

Machine learning approaches for the mortality risk assessment of patients undergoing hemodialysis

Cheng-Hong Yang¹, Yin-Syuan Chen, Sin-Hua Moi, Jin-Bor Chen², Lin Wang and Li-Yeh Chuang

Abstract

Introduction: Mortality is a major primary endpoint for long-term hemodialysis (HD) patients. The clinical status of HD patients generally relies on longitudinal clinical observations such as monthly laboratory examinations and physical examinations.

Methods: A total of 829 HD patients who met the inclusion criteria were analyzed. All patients were tracked from January 2009 to December 2013. Taken together, this study performed full-adjusted-Cox proportional hazards (CoxPH), stepwise-CoxPH, random survival forest (RSF)-CoxPH, and whale optimization algorithm (WOA)-CoxPH model for the all-cause mortality risk assessment in HD patients. The model performance between proposed selections of CoxPH models were evaluated using concordance index.

Results: The WOA-CoxPH model obtained the highest concordance index compared with RSF-CoxPH and typical selection CoxPH model. The eight significant parameters obtained from the WOA-CoxPH model, including age, diabetes mellitus (DM), hemoglobin (Hb), albumin, creatinine (Cr), potassium (K), Kt/V, and cardiothoracic ratio, have also showed significant survival difference between low- and high-risk characteristics in single-factor analysis. By integrating the risk characteristics of each single factor, patients who obtained seven or more risk characteristics of eight selected parameters were dichotomized as high-risk subgroup, and remaining is considered as low-risk subgroup. The integrated low- and high-risk subgroup showed greater discrepancy compared with each single risk factor selected by WOA-CoxPH model.

Conclusion: The study findings revealed WOA-CoxPH model could provide better risk assessment performance compared with RSF-CoxPH and typical selection CoxPH model in the HD patients. In summary, patients who had seven or more risk characteristics of eight selected parameters were at potentially increased risk of all-cause mortality in HD population.

Keywords: feature selection, hemodialysis, machine learning, risk assessment, survival analysis

Received: 6 November 2021; revised manuscript accepted: 27 July 2022.

Introduction

Patients with end stage kidney disease (ESKD) require long-term renal replacement therapy, such as hemodialysis (HD) and peritoneal dialysis. However, several clinical factors have been reported to influence long-term HD treatment outcomes such as serum albumin levels.¹

In addition, the HD treatment also becomes an economic burden to the reimbursement for medical care.^{2,3} Therefore, HD-related diagnosis, care planning, and prevention have become critical research issues. The all-cause mortality is commonly considered as a primary endpoint for patients undergoing long-term HD in the clinical

Ther Adv Chronic Dis

2022, Vol. 13: 1–17

DOI: 10.1177/
20406223221119617

© The Author(s), 2022.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Sin-Hua Moi
Center of Cancer Program
Development, E-Da
Cancer Hospital, I-Shou
University, Kaohsiung
82445.

moi9009@gmail.com

Jin-Bor Chen
Department of Neurology,
Kaohsiung Chang Gung
Memorial Hospital, Chang
Gung University College
of Medicine, Kaohsiung
83301.

chenjb1019@gmail.com

Lin Wang
Department of
Nephrology, Dalian
University Affiliated Xinhua
Hospital, Dalian, 116001,
China.

wanglinxiuhua@126.com

Li-Yeh Chuang
Biotechnology and
Chemical Engineering,
I-Shou University,
Kaohsiung 84004.
chuang@isu.edu.tw

Cheng-Hong Yang
Department of Information
Management, Tainan
University of Technology,
Tainan

Department of Electronic
Engineering, National
Kaohsiung University of
Science and Technology,
Kaohsiung

Biomedical Engineering,
Kaohsiung Medical
University, Kaohsiung

School of Dentistry,
Kaohsiung Medical
University, Kaohsiung

Drug Development and
Value Creation Research
Center, Kaohsiung Medical
University, Kaohsiung

Yin-Syuan Chen
Department of Electronic
Engineering, National
Kaohsiung University of
Science and Technology,
Kaohsiung

studies.^{4,5} Previous studies indicated all-cause mortality in HD patient is associated with multiple clinical factors, including comorbidity, medications, nutrition status, and others.^{6–9}

In current clinical medical research, it is no longer possible to consider only univariate association; there are complex interactions between multiple clinical factors and biomarkers for survival outcome in HD population.¹⁰ Therefore, it is necessary to explore the relevant risk factors that affect the disease in a multivariate and comprehensive manner.¹¹ In the era of big data, there may be limits to complete care for all clinical factors and biomarkers using typical statistical methods. Therefore, machine learning methods can be used to quickly find out more significant risk factors for diseases to achieve maintenance and slow down the occurrence of disease exacerbations.¹² In 2017, Ramspek *et al.*¹³ used Cox proportional hazards (CoxPH) model to predict risk of all-cause mortality in dialysis patients.

Multiple machine learning approaches, including artificial neural network,¹⁴ particle swarm optimization,¹⁵ biogeography-based optimization,¹⁶ and other hybrid methods have been widely used in the risk assessment of specific diseases.¹⁷ Comparing with typical statistical approaches, the combination use of machine learning could overcome several limitations faced by statistical methods, including the sample size restriction and computation complexity.¹⁸ In 2022, Radović *et al.*¹⁹ used kernel support vector machine and K-means to determine the expected mortality rate. Random survival forest (RSF) is an exploratory analysis method used to evaluate survival data.²⁰ RSF uses the survival splitting rules for growing survival tree to identify highly impacted risk factor of mortality. Garcia-Montemayor *et al.*²¹ evaluated the prediction performance of random forest and logistic regression in mortality of HD patients.

Whale optimization algorithm (WOA) is a meta-heuristic optimization algorithm, which is inspired by the hunting behavior of humpback whales.²² WOA uses exploration and exploitation capabilities to avoid the local optimum and accelerate the convergence in the optimization procedure of feature selection. WOA is highly compatible with other algorithms and statistical methods.^{23–25} WOA has the advantages of simple process and fast convergence speed, and it was found to be

enough competitive with other state-of-the-art meta-heuristic methods.²² WOA can be used to find the best feature combination in feature selection,²⁶ select the fewest features to obtain the maximum classification accuracy, and has excellent performance in solving optimization problems and applications widely.

The risk assessment of mortality plays important roles in long-term medical care for HD patients.⁵ The combined use of WOA, RSF, and typical survival analysis such as CoxPH model could provide more comprehensive risk assessment outcome.^{27,28} We selected three typical models in each field for comparison. The full-adjusted and stepwise selection model represents the typical statistical method, the RSF selection model represents the permutation method, and the WOA model represents the heuristic optimization algorithm. Therefore, this study aimed to apply the WOA, RSF, and typical statistical selection method in the mortality risk assessment, in purpose to identify the optimal risk characteristics combination for all-cause mortality in HD patients.

