Journal of Applied Biomedicine

Original research article

A decision support system for the prediction of mortality in patients with acute kidney injury admitted in intensive care unit

Selda Kayaaltı [1] *, Ömer Kayaaltı [2], Bekir Hakan Aksebzeci [3]

[1] *Develi Hatice-Muammer Kocatürk Public Hospital, Department of Anesthesiology and Reanimation, Develi, Kayseri, Turkey*
[2] *Kayseri University, Develi Hüseyin Şahin Vocational College, Department of Computer Technologies, Kayseri, Turkey*
[3] *Abdullah Gül University, Faculty of Engineering, Department of Biomedical Engineering, Kayseri, Turkey*

Abstract
Intensive care unit (ICU) is a very special unit of a hospital, where healthcare professionals provide treatment and, later, close follow-up to the patients. It is crucial to estimate mortality in ICU patients from many viewpoints. The purpose of this study is to classify the status of patients with acute kidney injury (AKI) in ICU as early mortality, late mortality, and survival by the application of Classification and Regression Trees (CART) algorithm to the patients' attributes such as blood urea nitrogen, creatinine, serum and urine neutrophil gelatinase-associated lipocalin (NGAL), alkaline phosphatase, lactate dehydrogenase (LDH), gamma-glutamyl transferase, laboratory electrolytes, blood gas, mean arterial pressure, central venous pressure and demographic details of patients. This study was conducted 50 patients with AKI who were followed up in the ICU. The study also aims to determine the significance of relationship between the attributes used in the prediction of mortality in CART and patients' status by employing the Kruskal–Wallis *H* test. The classification accuracy, sensitivity, and specificity of CART for the tested attributes for the prediction of early mortality, late mortality, and survival of patients were 90.00%, 83.33%, and 91.67%, respectively. The values of both urine NGAL and LDH on day 7 showed a considerable difference according to the patients' status after being examined by the Kruskal–Wallis *H* test.

Keywords: Acute kidney injury; Classification and regression trees; Lactate dehydrogenase; Mortality prediction; Neutrophil gelatinase-associated lipocalin

Highlights:
- A simple and straightforward decision support system with high classification accuracy has been developed.
- Urine neutrophil gelatinase-associated lipocalin is associated with early mortality in patients with acute kidney injury in intensive care unit.
- Lactate dehydrogenase is associated with late mortality in patients with acute kidney injury in intensive care unit.

Introduction

The estimation of mortality in the intensive care patients is beneficial from many viewpoints of patient care such as triage, treatment, resource planning, and providing information to patients and/or their relatives about the disease (Lee et al., 2015). Since the early 1980s, various scoring systems, such as The Acute Physiology and Chronicler Health Evaluation (APACHE), Sequential Organ Failure Assessment, and The Simplified Acute Physiology Score, have been employed for predicting the severity of disease in the intensive care patients. No definitive assessment criteria have yet been found because of the complexity of the data that need to be specific for each disease (Xu et al., 2017). Therefore, the mortality rate can be estimated approximately by the evaluation of essential demographic, physiological, clinical, and laboratory test data (Lee et al., 2015).

Multiple organ failure is the one of leading causes of deaths in the intensive care unit (ICU). For this reason, especially in order to follow the course of the disease, six organ systems, including cardiovascular, respiratory, renal, hepatic, hematologic, and central nervous system, are monitored closely during the treatment process (Mehta and Gupta, 2015).

Acute kidney injury (AKI) is known as a sudden or rapid decline in renal function. AKI is observed in up to 25% of patients undergoing intensive care treatment. It is known to have high mortality rates of up to 40–50% (Vincent, 2007). The mortality rate of AKI is still high despite various technical advances. AKI is also effective in the long-term prognosis of patients. In order to improve the prognosis in patients with AKI, it is vital to identify the various risk factors for AKI mortality. There are a number of parameters that are associated with short- and long-term mortality rates in AKI (Schmitz et al., 2017). Schmitz et al. (2017) showed that age and APACHE II scores affect the mortality rate, whereas cardio-

vascular risk factors, such as diabetes, hypertension, stroke, and so on, do not affect the mortality rate. Coca et al. (2014) demonstrated that high levels of urine IL8, neutrophil gelatinase-associated lipocalin (NGAL), kidney injury molecule 1, and liver-type fatty acid binding protein increased the risk of long-term mortality.

Medical/clinical databases that are stored in the hospital or laboratory systems include information about the patients' medical/surgical history along with physiological, and biochemical parameters. The discovery and understanding of the relationships between data in these databases is a vital issue in terms of obtaining new information. Especially in the analysis of large quantities of data, the development of new methods and tools has become possible with the increase of computing power of computers. Data mining provides prediction models by using medical/clinical data to support the doctors' decisions in the diagnosis and treatment of diseases. Furthermore, it also allows for the discovery of patterns (Koh and Tan, 2011). In addition to that, data mining has been successfully used to identify and model the relationship between a healthy person and a patient that has various diseases such as cancer, heart disorders, and hypertension by comparing the personal data such as demographic data, laboratory findings, and physiological data (Worachartcheewan et al., 2018).

