Original Research Article

Complementary frailty and mortality prediction models on older patients as a tool for assessing palliative care needs

Vicent Blanes-Selva [iD]

Biomedical Data Science Lab, Instituto Universitarios de Tecnologías de La Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia, Spain

Ascensión Doñate-Martínez

Polibienestar Research Institute, University of Valencia, Spain

Gordon Linklater [iD]

Highland Hospice and NHS Highland, Inverness, UK

Juan M García-Gómez

Biomedical Data Science Lab, Instituto Universitarios de Tecnologías de La Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia, Spain

Abstract

Palliative care (PC) has demonstrated benefits for life-limiting illnesses. Bad survival prognosis and patients' decline are working criteria to guide PC decision-making for older patients. Still, there is not a clear consensus on when to initiate early PC. This work aims to propose machine learning approaches to predict frailty and mortality in older patients in supporting PC decision-making. Predictive models based on Gradient Boosting Machines (GBM) and Deep Neural Networks (DNN) were implemented for binary 1-year mortality classification, survival estimation and 1-year frailty classification. Besides, we tested the similarity between mortality and frailty distributions. The 1-year mortality classifier achieved an Area Under the Curve Receiver Operating Characteristic (AUC ROC) of 0.87 [0.86, 0.87], whereas the mortality regression model achieved an mean absolute error (MAE) of 333.13 [323.10, 342.49] days. Moreover, the

Corresponding author:
Vicent Blanes-Selva, Biomedical Data Science Lab, Instituto Universitarios de Tecnologías de La Información y Comunicaciones (ITACA), Universitat Politècnica de València, Valencia 46022, Spain.
Email: viblasel@upv.es

1-year frailty classifier obtained an AUC ROC of 0.89 [0.88, 0.90]. Mortality and frailty criteria were weakly correlated and had different distributions, which can be interpreted as these assessment measurements are complementary for PC decision-making. This study provides new models that can be part of decision-making systems for PC services in older patients after their external validation.

Introduction

Palliative Care (PC) is a holistic approach that improves patients' quality of life with life-limiting diseases. It is recommended to incorporate early in the disease trajectory, even in conjunction with potentially curative treatments.[1] PC can improve quality of life,[2] mood,[3] symptom control,[4] reduce emergency department visits and hospitalisation,[5] and even increase 1-year survival.[6]

PC services have traditionally been mainly accessed by cancer patients, but there is growing consensus about the importance of promoting access for patients with non-malignant disease at earlier stages.[7,8,9] Patients' prognoses and functional decline are two crucial elements in decision-making to be considered by healthcare professionals in the introduction of PC need assessment and PC conversations with older people.

On the one hand, it is estimated that at least 75% of patients would benefit from access to PC during their terminal illness.[10] Nevertheless, uncertainty about prognostication is cited as a common barrier to PC referral, particularly for patients with non-malignant diseases.[11] On the other hand, frailty in older patients is defined as a state characterised by reduced physiological reserve and loss of resistance to stressors caused by accumulated age-related deficits.[12] Two of the most popular frailty dimensions are the frail phenotype by Fried et al.,[13] which describes frailty as a biological syndrome; and the Frailty Index (FI) by Mitnitski et al.,[14] which is based on health deficits accumulations, also, frailty has been defined since a more comprehensive approach taking into consideration a holistic understating of the person. In this sense, frailty can be experienced by a decrease in human functioning at the physical level and psychological and social domains.[15] Raudonis et al.[16] suggest in their study that frail older adults could benefit from involvement in PC programmes as frailty is associated with poor health outcomes and death.[17]

Different strategies have been used to try to aid prognostication. Clinical intuition was harnessed with the Surprise Question ('Would I be surprised if this patient died in the next year?') which, has been promoted as a tool to prompt clinicians to recognise patients with a limited prognosis.[18] However, in 2017 Downar et al.[19] published a systematic review of the surprise question, concluding that more accurate tools are required given its poor to modest performance as a mortality predictor. Also, it has been demonstrated that the risk of death increases with lower performance levels and with falling performance levels, but survival data varied across different healthcare systems.[20] In this line, the Supportive and Palliative Care Indicators Tool (SPICT) proposes a set of clinical indicators of poor prognosis developed through a consensus of expert opinion,[21] which has shown to have a predictive accuracy of up to 78%.[22] Other studies have used data analysis to propose alternative tools to predict short-term mortality. Bernabeu-Wittel in 2010 developed the PROFUND index,[23] a predictive model for patients with multimorbidity. Van Walraven et al. in 2015 proposed HOMR,[24] a tool for predicting 1-year mortality in adults (≥18 years and ≥ 20 years for the different cohorts). In 2018, Avati et al.[25] proposed a deep learning approach to identify patients with a survival between 3 and 12 months, in 2019 Wegier et al.[26] proposed a version of

Table 1. Patient demographic information.

Sex	N-individuals	Mean age (years)	STD age (years)
Female	9973	80.75	8.67
Male	9780	77.44	8.24
All	19,753	79.11	8.62

HOMR but using only variables available at the admission. In 2021, our team also presented a 1-year mortality model for adults.[27]

Additionally, and as stated before, quantifying frailty is important since as patients become frail, advance care planning conversations should be prioritised to establish patient goals and wishes in advancing serious illness,[28] which may include the involvement in PC programmes. A wide array of FI has been proposed to assess the health status of older adults. The FI has been used to predict mortality and poor health outcomes.[29] Some studies have tried to predict frailty status: Babič et al. in 2019[30] use a clustering approach to identify clusters considering the prefrail, non-frail and frail status using 10 numerical variables for adults over 60 years old. Sternberg et al.[31] in 2012 tried to identify frail patients with their methods against the VES frailty score[32] for patients over 65 years old. Bertini et al.[33] in 2018 created two predictive models for patients over 65 years old: one to assess frailty risk using the probability of hospitalisation or death within the year and a second one to assess worsening risk to each subject in the lower risk class.