Methods

Datasets and statistical analysis

This is a retrospective cohort study. All data were obtained from Kaohsiung Chang Gung Memorial Hospital. A total of 829 patients who received regular outpatient HD therapy (three times a week) before 1 January 2009 were retrospectively included under an approved data protocol (201800595B0). All patients were tracked from 1 January 2009 to 31 December 2013. This study was conducted in accordance with the Declaration of Helsinki. We enrolled all HD patients in the beginning, and then excluded those who have not met the inclusion criteria. A total of 874 patients who received regular outpatient HD treatment (three times a week) were enrolled in the initial phase of study. Then, we selected potential risk factors associated with the risk of mortality in HD patients, and then excluded 45 patients who had loss to follow-up, and 2 patients who had missing values for blood measurements. Finally, a total of 829 patients were included in the analysis, and the study population was divided into two groups according to their mortality status. A total of 633 patients who remained alive within the study period were considered as alive cohort, and 196

patients who expired within study period were considered as died cohort. Hence, the mortality rate of HD cohort was known to be approximately 15–20%. Assume the probability of type I error is 0.05 (α), the power is 80%, the population size of current study cohort is 829, the proportion of mortality of Taiwan population is 0.20, and the proportion of mortality of current study cohort is 0.295. By using to the power estimation formula, the estimated critical Z value for given α is approximately 4.279 (Φ), which is equal to power of 1. Thus, the post hoc power of sample size is 100%, which indicates the sample size for current study was appropriate for later analyses.

The clinical factors were collected at the initial phase of study, and the comorbidity and blood measurements data were collected for each HD therapy. The survival outcome was tracked using the death registry database of our institution. The baseline characteristics and laboratory measurements in HD patients including dialysis vintage, age, sex, and diabetes mellitus (DM) status were collected.²⁹ In addition, the baseline blood laboratory measurements, including Hb, blood urea

nitrogen (BUN), Cr, K, Ca, P, intact parathyroid hormone (iPTH), ferritin, HD adequacy index, Kt/V urea (Daugirdas), urea reduction ratio (URR), and cardiothoracic ratio by chest x-ray examination were collected at the initial phase of study. The distributions of baseline characteristics and mortality-related risk in HD patients were summarized into median (interquartile range), mean (standard deviation), or frequency (percentage) according to the survival categories. The independent two-sample t -test or chi-square test was used to estimate the difference between alive and died group. Univariate CoxPH regression analysis was used to evaluate the association between all-cause mortality and individual risk factors. In addition, we have sorted out the algorithm and application comparison of mortality risk assessment studies in related HD populations proposed by previous studies. Table 1 is as follows.

WOA feature selection

WOA is a novel nature-inspired meta-heuristic optimization algorithm, which was proposed by

Table 1. Comparison of algorithms and application of mortality risk assessment studies in hemodialysis population.

Study	Algorithms	Application	Model performance
Yang <i>et al.</i> (2022) (Proposed method)	1. CoxPH model 2. Stepwise selection 3. WOA selection 4. RSF selection 5. Kaplan–Meier	Identifying risk factors for mortality in hemodialysis patients using multiple feature selection, including stepwise, WOA, and RSF selection approaches to generate all-cause mortality risk assessment model	Optimal concordance: 1. full-adjusted-CoxPH: 0.7404 2. stepwise-CoxPH: 0.7388 3. RSF-CoxPH: 0.7406 4. WOA-CoxPH: 0.7409 Kaplan–Meier (log-rank test p value) 1. WOA-CoxPH model: Log-rank test p value < 0.001
Radović <i>et al.</i> ¹⁹	1. Kernel SVM algorithm 2. K-means clustering algorithm	Applied SVM method to assess the expected mortality of hemodialysis patients using nine relevant parameters provided by professional nephrologists	Mortality rate prediction is realized with accuracy up to 94.12% and up to 96.77%
Garcia-Montemayor <i>et al.</i> ²¹	1. Logistic regression analysis 2. Random forest 3. AUC	Prediction of mortality of hemodialysis patients at different time points using random forest algorithm and evaluate the prediction performance using AUC	AUC 1. random forest [Δ AUC 0.68–0.73] 2. logistic regression models [Δ AUC 0.007–0.046]
Ramspek <i>et al.</i> ¹³	1. CoxPH model 2. Kaplan–Meier	Predict risk of all-cause mortality in dialysis patients using typical Cox model and Kaplan–Meier methods	C -statistics ranging from 0.710 (interquartile range 0.708–0.711) to 0.752 (interquartile range 0.750–0.753)
AUC, area under curve; CoxPH, Cox proportional hazards; RSF, random survival forest; SVM, support vector machine; WOA, whale optimization algorithm.			

Algorithm 1. Pseudo-code of whale optimization algorithm (WOA)–based feature selection.

Input: N number of whales · T number of iteration with F size of dimension

Output: Optimal whale position.

Initialize X_i^j ($i=1, 2, \dots, N; j=1, 2, \dots, F$).

while ($t < T$)

for (each whale and dimension $\{X_i^j\}$)

if Sigmoid $\{X_i^j\} \geq 0.5$

$Y_i^j = 1$

else

$Y_i^j = 0$

end for

Convert each binary individual Y_i into a feature combination.

Calculating Cox proportional hazard using C-index for evaluation (fitness).

Update X^* if there is a better fit.

for (each whale $\{X_i\}$)

Calculate and Update a, A, C, p and l .

if1 $p < 0.5$ **then**

if2 ($|A| < 1$) **then**

Update position by Eq. (2)

else2 ($|A| \geq 1$)

Choose search agent randomly (X_{rand})

Update position by Eq. (10)

end if2

else if1 ($p \geq 0.5$)

Update position by Eq. (6)

else if1

end for

$t = t + 1$

end while

Mirjalili *et al.*²² This algorithm simulates the humpback whale predation behavior, including exploitation and exploration. The exploitation phase uses a spiral bubble-net attacking method to find a local optimum result, and the exploration phase simulates the prey searching behavior in order to find a global optimum result. WOA feature selection could accelerate convergence to find the optimal solution for the mortality-related risk assessment in HD patients.³⁰ WOA feature selection pseudocode is shown in Algorithm 1.

The algorithm first enters the stage of encircling prey, and subsequently the search in two stages, the first phase is the exploitation phase (spiral bubble-net attacking method), and the second phase is the exploration phase (search for prey). The behavior patterns in the two phases are described in detail below.

Encircling prey is the primary task, WOA assumes the current best solution or target prey, tries to advance toward the best search target and update

the position, and updates the current position with the current best solution. The following equations (1)–(4) describe this behavior:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| \quad (1)$$

$$\vec{X}(t+1) = \vec{x}^*(t+1) - \vec{A} \vec{D} \quad (2)$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (4)$$

where t is the current iteration, X is the current position vector solution, X^* is the best solution of current position vector, $|\cdot|$ is the absolute value, and A and C are coefficient vectors. A and C are defined in equations (3) and (4). r is random vector in the range $[0, 1]$. The current position vector is updated according to equation (2). The values of A affect the range of the area where the current position vector can be moved. The calculation of \vec{a} is shown in equation (5):

$$\bar{a} = 2 - t \frac{2}{\text{MaxIter}} \quad (5)$$

where MaxIter is the maximum number of allowed iterations. According to the above formula, \bar{a} is decreased linearly from 2 to 0 over the iterations. This will reduce the movable range of shrinking encircling mechanism with iteration. $\bar{X}(t)$ and $\bar{X}^*(t)$ can establish a spiral equation to simulate the spiral movement of humpback whales. The mathematical equation (6) is as follows:

$$\bar{X}(t+1) = \bar{D}^l e^{bl} \cdot \cos(2\pi l) + \bar{X}^*(t) \quad (6)$$

The local search capability is used in the exploitation phase. This phase is divided into two behavioral modes. The first is shrinking encircling mechanism, and the second is spiral updating procedure.

Shrinking encircling mechanism. This behavior pattern is achieved by decreasing the value of a in equation (3). The fluctuation range of A is decreased by linear reduction from 2 to 0 during the iteration of a , then A is a random value in the interval $[-a, a]$. A in $[-1, 1]$, and a new search position can be updated between the current position and the current best position.

