



# Article Entropy-Based Time Window Features Extraction for Machine Learning to Predict Acute Kidney Injury in ICU

Chun-Te Huang<sup>1</sup>, Rong-Ching Chang<sup>2</sup>, Yi-Lu Tsai<sup>3</sup>, Kai-Chih Pai<sup>2,3</sup>, Tsai-Jung Wang<sup>1</sup>, Chia-Tien Hsu<sup>1</sup>, Cheng-Hsu Chen<sup>1</sup>, Chien-Chung Huang<sup>2,4</sup>, Min-Shian Wang<sup>4</sup>, Lun-Chi Chen<sup>2,3</sup>, Ruey-Kai Sheu<sup>2,3</sup>, Chieh-Liang Wu<sup>1</sup> and Chun-Ming Lai<sup>2,3,\*</sup>

- <sup>1</sup> Nephrology and Critical Care Medicine, Department of Internal Medicine and Critical Care Medicine, Taichung Veterans General Hospital, Taichung 407, Taiwan; chunte@vghtc.gov.tw (C.-T.H.); tjwang@vghtc.gov.tw (T.-J.W.); jatenhsu@gmail.com (C.-T.H.); cschen@vghtc.gov.tw (C.-H.C.); clwu@vghtc.gov.tw (C.-L.W.)
- <sup>2</sup> Department of Computer Science, Tunghai University, Taichung 407, Taiwan; g08350003@thu.edu.tw (R.-C.C.); kcpai@thu.edu.tw (K.-C.P.); ccwhuang@vgtc.gov.tw (C.-C.H.); lunchi@thu.edu.tw (L.-C.C.); rickysheu@thu.edu.tw (R.-K.S.)
- <sup>3</sup> DDS-THU AI Center, Tunghai University, Taichung 407, Taiwan; yltsai@thu.edu.tw
- <sup>4</sup> Computer Center, Artificial Intelligence Studio, Taichung Veterans General Hospital, Taichung 407, Taiwan; minnshyan@vghtc.gov.tw
- \* Correspondence: cmlai@thu.edu.tw

# Featured Application: The proposed method has been applied and landed in the ICU at Taichung Veterans General Hospital since 2021.

Abstract: Acute kidney injury (AKI) refers to rapid decline of kidney function and is manifested by decreasing urine output or abnormal blood test (elevated serum creatinine). Electronic health records (EHRs) is fundamental for clinicians and machine learning algorithms to predict the clinical outcome of patients in the Intensive Care Unit (ICU). Early prediction of AKI could automatically warn the clinicians to review the possible risk factors and act in advance to prevent it. However, the enormous amount of patient data usually consists of a relatively incomplete data set and is very challenging for supervised machine learning process. In this paper, we propose an entropy-based feature engineering framework for vital signs based on their frequency of records. In particular, we address the missing at random (MAR) and missing not at random (MNAR) types of missing data according to different clinical scenarios. Regarding its applicability, we applied it to establish a prediction model for future AKI in ICU patients using 4278 ICU admissions from a tertiary hospital. Our result shows that the proposed entropy-based features are feasible to be used in the AKI prediction model and its performance improves as the data availability increases. In addition, we study the performance of AKI prediction model by comparing different time gaps and feature windows with the proposed vital sign entropy features. This work could be used as a guidance for feature windows selection and missing data processing during the development of a prediction model in ICU.

Keywords: acute kidney injury (AKI); machine learning; entropy

# 1. Introduction

Acute kidney injury (AKI) is a medical term to describe rapid decline of kidney function within seven days. The most broadly accepted definition of AKI is proposed by Kidney Disease Improving Global Outcomes (KDIGO) [1] using serum creatinine, a waste product of muscle considered to be an endogenous filtration marker to assess renal function, and urine output to diagnose and define the severity of AKI. The prevalence of AKI is estimated to be around 55–60% in intensive care units (ICU), and is associated with an increased risk of prolonged hospital stay, renal replacement therapy, mortality,



Citation: Huang, C.-T.; Chang, R.-C.; Tsai, Y.-L.; Pai, K.-C.; Wang, T.-J.; Hsu, C.-T.; Chen, C.-H.; Huang, C.-C.; Wang, M.-S.; Chen, L.-C.; et al. Entropy-Based Time Window Features Extraction for Machine Learning to Identify Acute Kidney Injury in ICU. *Appl. Sci.* **2021**, *11*, 6364.

https://doi.org/10.3390/app11146364

Academic Editors: Keun Ho Ryu and Larbi Boubchir

Received: 1 April 2021 Accepted: 6 July 2021 Published: 9 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and high medical cost [2]. Early prediction of AKI could help clinicians to provide timely intervention and probably avoid the grave prognosis of end-stage renal disease.

With the growth of computing power, artificial intelligence has been applied extensively in various fields from fingerprint recognition to arrhythmia prediction [3–5]. The clear and objective definition of AKI provide an ideal labelling outcome for machine learning. Tomašev et al. [6] developed a deep learning approach for continuous prediction of future AKI using more than 700,000 patients Electronic Health Records (EHRs) from a multi-site dataset of US Department of Veterans Affairs. Based on their prediction model, clinicians could be warned 48 h in advance before AKI occurrence. However, the prediction algorithm created by Tomašev et al. consists of more than 100 features and may not be easily applied to other hospitals where many features might have high proportion of missing data.