The purpose of this study is to develop a decision support system for the determination of early mortality, late mortality, and survival of patients with AKI undergoing treatment in the ICU. Therefore, this study also aims to examine the physiological parameters and/or laboratory tests of patients for the detection of early mortality, late mortality, and survival.

Material and methods

Dataset
Our study employed the dataset (Nader et al., 2018) of a prospective study conducted by Mahmoodpoor et al. (2018). The study of Mahmoodpoor et al. (2018) was approved by the Ethics Committee at the Tabriz University of Medical Sciences. Their study included data from 50 adult patients admitted to the ICU at a major university-affiliated teaching hospital. Written informed consents were obtained from the patients or their predetermined health decision-makers. The criterion for the study entry was determined as less than 200% increase in the creatinine level according to the KDIGO guidelines. Exclusion criteria of this study included individuals under the age of 18 years, individuals who were previously diagnosed with chronic renal failure (stage 3 or higher and glomerular filtration rate (GFR) less than 60 ml/min/1.73 m^2), hypovolemia, individuals with urinary tract obstruction, individuals with metabolic alkalosis, individuals with chronic diuretic use, pregnant women, individuals with corticosteroid use, and individuals with malignant and autoimmune diseases. The mortality cases of patients with AKI within 28 (early mortality) and 90 days (late mortality) were evaluated in this dataset.

The values of plasma NGAL (pNGAL), urinary NGAL (uNGAL), blood urea nitrogen (BUN), creatinine, lactate dehydrogenase (LDH), alkaline phosphatase (ALP), and gamma-glutamyl transferase (GGTP) were recorded on the day of admission. Afterwards, these values were recorded on both days 3 and 7 after the admission of the patients. In addition, the values of serum electrolytes (Na, K) and blood gas (pH, H$^+$) were also recorded on the day of admission. During the study, patients were followed up for the values of blood pressure and volume overload. Norepinephrine infusion was employed for supporting the blood pressure. Renal replacement therapy (RRT) was employed in the case of volume overload and acidosis that is resistant to medical treatment. Furthermore, this dataset includes variables such as gender, age, comorbid disease, mortality risk score (APACHE II), vasopressor requirement, diuretic use, central venous pressure, mean arterial pressure, GFR (Chronic Kidney Disease Epidemiology Collaboration and Modification of Diet Renal Disease), early mortality, late mortality, and some values calculated from these attributes. The values of attributes provided in the dataset are presented in Table 1.

Data mining (Classification)
Classification, one of data mining methods, models the relationship between the attributes of data (training data) and pre-defined class labels. Several methods/algorithms, such as decision trees, rule-based classifiers, Naive Bayes classifiers, neural networks, and SVMs, can be employed in the creation of a classification model. After the modeling process, the performance of the classifier is evaluated on the test data.

A decision tree is a diagram similar to a flowchart, which shows the various results of a series of decisions. Decision trees are preferred for solving the classification and regression problems because their rules are easily applicable and understandable. In addition, they can handle both numerical and categorical data, and have low computational cost. Decision trees have three main parts: root node, leaf nodes and branches. The root node is the starting point of the tree. Root and leaf nodes contain questions or criteria to be answered. The branches are the paths that connect the nodes and show the flow between the questions and answers (Han et al., 2012).

Training dataset, attributes, and attribute selection method are required for the basic tree-building algorithm. An attribute selection method is used for the determination of the attributes that best divide the given tuples into their classes (Han et al., 2012). Information gain, gain ratio, and Gini index are often used to determine the best separating attribute (Rokach and Maimon, 2005). The decision tree creation is a recursive process, in which the decision tree starts with a single root node where all the data have formed. It selects the best separating attribute with the help of an attribute selection method, thereby creating a node. The tree algorithm terminates the node as a leaf and defines the class label whether all the samples in the dataset belong to the same class or if there is no attribute to break. Otherwise, the attribute selection is done (node attribute and criterion are determined). Algorithm divides the dataset into subgroups. Unless the termination criterion is provided for each subgroup, the process recursively continues division and adding new nodes. If all the samples belong to the same class or there are no remaining attributes or there are no tuples for a given branch, the finishing criterion is provided (Han et al., 2012).

Although various decision tree building algorithms are in the literature, usually ID3, C4.5, and Classification and Regression Trees (CART) are widely used (Singh and Gupta, 2014). In our study, CART were chosen because the attributes of the dataset contain real numbers and the output value (class label) is discrete (early mortality, late mortality, and survival). The Gini index is used as the attribute selection method in CART.

The CART algorithm was implemented in (MATLAB R2018b, The MathWorks Inc., Natick, MA, USA). There was no restriction on the maximum tree depth. Maximum number of decision splits, minimum number of leaf nodes and minimum number of branch nodes were equal to "sample size"-1, 1 and

Table 1. Patients' characteristics and laboratory results of patients with AKI on the day of admission. (Normally distributed numerical variables are presented as "mean ± standard deviation," whereas non-normally distributed numerical variables are presented as "median [interquartile range]". Categorical variables are provided as frequency.)

