Informatics in Medicine Unlocked 19 (2020) 100351

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Mortality prediction enhancement in end-stage renal disease: A machine learning approach

Edwar Macias [a,*], Antoni Morell [a], Javier Serrano [a], Jose Lopez Vicario [a], Jose Ibeas [b]

[a] Wireless Information Networking (WIN) Group, Universitat Autònoma de Barcelona (UAB), Bellaterra, 08193, Spain
[b] Nephrology Department, Parc Taulí Hospital Universitari, Institut de Investigació i Innovació Parc Taulí I3PT. Universitat Autònoma de Barcelona, Sabadell, 08208, Spain

ARTICLE INFO

Keywords:
Mortality prediction
End-stage renal disease
Machine learning
LSTM
Random forest

ABSTRACT

In this work, we propose to combine massive variables collected during the evolution of patients in end-stage renal disease (ESRD), along with machine learning techniques to improve mortality prediction in ESRD. This work was carried out with a retrospective cohort of 261 patients, their evolution from diagnoses, laboratory tests, and variables recorded during haemodialysis sessions was combined. Random forest (RF) was used to explore the inference of the variables and define a base performance for long short-term memory (LSTM) recurrent neural networks. Then, LSTMs were trained with several groups of variables chosen by expert staff, the ones found by RF and all the available ones. The best performance was obtained using all the variables, but the ones found by RF had better predictive capacity than those chosen with expert knowledge. Integrating the three sources of information supposes an improvement in more than 4% in the area under the receiver operating characteristic curve. The approach is sufficiently robust to predict mortality at different time ranges. The massive integration of variables from patients in ESRD, together with the use of LSMTs, supposes an exceptional improvement in the predictive models of mortality. In conclusion, the machine learning approach can lead to a change in the paradigm in the analysis of predictive factors in mortality in ESRD.

1. Introduction

Chronic kidney disease (CKD) represents an epidemiological problem, USA 11% and Spain 9.2% in the adult population [1]. According to the World Health Organization (WHO), it has an indirect impact on the morbidity and mortality of the global population, increasing the mortality risk of the deadliest diseases [2,3]. CKD is closely related to cardiovascular (CV) risk, which is responsible for the highest mortality, especially on the end-stage renal disease (ESRD), where death from CV is one of the leading causes [1].

The most widely used way to detect the risk of suffering these kinds of pathologies is based on evidence-based medicine, which is translated into best practice guidelines, such as the American Heart Association/American College of Cardiology (ACC/AHA) [4], QRISK2 [5], Framingham [6] or Reynolds [7]. They are based on assuming linear relationships between risk factors and events. Nevertheless, the application of more sophisticated algorithms that use non-linear relationships, and can offer better performance in predictive models is still an open issue.

Thus, in the era of machine learning (ML), it is possible to generate complex models supported by large amounts of data [8–10]. Moreover, large-scale studies have begun to be described with ML to establish prognosis of mortality in the general population using routine clinical data [11–14]. However, those that exist in ESRD use approaches based on classical statistics [15–18] and some of them present a doubtful benefit [19].

There are few studies where ML techniques are applied to CKD. Salekin [20] and Abdullah [21] detect CKD using different classifiers (support vector machine, k-Nearest Neighbors, Random Forest (RF) and artificial neural networks (ANN)), Doi [22] trains logistic regression to predict mortality in patients starting with haemodialysis, and Titapiccolo [23] stratifies cardiovascular risk with RF. Predictive models of mortality using ML are even scarcer in the ESRD population, Akbilgic [24] used RF to predict mortality from one month to one year with an Area Under the Receiver Operating Characteristics (AUROC) of 0.736.

The aim of this manuscript is twofold, to present an exploratory

* Corresponding author.
E-mail addresses: edwar.macias@uab.cat (E. Macias), antoni.morell@uab.cat (A. Morell), javier.serrano@uab.cat (J. Serrano), jose.vicario@uab.cat (J.L. Vicario), jibeas@telefonica.net (J. Ibeas).

https://doi.org/10.1016/j.imu.2020.100351
Received 15 February 2020; Received in revised form 10 May 2020; Accepted 10 May 2020
Available online 30 May 2020

analysis of the potential by using massive variables and exploit temporal dependencies through long short-term memory (LSTM) recurrent neural networks (RNN) for improving predictive models of mortality in ESRD. It is evaluated the predictive capacity of several groups of variables. This study also takes advantage of the number of samples generated by the continuous monitoring of patients to propose predictive models of short-term mortality, which in our knowledge have not yet been achieved an AUROC higher than 0.736. This study points to the potential benefits of ML approaches to assess the medical staff with ESRD patients. It encourages the development of more powerful models using specialized ANNs as a predictive mechanisms.

2. Materials and methods

This retrospective study was carried out on a homogenous cohort of 1178 HD patients from a single centre with a reference population of almost half a million inhabitants. Of 1178, it was possible to extract information from 537 deceased patients, and of these, 261 provided the necessary massive data. These data were taken from the Information System of the Parc Tauli University Hospital, from the Haemodialysis (HD) Unit at the Nephrology Department from 2007 to 2018. This project passed through the ethics committee (Code 2018/508) and subsequently anonymised following the usual protocol. Inclusion criteria was being of legal age (> 18years). The available data include diagnoses, laboratory tests, and variables from haemodialysis sessions. The exposure period is from the moment the information of a patient is registered in ESRD in the information system of the hospital until the death of the patient.

To exploit the predictive capacity of variables and their temporal dependencies in the follow-up of patients in ESRD, data were first selected, then pre-processed and finally, the predictive models were generated in two stages. The first one using RF, due to its easy to tune and computational cost, to find the most important variables and for setting a baseline performance for more sophisticated algorithms. The second stage has the twofold purpose of exploiting temporal dependencies through LSTMs and analysing the impact of sets of variables including the ones found in the previous stage, groups of variables chosen by the expert staff and using all the available ones. All the necessary steps to carry out the prediction of mortality for patients in ESRD can be seen in Fig. 1 and are described below.

2.1. Data selection

Variables from the history of diagnoses, laboratory tests, HD sessions and demographics are taken as input for developing the predictive models. The outcome to predict is the mortality of patients. The variables are filtered based on their percentage of missing values (MV), variables with more than 43.2% of MVs are discarded. In Table 1 can be appreciated the selected features. Next are described the most relevant sources of information for this study.

2.1.1. Diagnoses

Refers to the historical hospital admissions that a patient has had. Each entry is associated with some particular diagnoses determined by examinations and evaluations of medical staff, which is encoded using the international classification of diseases (ICD9).

2.1.2. Laboratory

All the associated variables with samples from haematology, biochemistry, or some ESRD related hormones are stored as laboratory events. Some of them are taken with more or less periodicity. For instance, the most regular is the haemoglobin, which is measured every month, while proteins and PTH are measured every four months. Other measurements like immunology or tumour markers are taken more

Table 1
Selected variables from the data sources. The outcome is coded according to the death date of the patient. SBP and DBP refer to systolic and diastolic blood pressure, HR to heart rate and Temp to temperature.