Based on these previous results, our aim in this work is to propose a set of Machine Learning (ML) tools capable of making predictions about mortality and frailty for older patients, oncological and non-oncological, so healthcare professionals can benefit from quantitative approaches on data-driven evidence when deciding advance care planning. In this sense, we propose the creation of three different but complementary models: (a) a 1-year mortality classifier that will work as a binary predictor; (b) a survival regression model aimed to obtain a prediction in days from admission to death; and (c) a 1-year frailty classifier to predict the health status, assessed by the FI, of a patient 1 year after admission. The authors consider that the combination of both mortality and frailty criteria, working as complementary information sources, can positively impact detecting needs to start PC conversations.

Materials

Basic description

Data was extracted from the system on 1 November 2019. The dataset contained hospital admissions records for older patients (age ⩾ 65) from 1 January 2011 to 31 December 2018. Patients admitted to psychiatry and obstetrics services were excluded from the study.

Data contains a total of 39,310 hospitalisation episodes corresponding to 19,753 unique patients. The cohort was composed of 9780 males and 9973 females with a mean age of 80.75 years (see Table 1).

Mortality target variables

Mortality target variables were extracted from administrative admission data and the recorded death date of regional civil registration. Patients alive during data extraction were censored for the regression problem due to our inability to know their survival time from admission. However,
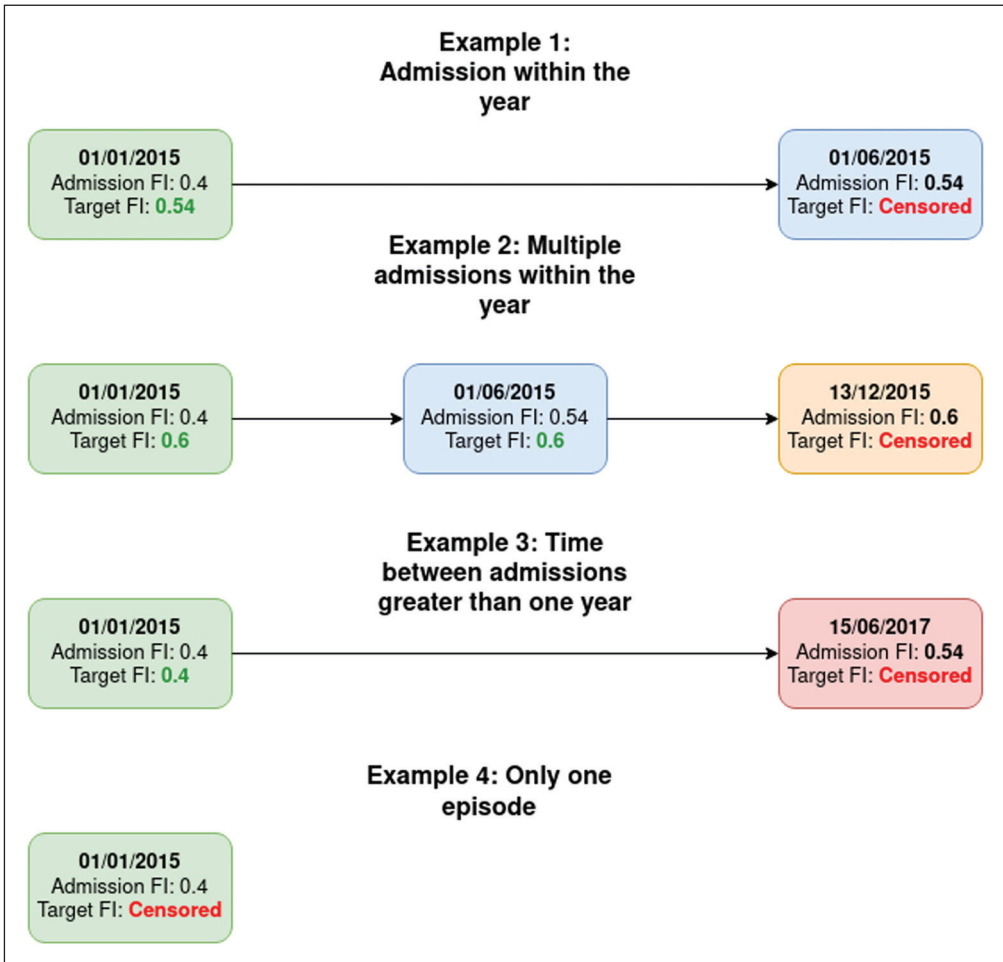
Figure 1. Visual representation of the algorithm to calculate the target FI in all four possible situations.

patients alive with an admission date prior to 1 November 2018 (1 year prior to the extraction) could be included since we could determine their mortality status within the year.

Frailty target variable

As for the frailty target, following the work of Searle et al.,[34] we calculated the FI of every episode (admission frailty) and sorted them chronologically. The target FI of a given episode was the admission frailty of the following episode if this next episode happened within the year. We used the most recent episode as the target if a patient had multiple admissions during the following year. Otherwise, target frailty was set to the same value as the current admission frailty. Most recent episodes and patients with only one episode were removed because no posterior data was available, so we considered them as censored data. Figure 1 presents an example of target FI calculation for each possible situation.

Table 2. List of variables included in the frailty index and their distribution. All variables are binary, and their distribution represents the condition's presence (Y) or absence (N).

Variable	Distribution	Variable	Distribution
Difficulties in dressing	Y: 3829, N: 35481	Difficulties in urinating	Y: 3683, N: 35627
Difficulties in bathing	Y: 5999, N: 33311	Difficulties in stooling	Y: 3121, N: 36189
Difficulties in grooming	Y: 5242, N: 34068	Difficulties in eating	Y: 2965, N: 36345
Difficulties in moving	Y: 3398, N: 35912	Hypertension	Y: 30975, N: 8335
COPD	Y: 9724, N: 29586	Heart failure	Y: 13228, N: 26082
Stroke	Y: 9828, N: 29482	Parkinson	Y: 1655, N: 37655
Thyroid disorders	Y: 4538, N: 34772	Diabetes mellitus	Y: 15910, N: 23400
Gastrointestinal or liver disease	Y: 27401, N: 11909	Musculoskeletal diseases	Y: 24330, N: 14980
Dementia	Y: 4479, N: 34831	Malnutrition	Y: 2718, N: 36592
Pressure ulcers	Y: 1886, N: 37424	Anaemia	Y: 12546, N: 26764
Hear impairment	Y: 6777, N: 32533	Gastrointestinal problems	Y: 12567, N: 26743
Chronic renal failure	Y: 8679, N: 30631	Depression	Y: 587, N: 38723
Cancer	Y: 16536, N: 22774	Constipation	Y: 5088, N: 34222
Atrial fibrillation	Y: 12434, N: 26876	Visual impairment	Y: 20100, N: 19210
Psychiatric disease	Y: 19436, N: 19874		

Finally, we stratified the FI into four categories according to the work of Hoover et al.[35] and aggregated the two less severe frailty conditions (Non-Frail + Vulnerable) and the two more frail statuses (Frail + Most Frail). Variables used in the FI are listed in Table 2 and were extracted as part of the original 147 variables.