Demographic data/ Laboratory results	Early mortality (death within 28 days) (N = 12)	Late mortality (death between 28 and 90 days) (N = 11)	Survival (after 90 days) (N = 27)
Gender	F = 5, M = 7	F = 4, M = 7	F = 9, M = 18
Age	67.67 ± 11.89	61.18 ± 14.69	61.67 ± 9.43
APACHE II score	30.58 ± 3.80	23.64 ± 4.03	21.67 ± 3.96
Central venous pressure (mm Hg)	17.17 ± 4.49	14.64 ± 2.84	12.56 ± 3.11
Glomerular filtration rate by CKD-EPI (ml/min)	21.11 ± 4.82	23.55 ± 10.05	22.95 ± 4.80
Glomerular filtration rate by MDRD (ml/min)	22.68 ± 4.51	24.62 ± 9.39	24.17 ± 4.90
Mean arterial pressure (mm Hg)	87.50 ± 24.05	89.24 ± 25.85	81.30 ± 21.98
Need for vasopressors	12	5	5
Need for diuresis	5	3	17
Blood urea nitrogen (mg/dl)	45.00 [10.50]	43.00 [13.25]	43.00 [9.75]
Serum creatinine (mg/dl)	2.75 [0.55]	2.90 [0.95]	2.70 [0.37]
Alkaline phosphatase (I.U.)	12.30 [0.80]	11.50 [1.70]	10.70 [1.10]
Gamma-glutamyl transferase (I.U.)	104.05 [25.20]	107.00 [46.00]	96.00 [17.30]
Lactate dehydrogenase (I.U.)	666.00 [142.00]	647.00 [236.00]	573.00 [199.50]
Plasma NGAL (ng/ml)	126.00 [12.00]	126.00 [21.75]	111.00 [39.00]
Urine NGAL (ng/ml)	61.00 [18.50]	65.00 [12.50]	61.00 [14.25]
Serum sodium (mEq/l)	137.00 [6.50]	137.00 [5.75]	140.00 [8.50]
Serum potassium (mEq/l)	3.62 [0.75]	4.00 [0.85]	4.23 [0.88]
pH	7.42 [0.10]	7.42 [0.14]	7.37 [0.10]
Plasma H+ (nmol/l)	37.60 [9.80]	38.00 [12.98]	42.70 [9.92]

10, respectively. After the decision tree was created, it was pruned one level from the leaf node.

The classification performance of the designed decision tree was evaluated according to accuracy, sensitivity, and specificity. Accuracy is the percentage of test samples that are correctly classified by the classifier. Sensitivity (true positive rate) is the proportion of true predicted positive samples over real positive samples. Specificity (true negative rate) is the proportion of true predicted negative samples over real negative samples (Han et al., 2012).

In the literature, while a classification model is built, generally 70–90% of the dataset is used as training set. In our study, we chose 80% of the dataset was for training and the rest part was for test. Decision tree was modelled in MATLAB software. Patients included in the training set were not included in the test set. The distribution according to the group of patients in the training and test sets are equal. In this way, different training and test data sets were created and decision trees were created. Correlation method was used to determine whether collinearity was found among the attributes used in the decision tree. Predictors/Attributes were considered to be collinear if the correlation coefficient exceeds 0.7 (Dormann et al., 2013). One of the decision trees that has highest test dataset classification accuracy rate and consists non-collinear attributes was selected and its results were presented.

Statistical approach/analysis

From the statistical perspective, in order to determine which tests can be applied to the data, the normality of the distribution of variables was examined by the Kolmogorov–Smirnov test. The Kruskal–Wallis H test, which is applicable to nonparametric and multi-class, was used to test whether there was a significant difference among the used attributes in the decision tree. These variables were also analyzed by using correlation analysis for the collinearity test. Statistical methods were applied by using SPSS 22.0 (Statistical Package of Social Sciences Inc., Chicago, IL, USA).

Results

The decision tree obtained by using the CART algorithm is presented in Fig. 1. Patients whose uNGAL score on day 7 were greater than or equal to 69 on the root (first) node of the decision tree were assigned to the "early mortality" group. On the first node, 12 of the 50 patients were assigned to the "early mortality" group. These patients were correctly classified because their real condition was in the "early mortality" group. If the patient's uNGAL value on day 7 is less than 69, the second node is passed. On the second node of the decision tree, patients with a LDH score on day 7 are greater than or equal to 627 were assigned to the late mortality group. In this node, eight patients were assigned to "late mortality" group. When the actual conditions of these patients were examined, seven patients were assigned in the late mortality group and one patient was assigned in the "survival" group. In total, 11 patients were in the late mortality group in the dataset and 7 of them were correctly classified in this node. At this stage, 1 out of 27 patients in the survival group was incorrectly classified as "late mortality" group. If the LDH value of the day 7 is less than 627, the third node is passed. On the third (final) node, patients with a pH value are greater than or equal to 7.484 were assigned to the "late mortality" group, otherwise, they were assigned to the survival group. At this node, 2 out of 11 patients

in the "late mortality" group and 26 out of 27 patients in the "survival" group were correctly classified. In summary, it was observed that one patient was assigned to the "survival" group on the second node and two patients assigned to the "late

Fig. 1. Decision tree model for patients with AKI in ICU

mortality" group on the third node were classified incorrectly. A total of three patients were assigned to different groups.