Laboratory tests	Haemodialysis	Diagnoses
Calcium	Acc weight	Arteriopathy
Creatinine	Average flow	Cardiopathy
Ferritin	Blood vol dia	Diabetes
Glucose	DBP post HD	Enteropathy
Haemoglobin	DBP pre HD	Fracture
Haemoglobin	Dry weight	Haemorrhage
HDL cholesterol	HD time	Hepatopathy
Hematocrit	HR post HD	Hypertension
Iron	HR pre HD	Infection
KTV	Hypotension	Neoplasia
LDL cholesterol	SBP post HD	Pneumopathy
leucocytes	SBP pre HD	
Lymphocytes	Temp post HD	
Monocyts	Temp pre HD	**Demographics**
Neutrophil	Vascular access	Age
Phosphorus		Sex
Platelets		
Potassium		
PTH		**Outcome**
Reticulocytes		Months to decease
Sodium		
Total cholesterol		
Total proteins		
Triglycerides		
Urea		

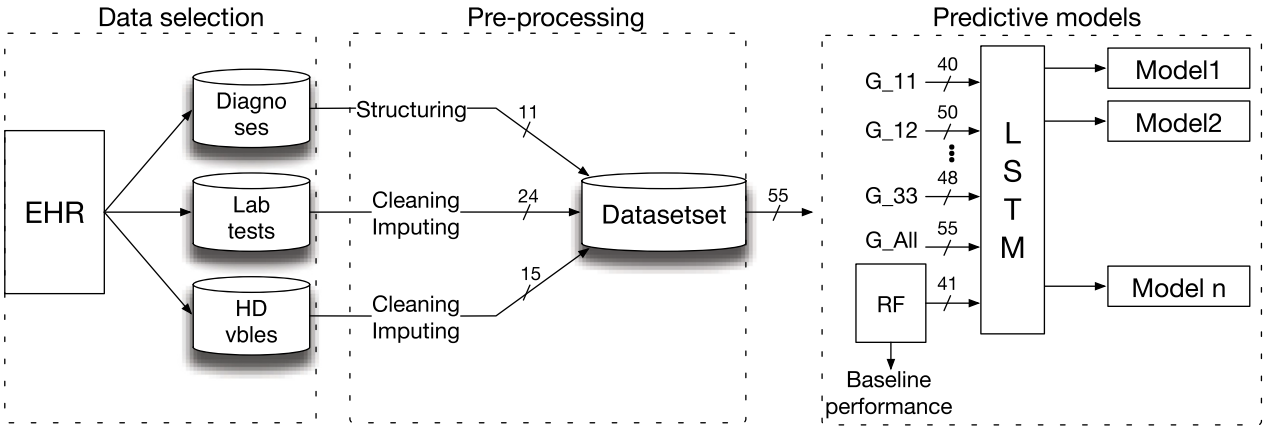

Fig. 1. Framework for developing predictive models in ESRD, G_{11} to G_{33} refer to set of variables ranked by their importance based on the expert staff experience. In pre-processing stage some features were generated based on 1-hot encoded for categorical features.

exceptionally.

2.1.3. HD variables

During the HD sessions, 3–4 per week, some variables are recorded at the beginning and the end of the HD session. Registered information includes the type of vascular access, duration of the session, episodes of hypotension and other variables taken from the haemodialysis machine, such as dry weight, temperature, systolic and diastolic blood pressure, heart rate, average flow, among others.

2.2. Data pre-processing

In general, raw data from electronic health records (EHR) are not appropriately structured to generate or test learning models. Thus, to prepare the information for predictive models, it is necessary to explore the variables, clean them and coherently correct wrong values. Below are the problems found in the samples.

- Data structure
- Incorrect values in variables
- Missing values

Initially, the information has to be structured. The diagnoses were grouped into 11 general ones based on expert knowledge. Then, they were structured in such a way that they became new variables for the final data set. In Fig. 2, this transformation can be appreciated. Later, the three information sources are combined based on their date they were measured. Finally, the follow-up of patients was summarized into one-month records, i.e., using the mean of variables in case of having more than one sample per month.

To correct the outliers and imputed the MVs, the ranges of the variables in laboratory tests and HD sessions were decided by the expert staff. The outliers of variables are identified and replaced with MVs to avoid losing the time stamp of variables. Then, the MVs are treated in two stages. The first one based on the individual imputation of the variables of each patient using second-order interpolations, to preserve trends in the evolution of the patient. In the second stage, MVs are imputed for patients without samples in some variables. Thus, from patients in training set, without MVs, the average value from each variable is extracted and used to impute MVs of the remaining patients.

2.3. Learning models

Initially, RF was used with the twofold purpose of establishing a baseline performance in terms of prediction, due to its easy set of parameters and computational cost, and quantifying the importance of the features for the final predictor. Then, due to the structure of the data, the different temporal dependencies are exploited using LSTMs.

2.3.1. Feature selection-Random forest

RF combines predictions based on decision trees [25]. They are trained with random subsets of data D_n. Branches of the decision trees are generated based on the calculation of the impurity of their features through the *Gini* index,

$$G(D_n) = 1 - \sum_{i=1}^{m} p_i^2 \quad (1)$$

where m is the number of classes (2 in our case, dead or alive), and p_i is the relative frequency of class i in a given branch of the tree. Initially, $G(D_n)$ is calculated for all the possible combinations of features and splitting thresholds. The combination which achieves the lowest value of $G(D_n)$ is chosen as far as it represents the best possible classification of D_n at this point of the tree. In subsequent branches, the same procedure is repeated up to the specified depth. In an RF approach, several trees are computed and fed with subsets of the data. Finally, the outcome produced by most of the trees is taken as the final decision (see Fig. 3).

On the other hand, the *Gini* index allows quantifying the importance of the features. This characteristic is used in this work to find the most relevant variables, more robustly, for predictors by combining a recursive feature elimination (RFE) [26] approach with RF. The traditional way to find the importance of features is to relate them individually with the outcome, without taking into account the interactions between variables. RFE solves this issue generating several predictors iteratively. Thus, in each iteration a predictor offer a performance measurement and the ranking of features. In the next iteration, the less important feature is

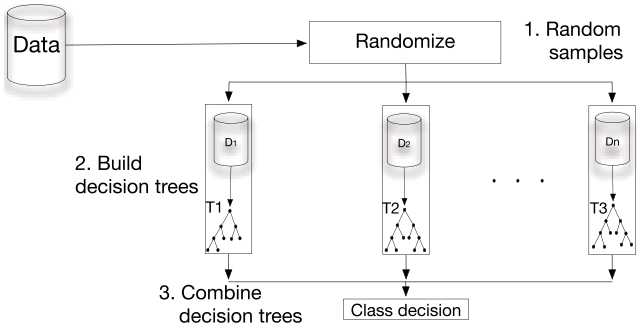

Fig. 3. Random forest data flow, at the end the class decision is made by voting of each tree.