Data censoring and distributions

After data censoring, the 1-year mortality target variable distribution was: 24,985 (65.83%) episodes were negative cases (time to exitus > 365 days) and 13,431 (34.17%) episodes were positive (time to exitus ⩽ 365 days) as shown in Figure 2(a). The survival regression target variable (20,959 episodes; mean 368.59; range [0 to 3033]) presents a right-skewed shape, as can be observed in its density plot in Figure 2(b).

The admission FI (mean 0.27; std 0.12) and the FI target variable (22,859 episodes; mean 0.32; std 0.14), resembled a slightly skewed normal distribution (plot in Figure 2(c) and Figure 2(d)), while the distributions of the different categories are: Non-Frail 986 (2.2%), Vulnerable 10,911 (24.34%), Frail 25,638 (57.19%), and Most Frail 7294 (16.27%). As aggrupation of two categories: Non-Frail + Vulnerable 11,897 (26.54%), Frail + Most Frail 32,932 (73.46%), data represented in Figure 2(e).

Methods

Predictive models

As the first approach for predictive models, we have selected the Gradient Boosting Machines (GBM),[36] which can be used for classification and regression. Gradient Boosting Machines are ensemble models composed of decision trees. This model follows an iterative training algorithm. In each step, the tree that minimises the selected loss function is added to the ensemble until the
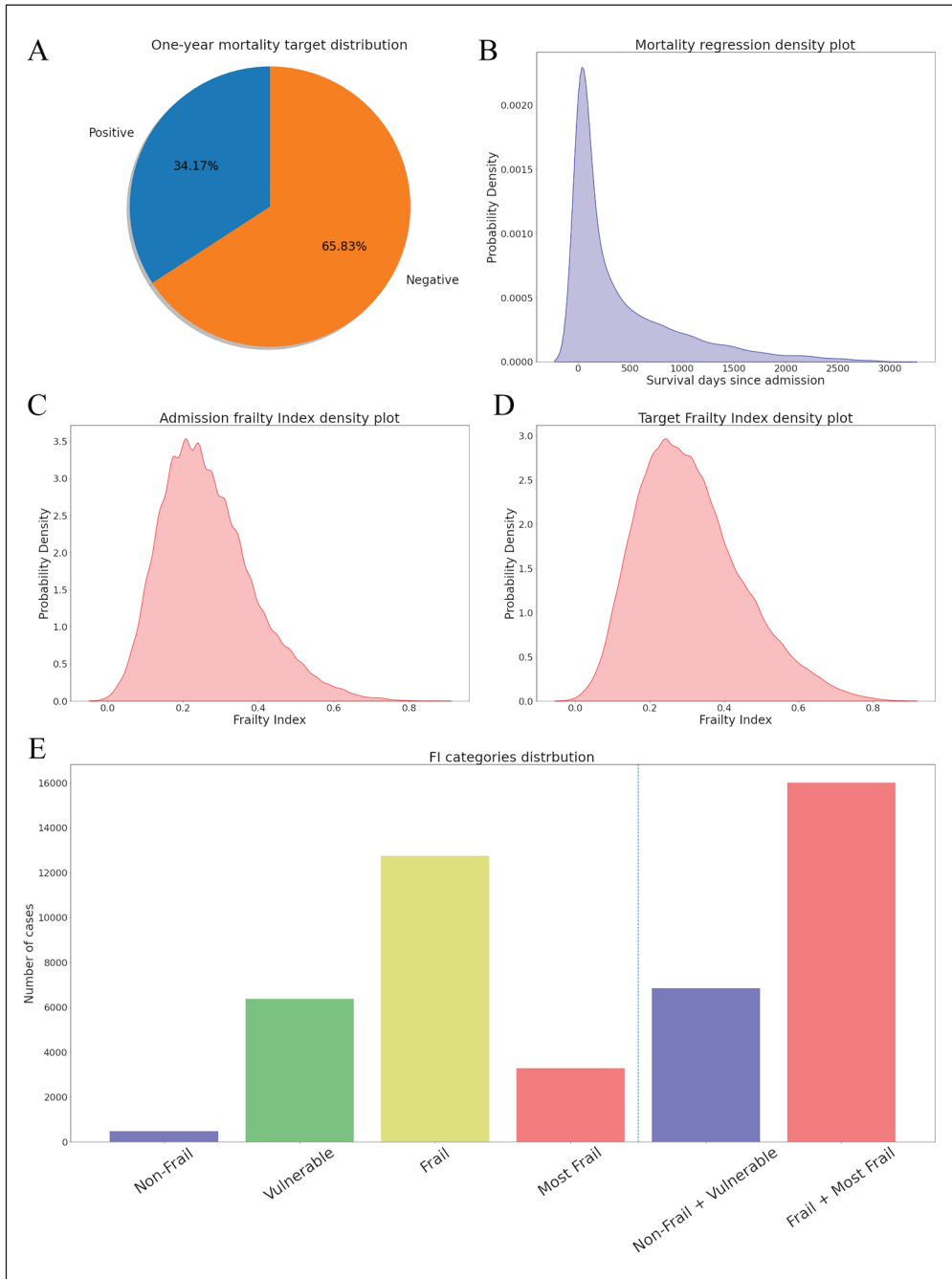
Figure 2. (A) One-year mortality target distribution; (B) Density plot from survival regression target variable; (C) Density plot from the FI target variable; (D) Density plot from the admission FI; (E) FI categories distribution.

hyperparameter setting the number of trees is reached. The GBM models are known for their notable performance on different problems.[37,38,39]

Our second approximation to the predictive models is through the Deep Neural Network (DNN).[40] Due to the tabular nature of the data, we are using a multilayer perceptron topology, which is composed of interconnected neurons. Weights connect the neurons, and their output is a function of the sum of the inputs to the neuron, applying a non-linear activation function afterwards [41]. Our models are using Batch Normalisation[42] and Dropout[43] as regularisation methods and the Leaky ReLU[44] function as activation function. Deep learning has been a trendy technology when dealing with the increasing amount of data, and its application to medicine is growing.[45]

Hyperparameters and variable selection

To select the hyperparameters and make the selection of variables, we split the datasets (80%/20%) into a design set and an evaluation set. Then, we used a recursive feature elimination process as a filter method on the design set. This process starts with the whole set of features, trains a tree-based model and calculates each variable's relevance using the Gini importance,[46] which measures the average gain of purity in the tree splits. Finally, less relevant variables are eliminated. The process is repeated until the desired number of features is obtained. The number of variables was set to 20 in each task, a number of variables able to be handled by a human operator, with two variables eliminated each iteration. Table 7.