The confusion matrix for training and test datasets with the criteria used in the CART is shown in Table 2. Classification accuracy, sensitivity, and specificity of the decision tree representation for training dataset were 95.00%, 94.71%, and 97.17% respectively. Classification accuracy, sensitivity, and specificity of it for test dataset were 90.00%, 83.33%, and 91.67% respectively.

The distribution of these attributes (uNGAL, LDH, and pH) for each patient group is given in Fig 2.

The Kruskal–Wallis H test was applied for the significance between the attributes (uNGAL, LDH, and pH) and patients' status (early mortality, late mortality, and survival). Test results are provided in Table 3. The uNGAL and LDH attributes on day 7 showed a significant difference according to the patient's condition ($p < 0.001$). In addition, it was found that there was also a tendency of high significance among the patients' status when the pH attribute is compared ($p = 0.097$).

There was no observed collinearity among the attributes used in CART because the correlation coefficients are less than 0.7 (Supplementary Table 1).
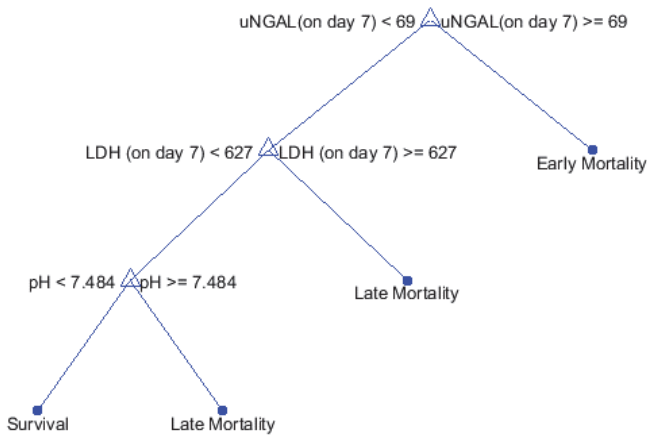
Table 2. The confusion matrix of the decision tree model for training and test datasets

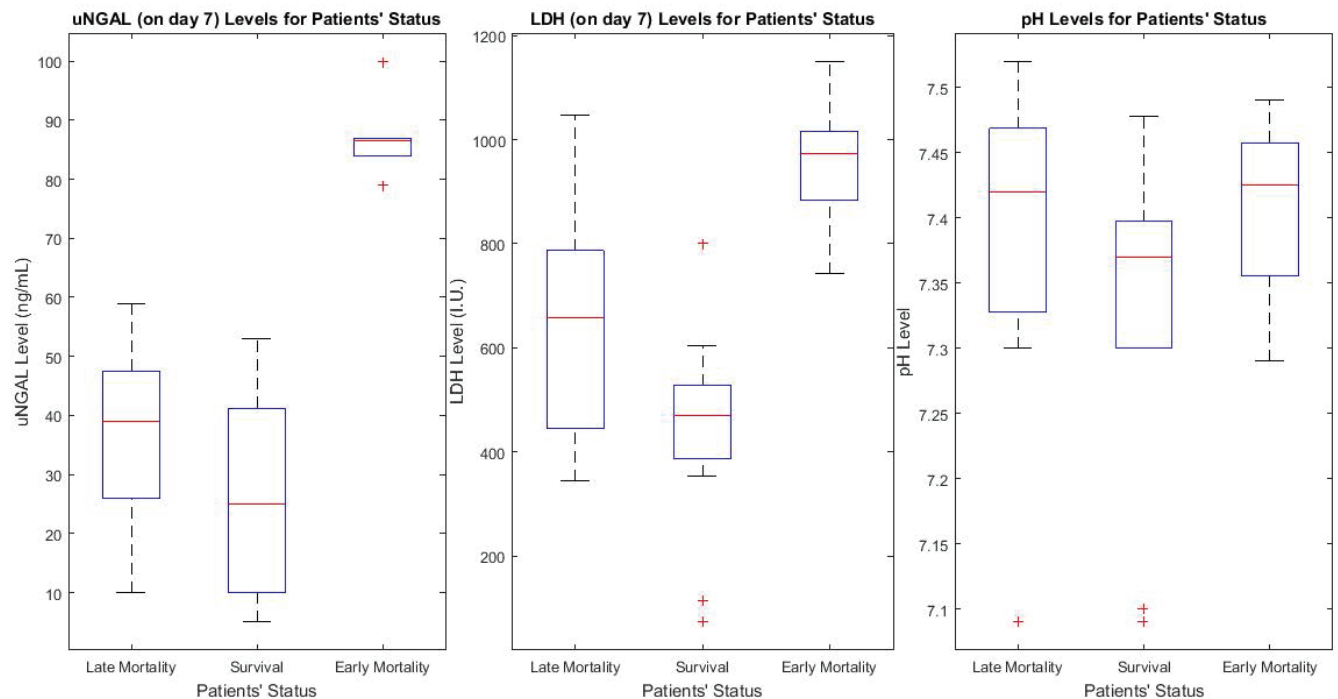|		Prediction					
		Training dataset (40 patients)			Test dataset (10 patients)		
		Survival	Early mortality	Late mortality	Survival	Early mortality	Late mortality
Actual	Survival	20	0	1	6	0	0
	Early mortality	0	10	0	0	2	0
	Late mortality	1	0	8	1	0	1