Patient_id	D_ad	C_Diag_1	C_Diag_2	C_Diag_3	C_Diag_4	...
14785899	13/12/11	I351	250.40	585.5	V45.1	...
14785899	02/06/08	I6529	428.0	I4891	250.40	...
14785899	07/01/14	707.19	858	V45.1	Z8611	...
14785899	17/10/14	I2510	I472	I120		...

Patient_id	D_ad	ARTERIOPATHY	CARDIOPATHY	DIABETES	...	INFECTION
14785899	13/12/11	0	1	0	...	0
14785899	02/06/08	0	0	1	...	0
14785899	07/01/14	1	0	1	...	1
14785899	17/10/14	0	1	1	...	0

Fig. 2. Initially, each entry is associated with a series of diagnoses. In the new scheme, the most important diagnoses are selected and coded using one-hot encoding.

eliminated and the new predictor will yield another performance and a new ranking, and so on.

2.3.2. Predictive model-LSTM

ANNs have been successfully applied in the past to classification problems. The goal is, given a set of N input examples $\mathbf{x_n}$, where $n = 1, \ldots, N$ and its corresponding classification targets $\mathbf{t_n} = [t_{1,n}, \ldots, t_{K,n}]$, to learn the best non-linear model that maps the input to its respective target. Note that $\mathbf{t_n}$ has all values equal to 0 except the class where $\mathbf{x_n}$ belongs to, which has a value of 1. In the case of the prediction of mortality of patients in ESRD, the objective is to classify data collected during a period of n_{data} months, to be able to determine if the patient will be alive or not after n_{pred} months. Thus, driven by a considerable number of training samples, an ANN can learn an optimised non-linear function in an iterative process so that the error between the input and the output is minimised. Fig. 4 shows the structure of a basic ANN.

An ANN is composed of L layers ($l = 0, \ldots, L$), where $l = 0$ represents the input layer, N_l neurons per layer and their corresponding interconnections. Samples are first presented through the input layer, whose neurons connect with one or more hidden layers, and they link to the output layer where the result of the model is obtained. The output of the i-th neuron at the l-th layer (the so-called activation) is the linear combination of the outputs at the previous layer, taking into account the weights learned, and modified by a specific non-linear function $f(\bullet)$, usually the sigmoid, the hyperbolic tangent or the Rectified Linear Unit (ReLU). In other words,

$$a_j^l = f\left(\sum_{i=0}^{N_{l-1}} w_{i,j}^{l-1} a_i^{l-1}\right) \tag{2}$$

where $w_{i,j}^{l-1}$ is the weight that connects the i-th activation at layer $l-1$ to the input of the j-th neuron at layer l. Note that $a_0^l = 1$ in all layer except the output layer to take into account the bias term.

At the output layer, activations are usually normalised (e.g., softmax) so that the resulting values take a value between 0 and 1. They can be interpreted as a probability estimation; for example, y_k represents the probability that the input example belongs to the k-th class.

The network is trained in order to achieve a minimum of a given cost function that measures the error between predictions and the corresponding true values. A common function is cross-entropy,

$$C\left(\left\{w_{i,j}^l\right\}\right) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{k,n}\log\left(y_k\left(\mathbf{x_n}, \left\{w_{i,j}^l\right\}\right)\right) \tag{3}$$

A known problem using ANN is overfitting, which happens when the network does not learn a model from the underlying data but memorizes the individual examples. A common way to reduce this effect is applying L2 weight regularization [27], a quadratic penalty function is added to the weight, i.e. (3) is modified to

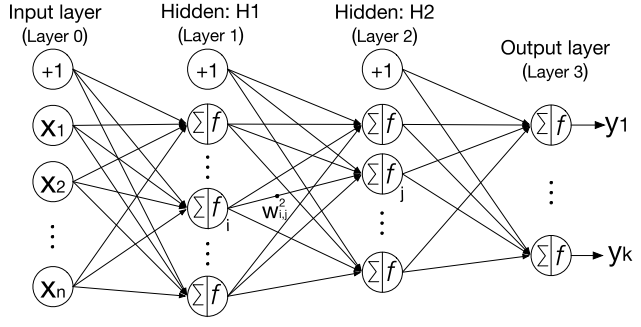

Fig. 4. Structure of a feed-forward artificial neural network with two hidden layers.

$$C'\left(\left\{w_{i,j}^l\right\}\right) = C\left(\left\{w_{i,j}^l\right\}\right) + \frac{\lambda}{2N}\sum_{l=1}^{L}\sum_{i=1}^{N_{l-1}}\sum_{j}^{N_l}\left(w_{i,j}^l\right)^2 \tag{4}$$

With L2 regularization, controlled by λ, we limit the adaptation capacity of the network by penalizing large weights.

Training is carried out with a gradient descent approach and the so-called back-propagation algorithm [28]. Thus, weights of the network are updated iteratively towards the opposite of the gradient, being the step controlled by the learning rate (LR). Currently, some algorithms accelerate the learning process by dynamically changing the LR. In this work we consider adaptive moment estimation (ADAM) [29].

On the other hand, RNNs are a variation of the networks shown in Fig. 4. Unlike feed-forward ANNs, RNNs use feedback connections to retain information about past events. In recent years one of the RNN implementations that have been successful is LSTM. To carry out mortality prediction in ESRD, LSTMs are used to exploit temporal dependencies in the follow-up of the patients. Fig. 5 shows the component of an LSTM cell. The memory mechanism is controlled by the gates, made up of ANNs with a specific activation function at the output layer. Each cell is responsible for filtering relevant information. The core idea is to combine the information from the gates in the cell gate, c_t. The forget gate, f_t, indicates which information from the combination of the previous state, h_t, and the input, x_t, is discarded. Then new information is added to c_t through the combination of two gates, the input gate, i_t, which decides the information to updated and the candidate values, $\tilde{c}_t$. Finally, c_t is updated and the output is a filtered version of the cell gate modulated.

In the case of mortality in ESRD, LSTMs are fed with concatenated vectors that contain the evolution of n months and the prediction is carried out to p months. For instance, Fig. 6 illustrates the follow-up of a patient during m months, from the first encounter with the hospital's system to the death event. The follow-up is structured into samples, taking information of n months of evolution. Then, using the timestamp of the samples and date of death, d in the figure, the moths to the death event of the structured samples are computed. Thus, the binary target of the generated data depends on the prediction range using the rule,

$$f(t_d) = \begin{cases} 0, & \text{if } t_d > p \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

where p is the prediction range, t_d is the time to the death event. '0′ and '1′ indicates the class sample, alive and deceased respectively.