Spiral updating procedure. $\bar{X}(t)$ and $\bar{X}^*(t)$ are established as a spiral equation to simulate the spiral movement of humpback whales to update the current best position. The spiral updating procedure are shown in equations (6)–(8):

$$\bar{D}^l = \left| \bar{X}^*(t) - \bar{X}(t) \right| \quad (7)$$

$$\bar{X}(t+1) = \begin{cases} \bar{X}^*(t) - \bar{A}\bar{D} & \text{if } p < 0.5 \\ \bar{D}^l \cdot e^{bl} \cdot \cos(2\pi l) + \bar{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (8)$$

where l is a random number in the range $[-1, 1]$, b is a constant defining the spiral's shape, and D is the distance between the whale X and a prey. In order to the behavior pattern of whale shrinking

the enclosing mechanism is simulated, the trigger probability of the two behavior patterns (2) and (6) is set to 50% each. The mathematical equation is as follows:

When $|\bar{A}| < 1$, we use equation (7) to update the position, where p is a random number in $[0, 1]$

The exploration phase is used in global search capability, and the search prey behavior pattern is used. When $|\bar{A}| \geq 1$, it enters the exploration stage, and uses the randomly selected whale method to update the current individual's position vector. Since $|\bar{A}|$ must be greater than 1 at this stage, according to equation (10), the updated position will deviate from the reference whale, thereby the purpose of global search is achieved:

$$\bar{D} = \left| \bar{C} \cdot \bar{X}_{rand} - \bar{X}(t) \right| \quad (9)$$

$$X(\bar{t} + 1) = \bar{X}_{rand} - \bar{A} \cdot \bar{D} \quad (10)$$

\bar{X}_{rand} is a random whale position vector selected from the current population.

WOA was used to solve the feature selection problem which must be converted into binary classification.²⁶ Therefore, the sigmoid function is added to fit the binary trait of mortality-related risk category in HD patients. The position vector of each feature is converted into binary by S function to search for the best feature combination.²⁶ The sigmoid function is shown in equation (11):

$$y^k = \frac{1}{1 + e^{-v_i^k(t)}} \quad (11)$$

$$X_i^d = \begin{cases} 1 & \text{if } S(x_i^k(t+1)) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where S indicates the sigmoid conversion. This study uses 0.5 to make the threshold for binary. $x_i^k(t+1)$ in equation (12) represents that the i th whale is the vector of the k th dimension at the $(t+1)$ iteration, and the value mapped by the S function is converted to an integer of 1 or 0 by

Algorithm 2. Pseudo-code of random survival forest (RSF)-based feature selection.

```

Input: Dataset:  $D = (\tau_i, \delta_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$ ,  $N$  number of trees
Output: Random Survival Forest (RSF)
Initialize: RSF is empty, all  $p$  covariates, mtry: number of variables randomly selected as candidates for
splitting a node,  $B$ : Ensemble size
for  $i$  to 1:  $B$  do
    Draw a bootstrap sample with size  $N$  from  $D$ 
    while node  $d_0 > 0$  unique deaths do
        randomly select  $mtry$  from  $p$ 
        for  $j$  to 1 to  $mtry$  do
            if  $j$ -th survival splitting criterion then
                split internal node into two child nodes
                break;
            end
        end
    end
end
return the ensemble tree of all  $B$  sub-trees grown in the for loop;

```

equation (12). When the dimension k is 1, it means that the feature is selected. Conversely, when k is 0, this feature is not used.

RSF feature selection

Random forest was proposed by Breiman *et al.* (2001)³¹, and mainly applied to classification and regression.²⁰ In 2008, Ishwaran *et al.*³² applied random forest extension to survival analysis, and developed the RSF, which belongs to integrated learning. RSF is a nonparametric tree-based survival analysis algorithm based on the random forest algorithm.³³ RSF feature selection approach evaluates variable importance (VIMP) by simultaneously considering survival time and censor status of study population. RSF uses the survival splitting rules for growing survival tree to identify highly impacted risk factor of all-cause mortality in study population. The pseudocode of RSF feature selection algorithm is shown in Algorithm 2.

In this study, RSF is used in feature selection, and mortality-related risk in HD patients was converted from continuous variables to categorical variable. The purpose is to improve the accuracy of RSF variable selection. VIMP, which was calculated to filter select variables, can be ranked and screened. A larger VIMP indicates that the variable has predictive power. Otherwise, a VIMP of zero or negative indicates that the variable's unpredictable power needs to be considered for filtering.

Fitness

In each iteration, the CoxPH model was used as the objective function to evaluate and update the whale position.³⁴ In survival analysis, the CoxPH model is a commonly used statistic that uses medical research patients and univariate or multivariate variables to predict associations between survival times. The CoxPH model can be used to assess how specified factors influence the rate of a particular event happening (e.g. infection, death).

The CoxPH model was used to establish the survival objects of HD patients as dependent variables, and other HD influencing factors were used as independent variables to assess whether it was related to the mortality of HD patients, and the hazard ratio (HR) value was calculated for the death risk determination.³⁵ The model can be written as follows:

$$\ln h(t) = \ln h_0(t) + b_1x_1 + \dots + b_px_p \quad (13)$$

or
$$\ln \frac{h(t)}{h_0(t)} = b_1x_1 + \dots + b_px_p \quad (14)$$

where $h(t)$ is the risk of time (t); $x_1, x_2 \dots x_p$ are explanatory variables. When HR value is less than 1, more than 1, and equal to 1, the risk of death is reduced, increased, and invalidated, respectively.

Concordance index is used to evaluate the predictive ability of the model.³⁶ It was first proposed by

Frank E Harrell Jr, a professor of biostatistics at Vanderbilt University in 1996. It is mainly used to calculate the discrimination between the predicted value and the true value of the CoxPH model in survival analysis. It is also called Harrell's concordance index. The C-index calculation method randomly pairs study objects. For a pair of patients, if A's actual survival time is longer than B, and the predicted result is the same, it is called concordance. The calculation formula of concordance index c is as follows:

$$c = \frac{K}{M} \quad (15)$$

where M is the number of remaining matches.

Optimal CoxPH regression model

This study aimed to assess of mortality-related risk in HD patients using machine learning approach to select features. The typical CoxPH model selection approach including all associated factors in known as full-adjusted-CoxPH, and the stepwise-CoxPH model includes the associated factors which have met the critical p value (<0.2). In addition to the typical statistic selection approaches, RSF and WOA methods were used to assess the optimal risk factors combination for all-cause mortality in CoxPH regression model. RSF model uses VIMP to identify associated risk factors, and WOA model uses local and global optimum search to identify associated risk factors. Taken together, this study performed full-adjusted-CoxPH, stepwise-CoxPH, RSF-CoxPH, and WOA-CoxPH model for the all-cause mortality risk assessment in HD patients, and compared the model performance using concordance index. Then, the survival interval and all-cause mortality status of HD patients were used to create a survival object, and concordance index was used as the objective function. Model comparison between four feature selection CoxPH model was performed to clarify the efficiency of proposed feature selection approaches for mortality risk assessment in HD patients. Harrell's concordance index value was used to compare the model performance; the higher concordance index represents better combination solution. Finally, the identified associated risk factors will be assessed for relevance to all-cause mortality in HD patients. The HR and 95% confidence interval (CI) were

computed. The significant associated risk factors identified by optimal CoxPH model were illustrated using Kaplan–Meier survival curves, and the survival difference between subgroups was tested using log-rank test. Moreover, the significant factors in optimal model were further used to generate an integrated risk score for mortality risk assessment of study population. Receiver operating characteristics (ROC) analysis was performed to further dichotomize the study population into low- and high-risk subgroups, and the survival difference between integrated risk subgroup was also tested using log-rank test. All p values were two-sided, and a p value less than 0.5 was considered statistically significant. All analyses were performed by R software (R Development Core Team 2020, version 4.0.2).