Fig. 2. Box plot representations of the uNGAL (on day 7), LDH (on day 7), and pH values with respect to patients' status/group

Table 3. Kruskal–Wallis *H* test statistics for uNGAL (on day 7), LDH (on day 7), and pH values with respect to the patients' status/group

	uNGAL (on day 7)	LDH (on day 7)	pH
Chi-Square	29.401	27.792	4.661
Df	2	2	2
Asymp. Sig.	<0.001	<0.001	0.097

Discussion

The purpose of this study was to determine the parameters/attributes that can be employed for prediction of mortality in patients who develop AKI after being admitted to an ICU. We tried to estimate the mortality of the patients by the application of data mining methods to patient's attributes, such as NGAL, APACHE II score, vasopressor, and diuretic need, which are some of the early markers of AKI and the data obtained from the routine laboratory tests in ICU. A decision tree was created with the CART algorithm to achieve this purpose.

We developed a simple decision tree, and the training and test accuracy rates of that tree were 95.00% and 90.00%, respectively. If the decision tree of three nodes is finished with two nodes, two patients in the "late mortality" group at node 2 will be assigned to the "survival" group and there will be five errors in total. If the decision tree is designed with two nodes, then it classifies the entire dataset with 90.00% accuracy. In each case, the decision tree can predict the condition of the patients with a high accuracy rate. In decision trees, overfitting occurs when the tree is designed to perfectly fit all the samples in the training dataset. In this situation, there are many branches to categorize each sample in the classes. A tree that overfits the training data provides high classification accuracy for training data but the predictive performance is poor on the unseen test data. The proposed decision tree model has very few nodes and high classification accuracy in both training and test datasets. In this respect, the decision tree is a good classification model without overfitting. In the second part of the study, the Kruskal–Wallis *H* test was applied to evaluate whether the attributes used in the decision tree differ according to the patient's condition (early mortality, late mortality, and survival). Both the proposed decision tree model and the statistical method were used to determine that the uNGAL and LDH values are the parameters/attributes that can be used in the prediction of mortality in the patients with AKI. NGAL, normally secreted at a lower level from the epithelial cells of kidneys, is an early biomarker for the determination of ischemic, septic, or nephrotoxic renal damage. In the case of tubular epithelial damage, its secretion begins within three hours. Both uNGAL and pNGAL levels increase rapidly in proportion to the duration and severity of the injury (Nickolas et al., 2008). If the patient has prerenal or transient azotemia without kidney damage, then the NGAL levels do not exceed the normal levels in this case (Belcher et al., 2014). In this regard, both early and discriminative diagnosis with AKI and NGAL is better than serum Cr (Au et al., 2016). Shyam et al. (2017) also reported that the use of serum Cr levels in conjunction with uNGAL, instead of only using serum Cr, can improve the patient management and reduce the expected costs in the diagnosis of AKI.

Various studies (Gungor et al., 2014; Haase et al., 2009; Ralib et al., 2014; Shyam et al., 2017; Wang et al., 2014; Yang et al., 2014) have demonstrated that NGAL could be used in the prediction of mortality. Parikh et al. (2017) showed that there is a considerable increase in the NGAL levels after shock (circulatory) in the patients undergoing intensive care. It was stated that NGAL could be used for predicting the development of AKI before one to three days according to creatinine level; therefore, it would serve as a suitable early descriptive biomarker for AKI. Gungor et al. (2014) prospectively monitored the uNGAL and pNGAL levels of 64 patients [30 patients with HRS (+)] with cirrhosis. Their study concluded that NGAL could be used independently in the prediction of mortality for other commonly used risk factors in patients with hepatorenal syndrome. Yang et al. (2014) conducted their study in patients with burn (burn area more than 20%), in which they collected the values of serum Cr, serum cystatin C, uNGAL, and pNGAL from the samples at 0, 3, 6, 12, 24, and 48 hours. The data of these samples were evaluated in order to determine the predictive characteristics of these biomarkers in terms of the risks of AKI and mortality. The maintained high levels of uNGAL and pNGAL until the 12th hour were considered as high risk factors in terms of early development of AKI and early mortality. However, early uNGAL levels have failed to predict late AKI or late mortality. The maintained high levels of uNGAL and pNGAL at the 48th hour were found to be related to AKI and mortality independently from other factors. In addition to these studies, in many of the studies on patients in shock, NGAL has been shown to be a good prognostic agent in predicting the need for RRT and mortality (Haase et al., 2009; Ralib et al., 2014; Wang et al., 2014). By using the same dataset, our previous study (Kayaaltı et al., 2018) demonstrated that uNGAL levels are an important parameter/attribute for the prediction of an early mortality rate. However, our previous study found that only statistically significant parameters were associated with early mortality, whereas different data mining methods were used in this study. In this way, we found that in addition to the relationship between uNGAL and early mortality, LDH, and pH were also associated with late mortality.