The selection of hyperparameters for each model was performed using the Optuna optimisation library.[47] Using this approach, we selected the most relevant hyperparameters for the GBM and the DNN and provided feasible ranges. During the process, the method selects a value for each hyperparameter, trains the model with 80% of the design set, and evaluates it with the remaining 20% and the appropriate metric. As more iterations occur, Optuna makes a smarter selection of the hyperparameters until the algorithm reaches a selected number of iterations. The hyperparameters used in each model can be consulted in Table 8.

Evaluation

We used the bootstrap technique[48] to evaluate the models with 1000 resamples on the unseen evaluation set. To evaluate the performance of the 1-year mortality and the frailty binary classifier, we selected the following metrics: area under Receiver Operating Characteristic curve (AUC ROC), accuracy, sensitivity (or True Positive Rate) and specificity (or True Negative Rate). We selected the mean absolute error (MAE) for the survival regression model. In addition, we repeated the regression experiments using only those cases where the prediction is < 500 days. In addition, since the GBM is an explicable model, we reported the contribution of each variable in percentage.

Comparison with baseline models

To compare our mortality regression model with state of the art, we have performed survival analysis over the data processed with the same pipelines described above. We chose the Cox regression model,[49] from which we obtained survival estimations for patients by calculating the survival expected time. We trained a binary Logistic Regression to compare the classification models for both mortality and frailty.

Software

The whole experimentation described in this work has been carried using the python 3 programming language,[50] and the following scientific libraries and packages: numpy as the main mathematical library,[51] pandas' data frames to handle the data representation,[52] scikit-learn's implementation of GBM,[53] Pytorch's DNN implementation,[54] Optuna as hyperparameters selection[47] and lifelines' implementation of the Cox model.[55]

Results

Associations between distributions

The Spearman's correlation coefficient between the survival target in days and the admission FI was −0.10 while the correlation between survival and the target FI was −0.16; both correlations were statistically significant ($p < .001$). The similarity between the binary 1-year mortality target and the binary FI target was studied using the Chi-Squared test. However, we had to reject the null independence hypothesis ($p < .001$), and therefore it exists a similarity between both binary variables.

One-year mortality classifier

Gradient Boosting Machine and DNN performed very closely (0.87 CI 95% [0.86, 0.87] and 0.86 CI 95% [0.85, 0.86] AUC ROC), both outperforming the logistic regression baseline, complete results and metrics on Table 3.

Survival regression

The cox regression produced a MAE of 444.8 days while the GBM and the DNN model achieved an MAE of 333.13 and 338.88 days, respectively. The GBM outperformed the other models when using only samples with survival < 500, complete performance for survival regression models on Table 4.

Table 3. One-year mortality classifier evaluation. Reporting the mean and the 95% confidence interval.

Model	AUC ROC	Sensitivity (TPR)	Specificity (TNR)	Accuracy
GBM	0.87 [0.86, 0.88]	0.78 [0.76, 0.82]	0.79 [0.75, 0.81]	0.79 [0.77, 0.80]
DNN	0.86 [0.85, 0.86]	0.79 [0.74, 0.83]	0.76 [0.71, 0.81]	0.77 [0.75, 0.79]
Log. Reg.	0.80 [0.79, 0.81]	0.75 [0.63, 0.8]	0.69 [0.64, 0.81]	0.71 [0.69, 0.75]

Table 4. Mortality regressor evaluation. Reporting the mean and the 95% confidence interval.

Model	MAE	MAE (<500d)
GBM	333.13 [323.10, 342.49]	94.67 [92.02, 97.49]
DNN	338.88 [329.07, 349.37]	103.21 [100.47, 106.08]
Cox	444.8 [438.9, 450.9]	116.71 [115.23, 118.08]

Table 5. Frailty classifier evaluation.

Model	AUC ROC	Sensitivity (TPR)	Specificity (TNR)	Accuracy
GBM	0.89 [0.88, 0.90]	0.77 [0.73, 0.81]	0.85 [0.81, 0.89]	0.79 [0.78, 0.81]
DNN	0.89 [0.88, 0.90]	0.76 [0.72, 0.83]	0.85 [0.78, 0.89]	0.79 [0.77, 0.82]
Logistic reg.	0.84 [0.83, 0.85]	0.74 [0.70, 0.78]	0.78 [0.73, 0.83]	0.75 [0.73, 0.77]

One-year frailty classifier

The classification model based on the logistic regression achieved an AUC ROC of 0.84, while the GBM and DNN outperformed it with an AUC ROC of 0.89. Complete metrics for the frailty classification are available in Table 5.

Gini Importances

Following the previous methodology, we have calculated the Gini importance for each of the GBM predictive models. For the 1-year mortality model, the most important variables were: Number of Active Groups, Charlson Index and Age. In the regression task: Number of Active Groups, Charlson Index and Service whereas in the model version including only cases with survival < 500 days were: Leukocytes, C-reactive protein and Urea. Finally, the most relevant features in the frailty model were the Charlson Index, Number of previous Emergency Room visits and Hypertension. Complete details are in Table 6.

Discussion

The overall aim of this study was to develop machine learning models capable of making predictions about mortality and frailty focussed on older adults so that health professionals can benefit from quantitative approaches based on data-driven evidence. We have developed an ML model to predict frailty status within the year without using other problems as proxies. Regarding the mortality criterion, and despite different approximations to this task in the literature, we decided to focus on older patients to be more specific within this age group.

