able to identify key predictors of readmission and mortality at 30- and 180-day time frames.

Our approach introduces the use of Shapley values, a concept borrowed from cooperative game theory, to model these complex relationships. This revealed several notable predictors for both outcomes and time frames. Red cell distribution width (RDW), age, troponin, urea level, albumin level, and the number of medical progress notes were among the most predictive factors across all four models.

Variables related to immune response, such as neutrophils, c-reactive proteins, eosinophils, monocytes, and lymphocytes, were crucial in the prediction of mortality. This showed the link between the immune system and the progression and outcomes of cardiovascular disease, underlining the importance of considering a patient's overall health status in cardiovascular disease management.

Model performance for predicting mortality was superior to readmission, suggesting that while many cardiac patients are at risk of readmission, these risks may not necessarily translate into higher mortality rates within the 30- or 180-day time frames.

The use of more specific cohorts of cardiovascular patients with heart failure, ischaemic heart disease, etc. [44–47] may improve machine learning performances and also lead to more clinically meaningful results. We leave this for future studies when a larger sample size of patients becomes available.

In the future, further detailed exploration of other potential predictors, as well as the development of more nuanced prediction models, may aid to improve the quality of care and prognosis of patients with cardiovascular diseases. Our results reinforce the use of machine learning applications in the management of cardiovascular disease.

Hardware and software specification

Python 3.9.7 with packages NumPy 1.20.3, pandas 1.3.4, scikit-learn 0.24.2, lightgbm 3.3.2, optuna 2.10.1, shap 0.41.0 on Amazon WorkSpaces (Windows Server 2016).

Ethics approval

Approval was granted by the South Eastern Sydney Local Health District HREC Executive Committee (2019/ETH12625) & NSW Population Health Services Research Ethics Committee (ETH01614).

CRediT authorship contribution statement

Leo C.E. Huberts: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sihan Li:** Conceptualization, Writing – original draft. **Victoria Blake:** Project administration, Resources, Software, Supervision, Validation. **Louisa Jorm:** Funding acquisition, Methodology, Resources, Software, Supervision. **Jennifer Yu:** Funding acquisition, Resources, Conceptualization, Writing – review & editing. **Sze-Yuan Ooi:** Conceptualization, Funding acquisition, Resources, Validation, Writing – review & editing. **Blanca Gallego:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Blanca Gallego reports financial support was provided by Australian Government Department of Health and Aged Care. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the South Eastern Sydney (SESLHD) and Illawarra Shoalhaven (ISLHD) Local Health Districts for their continued participation in and support for the CardiacAI project whose data were used for this project. We also acknowledge the contribution of the Centre for Health Record Linkage in facilitating data linkage with population health datasets. Other financial support has been provided by the 2020 Medical Research Future Fund (MRFF) Cardiovascular Health Mission Grant program, and the 2021 UNSW Medicine's Cardiac Vascular and Metabolic Medicine Big Ideas Seed Grant program. The views expressed in this article represent those of the authors and not the funding organisations.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108321.

References

[1] World Health Organization, Cardiovascular diseases (CVD), 2023, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed: June 22, 2023.

[2] P. Ponikowski, S.D. Anker, K.F. AlHabib, M.R. Cowie, T.L. Force, S. Hu, T. Jaarsma, H. Krum, V. Rastogi, L.E. Rohde, et al., Heart failure: preventing disease and death worldwide, ESC Heart Fail. 1 (2014) 4–25.

[3] S.S. Virani, A. Alonso, E.J. Benjamin, M.S. Bittencourt, C.W. Callaway, A.P. Carson, A.M. Chamberlain, A.R. Chang, S. Cheng, F.N. Delling, et al., Heart disease and stroke statistics-2020 update: A report from the American Heart Association, Circulation 141 (2020) e139–e596.

[4] Australian Commission on Safety and Quality in Health Care, Avoidable Hospital Readmissions: Report on Australian and International Indicators, Their Use and the Efficacy of Interventions to Reduce Readmissions, Sydney, Australia, 2019.

[5] S. Jencks, M. Williams, E. Coleman, Rehospitalizations among patients in the medicare fee-for-service program, N. Engl. J. Med. 360 (2009) 1418–1428.

[6] E. Atkins, E. Geelhoed, M. Knuiman, T. Briffa, One third of hospital costs for atherothrombotic disease are attributable to readmissions: a linked data analysis, BMC Health Serv. Res. 14 (338) (2014) 1–9.

[7] Australian Institute of Health and Welfare, Heart, stroke and vascular disease: Australian facts, 2023, https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts.