Results

Baseline characteristics and laboratory measurements

Table 2 showed the distribution and comparison of clinicopathological characteristics between alive and died groups. A total of 633 alive patients and 196 dead patients were analyzed. The died group had significant higher proportion in age ≥ 65 (died *versus* alive: 57.65% *versus* 28.28%, $p < 0.001$) and DM (38.27% *versus* 22.27%, $p < 0.001$). In laboratory measurements, the died group had significant higher proportion in hemoglobin (Hb) < 10.64 g/dl (59.18% *versus* 46.76%, $p = 0.003$), albumin < 3.87 g/dl (62.24% *versus* 36.49%, $p < 0.001$), creatinine (Cr) < 10.52 mg/dl (67.86% *versus* 46.29%, $p < 0.001$), potassium (K) < 4.95 meq/l (59.18% *versus* 47.24%, $p = 0.004$), Kt/V urea < 1.7 (65.31% *versus* 52.61%, $p = 0.002$), and cardiothoracic ratio ≥ 0.5 (72.96% *versus* 46.45%, $p < 0.001$) compared with the alive group.

Individual risk factors of all-cause mortality

Table 3 showed the univariate CoxPH regression analysis results for all-cause mortality in HD patients. The univariate analysis results showed patients with age ≥ 65 years (HR = 2.91, 95% CI = 2.19–3.87, $p < 0.001$), DM (HR = 1.98, 95% CI = 1.49–2.65, $p < 0.001$), and cardiothoracic ratio ≥ 0.50 (HR = 2.75, 95% CI = 2.01–3.77, $p < 0.001$) were more likely to achieve higher mortality risk. In addition, HD patients

Table 2. Baseline characteristics of mortality categories ($n=829$) with two categories.

Characteristics	Total ($n=829$)	Alive ($n=633$)	Died ($n=196$)	<i>p</i>
Dialysis vintage (years)				0.213
≥ 5.60	415 (50.06)	325 (51.34)	90 (45.92)	
< 5.60	414 (49.94)	308 (48.66)	106 (54.08)	
Age (years)				<0.001
≥ 65	292 (35.22)	179 (28.28)	113 (57.65)	
< 65	537 (64.78)	454 (71.72)	83 (42.35)	
Sex				0.819
Male	376 (45.36)	289 (45.66)	87 (44.39)	
Female	453 (54.64)	344 (45.34)	109 (55.61)	
DM	216 (26.06)	141 (22.27)	75 (38.27)	<0.001
Laboratory measurements				
Hb (g/dl)				0.003
≥ 10.64	417 (50.30)	337 (53.24)	80 (40.82)	
< 10.64	412 (49.70)	296 (46.76)	116 (59.18)	
Albumin (g/dl)				<0.001
≥ 3.87	476 (57.42)	402 (63.51)	74 (37.76)	
< 3.87	353 (42.58)	231 (36.49)	122 (62.24)	
BUN (mg/dl)				0.180
≥ 68.91	409 (49.33)	321 (50.71)	88 (44.90)	
< 68.91	420 (50.66)	312 (49.29)	108 (55.10)	
Cr (mg/dl)				<0.001
≥ 10.52	403 (48.61)	340 (53.71)	63 (32.14)	
< 10.52	426 (51.39)	293 (46.29)	133 (67.86)	
K (meq/l)				0.004
≥ 4.95	414 (49.94)	334 (52.76)	80 (40.82)	
< 4.9	415 (50.06)	299 (47.24)	116 (59.18)	
Ca (mg/dl)				0.064
≥ 9.22	384 (46.32)	304 (48.18)	79 (40.31)	
< 9.22	445 (53.68)	328 (51.82)	117 (59.69)	

(Continued)

Table 2. (Continued)

Characteristics	Total (n=829)	Alive (n=633)	Died (n=196)	p
Phosphate (mg/dl)				0.588
≥4.84	401 (48.37)	310 (48.97)	91 (45.43)	
<4.84	428 (51.63)	323 (51.03)	105 (53.57)	
iPTH (pg/ml)				0.919
≥205.60	415 (50.06)	318 (50.24)	97 (49.49)	
<205.60	414 (49.94)	315 (49.76)	99 (50.51)	
Ferritin (ng/ml)				0.171
≥412.70	415 (50.06)	308 (48.66)	107 (54.59)	
<412.70	414 (49.94)	325 (51.34)	89 (45.41)	
Kt/V urea				0.002
≥1.70	368 (44.39)	300 (47.39)	68 (34.69)	
<1.70	461 (55.61)	333(52.61)	128 (65.31)	
URR				1.000
≥0.65	786 (94.81)	600 (94.79)	186 (94.90)	
<0.65	43 (5.19)	33 (5.21)	10 (5.10)	
Cardiothoracic ratio				<0.001
≥0.50	437 (52.71)	294 (46.45)	143 (72.96)	
<0.50	392 (47.29)	339 (53.55)	53 (27.04)	

BUN, blood urine nitrogen; DM, diabetes mellitus; iPTH, intact parathyroid hormone; TSA, time-averaged serum albumin; URR, urea reduction ratio.
The *p* value is estimated using independent two-sampled *t*-test.
p values less than 0.5 were considered statistically significant.

with Hb ≥ 10.64 (HR = 0.62, 95% CI = 0.47–0.83, *p* = 0.001), albumin ≥ 3.87 g/dl (HR = 0.39, 95% CI = 0.29–0.52, *p* < 0.001), Cr ≥ 10.52 mg/dl (HR = 0.44, 95% CI = 0.33–0.60, *p* < 0.001), K ≥ 4.95 meq/l (HR = 0.66, 95% CI = 0.50–0.88, *p* = 0.004), Ca ≥ 9.22 mg/dl (HR = 0.74, 95% CI = 0.56–0.98, *p* = 0.037), and Kt/V urea ≥ 1.70 (HR = 0.63, 95% CI = 0.47–0.84, *p* = 0.002) were more likely to achieve lower mortality risk.

Risk assessment model of all-cause mortality

The analysis results of full-adjusted-CoxPH, stepwise-CoxPH, RSF-CoxPH, and WOA-CoxPH model for all-cause mortality in HD

patients are summarized in Table 4. The full-adjusted-CoxPH model showed patients with age ≥ 65 years (HR = 1.19, 95% CI = 1.46–2.68, *p* < 0.001), DM (HR = 1.64, 95% CI = 1.19–2.27, *p* = 0.002), Hb < 10.64 (HR = 0.66, 95% CI = 0.49–0.89, *p* = 0.006), albumin < 3.87 g/dl (HR = 0.62, 95% CI = 0.45–0.85, *p* = 0.003), Cr < 10.52 mg/dl (HR = 0.61, 95% CI = 0.43–0.88, *p* = 0.008), K < 4.95 meq/l (HR = 0.73, 95% CI = 0.55–0.99, *p* = 0.041), Kt/V urea < 0.71 (HR = 0.61, 95% CI = 0.44–0.86, *p* = 0.005), and cardiothoracic ratio ≥ 0.5 (HR = 1.98, 95% CI = 1.42–2.75, *p* < 0.001) were more likely to associate with the increasing risk of all-cause mortality in HD patients.

Table 3. Univariate Cox proportional hazard regression analysis results for all-cause mortality.