3. Results

The samples for this analysis were extracted from 261 patients who had the three types of variables, as described in Section 2.1. Table 2 shows the description of the population. Because the duration of HD treatment varies across the cohort, each patient generates a different number of monthly samples. In total, 8394 monthly samples were extracted. In this work mortality is predicted to 1, 2, 3, 6 and 12 months. Thus, five datasets with the same data but different targets, after applying the transformation in Eq. (5), are generated. Fig. 7 shows the mortality trajectories for patients in the training and test sets.

For models development, patients were split into training and test sets (80-20%). The training set was divided into 5-folds for cross-validation (CV), see Fig. 8. With this approach, it is possible to find the hyperparameters for RFE-RF and LSTMs. Such parameters are the ones that can be calibrated manually. For RFE-RF the number of trees, depth of the decision trees and splitting criteria. For LSTMs the number of cells, neurons per cell, LR, among others. Then, with the hyperparameters fixed, the parameters of the network (weights of the LSTMs) are computed and five different models, $M1, M2, \ldots, M5$ in Fig. 8, from the 5-folds are obtained as a result. The evaluation is done in the initial test set.

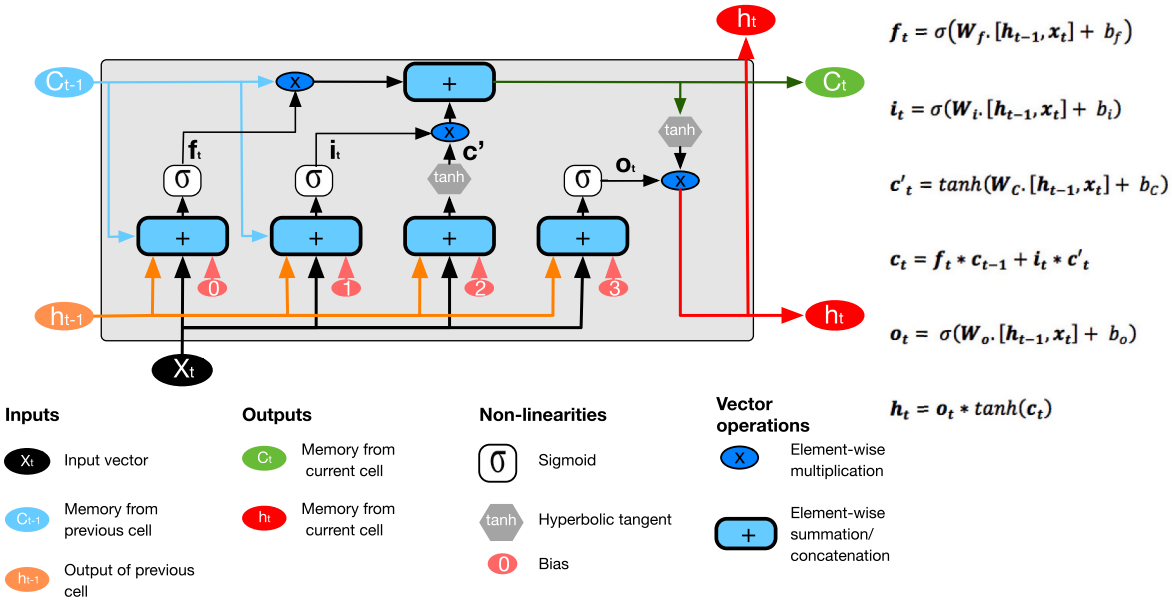

Fig. 5. LSTM cell, σ is the sigmoid activation function.

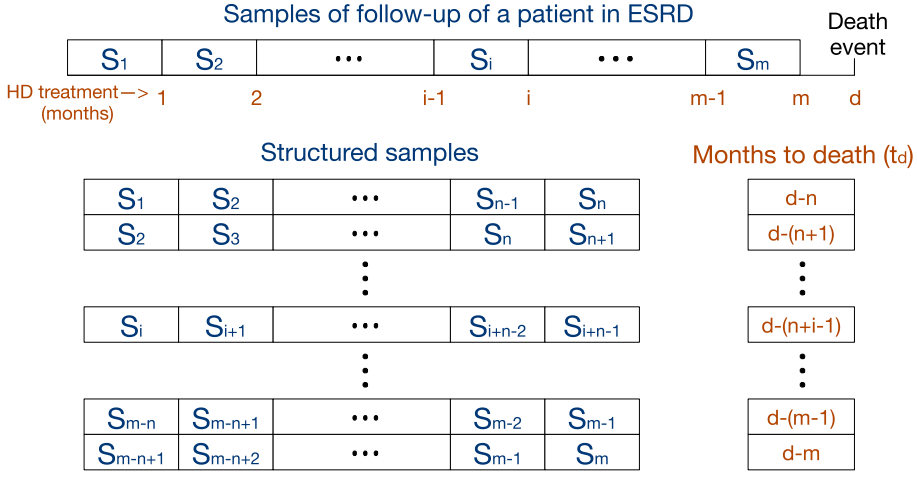

Fig. 6. Sample structuring from follow-up of a patient with *m* months in HD treatment to the death event, *d*.

To estimate the performance of the classifiers in the test set, AUROC was used. It measures the area under the graphic representation of the general accuracy, showing the variation of the sensitivity and specificity of a binary classifier when the decision threshold varies. The metric takes values between 0 and 1, with 1 corresponding to the perfect classifier.

3.1. Feature selection-RF

Our first experiment studies the importance of the available features automatically, using an RF approach, together with an RFE approach. The optimal hyperparameters for RF were 103 trees, maximum depth of 3, using the *Gini* index for splitting the nodes and calculate the importance of features. For RFE approach, the 5-folds were used to find the best features more robustly. With the approach, it was found that 42 features offered the best performance for all the predictors. AUROCs of 0.737, 0.714, 0.712, 0.668 and 0.615 were the baseline performance predicting mortality to 1, 2, 3, 6 and 12 months, respectively. Fig. 9 illustrate the AUROC as a function of the number of considered features for the prediction of mortality to one month.

The features not considered by RF-RFE were: cardiopathy, enteropathy, haemorrhage, hepatopathy, hypertension, neoplasia, pneumopathy, fracture, infection and the type of vascular access.

3.2. Predictive model-LSTM

In the second experiment, we consider a more powerful model based on LSTMs. After parameter optimization, we found that the best configuration was using an LSTM with two cells and with 750 and 500 units, respectively. We used ADAM optimizer with LR = 0.001 and L2 regularization with $\lambda = 0.001$. Then, the LSTM approach is evaluated in several groups of variables chosen by the experience of the expert staff, the group of variables found by RFE-RF and all the available ones. Table 3 shows the importance level of both laboratory and HD variables as determined by the experience of the hospital expert staff.