[8] C. Labrosciano, T. Air, R. Tavella, J. Beltrame, I. Ranasinghe, Readmissions following hospitalizations for cardiovascular disease: a scoping review of the Australian literature, Aust. Health Rev. 44 (2020) 93–103.

[9] H. Hemingway, F.W. Asselbergs, J. Danesh, R. Dobson, N. Maniadakis, A. Maggioni, G.J. Van Thiel, M. Cronin, G. Brobert, P. Vardas, et al., Big data from electronic health records for early and late translational cardiovascular research: challenges and potential, Eur. Heart J. 39 (2018) 1481–1495.

[10] V.V. Mišić, E. Gabel, I. Hofer, K. Rajaram, A. Mahajan, Machine learning prediction of postoperative emergency department hospital readmission, Anesthesiology 132 (5) (2020) 968–980.

[11] V.V. Mišić, K. Rajaram, E. Gabel, A simulation-based evaluation of machine learning models for clinical decision support: application and analysis using hospital readmission, NPJ Digit. Med. 4 (1) (2021) 98.

[12] C.K. Lee, I. Hofer, E. Gabel, P. Baldi, M. Cannesson, Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality, Anesthesiology 129 (4) (2018) 649–662.

[13] J. Donzé, D. Aujesky, D. Williams, J. Schnipper, Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model, JAMA Intern. Med. 173 (2013) 632–638.