Our 1-year mortality model ranked among the best general admission models in terms of AUC ROC (0.87 CI 95% [0.86, 0.88]). Outperforming PROFUND (0.77),[23] scoring slightly below HOMR (0.89–0.92),[24] mHOMR (0.89)[26] and our previous work.[27] However, the results are in the same range as Avati's deep learning approach (0.93, 0.87 for admitted only patients).[25] However, our model is not fully comparable since it targeted older adults ($\geqslant$ 65 years old); meanwhile, all the mentioned studies use inclusion criteria of $\geqslant$ 18, except Avati, which includes paediatric records. Yourman et al.[56] reviewed prognosis indices for older patients, where the better AUC ROC for the 1-year index was 0.83, which is below our lower 95% CI bound. The authors believe that excluding younger and possibly healthier patients from the sample made the problem more difficult and negatively affected the metrics. This is the case of our previous work[27] which used data from the same hospital but reported better results using the whole adult population. As expected, the GBM model performed significantly better than the Logistic Regression counterpart and slightly better than the DNN model.

Our survival regression model scored a MAE of 329.97 days, outperforming the 444.8 days scored by the cox model. Despite obtaining better predictions than one of the most used models when dealing with survival time, a mean error of almost a year does not adequately meet this

Table 6. Gini importance of the GBM for mortality and frailty tasks. Variables are sorted using the sum of the Gini importances in all tasks.

Variable	Gini 1YM (%)	Gini Reg. (%)	Gini Reg. < 500 (%)	Gini frailty (%)
Charlson Index	**14.45**	**8.40**	1.65	**29.86**
Number active groups	**16.28**	**12.97**	3.24	-
Service	7.66	**10.46**	8.37	2.05
Leukocytes	5.58	6.78	**14.41**	0.60
Age	**9.36**	5.83	3.25	4.51
Barthel Index	6.23	6.09	5.43	5.10
Urea	5.28	4.62	**9.62**	0.83
Number previous ER	1.04	4.42	2.25	**10.9**
C-reactive protein	2.68	4.45	**10.31**	-
RDW-SD	4.80	3.20	4.77	2.28
DRG	3.89	4.05	4.92	1.16
Admission diagnose code	1.78	7.26	2.96	0.63
Glucose	2.20	2.18	5.89	1.63
RDW-CV	2.80	2.88	3.60	2.00
Creatinine	2.42	2.57	3.49	1.65
Number of previous stays	1.60	2.56	5.20	0.72
Hypertension	-	-	-	**9.67**
Haematocrit	2.15	1.56	3.61	1.85
Filtered glomerular CKD	-	7.41	1.24	-
Psychiatric disease	-	-	-	8.29
Atrial fibrillation	-	-	-	7.92
Gastrointestinal or liver disease	-	-	-	7.59
Potassium	1.38	1.27	4.12	0.76
Metastatic tumour	5.35	-	-	-
Sodium	3.07	-	-	-
Number previous ER 365d	-	1.04	1.67	-

model's original purpose. When removing cases where survival time is longer than 500 days, the GBM performs better than the other models achieving a mean error of 94.67 days; this improves the prediction error and will be likely better accepted by the healthcare professionals. This improvement in the predictive power is likely due to removing the long tail in the distribution that includes infrequent values and outliers. It would also be possible to train a model using cases where survival was less than 365 days. In this case, the model would be used only when the 1-year mortality produces a positive result; a preliminary result using the GBM configuration produced an MAE of 69.89 CI 95% [67.83, 72.08]. A further study concerning healthcare experts' preferences is needed to know if this alternative is preferred over the standard approximation.

The 1-year frailty model scored a 0.89 AUC ROC on GBM and DNN, outperforming the logistic regression version (0.84 AUC ROC). These results demonstrate a significant predictive power for assessing a patient's FI category 1 year from admission. As far as the authors know, this is the first study where a model is used to predict a future frailty status without using proxies such as mortality or disability. These models use variables containing information about the current frailty status combined with other factors such as the previous stays in the emergency room or the age to determine the future frailty status. Since most of the variables are shared with the other two

Table 7. Variables used in the predictive models and their descriptions.

Variable	Description
Admission diagnose code	ICD9 code representing the main reason for the admission
Age	Patient's age
Atrial fibrillation	ICD9 diagnosis code: Atrial fibrillation (no/yes)
Barthel index	Barthel Index is an ordinal scale used to measure performance in activities of daily living (ADL). Ten variables describing ADL and mobility are scored, a higher number being a reflection of greater ability to function independently following hospital discharge.
Charlson index	The charlson comorbidity index predicts the 1-year mortality for a patient who may have a range of comorbid conditions, such as heart disease, AIDS, or cancer (a total of 17 conditions: Acute myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic lung disease, mild liver disease, mild to moderate diabetes, diabetes with chronic complications, hemiparaplegia or paraplegia, kidney disease, malignant tumours, moderate to serious liver disease, solid, metastatic tumour and AIDS). Each condition is assigned a score of 1, 2, 3 or 6, depending on the risk of dying associated with each one.
Creatinine	Lab result expressed in mg/dL
DRG	Diagnosis-related group (DRG) is a system to classify hospital cases into one of originally 467 groups
Filtered glomerular CKD	Filtered glomerular CKD lab result in ml/min/1,73 m^2
Gastrointestinal or liver disease	ICD9 diagnosis code: Gastrointestinal or liver disease (no/yes)
Glucose	Lab result expressed in mg/dL
Haematocrit	Lab result expressed in %
Hypertension	ICD9 diagnosis code: Hypertension (no/yes)
Leucocyte	10^3/microL
Number active groups	Number of active groups (medications) in each episode
Number of previous stays	Number of previous hospital admissions
Number previous ER 365d	Number of previous emergency room visits (last 365 days)
Number previous ER	Number of previous emergency room visits
Metastatic tumour	ICD9 diagnosis code: Metastatic tumour (no/yes)
PCR	C-reactive protein lab result expressed in mg/L
Potassium	Lab result expressed in mEq/L
Psychiatric disease	ICD9 diagnosis code: Psychiatric disease (no/yes)
RDW-CV	The red cell distribution width (RDW) blood test measures the amount of red blood cell variation in volume and size. This values is the coefficient of variation of RDW
RDW-SD	Standard deviation of RDW measure
Service	Last service updated during the stay
Sodium	Lab result expressed in mEq/L
Urea	Lab result expressed in mg/dL

Table 8. Hyperparameters selected by Optuna. The non-specified hyperparameters have the default value defined in their libraries: scikit-learn v1.0 for the GBM and Pytorch v1.9.1 for the DNN.