Fig. 10 shows ROC curves comparing the groups of variables using a 4 months to feed the LSTM and predicting mortality to 1 month. As illustration, Group_12 considers laboratory variables with an importance label of 1 and HD session variables with an importance level of 2 and Group_RFE refers to the ones found by RFE-RF. Note that diagnosis

Table 2
Cohort description. Variable Samples/Patient contains the information about the number of samples that the patients generate. For diagnoses, Number of patients column represents the total of patients with a specific diagnose. VA refers to vascular access, SBP and DBP to systolic and diastolic blood pressure.

Feature	Units	Number of patients	MV (%)	Mean	Std	Min	Max
Age	Years	–	0.0	71.41	10.69	24.00	91.00
Sex (Women)	–	104	0.0	–	–	–	–
Sex (Men)	–	157	0.0	–	–	–	–
Samples/Patient	–	–	–	26	22	1	116
Calcium	mg/dL	261	10.8	9.10	0.69	6.30	13.00
Creatinine	mg/dL	261	25.0	6.80	2.30	0.30	15.50
Ferritin	ng/mL	261	28.1	472.1	368.32	8.10	6590.00
Glucose	mg/dL	261	25.7	123.30	67.85	13.00	1370.00
Haemoglobin	g/L	261	29.2	6.21	1.26	4.10	13.60
HDL cholesterol	mg/dL	261	18.3	43.73	14.60	4.40	115.60
Hematocrit	L/L	261	1.0	349.990	0.04	0.17	0.49
Hemoglobin	g/L	261	1.1	111.69	14.21	46.00	161.00
Iron	μg/dL	261	38.1	59.44	26.70	10.00	340.00
KTV	mL/min	261	17.3	1.43	0.28	0.42	02.09
LDL cholesterol	mg/dL	261	18.9	83.40	33.20	8.00	240.00
Leucocytes	$x10^9/L$	261	1.0	7.63	4.97	1.25	11.3
Lymphocytes	$x10^9/L$	261	5.8	1.50	0.76	0.22	12.74
Monocyts	$x10^9/L$	261	5.8	0.56	0.22	0.03	2.69
Neutrophil	$x10^9/L$	261	5.8	5.24	2.29	0.22	7.25
Phosphorus	mg/dL	261	26.1	4.33	1.39	0.20	11.80
Platelets	$x10^9/L$	261	1.1	223.37	83.17	14.40	1067.00
Potassium	mEq/L	261	35.0	4.95	0.80	0.30	8.90
PTH	pg/mL	261	28.3	228.05	189.17	6.00	3264.00
Reticulocytes	$x10^9/L$	261	28.4	5.37	2.69	0.23	35.23
Sodium	mEq/L	261	31.5	138.66	3.59	121.00	198.00
Total cholesterol	mg/dL	261	38.1	149.98	39.41	45.00	432.00
Total proteins	g/L	261	27.6	66.02	6.84	28.5	96.00
Triglycerides	mg/dL	261	18.1	140.49	107.92	20.00	2673.00
Urea	mg/dL	261	43.2	102.40	51.12	20.20	317.20
Accumulative weight	Kg	261	21.7	1.95	0.77	−3.05	4.44
Average flow	mL/min	261	16.2	290.28	34.48	200.00	414.55
Blood vol dia	mL/min	261	12.0	65.08	10.52	40.00	98.43
DBP post HD	mmHg	261	10.4	65.34	10.22	40.00	105.61
DBP pre HD	mmHg	261	10.5	64.44	10.62	40.00	106.08
Dry weight	Kg	261	0.9	66.78	15.24	31.29	149.63
HD session time	Hours	261	0.0	3.73	0.35	3.50	7.30
HR post HD	BPM	261	10.6	75.59	11.49	41.00	122.00
HR pre HD	BPM	261	6.6	73.19	10.57	42.00	121.17
Hypotension	Cases/month	261	0.0	2	4	0	24
SBP post HD	mmHg	261	13.3	138.11	22.7	57.00	205.00
SBP pre HD	mmHg	261	6.3	137.31	22.19	56.07	218.60
Temp post HD	°C	261	16.9	35.58	0.33	33.00	38.20
Temp pre HD	°C	261	11.6	35.52	0.34	33.85	38.00
Arteriopathy	–	177	0.0	–	–	–	–
Cardiopathy	–	241	0.0	–	–	–	–
Diabetes	–	204	0.0	–	–	–	–
Enteropathy	–	94	0.0	–	–	–	–
Fracture	–	9	0.0	–	–	–	–
Hemorrhague	–	6	0.0	–	–	–	–
Hepatopathy	–	18	0.0	–	–	–	–
Hypertension	–	223	0.0	–	–	–	–
Infection	–	102	0.0	–	–	–	–
Neoplasia	–	79	0.0	–	–	–	–
Pneumopathy	–	115	0.0	–	–	–	–
VA (AVF)	–	168	0.0	–	–	–	
VA (Catheter)	–	164	0.0	–	–	–	–
VA (Graft)	–	6	0.0	–	–	–	–
Mortality	Months	–	0.0	25.52	21.89	1.00	116.00

variables (11 in total) are included in all cases.

Finally, in Fig. 11 we test the performance of our algorithm by considering: i) all variables; ii) HD data only and iii) diagnosis and laboratory data only.

4. Discussion

This work explored how deep learning can help in the study of ESRD. After the experiments conducted, in this case, focused on the evaluation of mortality, the lessons learned are: i) we can improve model accuracy w.r.t. the other works in the literature; ii) including knowledge expert not always leads to better models and iii) solutions can guide the research in a specific field by revealing possible causal relations not explored before, possibly far from human intuition. Table 4, includes a performance comparison in terms of AUROC with the existing solutions in the literature. Although one-year mortality does not exceed that stated in the literature, the improvement in short-term mortality grows to 4% if we reduce the prediction time to 3 months. When we compared our approach to these other works, we realised that we combined three sources of data, i.e., diagnosis, laboratory and HD data, being not the

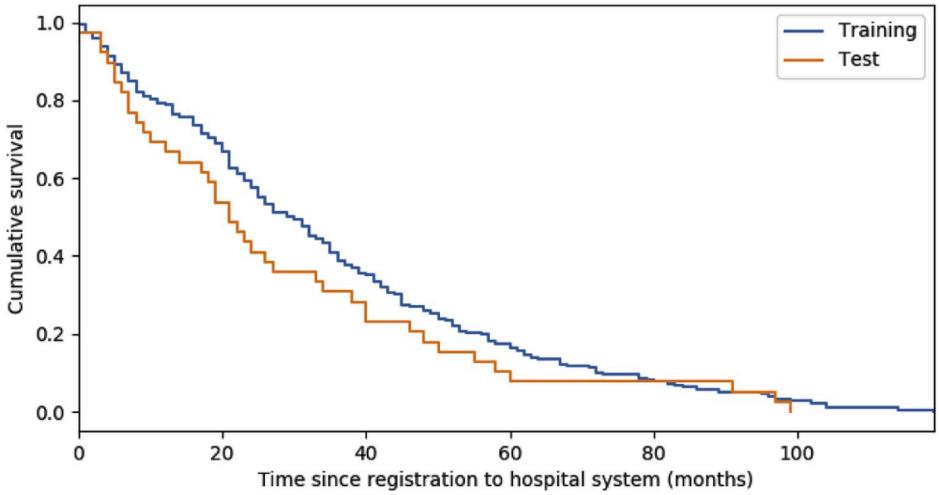

Fig. 7. Kaplan Meier mortality model for training and test set. p = 0.17.