Characteristics	Comparison	Unadjusted		
		HR	95% CI	<i>p</i>
Dialysis vintage(years)	≥5.60 versus <5.60	0.82	0.62–1.09	0.167
Age	≥65 versus <65	2.91	2.19–3.87	<0.001
Sex	Female versus male	1.05	0.79–1.39	0.735
DM	Yes versus no	1.98	1.49–2.65	<0.001
Laboratory measurements				
Hb (g/dl)	≥10.64 versus <10.64	0.62	0.47–0.83	0.001
Albumin (g/dl)	≥3.87 versus <3.87	0.39	0.29–0.52	<0.001
BUN (mg/dl)	≥68.91 versus <68.91	0.80	0.60–1.06	0.116
Cr (mg/dl)	≥10.52 versus <10.52	0.44	0.33–0.60	<0.001
K (meq/l)	≥4.95 versus <4.95	0.66	0.50–0.88	0.004
Ca (mg/dl)	≥9.22 versus <9.22	0.74	0.56–0.98	0.037
Phosphate (mg/dl)	≥4.84 versus <4.84	0.90	0.68–1.19	0.465
iPTH (pg/ml)	≥205.60 versus <205.60	0.96	0.73–1.27	0.795
Ferritin (ng/ml)	≥412.70 versus <412.70	1.27	0.96–1.68	0.095
Kt/V urea	≥1.70 versus <1.70	0.63	0.47–0.84	0.002
URR	≥0.65 versus <0.65	0.97	0.51–1.83	0.927
Cardiothoracic ratio	≥0.50 versus <0.50	2.75	2.01–3.77	<0.001
Optimal concordance		0.6290		
BUN, blood urea nitrogen; CI, confidence interval; DM, diabetes mellitus; HR, hazard ratio; iPTH, intact parathyroid hormone; URR, urea reduction ratio; WOA, whale optimization algorithm. <i>p</i> values less than 0.5 were considered statistically significant.				

In the stepwise-CoxPH model, the risk factors that met critical *p* value < 0.2 were included. The stepwise-CoxPH model indicates patients with age ≥ 65 years (HR = 1.92, 95% CI = 1.43–2.59, *p* < 0.001), DM (HR = 1.55, 95% CI = 1.15–2.10, *p* = 0.004), Hb < 10.64 g/dl (HR = 0.65, 95% CI = 0.49–0.87, *p* = 0.004), albumin < 3.87 g/dl (HR = 0.62, 95% CI = 0.46–0.84, *p* = 0.002), Cr < 10.52 mg/dl (HR = 0.64, 95% CI = 0.46–0.88, *p* = 0.007), K < 4.95 meq/l (HR = 0.75, 95% CI = 0.56–1.00, *p* = 0.048), Kt/V urea < 0.71 (HR = 0.61, 95% CI = 0.45–0.84, *p* = 0.002), and cardiothoracic ratio ≥ 0.5 (HR = 1.99, 95%

CI = 1.44–2.75, *p* < 0.001) were more likely to increased all-cause mortality risk.

The RSF-CoxPH model included dialysis vintage, age, sex, DM, Hb, albumin, BUN, creatinine, K, Ca, phosphate, Kt/V urea, and cardiothoracic ratio. Patients characterized with age ≥ 65 years (HR = 1.97, 95% CI = 1.46–2.66, *p* < 0.001), with DM (HR = 1.63, 95% CI = 1.18–2.25, *p* = 0.003), Hb < 10.64 g/dl (HR = 0.66, 95% CI = 0.49–0.89, *p* = 0.006), albumin < 3.87 g/dl (HR = 0.62, 95% CI = 0.45–0.85, *p* = 0.003), Cr < 10.52 mg/dl (HR = 0.62, 95% CI = 0.46–0.88,

Table 4. Multivariate Cox proportional hazard for all-cause mortality based on different feature selection approaches.

Characteristics	Comparison	Full-adjusted			Stepwise ($p < 0.2$)			RSF			WOA		
		HR	95% CI	P	HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Dialysis vintage(years)	≥ 5.60 versus < 5.60	1.19	0.86–1.63	0.289	-	-	-	1.19	0.87–1.63	0.278	1.20	0.88–1.64	0.256
Age	≥ 65 versus < 65	1.98	1.46–2.68	<0.001	1.92	1.43–2.59	<0.001	1.97	1.46–2.66	<0.001	1.99	1.47–2.69	<0.001
Sex	Female versus male	0.95	0.68–1.33	0.779	-	-	-	0.96	0.69–1.35	0.829	-	-	-
DM	Yes versus no	1.64	1.19–2.27	0.002	1.55	1.15–2.10	0.004	1.63	1.18–2.25	0.003	1.63	1.18–2.24	0.003
Laboratory measurements													
Hb (g/dl)	≥ 10.64 versus < 10.64	0.66	0.49–0.89	0.006	0.65	0.49–0.87	0.004	0.66	0.49–0.89	0.006	0.65	0.48–0.88	0.005
Albumin (g/dl)	≥ 3.87 versus < 3.87	0.62	0.45–0.85	0.003	0.62	0.46–0.84	0.002	0.62	0.45–0.85	0.003	0.62	0.45–0.85	0.003
BUN (mg/dl)	≥ 68.91 versus < 68.91	1.12	0.82–1.53	0.460	-	-	-	1.12	0.82–1.52	0.482	1.06	0.78–1.43	0.730
Cr (mg/dl)	≥ 10.52 versus < 10.52	0.61	0.43–0.88	0.008	0.64	0.46–0.88	0.007	0.62	0.43–0.88	0.008	0.62	0.44–0.87	0.006
K (meq/l)	≥ 4.95 versus < 4.95	0.73	0.55–0.99	0.041	0.75	0.56–1.00	0.048	0.74	0.55–0.99	0.042	-	-	-
Ca (mg/dl)	≥ 9.22 versus < 9.22	0.90	0.65–1.23	0.500	-	-	-	0.92	0.67–1.25	0.580	0.90	0.66–1.23	0.508
Phosphate (mg/dl)	≥ 4.84 versus < 4.84	1.00	0.73–1.37	0.993	-	-	-	1.02	0.75–1.39	0.905	1.00	0.73–1.35	0.978
iPTH (pg/ml)	≥ 205.60 versus < 205.60	1.12	0.83–1.51	0.452	-	-	-	-	-	-	-	-	-
Ferritin (ng/ml)	≥ 412.70 versus < 412.70	0.99	0.74–1.32	0.940	-	-	-	-	-	-	0.99	0.74–1.32	0.946
Kt/V urea	≥ 1.70 versus < 1.70	0.61	0.44–0.86	0.005	0.61	0.45–0.84	0.002	0.62	0.44–0.86	0.005	0.60	0.44–0.83	0.002
URR	≥ 0.65 versus < 0.65	1.10	0.57–2.10	0.784	-	-	-	-	-	-	1.10	0.57–2.11	0.772
Cardiothoracic ratio	≥ 0.50 versus < 0.50	1.98	1.42–2.75	<0.001	1.99	1.44–2.75	<0.001	1.99	1.43–2.76	<0.001	1.94	1.40–2.70	<0.001
Optimal concordance		0.7404			0.7388			0.7406			0.7409		

BUN, blood urea nitrogen; CI, confidence interval; DM, diabetes mellitus; HR, hazard ratio; iPTH, intact parathyroid hormone; RSF, Random survival forest; URR, urea reduction ratio; WOA, whole optimization algorithm.
p values less than 0.5 were considered statistically significant.