Task	Model	Parameters	
1ym	GBM	Criterion	Friedman MSE
		Max depth	5
		Max features	Auto
		N Estimators	291
	DNN	Learning Rate	0.017324716287571128
		Epochs	50
		Activation Function(s)	Leaky ReLU
		Final function	Softmax
		Batch norm	Yes, every layer
		Layer 1 size	512
		Layer 1 dropout	0.45
		Layer 2 size	256
		Layer 2 dropout	0.40
		Layer 3 size	512
		Layer 3 dropout	0.25
		Layer 4 size	512
		Layer 4 dropout	0.34
		Layer 5 size	256
		Layer 5 dropout	0.3
Regression	GBM	Criterion	MSE
		Max depth	5
		Max features	Auto
		N Estimators	286
	DNN	Learning Rate	0.0009571160666083575
		Epochs	30
		Activation Function(s)	Leaky ReLU
		Final function	ReLU
		Batch norm	Yes, every layer
		Layer 1 size	256
		Layer 1 dropout	0.23
		Layer 2 size	64
		Layer 2 dropout	0.35
		Layer 3 size	256
		Layer 3 dropout	0.29
		Layer 4 size	64
		Layer 4 dropout	0.
		Layer 5 size	128
		Layer 5 dropout	0.49
		Layer 6 size	512
		Layer 6 dropout	0.25

Table 8. (Continued)

Task	Model	Parameters	
Frailty	GBM	Criterion	MSE
		Max depth	4
		Max features	SQRT
		N Estimators	149
		Learning Rate	1.301440136399707e-05
		Epochs	100
		Activation Function(s)	Leaky ReLU
		Final function	Softmax
		Batch norm	Yes, every layer
		Layer 1 size	512
		Layer 1 dropout	0.5
		Layer 2 size	128
		Layer 2 dropout	0.284
		Layer 3 size	64
		Layer 3 dropout	0.21
		Layer 4 size	16
		Layer 4 dropout	0.44

mortality models, the addition of a few extra variables means that we can obtain a prediction regarding the patient's health decay with a low extra effort.

Each model was set with the 20 most relevant variables from a total of 147, a number that was arguably too high to be used by a human operator. This selection was performed using the Random Forest's Gini importance criteria with recursive feature elimination as a data-driven method. This method is known to have a favourable bias towards categorical variables with many categories and continuous variables. However, it is widely used because it is fast and straightforward to compute.[46] In the end, all three models share a great number of variables (Table 6), being only 26 different variables. The selected variables by the recursive feature elimination algorithm are coherent with the different mortality works in the literature.[23,24] In addition, this final set of variables can be obtained easily a few hours after admission, where the first diagnosis and laboratory tests are performed.

These results provide a complementary perspective based on an objective measure of frailty to initiate early PC. The mean admission FI was 0.27 ± 0.12, and its shape resembles a normal distribution. This is a coherent behaviour with the findings in the Mitnitski et al. study,[14] where the most impaired groups have a bigger FI mean, and the distribution is shaped like a normal distribution, as opposed to the less impaired groups, which had a smaller mean FI and can be approximated using a gamma distribution. The correlation between our admission FI and MR target in days is -0.10, lower than the one reported in ref. 14, which was -0.234. This means that the FI used in this work for this sample is less associated with mortality. However, the Chi-Squared test performed on both binary targets discarded the hypothesis of independence, so in our sample, we can confirm a weak association between both criteria.

The relationship between frailty and mortality have been studied previously,[29] pointing to the association between both. Despite the similarity in the input variables, the target variable distributions are poorly correlated and have different shapes. Both criteria have been highlighted as

important for accessing PC in previous studies and are related. However, they reflect two different distributions, and the authors think of them as two complementary criteria. Therefore, we conclude that the best approximation for taking advantage of both mortality and frailty criteria is to have different predictive models working simultaneously, increasing the information to support the decision-making process. The incorporation of the frailty criterion may represent an added value for those health professionals deciding about inclusion in PC services. This is in line with Almagro et al. (2017),[57] showing that poor vital prognosis as the sole criterion for initiating PC among COPD patients should be critically appraised.

This study's clinical impact resides in the potential to predict adverse outcomes for hospital admitted patients within the following year. First, we choose 1 year as a horizon to make the mortality prediction; as stated elsewhere,[25] longer than 12 months is not desirable due to the difficulty in the predictions and the limited resources of the programmes, which are better to focus on immediate needs. Thus, referral to PC may be focused on immediate needs. Also, despite being more difficult to predict, the information provided by the survival regression model may help contextualise the 1-year mortality model results. Therefore, healthcare professionals would be supported with additional information such as the magnitude of the remaining time until death in days, weeks or months. Including these models into clinical practice could help anticipate the decline in admitted patients, allowing healthcare professionals to allocate scarce resources to patients who will need them the most.

The main contribution of this work is the development of the frailty predictive model, which is a novel approach to try to identify patients in need of ACP. This frailty approach complements the more traditional mortality approach, which we also tried to enrich by adding 1-year mortality classification and regression to provide more information to healthcare experts during the decision-making process without providing excessive extra information burden. The three models were implemented as an online Clinical Decision Support System[58] available to any healthcare expert for academic use until further validations at.[59] Besides, we have demonstrated the complementariness of the mortality and frailty models testing the low correlation between both factors in our dataset, so we should treat them as complementary criteria.

The main limitation of this study is the use of data from only one hospital. Therefore, internal validation only assures the performance of the models with similar data. We cannot ensure the reported efficiency in other hospitals and other patient populations.[60] Also, data from the same centres can change over time for various reasons, such as a change in protocols or external agents such as a pandemic.[61,62] Additional external validations are needed for future work. Broader populations can be approached by implementing predictive models using Electronic Health Records (EHR), supporting an effective identification of patients needing further specialised care.[63] Thus, besides external validation of the models, future authors' work will require significant software development and implementation project to connect these systems with hospital EHR and avoid manual input by professionals. Also, the maturity of the models and the software wrapping them needs to be *field-tested* before their inclusion as a standard tool to the hospital information system.