In our study, the LDH parameter on day 7 was a decisive factor on the second node of our decision tree model. LDH is a fast and universally accessible biomarker and a widely released intracellular enzyme, which is also used in the routine care of patients admitted in ICU due to its low cost. During hypoxia, it is transformed into the first pyruvate molecules and then to lactate molecules (Morello et al., 2016). High LDH levels are observed in the cases of serious diseases, such as tissue damage, necrosis, hypoxia, hemolysis, or malignancy (Eigentler et al., 2011; Karlsson et al., 2010; Kato et al., 2006). Various studies (Cui et al., 2017; Hu et al., 2015; Leite et al., 2013; Morello et al., 2016; Sirikutt and Kalayanarooj, 2014; Wu et al., 2016; Zein et al., 2004) have detected high serum levels of LDH in several diseases (acute pancreatitis, acute aortic syndrome, sepsis, pulmonary embolism, metabolic syndrome and pulmonary arterial hypertension), and these studies concluded that LDH can be used for prognosis.

Our study used the proposed LDH values of the day 7 (> = 627) on the decision tree model to decide for late mortality. Mahmoodpoor et al. (2018) pointed out that there is a relationship between LDH levels along with progression of AKI and early mortality in their studies. Our study also demonstrated that LDH can be used to predict late mortality by both decision tree and statistical methods. Cui et al. (2017) investigated acute pancreatitis with the association between LDH and persistent organ damage. Their study further confirmed

that the LDH values measured in the admission of patients are a potential prognostic factor. Zein et al. (2004) investigated the utility of high LDH levels in determining the severity of tissue damage in sepsis and septic shock. They also concluded that LDH is a marker that reflects the degree of tissue damage. Leite et al. (2013) demonstrated in their study, which evaluated 165 pulmonary embolism patients retrospectively, that LDH is a crucial parameter in the prediction of all-cause mortality in the cases of both hospital and non-hospital. Wu et al. (2016) conducted a comprehensive retrospective study to determine the association of LDH with mortality in patients with metabolic syndrome. As a result, this observation shows that there is a positive correlation between high LDH levels and all-cause mortality in patients with metabolic syndrome. Sirikutt and Kalayanarooj (2014) examined LDH levels in patients with dengue hemorrhagic fever with shock and hepatic injury. Their study found high LDH levels in these patients. Deterioration in the function of the liver provides information about the severity of heart failure. For this reason, high LDH levels are expected in the patients with heart failure and liver dysfunction. Hu et al. (2015) aimed to determine the prognostic significance of LDH in patients who have a severe heart failure with idiopathic pulmonary arterial hypertension (IPAH). They determined that high LDH levels were a risk factor for mortality in patients with IPAH. In most of these studies, a single LDH value was examined. Only Sirikutt and Kalayanarooj (2014) considered the LDH values at three different times. In most of these studies, the LDH values at the time of admission were used to determine mortality or prognosis. Our study showed that LDH levels on the day 7 were associated with late mortality in AKI. Alzahri et al. (2015) and Zager et al. (2013) have demonstrated the relationship between LDH and kidney damage. Our study found that LDH can also be used in the prediction of mortality in patients with AKI.

We think that the correlation of mortality with the day 7 values of uNGAL and LDH is also essential in terms of evaluating the response to treatment in patients with AKI. We think that the fact that the last day data (LDH, uNGAL) of the patients showed a high success rate in the prediction of mortality would be very helpful in clinical practice in the continuation of the current treatment or switching to another treatment option. The limitation of our study is the small sample size of patients included in the study. There is a need for further studies with more patients.

Conclusions

In the proposed decision tree model, this study determined that the uNGAL levels on day 7 were associated with early mortality, whereas the LDH levels on day 7 and pH value were associated with late mortality. After performing the statistical test, these attributes were observed to be related to mortality. Often, single measurements at the time of diagnosis or at the time of admission to hospital may be insufficient for the prediction of mortality. Further studies are needed to answer the question of whether more accurate results/higher success rates can be achieved with the values of uNGAL and LDH (which we have shown that these can be used as follow-up parameters) before or after day 7.

Conflict of interests
The authors have no conflict of interests to declare.

References

Alzahri MS, Mousa SA, Almomen AM, Hasanato RM, Polimeni JM, Racz MJ (2015). Lactate dehydrogenase as a biomarker for early renal damage in patients with sickle cell disease. Saudi J Kidney Dis Transpl 26(6): 1161–1168. DOI: 10.4103/1319-2442.168596.