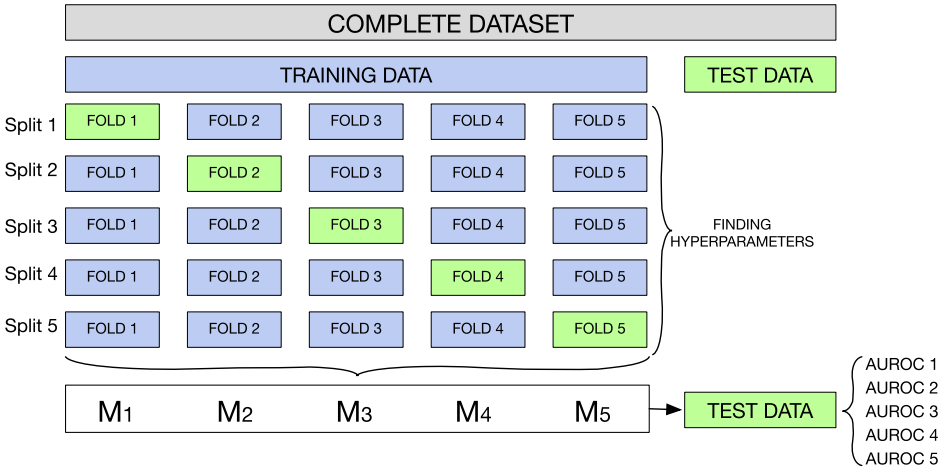

Fig. 8. Cross validation with 5-folds. Test Data is only used when the hyperparameters are found.

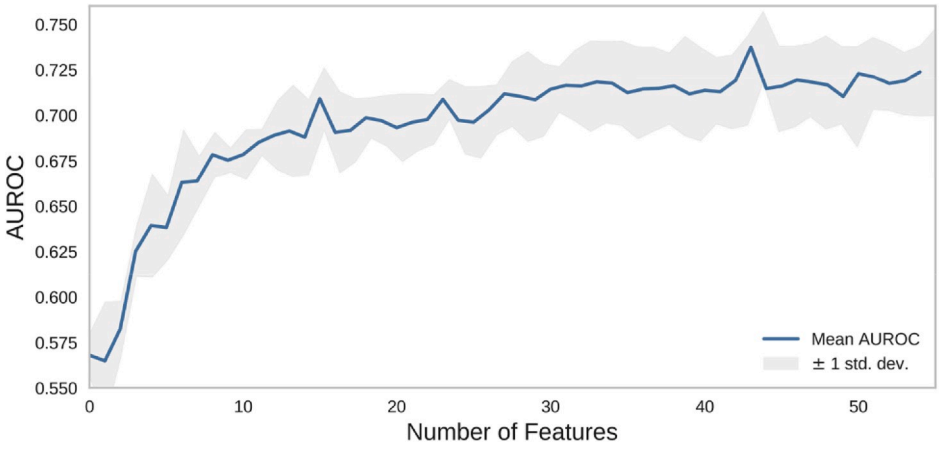

Fig. 9. Recursive feature selection, with 5-folds cross-validation, using RF as learning model.

case in the available works. Most of them use either laboratory and diagnosis data or HD session data. Fig. 11 shows that the inclusion of all variables improves the AUROC at least by 11% in AUROC.

In order to study how considering knowledge expert influences the performance of algorithms, expert staff labelled HD and laboratory data according to their importance level, being 1 the highest level and 3 the lowest (see Table 3). Accordingly, in Fig. 10 we tested our model with several combinations of the subsets of variables. We could expect to achieve the best possible performance by using level 1 laboratory data together with level 1 HD variables, i.e., Group_11 (recall that diagnostic

Table 3
Ranking of features chosen by the experience of the expert. Their importance are marked from 1 to 3, being 3 the less important features. VA refers to vascular access.

Laboratory	Importance	HD variables	Importance
Calcium	1	HD time	3
Creatinine	3	HR post HD	1
Ferritin	2	HR pre HD	1
Glucose	3	Hypotension	1
Haemoglobin	1	SBP post HD	1
HDL cholesterol	2	SBP pre HD	1
Iron	3	Temp post HD	3
KTV	1	Temp pre HD	3
LDL cholesterol	2	VA (AVF)	1
Leucocytes	2	VA (Catheter)	1
Lymphocytes	2	VA (Graft)	1
Monocytes	2		
Neutrophil	2		
Phosphorus	1		
Platelets	3		
Potassium	2		
PTH	1		
Reticulocytes	2		
Sodium	1		
Total cholesterol	2		
Total proteins	3		
Triglycerides	2		
Urea	1		

data is included in all cases). However, the performance achieved is similar to Group_33, and the inclusion of all variables boosts the AUROC value in 9%. In other words, expert knowledge is undoubtedly relevant, but it is also important to explore beyond it.

Finally, machine learning approaches can also help the research of physicians by revealing causal relations possibly not explored before. In Fig. 9, we tested how an automatic feature selection tool such as RFE may help. In this case, 42 features gave us the best classification performance using an RF approach, and when we consider this selection as input to our LSTM solution, the performance is close to best one, which is obtained with all the features. Therefore, physicians can explore the subset of variables selected, reduce or increase it as far as performance is sustained (see Fig. 9) and investigate the importance and effects of the chosen features. However, it should be noted that RF-RFE did not take into account most of the diagnoses or the type of vascular access. This evidence could suggest that those variables that were not considered important could lead to new medical research.

5. Conclusion

In this work, we demonstrate the potential of the massive use of variables together with machine learning techniques for the improvement of mortality predictive models in ESRD. We designed a baseline predictor and feature selector using an RFE-RF approach. Then we improve it using LSTM strategy that exploits temporal dependencies in the data. We conclude that thanks to considering diagnostic variables along with laboratory and HD session data, we could improve performance in the prediction of mortality in the ESRD patient by at least 4% w.r.t. existing works for short-term mortality. Furthermore, results show that expert knowledge has to contribute to the analysis, but we shall not limit our algorithms to it. In our experiment, the best performance achieved by the groups chosen does not exceed the RF-RFE. Therefore, machine learning methods like the ones explored here can provide feedback to the experts, improve our knowledge and can lead to a change in the paradigm in the analysis of predictive factors in mortality in ESRD.