Conclusion

This work proposes using three different machine learning models based on hospital admission data to assess the PC needs of older adults and help healthcare professionals in the decision-making process. The authors constructed three different but complementary predictive systems: a 1-year mortality model, a regression mortality model to provide more information about the first prediction and a 1-year frailty model. Previous modern mortality models are using machine learning

methods available elsewhere, but they are not specifically focused on older populations. Also, to our knowledge, this is the first study predicting 1-year frailty status based on an FI. As previous studies have shown, mortality and frailty could be relevant criteria to admit patients to PC programmes. Therefore, health professionals could benefit from using data-driven accurate predictions of these two dimensions on patients over 65. In addition to the benefits experienced by patients and their families, the early identification of these patients' needs can help better manage the available health and social care resources and reduce costs overall. Consequently, the authors propose using predictions in both mortality and frailty as complementary predictions to help assess PC needs due to its relevance but weak correlation, reliability and great predictive power. The described models have been implemented and publicly available for the academic purpose at.[59]

Declaration of conflicting interests

Ethics

The data used in this study comes from the University and Polytechnic La Fe Hospital of Valencia and was retrospectively collected from the Electronic Health Records (EHR) of the hospital. This procedure was assessed and approved by the Ethical Committee of the University and Polytechnic La Fe Hospital of Valencia (registration number: 2019-88-1). Required patient patient-informed consent was waived. All methods were performed in accordance with the relevant guidelines and regulations.

ORCID iDs

Vicent Blanes-Selva https://orcid.org/0000-0002-0056-0329
Gordon Linklater https://orcid.org/0000-0003-3907-2602

References

1 Callaway MV, Connor SR and Foley KM. World health organization public health model: a roadmap for palliative care development. *J Pain Symptom Management* 2018; 55(2): S6–S13.

2 Temel JS, Greer JA and Muzikansky A, et al. Early palliative care for patients with metastatic non–small-cell lung cancer. *New Engl J Med* 2010; 363(8): 733–742.

3 Bakitas M, Lyons KD and Hegel MT, et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomised controlled trial. *JAMA* 2009; 302(7): 741–749.

4 Yennurajalingam S, Urbauer DL and Casper KLB, et al. Impact of a palliative care consultation team on cancer-related symptoms in advanced cancer patients referred to an outpatient supportive care clinic. *J Pain Symptom Manage* 2011; 41(1): 49–56.

5 Quinn KL, Stukel T and Stall NM, et al. Association between palliative care and healthcare outcomes among adults with terminal non-cancer illness: population based matched cohort study. *BMJ* 2020; 370: m2257.

6   Bakitas MA, Tosteson TD and Li Z, et al. Early versus delayed initiation of concurrent palliative oncology care: patient outcomes in the ENABLE III randomised controlled trial. *J Clin Oncol* 2015; 33(13): 1438–1445.

7   McIlfatrick S. Assessing palliative care needs: views of patients, informal carers and healthcare professionals. *J Adv Nurs* 2007; 57(1): 77–86.

8   Addington-Hall JM and Higginson IJ. *Palliative care for non-cancer patients*. Oxford: Oxford University Press, 2001.

9   Kingston AE, Kirkland J and Hadjimichalis A. Palliative care in non-malignant disease. *Medicine* 2020; 48(1): 37–42.

10   Etkind SN, Bone AE and Gomes B, et al. How many people will need palliative care in 2040? Past trends, future projections and implications for services. *BMC Med* 2017; 15(1): 102.

11   Murray SA, Firth A and Schneider N, et al. Promoting palliative care in the community: production of the primary palliative care toolkit by the European association of palliative care taskforce in primary palliative care. *Palliat Med* 2015; 29(2): 101–111.

12   Clegg A, Young J and Iliffe S, et al. Frailty in elderly people. *Lancet* 2013; 381(9868): 752–762.

13   Fried LP, Tangen CM and Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol Ser A: Biol Sci Med Sci* 2001; 56(3): M146–M157.

14   Mitnitski AB, Mogilner AJ and Rockwood K. Accumulation of deficits as a proxy measure of aging. *Scientific World J* 2001; 1: 323–336.

15   Gobbens RJJ, Luijkx KG and Wijnen-Sponselee MT, et al. In search of an integral conceptual definition of frailty: opinions of experts. *J Am Med Directors Assoc* 2010; 11(5): 338–343.

16   Raudonis BM and Daniel K. Frailty: an indication for palliative care. *Geriatr Nurs* 2010; 31(5): 379–384.

17   Koller K and Rockwood K. Frailty in older adults: implications for end-of-life care. *Cleveland Clinic J Med* 2013; 80(3): 168–174.

18   Moss AH, Ganjoo J and Sharma S, et al. Utility of the "surprise" question to identify dialysis patients with high mortality. *Clin J Am Soc Nephrol* 2008; 3(5): 1379–1384.

19   Downar J, Goldman R and Pinto R, et al. The "surprise question" for predicting death in seriously ill patients: a systematic review and meta-analysis. *Can Med Assoc J* 2017; 189(13): E484–E493.

20   Linklater G, Lawton S and Fielding S, et al. Introducing the palliative performance scale to clinicians: the Grampian experience. *BMJ Support Palliat Care* 2012; 2(2): 121–126.

21   Highet G, Crawford D and Murray SA, et al. Development and evaluation of the supportive and palliative care indicators tool (SPICT): a mixed-methods study. *BMJ Support Palliat Care* 2014; 4(3): 285–290.

22   Woolfield A, Mitchell G and Kondalsamy-Chennakesavan S, et al. Predicting those who are at risk of dying within six to twelve months in primary care: a retrospective case control general practice chart analysis. *J Palliative Med* 2019; 22(11): 1417–1424.

23   Bernabeu-Wittel M, Ollero-Baturone M and Moreno-Gaviño L, et al. Development of a new predictive model for polypathological patients. The PROFUND index. *Eur J Int Med* 2011; 22(3): 311–317.