$p=0.008$), $K < 4.95$ meq/l (HR = 0.74, 95% CI = 0.55–0.99, $p=0.042$), Kt/V urea < 0.71 (HR = 0.62, 95% CI = 0.44–0.86, $p=0.005$), and cardiothoracic ratio ≥ 0.5 (HR = 1.99, 95% CI = 1.43–2.76, $p < 0.001$) were more likely to increased all-cause mortality risk.

The WOA-CoxPH model included dialysis vintage, age, DM, Hb, albumin, BUN, Cr, Ca, phosphate, ferritin, Kt/V urea, URR, and cardiothoracic ratio. Patients characterized with age ≥ 65 years (HR = 1.99, 95% CI = 1.47–2.69, $p < 0.001$), DM (HR = 1.63, 95% CI = 1.18–2.24, $p=0.003$), Hb < 10.64 g/dl (HR = 0.65, 95% CI = 0.48–0.88, $p=0.005$), albumin < 3.87 g/dl (HR = 0.62, 95% CI = 0.45–0.85, $p=0.003$), Cr < 10.52 mg/dl (HR = 0.62, 95% CI = 0.44–0.87, $p=0.006$), Kt/V urea < 0.71 (HR = 0.60, 95% CI = 0.44–0.83, $p=0.002$), and cardiothoracic ratio ≥ 0.5 (HR = 1.94, 95% CI = 1.40–2.70, $p < 0.001$) were more likely to increased all-cause mortality risk.

RSF-CoxPH and WOA-CoxPH model obtained similar analysis results. Both models selected 13 out of the 16 factors but differ in three factors. The concordance index for full-adjusted-CoxPH, stepwise-CoxPH, RSF-CoxPH, and WOA-CoxPH model were 0.7404, 0.7388, 0.7406, and 0.7409, respectively. The comparison results showed that WOA-CoxPH model obtained the highest C-index among all models, which indicates that the WOA model could achieve better concordance in all-cause mortality risk estimation for HD patients. Furthermore, the Kaplan–Meier curve of significant risk factors identified in both RSF and WOA model were illustrated in Figure 1. Figure 1(a) presented the overall survival probability of all patients, the overall survival rate of all patients was 76.9% (95% CI = 74.0–79.8). Figure 1(b)–(i) illustrated overall survival probability analysis results of the eight significant parameters obtained from the WOA-CoxPH model including age (log-rank test $p < 0.001$), DM (log-rank test $p < 0.001$), Hb (log-rank test $p=0.001$), albumin (log-rank test $p < 0.001$), Cr (log-rank test $p < 0.001$), K (log-rank test $p=0.004$), Kt/V (log-rank test $p=0.016$), and cardiothoracic ratio (log-rank test $p < 0.001$), which have also showed significant survival difference between low- and high-risk characteristics in single-factor analysis. The red solid line indicates the high-risk characteristics, and the blue solid

line indicates the low-risk characteristics. The high-risk characteristics of eight selected parameters were scored 1, and the low-risk characteristics were scored 0. An integrated risk score for overall survival was generated based on the summation of eight selected parameters derived using WOA-selection model. An optimal cutoff point for the integrated risk score was determined using ROC analysis according to the mortality status. Thus, the patients were dichotomized into low- and high-risk subgroup by the optimal cutoff points of seven. Figure 1(j) showed that the patients who obtained seven or more risk characteristics of eight selected parameters will achieve significant worse overall survival probability compared with those who obtained less or equal to six risk characteristics (log-rank test $p < 0.001$). Overall, the integrated low- and high-risk subgroup showed greater discrepancy compared with each single risk factor selected by WOA-CoxPH model. In summary, patients who obtained seven or more risk characteristics of eight selected parameters could have potentially increased risk of all-cause mortality in HD population.

Discussion

In this study, we identified optimal risk factors combination of all-cause mortality in HD patients using WOA-CoxPH model. The identified risk factors were age, diabetes, hemoglobin, serum albumin, serum creatinine, serum potassium, Kt/V urea, and cardiothoracic ratio. The study results are consistent with previous studies in HD patients.^{1,37–46} Based on clinical practice point, aging is characterized by the progressive decline in function of organs. Consequently, aging contributes to the disease occurrence and disease-related mortality. In our previous investigations, we found that elder age was associated with decline in physical functional performance and cognitive function in HD patients.^{12,47} We believe that these disability circumstances predispose HD patients to death risk. It is well recognized that protein-energy-wasting is strongly associated with mortality in dialysis patients.³⁸ Serum albumin, creatinine, and hemoglobin are commonly applied to be as surrogates for nutritional status. Our previous observational studies have validated that these clinical factors are associated with poor quality of life and death risk in dialysis patients.^{1,12,48} Diabetes has become the major etiology of chronic kidney disease. This disease