Ethical statement

The authors of the manuscript certify that the manuscript entitle "*Mortality prediction enhancement in end-stage renal disease: A machine learning approach*" has not been and will not be submitted to or published in any other publication before its appearance in the "*Informatics in Medicine Unlocked*" journal.

The data used for this study were taken from the information system of the Parc Taulí University Hospital, from the Haemodialysis unit at the Nephrology Department. This project passed through the ethics committee (Code 2018/508).

Next the contributions of the authors are listed.

Author contribution

Conception and design of the study: E Macias, A Morell, J Serrano, JL Vicario and J Ibeas.

Adquisition of data: J Ibeas.

Analysis and/or interpretation of data: E Macias, A Morell, J Serrano, JL Vicario and J Ibeas.

Drafting the manuscript: E Macias, A Morell, J Serrano, JL Vicario and J Ibeas.

Revising the manuscript critically for important intellectual content: A Morell, J Serrano, JL Vicario and J Ibeas.

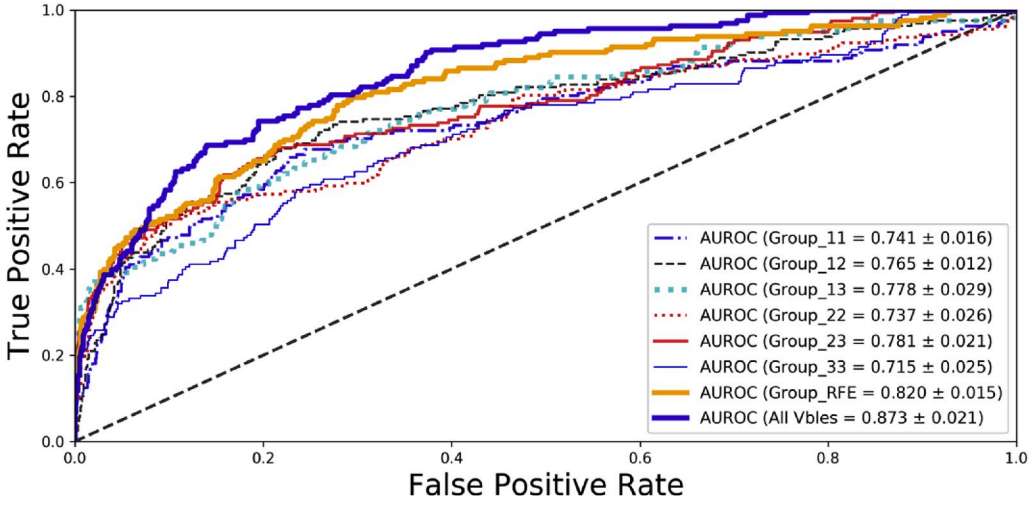

Fig. 10. Performance comparison between best features found by RFE-RF, the combinations of features chosen by expert staff and using all the available information. The groups are generated based on the experience of the medical staff, from Group_11 which can be inferred from the combination of the most important analytics with the most important HD variables to Group_33, the least significant ones.

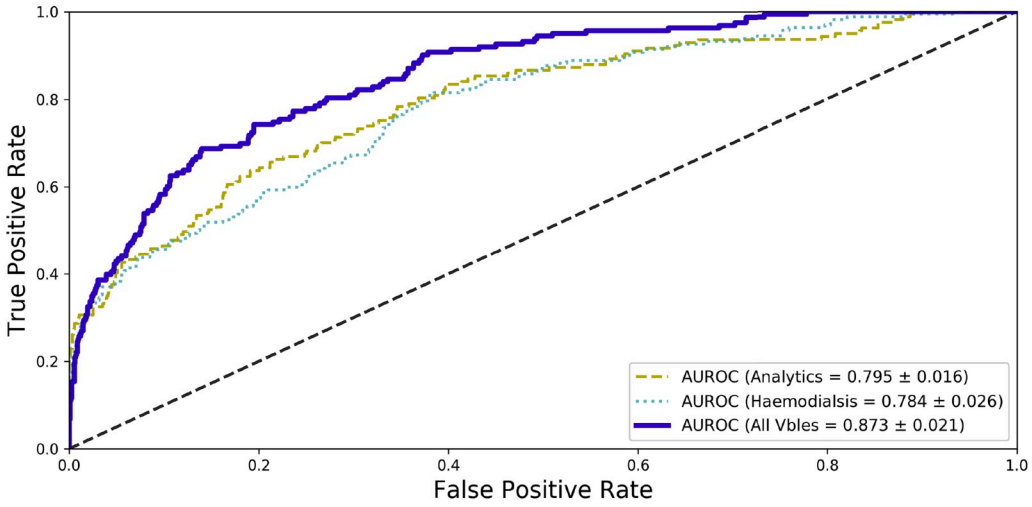

Fig. 11. Comparison of performance using laboratory tests and diagnoses, all the available variables and just variables taken during haemodialysis sessions.

Table 4
Comparison methods in literature with proposed one.

Reference	Mortality prediction	Study population	Prediction algorithm	AUROC (CI 95%)
Mauri et at [15]	1 year	5738	Logistic regression	0.670 (0.668–0.675)
Akbilgic et al. [24]	1 month	27615	Random forest	0.736 (0.715–0.757)
	3 months			0.764 (0.754–0.774)
	6 months			0.760 (0.747–0.775)
	1 year			0.757 (0.746–0.769)
Wagner et al. [30]	3 years	5447	Cox proportional hazards	0.730 (0.700–0.760)
	7–12 months			0.698 (0.679–0.717)
	13–18 months			0.717 (0.696–0.737)
	19–24 months			0.670 (0.646–0.694)
Isshiki et al. [31]	3–36 months	62	Hazard ratio	0.696
Proposed approach	1 month	261	LSTM	0.873 (0.871–0.876)
	2 months			0.813 (0.811–0.815)
	3 months			0.798 (0.796–0.800)
	6 months			0.752 (0.751–0.753)
	1 year			0.720 (0.703–0.737)

Approval of the manuscript to be published: A Morell, J Serrano, JL Vicario and J Ibeas.

Declaration of competing interest

The authors confirm that there are no known conflicts of interest associated with this publication.

Acknowledgements

This work is supported by the Spanish Government under Project TEC2017-84321-C4-4-R co-funded with European Union ERDF funds and also by the Catalan Government under Project 2017 SGR 1670.