Au V, Feit J, Barasch J, Sladen RN, Wagener G (2016). Urinary neutrophil gelatinase–associated lipocalin (NGAL) distinguishes sustained from transient acute kidney injury after general surgery. Kidney Int Rep 1(1): 3–9. DOI: 10.1016/j.ekir.2016.04.003.

Belcher JM, Sanyal AJ, Peixoto AJ, Perazella MA, Lim J, Thiessen-Philbrook H, et al. (2014). Kidney biomarkers and differential diagnosis of patients with cirrhosis and acute kidney injury. Hepatology 60(2): 622–632. DOI: 10.1002/hep.26980.

Coca SG, Garg AX, Thiessen-Philbrook H, Koyner JL, Patel UD, Krumholz HM, et al. (2014). Urinary biomarkers of AKI and mortality 3 years after cardiac surgery. J Am Soc Nephrol 25(5): 1063–1071. DOI: 10.1681/ASN.2013070742.

Cui J, Xiong J, Zhang Y, Peng T, Huang M, Lin Y, et al. (2017). Serum lactate dehydrogenase is predictive of persistent organ failure in acute pancreatitis. J Crit Care 41: 161–165. DOI: 10.1016/j.jcrc.2017.05.001.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1): 27–46. DOI: 10.1111/j.1600-0587.2012.07348.x.

Eigentler T, Figl A, Krex D, Mohr P, Mauch C, Rass K, et al. (2011). Number of metastases, serum lactate dehydrogenase level, and type of treatment are prognostic factors in patients with brain metastases of malignant melanoma. Cancer 117(8): 1697–1703. DOI: 10.1002/cncr.25631.

Gungor G, Ataseven H, Demir A, Solak Y, Gaipov A, Biyik M, et al. (2014). Neutrophil gelatinase-associated lipocalin in prediction of mortality in patients with hepatorenal syndrome: a prospective observational study. Liver Int 34(1): 49–57. DOI: 10.1111/liv.12232.

Haase M, Bellomo R, Devarajan P, Schlattmann P, Haase-Fielitz A (2009). Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis 54(6): 1012–1024. DOI: 10.1053/j.ajkd.2009.07.020.

Han J, Pei J, Kamber M (2012). Data mining: concepts and techniques (3rd edition). Morgan Kaufmann.

Hu EC, He JG, Liu ZH, Ni XH, Zheng YG, Gu Q, et al. (2015). High levels of serum lactate dehydrogenase correlate with the severity and mortality of idiopathic pulmonary arterial hypertension. Exp Ther Med 9(6): 2109–2113. DOI: 10.3892/etm.2015.2376.

Karlsson M, Wiberg-Itzel E, Chakkarapani E, Blennow M, Winbladh B, Thoresen M (2010). Lactate dehydrogenase predicts hypoxic ischaemic encephalopathy in newborn infants: a preliminary study. Acta Paediatr 99(8): 1139–1144. DOI: 10.1111/j.1651-2227.2010.01802.x.

Kato GJ, Mcgowan V, Machado RF, Little JA, Taylor J, Morris CR, et al. (2006). Lactate dehydrogenase as a biomarker of hemolysis-associated nitric oxide resistance, priapism, leg ulceration, pulmonary hypertension, and death in patients with sickle cell disease. Blood 107(6): 2279–2285. DOI: 10.1182/blood-2005-06-2373.

Kayaaltı S, Kayaaltı Ö, Aksebzeci BH (2018). Relationship between neutrophil gelatinase–associated lipocalin and mortality in acute kidney injury. Turk J Intense Care 16(3): 101–108. DOI: 10.4274/tybd.86158.

Koh HC, Tan G (2011). Data mining applications in healthcare. J Healthc Inf Manag 19(2): 65. PMID: 15869215.

Lee J, Maslove DM, Dubin JA (2015). Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One 10(5): e0127428. DOI: 10.1371/journal.pone.0127428.

Leite L, Moura J, Ferreira R, Lazaro S, Madaleno J, Moreira N, et al. (2013). LDH as a predictor of in-hospital and late mortality in acute pulmonary embolism. Eur Heart J 34(suppl_1): p1154. DOI: 10.1093/eurheartj/eht308.P1154.

Mahmoodpoor A, Hamishehkar H, Fattahi V, Sanaie S, Arora P, Nader ND (2018). Urinary versus plasma neutrophil gelatinase-associated lipocalin (NGAL) as a predictor of mortality for acute kidney injury in intensive care unit patients. J Clin Anesth 44: 12–17. DOI: 10.1016/j.jclinane.2017.10.010.

Mehta Y, Gupta M (2015). Textbook of critical care: two volume set, JP Medical Ltd.

Morello F, Ravetti A, Nazerian P, Liedl G, Veglio MG, Battista S, et al. (2016). Plasma lactate dehydrogenase levels predict mortality in acute aortic syndromes: a diagnostic accuracy and observational outcome study. Medicine (Baltimore) 95(6): e2776. DOI: 10.1097/MD.0000000000002776.