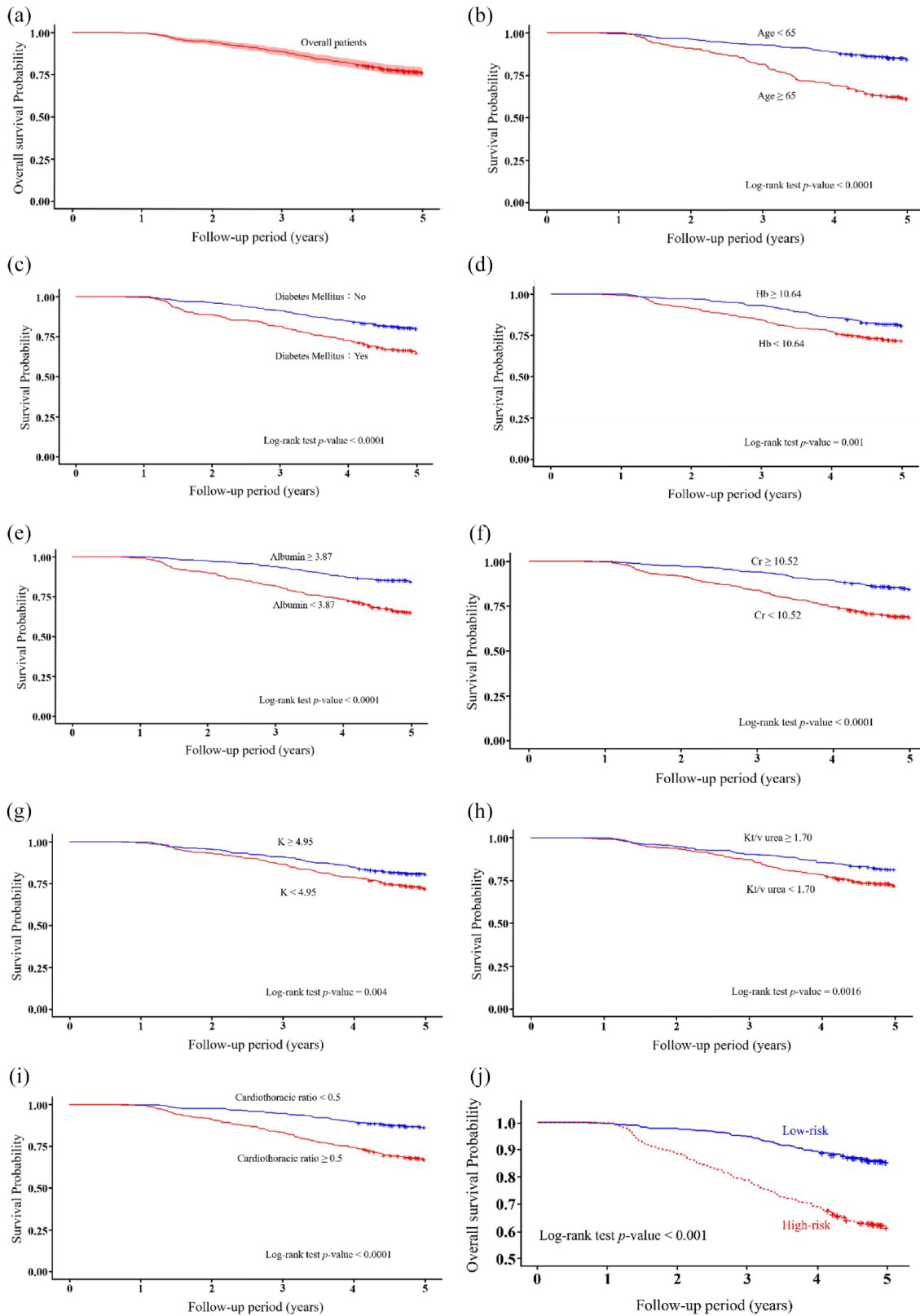


Figure 1. Kaplan–Meier curve for all-cause mortality in (a) all patients and difference subgroup of (b) age, (c) DM status, (d) hemoglobin, (e) albumin, (f) creatinine, (g) potassium, (h) Kt/V, (i) cardiothoracic ratio group, and (j) integrated risk subgroup derived using WOA-CoxPH model.

involves endothelial dysfunction and eventually organ failure. The association between diabetes and risk of mortality in HD patients also has evidenced *via* systematic review and meta-analysis.⁴⁹ Chronic kidney disease patients commonly have elevated serum potassium levels due to reduced renal clearance. Cardiac conductive activity is influenced by serum potassium levels. Elevated serum potassium levels could increase cardiac arrhythmia events in dialysis patients. Consequently, death risk would increase in the event of serum potassium increment especially in HD patients.⁴⁵ Considering HD adequacy, Kt/V-urea calculation is commonly used.⁴⁴ This calculation reflects the clearance of small molecular weight uremic toxins. This index has been applied as standard surrogate to evaluate dialysis adequacy either on HD or on peritoneal dialysis for years. Moreover, this adequacy index has been reported with survival in HD patients.⁵⁰ Taken together, selected risk factors for death in HD patients *via* WOA and RSF feature are reasonable *via* clinical practice point and literature review.

The study results indicate WOA-CoxPH model could perform as better survival model compared with typical statistical survival selection model.⁵¹ WOA used a group-based approach combined with exploration and exploitation procedures to accelerate the optimal search process. WOA algorithm could maintain an appropriate balance between exploration and exploitation by considering both local and global search for optimal combination. The mentioned search strategy made WOA more competitive than other methods, especially in terms of complex and interactive feature issues. Moreover, RSF-CoxPH also provides a similar performance with WOA-CoxPH compared with typical CoxPH models. RSF calculates the VIMP of each variable and ranked the value in purpose to filter the high-impact variables for interest outcome, which could provide lower prediction errors.⁵² Both WOA-CoxPH and RSF-CoxPH showed the addition machine learning algorithm in feature selection procedure could improve the performance of CoxPH model.

Several limitations of current study should be noted. First, the inclusion of certain covariates was limited due to the retrospective nature of current study. Second, this is a single institution study. Third, the study findings are restricted in

HD population. Although the mentioned limitations could limit the generalisability of current findings, this study still provides an optimal feature selection approach by using RSF and WOA in survival model, which could provide better concordance model performance compared with typical feature selection methods in survival model estimation.

Conclusion

The all-cause mortality is commonly considered as primary endpoints for patients undergoing long-term HD in the clinical studies. Previous studies indicated all-cause mortality in HD patient is associated with multiple clinical factors, including comorbidity, medications, nutrition status, and others. This study performed the all-cause mortality risk assessment model of HD patients based on machine learning feature selection procedure. Compared with the typical statistical selection model, RSF-CoxPH and WOA-CoxPH model showed better model performance. Therefore, RSF-CoxPH and WOA-CoxPH model identified risk factors combination might contribute to the more precise risk assessment of all-cause mortality in HD patients, and those identified risk factors could be considered as an important monitoring index in further management for HD patients in order to provide better survival outcome in maintenance HD population.

Declarations

Ethics approval and consent to participate

The study was employed in accordance with the Declaration of Helsinki and approved by the Chuang Gung Medical Foundation Institutional Review Board (201800595B0). Written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Author contributions

Cheng-Hong Yang: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing – review & editing.

Yin-Syuan Chen: Conceptualization; Data curation; Software.

Sin-Hua Moi: Formal analysis; Validation; Writing – original draft.

Jin-Bor Chen: Data curation; Formal analysis; Investigation; Resources; Visualization.

Lin Wang: Conceptualization; Formal analysis; Validation; Writing – review & editing.

Li-Yeh Chuang: Formal analysis; Supervision; Validation; Writing – original draft.

Acknowledgements

None

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partly supported by the National Science and Technology Council, R.O.C. (111-2221-E-165-001-MY3), Taiwan.

Competing interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Availability of data and materials

Not applicable.

ORCID iDs

Cheng-Hong Yang  <https://orcid.org/0000-0002-2741-0072>

Jin-Bor Chen  <https://orcid.org/0000-0003-4007-1455>

References

- Chen J-B, Cheng B-C, Yang C-H, *et al.* An association between time-varying serum albumin level and the mortality rate in maintenance haemodialysis patients: a five-year clinical cohort study. *BMC Nephrol* 2016; 17: 117.
- Eckardt K-U, Coresh J, Devuyst O, *et al.* Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* 2013; 382: 158–169.
- Luyckx VA, Tonelli M and Stanifer JW. The global burden of kidney disease and the sustainable development goals. *Bull World Health Organ* 2018; 96: 414–422.
- Lazarus J, Lowrie E, Hampers C, *et al.* Cardiovascular disease in uremic patients on hemodialysis. *Kidney Int Suppl* 1975; 2: 167–175.
- do Sameiro-Faria M, Ribeiro S, Costa E, *et al.* Risk factors for mortality in hemodialysis patients: two-year follow-up study. *Dis Markers* 2013; 35: 791–798.
- Lok CE and Mokrzycki MH. Prevention and management of catheter-related infection in hemodialysis patients. *Kidney Int* 2011; 79: 587–598.
- Kang SS, Chang JW and Park Y. Nutritional status predicts 10-year mortality in patients with end-stage renal disease on hemodialysis. *Nutrients* 2017; 9: 399.
- Snaedal S, Qureshi AR, Lund SH, *et al.* Dialysis modality and nutritional status are associated with variability of inflammatory markers. *Nephrol Dial Transplant* 2016; 31: 1320–1327.
- Goldwasser P, Mittman N, Antignani A, *et al.* Predictors of mortality in hemodialysis patients. *J Am Soc Nephrol* 1993; 3: 1613–1622.
- Lee SY, Yang DH, Hwang E, *et al.* The prevalence, association, and clinical outcomes of frailty in maintenance dialysis patients. *J Ren Nutr* 2017; 27: 106–112.
- Chandrashekar A, Ramakrishnan S and Rangarajan D. Survival analysis of patients on maintenance hemodialysis. *Indian J Nephrol* 2014; 24: 206–213.
- Chen J-B, Lee W-C, Cheng B-C, *et al.* Impact of risk factors on functional status in maintenance hemodialysis patients. *Eur J Med Res* 2017; 22: 1–8.
- Ramspek CL, Voskamp PW, van Ittersum FJ, *et al.* Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study. *Clin Epidemiol* 2017; 9: 451–464.
- Al-Shayea QK. Artificial neural networks in medical diagnosis. *Int J Comput Sci Issues* 2011; 8: 150–154.
- Qasem SN and Shamsuddin SM. Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. *Appl Soft Comput* 2011; 11: 1427–1438.
- Chuang LY, Chen GY, Moi SH, *et al.* Relationship between clinicopathologic variables in breast cancer overall survival using