24   van Walraven C, McAlister FA and Bakal JA, et al. External validation of the hospital-patient one-year mortality risk (HOMR) model for predicting death within 1 year after hospital admission. *CMAJ* 2015; 187(10): 725–733.

25   Avati A, Jung K and Harman S, et al. Improving palliative care with deep learning. *BMC Med Informat Decision Mak* 2018; 18(4): 122.

26   Wegier P, Koo E and Ansari S, et al. mHOMR: a feasibility study of an automated system for identifying inpatients having an elevated risk of 1-year mortality. *BMJ Quality Safety* 2019; 28(12): 971–979.

27   Blanes-Selva V, Ruiz-García V and Tortajada S, et al. Design of 1-year mortality forecast at hospital admission: a machine learning approach. *Health Informat J* 2021, DOI: 10.1177/1460458220987580.

28   Porter AS, Harman S and Lakin JR. Power and perils of prediction in palliative care. *Lancet* 2020; 395(10225): 680–681.

29   Shamliyan T, Talley KMC and Ramakrishnan R, et al. Association of frailty with survival: a systematic literature review. *Ageing Res Rev* 2013; 12(2): 719–736.

30 Babič F, Majnarić LT and Bekić S, et al. Machine learning for family doctors: a case of cluster analysis for studying aging associated comorbidities and frailty. In: *International cross-domain conference for machine learning and knowledge extraction*. Cham: Springer, 2019, pp. 178–194.

31 Sternberg SA, Bentur N and Abrams C, et al. Identifying frail older people using predictive modeling. *Am J Managed Care* 2012; 18(10): e392–e397.

32 Saliba D, Elliott M and Rubenstein LZ, et al. The vulnerable elders survey: a tool for identifying vulnerable older people in the community. *J Am Geriatr Soc* 2001; 49(12): 1691–1699.

33 Bertini F, Bergami G and Montesi D, et al. Predicting frailty condition in elderly using multidimensional socioclinical databases. *Proc IEEE* 2018; 106(4): 723–737.

34 Searle SD, Mitnitski A and Gahbauer EA, et al. A standard procedure for creating a frailty index. *BMC Geriatrics* 2008; 8(1): 24.

35 Hoover M, Rotermann M and Sanmartin C, et al. Validation of an index to estimate the prevalence of frailty among community-dwelling seniors. *Health Rep* 2013; 24(9): 10–17.

36 Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist* 2001; 29(5): 1189–1232.

37 Touzani S, Granderson J and Fernandes S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energ Build* 2018; 158: 1533–1543.

38 Chen X, Huang L and Xie D, et al. EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. *Cel Death Dis* 2018; 9(1): 1–12.

39 Zhou J, Li E and Yang S, et al. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Saf Sci* 2019; 118: 505–518.

40 Hinton GE. Connectionist learning procedures. In: *Machine learning*. San Mateo, CA: Morgan Kaufmann, 1990, pp. 555–610.

41 Gardner MW and Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ* 1998; 32(14–15): 2627–2636.

42 Ioffe S and Szegedy C. Batch normalisation: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, Lille, France, 6 Jul–11 Jul, 2015, PMLR; 2015, pp. 448–456.

43 Srivastava N, Hinton G and Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; 15(1): 1929–1958.

44 Maas AL, Hannun AY and Ng AY. Rectifier nonlinearities improve neural network acoustic models. *Proc Icml* 2013; 30(1): 3.

45 Piccialli F, Somma VD and Giampaolo F, et al. A survey on deep learning in medicine: why, how and when? *Inf Fusion* 2021; 66: 111–137.

46 Nembrini S, König IR and Wright MN. The revival of the Gini importance? *Bioinformatics* 2018; 3421: 3711–3718.

47 Akiba T, Sano S and Yanase T, et al. Optuna: a next-generation hyperparameter optimisation framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, Anchorage, AK, USA, 4–8 August, 2019, 2019.

48 Efron B and Robert J. *Tibshirani. An introduction to the bootstrap*. Boca Raton: CRC Press, 1994.

49 Cox DR. Regression models and life-tables. *J R Stat Soc Ser B* 1972; 34(2): 187–202.

50 Van Rossum G and Drake FL. *The python language reference manual*. Bristol: Network Theory Ltd., 2011.

51 van der Walt S, Colbert SC and Varoquaux G. The NumPy array: a structure for efficient numerical computation. *Comput Sci Eng* 2011; 13(2): 22–30.

52 McKinney W. Data structures for statistical computing in python. *Proc 9th Python Sci Conf* 2010; 445: 51–56.

53 Pedregosa F, Varoquaux G and Gramfort A, et al. Scikit-learn: machine learning in Python. *The J Machine Learn Research* 2011; 12: 2825–2830.

54 Paszke A, Gross S and Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inform Processing Sys* 2019; 32: 8026–8037.

55 Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw* 2019; 4(40): 1317.

56 Yourman LC, Lee SJ and Schonberg MA, et al. Prognostic indices for older adults: a systematic review. *JAMA* 2012; 307(2): 182–192.

57 Almagro P, Yun S and Sangil A, et al. Palliative care and prognosis in COPD: a systematic review with a validation cohort. *Int J Chron Obstructive Pulmon Dis* 2017; 12: 1721–1729.

58 Hajioff S. Computerized decision support systems: an overview. *Health Informat J* 1998; 4(1): 23–28.

59 *Palliative Care Models Webapp (Demo Aleph)*. http://demoiapc.upv.es/ (Accessed 27 August, 2021)

60 Sáez C, Romero N and Conejero JA, et al. Potential limitations in COVID-19 machine learning due to data source variability: a case study in the nCov2019 dataset. *J Am Med Inform Assoc* 2020; 28: 360–364.

61 Sáez C and García-Gómez JM. Kinematics of big biomedical data to characterise temporal variability and seasonality of data repositories: functional data analysis of data temporal evolution over non-parametric statistical manifolds. *Int J Med Informat* 2018; 119: 109–124.

62 Sáez C, Gutiérrez-Sacristán A and Kohane I, et al. EHRtemporalVariability: delineating temporal dataset shifts in electronic health records. *GigaScience* 2020; 9(8): giaa079.

63 Jung K, Sudat SEK and Kwon N, et al. Predicting need for advanced illness or palliative care in a primary care population using electronic health record data. *J Biomed Informat* 2019; 92: 103115.