Nader N, Hamishehkar H, Mahmoodpoor A, Sanaie S, Fattahi V, Arora P (2018). [Dataset] Data for: Neutrophil gelatinase-associated lipocalin (NGAL); An early urinary vs. plasma marker of acute kidney injury in intensive care units. Mendeley Data, v1. DOI: 10.17632/579pxjyjz8.1.

Nickolas TL, O'Rourke MJ, Yang J, Sise ME, Canetta PA, Barasch N, et al. (2008). Sensitivity and specificity of a single emergency department measurement of urinary neutrophil gelatinase–associated lipocalin for diagnosing acute kidney injury. Ann Intern Med 148(11): 810–819. DOI: 10.7326/0003-4819-148-11-200806030-00003.

Parikh A, Rizzo JA, Canetta P, Forster C, Sise M, Maarouf O, et al. (2017). Does NGAL reduce costs? A cost analysis of urine NGAL (uNGAL) & serum creatinine (sCr) for acute kidney injury (AKI) diagnosis. PLoS One 12(5): e0178091. DOI: 10.1371/journal.pone.0178091.

Ralib AM, Pickering JW, Shaw GM, Than MP, George PM, Endre ZH (2014). The clinical utility window for acute kidney injury biomarkers in the critically ill. Crit Care 18(6): 601. DOI: 10.1186/s13054-014-0601-2.

Rokach L, Maimon O (2005). Top-down induction of decision trees classifiers-a survey. IEEE Trans Syst Man Cybern C Cybern 35(4): 476–487. DOI: 10.1109/TSMCC.2004.843247.

Schmitz M, Tillmann FP, Paluckaite A, Laufer EA, Rayner B, Rump LC, et al. (2017). Mortality risk factors in intensive care unit patients with acute kidney injury requiring renal replacement therapy: a retrospective cohort study. Clin Nephrol 88(1): 27–32. DOI: 10.5414/CN109078.

Shyam R, Patel ML, Sachan R, Kumar S, Pushkar DK (2017). Role of urinary neutrophil gelatinase-associated lipocalin as a biomarker of acute kidney injury in patients with circulatory shock.

Indian J Crit Care Med 21(11): 740–745. DOI: 10.4103/ijccm.IJCCM_315_17.

Singh S, Gupta P (2014). Comparative study ID3, cart and C4.5 decision tree algorithm: a survey. Int J Adv Sci Eng Inf Technol 27(27): 97–103.

Sirikutt P, Kalayanarooj S (2014). Serum lactate and lactate dehydrogenase as parameters for the prediction of dengue severity. J Med Assoc Thai 97(6): S220–231. PMID: 25391197.

Vincent JL (2007). Critical care nephrology: a multidisciplinary approach. Contrib Nephrol 156: 24–31. DOI: 10.1159/000102012.

Wang M, Zhang Q, Zhao X, Dong G, Li C (2014). Diagnostic and prognostic value of neutrophil gelatinase-associated lipocalin, matrix metalloproteinase-9, and tissue inhibitor of matrix metalloproteinases-1 for sepsis in the Emergency Department: an observational study. Crit Care 18(6): 634. DOI: 10.1186/s13054-014-0634-6.

Worachartcheewan A, Schaduangrat N, Prachayasittikul V, Nantasenamat C (2018). Data mining for the identification of metabolic syndrome status. EXCLI J 17: 72–78. DOI: 10.17179/excli2017-911.

Wu LW, Kao TW, Lin CM, Yang HF, Sun YS, Liaw FY, et al. (2016). Examining the association between serum lactic dehydrogenase and all-cause mortality in patients with metabolic syndrome: a retrospective observational study. BMJ Open 6(5): e011186. DOI: 10.1136/bmjopen-2016-011186.

Xu J, Zhang Y, Zhang P, Mahmood A, Li Y, Khatoon S (2017). Data mining on ICU mortality prediction using early temporal data: a survey. Int J Inf Technol Decis Mak 16(01): 117–159. DOI: 10.1142/S0219622016300020.

Yang HT, Yim H, Cho YS, Kym D, Hur J, Kim JH, et al. (2014). Assessment of biochemical markers in the early post-burn period for predicting acute kidney injury and mortality in patients with major burn injury: comparison of serum creatinine, serum cystatin-C, plasma and urine neutrophil gelatinase-associated lipocalin. Crit Care 18(4): R151. DOI: 10.1186/cc13989.

Zager RA, Johnson AC, Becker K (2013). Renal cortical lactate dehydrogenase: a useful, accurate, quantitative marker of *in vivo* tubular injury and acute renal failure. PLoS One 8(6): e66776. DOI: 10.1371/journal.pone.0066776.

Zein JG, Lee GL, Tawk M, Dabaja M, Kinasewitz GT (2004). Prognostic significance of elevated serum lactate dehydrogenase (LDH) in patients with severe sepsis. Chest 126(4): 873S. DOI: 10.1378/chest.126.4_MeetingAbstracts.873S.