- biogeography-based optimization algorithm. *Biomed Res Int* 2019; 2019: 2304128.
17. Sharma S, Sharma V and Sharma A. Performance based evaluation of various machine learning classification techniques for chronic kidney disease diagnosis. *arXiv Preprint arXiv* 2016: 160609581, <https://arxiv.org/abs/1606.09581>
 18. Chang S-W, Abdul-Kareem S, Merican AF, *et al.* Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. *BMC Bioinformatics* 2013; 14: 170.
 19. Radović N, Prelević V, Erceg M, *et al.* Machine learning approach in mortality rate prediction for hemodialysis patients. *Comput Methods Biomech Biomed Engin* 2022; 25: 111–122.
 20. Ishwaran H, Kogalur UB, Blackstone EH, *et al.* Random survival forests. *Ann Appl Stat* 2008; 2: 841–860.
 21. Garcia-Montemayor V, Martin-Malo A, Barbieri C, *et al.* Predicting mortality in hemodialysis patients using machine learning analysis. *Clin Kidney J* 2021; 14: 1388–1395.
 22. Mirjalili S and Lewis A. The whale optimization algorithm. *Adv Eng Softw* 2016; 95: 51–67.
 23. Mafarja MM and Mirjalili S. Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* 2017; 260: 302–312.
 24. Gharehchopogh FS and Gholizadeh H. A comprehensive survey: whale optimization algorithm and its applications. *Swarm Evol Comput* 2019; 48: 1–24.
 25. Suzuki MN, Fregonesi BM, Machado CS, *et al.* Hemodialysis water parameters as predisposing factors for anemia in patients in dialytic treatment: application of mixed regression models. *Biol Trace Elem Res* 2019; 190: 30–37.
 26. Hussien AG, Hassanien AE, Houssein EH, *et al.* S-shaped binary whale optimization algorithm for feature selection. In: Bhattacharyya S, Mukherjee A, Bhaumik H, *et al.* (eds) *Recent trends in signal and image processing*. Singapore: Springer, 2019, pp. 79–87.
 27. Mafarja M and Mirjalili S. Whale optimization approaches for wrapper feature selection. *Appl Soft Comput* 2018; 62: 441–453.
 28. Tawhid MA and Ibrahim AM. Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. *Int J Mach Learn Cybern* 2020; 11: 573–602.
 29. Yang CH, Moi SH, Chuang LY, *et al.* Higher-order clinical risk factor interaction analysis for overall mortality in maintenance hemodialysis patients. *Ther Adv Chronic Dis* 2020; 11: 2040622320949060.
 30. Sharawi M, Zawbaa HM and Emary E. Feature selection approach based on whale optimization algorithm. In: *2017 ninth international conference on advanced computational intelligence (ICACI)*, Doha, Qatar, 4–6 February 2017, pp. 163–168. New York: IEEE.
 31. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
 32. Ishwaran H, Kogalur UB, Blackstone EH, *et al.* Random survival forests. *Ann Appl Stat* 2008; 2: 841–860.
 33. Ishwaran H and Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* 2019; 38: 558–582.
 34. McGregor DE, Palarea-Albaladejo J, Dall PM, *et al.* Cox regression survival analysis with compositional covariates: application to modelling mortality risk from 24-h physical activity patterns. *Stat Methods Med Res* 2020; 29: 1447–1465.
 35. Du M, Haag DG, Lynch JW, *et al.* Comparison of the tree-based machine learning algorithms to Cox regression in predicting the survival of oral and pharyngeal cancers: analyses based on SEER database. *Cancers* 2020; 12: 2802.
 36. Qiu X, Gao J, Yang J, *et al.* A comparison study of machine learning (random survival forest) and classic statistic (Cox proportional hazards) for predicting progression in high-grade glioma after proton and carbon ion radiotherapy. *Front Oncol* 2020; 10: 551420.
 37. Cooper BA, Penne EL, Bartlett LH, *et al.* Protein malnutrition and hypoalbuminemia as predictors of vascular events and mortality in ESRD. *Am J Kidney Dis* 2004; 43: 61–66.
 38. Rambod M, Bross R, Zitterkoph J, *et al.* Association of Malnutrition-Inflammation Score with quality of life and mortality in hemodialysis patients: a 5-year prospective cohort study. *Am J Kidney Dis* 2009; 53: 298–309.
 39. Asakawa T, Joki N, Tanaka Y, *et al.* Association between the hemoglobin level and cardiothoracic ratio in patients on incident dialysis. *Cardiorenal Med* 2014; 4: 189–200.
 40. Park JI, Bae E, Kim YL, *et al.* Glycemic control and mortality in diabetic patients undergoing

- dialysis focusing on the effects of age and dialysis type: a prospective cohort study in Korea. *PLoS ONE* 2015; 10: e0136085.
41. Dekker MJE, Konings C, Canaud B, *et al.* Interactions between malnutrition, inflammation, and fluid overload and their associations with survival in prevalent hemodialysis patients. *J Ren Nutr* 2018; 28: 435–444.
 42. Hanafusa N, Tsuchiya K and Nitta K. Malnutrition-wasting conditions in older dialysis patients: an individualized approach. *Contrib Nephrol* 2019; 198: 12–20.
 43. Staibano P, Perelman I, Lombardi J, *et al.* Patient-centred outcomes in anaemia and renal disease: a systematic review. *Kidney Dis* 2020; 6: 74–84.
 44. Churchill BM and Patri P. The Nitty-Gritties of Kt/V(urea) calculations in hemodialysis and peritoneal dialysis. *Indian J Nephrol* 2021; 31: 97–110.
 45. Karaboyas A, Zee J, Brunelli SM, *et al.* Dialysate potassium, serum potassium, mortality, and arrhythmia events in hemodialysis: results from the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Am J Kidney Dis* 2017; 69: 266–277.
 46. Dashputre AA, Sumida K, Potukuchi PK, *et al.* Potassium trajectories prior to dialysis and mortality following dialysis initiation in patients with advanced CKD. *Nephron* 2021; 145: 265–274.
 47. Chen JB, Chang CC, Li LC, *et al.* Mutual interaction of clinical factors and specific microRNAs to predict mild cognitive impairment in patients receiving hemodialysis. *Cells* 2020; 9: 2303.
 48. Chen JB, Lam KK, Su YJ, *et al.* Relationship between Kt/V urea-based dialysis adequacy and nutritional status and their effect on the components of the quality of life in incident peritoneal dialysis patients. *BMC Nephrol* 2012; 13: 39.
 49. Ma L and Zhao S. Risk factors for mortality in patients undergoing hemodialysis: a systematic review and meta-analysis. *Int J Cardiol* 2017; 238: 151–158.
 50. Collins AJ, Ma JZ, Umen A, *et al.* Urea index and other predictors of hemodialysis patient survival. *Am J Kidney Dis* 1994; 23: 272–282.
 51. Tapak L, Sheikh V, Jenabi E, *et al.* Predictors of mortality among hemodialysis patients in Hamadan province using random survival forests. *J Prev Med Hyg* 2020; 61: E482–E488.
 52. Hsich E, Gorodeski EZ, Blackstone EH, *et al.* Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circ Cardiovasc Qual Outcomes* 2011; 4: 39–45.

Visit SAGE journals online
[journals.sagepub.com/
 home/taj](https://journals.sagepub.com/home/taj)

 SAGE journals