[14] L.N. Smith, A.N. Makam, D. Darden, H. Mayo, S.R. Das, E.A. Halm, O.K. Nguyen, Acute myocardial infarction readmission risk prediction models: a systematic review of model performance, Circ.: Cardiovasc. Qual. Outcomes 11 (1) (2018) e003885.
[15] J. Brown, S. Conley, N. Niles, Predicting readmission or death after acute ST-elevation myocardial infarction, Clin. Cardiol. 36 (2013) 570–575.
[16] V. Blake, L. Jorm, J. Yu, A. Lee, B. Gallego, S.-Y. Ooi, The Cardiac Analytics and Innovation (CardiacAI) Data Repository: An Australian data resource for translational cardiovascular research, 2023, arXiv preprint arXiv:2304.09341.
[17] Australian Bureau of Statistics, Socio-economic indexes for areas (SEIFA), Australia, 2021, URL https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/latest-release.
[18] J. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Statist. 29 (2001) 1189–1232.
[19] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, 2019, arXiv preprint arXiv:1907.10902.
[20] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[21] K. Meelis, M. Telmo, F. Peter, Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration, Electron. J. Stat. 11 (2017) 5052–5080.
[22] C. Dormann, Calibration of probability predictions from machine-learning and statistical models, Global Ecol. Biogeogr. 29 (2020) 760–765.
[23] W. Youden, Index for rating diagnostic tests, Cancer 3 (1950) 32–35.
[24] J. Donzé, D. Aujesky, D. Williams, J.L. Schnipper, Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model, JAMA Intern. Med. 173 (8) (2013) 632–638, http://dx.doi.org/10.1001/jamainternmed.2013.3023.
[25] C. Van Walraven, I.A. Dhalla, C. Bell, E. Etchells, I.G. Stiell, K. Zarnke, P.C. Austin, A.J. Forster, Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community, CMAJ 182 (6) (2010) 551–557, http://dx.doi.org/10.1503/cmaj.091117.
[26] W.S. Hong, A. Rudas, E.J. Bell, J.N. Chiang, Association of red blood cell distribution width with hospital admission and in-hospital mortality across all-cause adult emergency department visits, JAMIA Open 6 (3) (2023) ooad053.
[27] R. Dhingra, R.S. Vasan, Age as a risk factor, Med. Clin. 96 (1) (2012) 87–91.
[28] Y. Fan, M. Jiang, D. Gong, C. Man, Y. Chen, Cardiac troponin for predicting all-cause mortality in patients with acute ischemic stroke: a meta-analysis, Biosci. Rep. 38 (2) (2018).
[29] J. Rangaswami, V. Bhalla, J.E. Blair, T.I. Chang, S. Costa, K.L. Lentine, E.V. Lerma, K. Mezue, M. Molitch, W. Mullens, et al., Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the American Heart Association, Circulation 139 (16) (2019) e840–e878.
[30] D.J. Rubin, S.H. Golden, M.E. McDonnell, H. Zhao, Predicting readmission risk of patients with diabetes hospitalized for cardiovascular disease: a retrospective cohort study, J. Diabetes Complicat. 31 (8) (2017) 1332–1339.
[31] E.G. Zouridakis, X. Garcia-Moll, J.C. Kaski, Usefulness of the blood lymphocyte count in predicting recurrent instability and death in patients with unstable angina pectoris, Am. J. Cardiol. 86 (4) (2000) 449–451.
[32] A.C. Sawant, P. Adhikari, S.R. Narra, S.S. Srivatsa, P.K. Mills, S.S. Srivatsa, Neutrophil to lymphocyte ratio predicts short-and long-term mortality following revascularization therapy for ST elevation myocardial infarction, Cardiol. J. 21 (5) (2014) 500–508.
[33] F.K. Swirski, M. Nahrendorf, Leukocyte behavior in atherosclerosis, myocardial infarction, and heart failure, Science 339 (6116) (2013) 161–166.
[34] E.F. Philbin, G.W. Dec, P.L. Jenkins, T.G. DiSalvo, Socioeconomic status as an independent risk factor for hospital readmission for heart failure, Am. J. Cardiol. 87 (12) (2001) 1367–1371.
[35] L. Vizioli, S. Muscari, A. Muscari, The relationship of mean platelet volume with the risk and prognosis of cardiovascular diseases, Int. J. Clin. Pract. 63 (10) (2009) 1509–1515.
[36] L. Paul, P. Jeemon, J. Hewitt, L. McCallum, P. Higgins, M. Walters, J. McClure, J. Dawson, P. Meredith, G.C. Jones, et al., Hematocrit predicts long-term mortality in a nonlinear and sex-specific manner in hypertensive adults, Hypertension 60 (3) (2012) 631–638.
[37] E. Jahangir, A. De Schutter, C.J. Lavie, Low weight and overweightness in older adults: Risk and clinical management, Prog. Cardiovasc. Dis. 57 (2) (2014) 127–133.
[38] R. Kang, T. Nagoshi, H. Kimura, T.D. Tanaka, A. Yoshii, Y. Inoue, S. Morimoto, K. Ogawa, K. Minai, T. Ogawa, et al., Possible association between body temperature and B-type natriuretic peptide in patients with cardiovascular diseases, J. Card. Fail. 27 (1) (2021) 75–82.
[39] A. Ahmed, I. Aboshady, S.M. Munir, S. Gondi, A. Brewer, S.D. Gertz, D. Lai, N.A. Shaik, K. Shankar, A. Deswal, et al., Decreasing body temperature predicts early rehospitalization in congestive heart failure, J. Card. Fail. 14 (6) (2008) 489–496.
[40] B. Goldstein, A. Navar, M. Pencina, J. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, J. Am. Med. Inform. Assoc.: JAMIA 24 (2017) 198–208.
[41] Bureau of Health Information, Spotlight on Measurement: Return to acute care following hospitalisation: spotlight on Readmissions, 2015.
[42] S. Shin, P.C. Austin, H.J. Ross, H. Abdel-Qadir, C. Freitas, G. Tomlinson, D. Chicco, M. Mahendiran, P.R. Lawler, F. Billia, et al., Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality, ESC Heart Fail. 8 (2021) 106–115.
[43] S. Mahajan, R. Ghani, Predicting risk of readmission for heart failure using ensemble machine learning methods, Stud. Health Technol. Inform. 264 (2019) 243–247.
[44] V. Sharma, V. Kulkarni, F. McAlister, D. Eurich, S. Keshwani, S. Simpson, D. Voaklander, S. Samanani, Predicting 30-day readmissions in patients with heart failure using administrative data: A machine learning approach, J. Card. Fail. (2022).
[45] A. Okere, V. Sanogo, H. Alqhtani, V. Diaby, Identification of risk factors of 30-day readmission and 180-day in-hospital mortality, and its corresponding relative importance in patients with Ischemic heart disease: a machine learning approach, Expert Rev. Pharmacoecon. Outcomes Res. 21 (2021) 1043–1048.
[46] M.E. Matheny, I. Ricket, C.A. Goodrich, R.U. Shah, M.E. Stabler, A.M. Perkins, C. Dorn, J. Denton, B.E. Bray, R. Gouripeddi, et al., Development of electronic health record-based prediction models for 30-day readmission risk among patients hospitalized for acute myocardial infarction, JAMA Netw. Open 4 (2021) e2035782.
[47] R. Najafi-Vosough, J. Faradmal, S. Hosseini, A. Moghimbeigi, H. Mahjub, Predicting hospital readmission in heart failure patients in Iran: A comparison of various machine learning methods, Healthc. Inform. Res. 27 (2021) 307–314.