References

[1] Martínez-Castelao Alberto, Górriz José L, Bover Jordi, Segura-de la Morena Julián, Morales-Olivas Francisco, Cebollada Jesús, Suárez Carmen, Escalada Javier, Muñoz-Torres Manuel, Esmatjes Enric, De Pablos-Velasco Pedro, Fácila Lorenzo, Mazón Pilar, Gamarra Javier, Hernánd-Moreno Julio, Gràcia Silvia, Montañés Rosario, Llisterri-Caro José L, De Santiago Ana, Sánchez-Celaya Marta, Tranche Salvador. Consensus document for the detection and management of chronic kidney disease. Endocrinol Nutr 2014.
[2] Luyckx Valerie A, Tonelli Marcello, Stanifer John W. The global burden of kidney disease and the sustainable development goals. Bull World Health Organ 2018.
[3] Rhee Connie M, Kovesdy Csaba P. Spotlight on CKD deaths—increasing mortality worldwide. Nat Rev Nephrol 2015.
[4] Goff David C, Lloyd-Jones Donald M, Bennett Glen, Coady Sean, D'Agostino Ralph B, Gibbons Raymond, Greenland Philip, Lackland Daniel T, Levy Daniel, O'Donnell Christopher J, Robinson Jennifer G, Sanford Schwartz J, Shero Susan T, Smith Sidney C, Paul Sorlie, Stone Neil J, Wilson Peter WF. ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. J Am Coll Cardiol 2013;2014.
[5] Hippisley-Cox Julia, Coupland Carol, Vinogradova Yana, Robson John, Rubin Minhas, Aziz Sheikh, Brindle Peter. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008.
[6] D'Agostino Ralph B, Vasan Ramachandran S, Pencina Michael J, Wolf Philip A, Cobain Mark, Massaro Joseph M, Kannel William B. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation 2008.
[7] Ridker Paul M, Buring Julie E, Rifai Nader, Cook Nancy R. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. J Am Med Assoc 2007.
[8] Dreiseitl Stephan, Ohno-Machado Lucila. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inf 2002.
[9] Macias E, Morell A, Serrano J, Vicario JL. Knowledge extraction based on wavelets and dnn for classification of physiological signals: arousals case. In: Computing in Cardiology conference (CinC). vol. 45. IEEE; 2018. p. 1–4. 2018.
[10] Alanazi Hamdan O, Abdullah Abdul Hanan, Qureshi Kashif Naseer. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. J Med Syst 2017.
[11] Miller Rebecca, Tumin Dmitry, Cooper Jennifer, Hayes Don, Joseph D, Tobias. Prediction of mortality following pediatric heart transplant using machine learning algorithms. Pediatr Transplant 2019.
[12] Ross Elsie Gyang, Shah Nigam H, Dalman Ronald L, Nead Kevin T, Cooke John P, Leeper Nicholas J. The use of machine learning for the identification of peripheral artery disease and future mortality risk. J Vasc Surg 2016;64(5):1515–22.
[13] Çelik Güner, Baykan Ömer K, Kara Yakup, Tireli Hülya. Predicting 10-day mortality in patients with strokes using neural networks and multivariate statistical methods. J Stroke Cerebrovasc Dis 2014;23(6):1506–12.
[14] Wang Guanjin, Lam Kin Man, Deng Zhaohong, Choi Kup Sze. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. Comput Biol Med 2015.
[15] Mauri Joan M, Clèries Montse, Vela Emili, Catalan Renal Registry. Design and validation of a model to predict early mortality in haemodialysis patients. Nephrol Dial Transplant 2008.
[16] Srinivasan Beddhu, Bruns Frank J, Saul Melissa, Patricia Seddon, Mark L, Zeidel. A simple comorbidity scale predicts clinical outcomes and costs in dialysis patients. Am J Med 2000.

[17] Liu Jiannong, Huang Zhi, Gilbertson David T, Foley Robert N, Collins Allan J. An improved comorbidity index for outcome analyses among dialysis patients. Kidney Int 2010.
[18] Couchoud Cécile, Labeeuw Michel, Olivier Moranne, Vincent Allot, Esnault Vincent, Frimat Luc, Bénédicte Stengel. A clinical score to predict 6-month prognosis in elderly patients starting dialysis for end-stage renal disease. Nephrol Dial Transplant 2009.
[19] Otero-López M, Martínez-Ocaña Juan C, Betancourt-Castellanos Loreley, Rodríguez-Salazar Eleonora, García-García Manuel. Two prognostic scores for early mortality and their clinical applicability in elderly patients on haemodialysis: poor predictive success in individual patients. Nefrologia 2012;32(2):213–20.
[20] Salekin Asif, Stankovic John. Detection of chronic kidney disease and selecting important predictive attributes. In: IEEE international conference on healthcare informatics (ICHI), vol. 2016. IEEE; 2016. p. 262–70.
[21] Almansour Njoud Abdullah, Syed Hajra Fahim, Khayat Nuha Radwan, Altheeb Rawan Kanaan, Juri Renad Emad, Alhiyafi Jamal, Alrashed Saleh, Sunday O, Olatunji. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. Comput Biol Med 2019;109:101–11.
[22] Doi Toshiki, Yamamoto Suguru, Morinaga Takatoshi, Ken-ei Sada, Kurita Noriaki, Onishi Yoshihiro. Risk score to predict 1-year mortality after haemodialysis initiation in patients with stage 5 chronic kidney disease under predialysis nephrology care. PloS One 2015;10(6).
[23] Ion Titapiccolo Jasmine, Ferrario Manuela, Cerutti Sergio, Barbieri Carlo, Mari Flavio, Gatti Emanuele, Gabriella Signorini Maria. Artificial intelligence models to stratify cardiovascular risk in incident hemodialysis patients. Expert Syst Appl 2013.
[24] Oguz Akbilgic, Obi Yoshitsugu, Potukuchi Praveen K, Karabayir Ibrahim, Nguyen Danh V, Soohoo Melissa, Streja Elani, Molnar Miklos Z, Rhee Connie M, Kalantar-Zadeh Kamyar, Kovesdy Csaba P. Machine learning to identify dialysis patients at high death risk. Kidney Int. Rep. 2019;4(9):1219–29.
[25] Breiman Leo, Leo. Random forests. Mach Learn 2001;45(1):5–32.
[26] Darst Burcu F, Malecki Kristen C, Engelman Corinne D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet 2018;19(1):65.
[27] Anders Krogh, Hertz John A. A simple weight decay can improve generalization. In: Advances in neural information processing systems; 1992. p. 950–7.
[28] Rumelhart David E, Hinton Geoffrey E, Williams Ronald J. Learning representations by back-propagating errors. Nature 1986;323(6088):533–6.
[29] Kingma Diederik P, Jimmy Ba. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv, 1412.6980.
[30] Wagner Martin, Ansell David, Kent David M, Griffith John L, Naimark David, Wanner Christoph, Tangri Navdeep. Predicting mortality in incident dialysis patients: an analysis of the United Kingdom renal registry. Am J Kidney Dis 2011.
[31] Isshiki Keiji, Nishio Toshiki, Isono Motohide, Makiishi Tetsuya, Shikano Tsutomu, Tomita Koubin, Nishio Toshiji, Kanasaki Masami, Maegawa Hiroshi, Uzu Takashi. Glycated albumin predicts the risk of mortality in type 2 diabetic patients on hemodialysis: evaluation of a target level for improving survival. Ther Apher Dial 2014;18(5):434–42.