Entropy, also known as Shannon's Entropy or self-information of an event, denotes the uncertainty and contributed information about the state of the system or data [7]. Entropy can be interpreted as a measure of the probability distribution for the amount of missing information [8]. The occurrence of the random event has higher information contribution than the common event observed. Entropy has been applied in numerous research fields, including medical, biometrics, and other real-world applications. Li et al. [9] applied entropy analysis to Electroencephalography (EEG) data to examine its performance on epilepsy detection. Chicote et al. [10] used fuzzy entropy and sample entropy as predictors for Out-of-hospital cardiac arrest (OHCA). Their study showed that entropy is a reliable predictor and outperformed other predictors. Chen et al. [11] concluded that multi-scale entropy could potentially be an early predictor of stroke-in-evolution in ICU-admitted non-atrial fibrillation stroke patients.

In this study, we use Random Forest (RF) machine learning algorithm to build the AKI prediction model [12]. RF has become one of the major machine learning methods, due to its robustness and non-linear ensemble nature and has been used in ICU early warning system [13,14]. RF algorithm can use cross-entropy and entropy-as-loss-functions to find the best split in a classification tree. However, directly applying it without considering the potential impact of missing data may result in poor model performance or bias. In order to conquer the challenges of missing data in EHRs, we take the advantages of entropy in developing early prediction model for AKI, and focus our research question on which data in which time windows should be included to build AKI prediction model. We propose a novel entropy-based feature engineering framework for vital signs and navigate the research question through finding an appropriate feature window for clinical feature inclusion and time gap for subsequent AKI prediction.

The rest of this paper is organized as follows: Section 2 describes the methodology, research material, and the proposed entropy-based feature engineering framework. Section 3 presents the machine learning experimental results using the proposed entropy-based feature engineering framework, time gap variations, and feature window variations. Section 4 provides a discussion and details the limitations of the proposed framework. Section 5 concludes this study.

#### 2. Materials and Methods

In this section, we explain the data source, problem definition, and the proposed entropy-based feature engineering framework.

#### 2.1. Study Population

Taichung Veterans General Hospital (TVGH) is a tertiary teaching hospital with 112 adult ICU beds in central part of Taiwan. 24,518 adult ICU admissions between July 2015 and December 2019 in the TVGH EHRs database were extracted for analysis. The inclusion criteria is age over 20 years old adults. Exclusion criteria are (1) End stage renal disease with the International Classification of Diseases Tenth Revision (ICD-10) code assignment of N18.6. (2) Had been treated with renal replacement therapy before the

index ICU admission. (3) Received renal replacement therapy within 24 h after index ICU admission. (4) Stayed less than 24 h in ICU. (5) Had an invalid or missing urine output (UO) record. After selection, 15,702 adult ICU admissions were included for final analysis.

AKI was labelled according to Section 2.2, where 9589 AKI patients and 6067 non-AKI patients were identified, respectively. To conduct research up to 72 h before AKI occurrence, we set additional filtering criteria for both AKI patients and non-AKI patients to ensure EHRs are longer than 72 h of time frame. The advantage of such filtering is that it gave us more data and longer period to study patient conditions. Additionally, by choosing a minimal 72-h time frame in ICU for both AKI and non-AKI group, we alleviate the possible bias in non-AKI group selection. Celi et al. [15] performed AKI mortality prediction on MIMIC III data with the Simplified Acute Physiology Score (SAPS), also focused on patients who survived in ICU for more than 72 h using multi-variable regression models.

We obtained a final set of 4278 patients from both groups after filtering with greater than or equal to 72 h admission time, containing 1631 AKI patients and 2647 non-AKI patients. Figure 1 shows the data cohort workflow.



Figure 1. Data cohort workflow.

#### 2.2. AKI Definition

AKI was defined, identified, and labeled according to the internationally accepted 'Kidney Disease: Improving Global Outcomes' (KDIGO) criteria [16], using both serum creatinine (SCr) and UO under three conditions: (1) SCr increased by 0.3 mg/dl within 48 h; (2) SCr increased more than 1.5 times baseline within 7 days; and (3) UO was less than 0.5 mL/kg/h over 6 h.

#### 2.3. Problem Formulation

For this study, we formulated the primary objective to be the early prediction of AKI as a binary classification task. Patients who developed AKI according to the KDIGO criteria were labeled as the positive class, while patients who did not develop AKI were labeled as the negative class. Samples were patient EHR records, with a series of vital sign values with time stamps. Each patient had a unique patient ID, *i*, and unique admission ID, *j*, for every ICU admission.

Formally, y = f(F) was our target function. Our goal in early AKI prediction was to learn a function f(F) that maps a set of features F to predict the binary outcome of  $y \in \{-1, 1\}$ , where y is equal to 1 if the patient has a positive AKI label and y with non-AKI labels is indicated by -1. Patient i was admitted to the ICU with j as their unique admission ID, while t denotes the timestamps. Time gap is the time between AKI onset

and how many hours in advance before the onset do we want to predict the onset event without available data during the gap but using the data from feature window instead. In our study, we denote time gap as  $T_{TP} = [0, 48]$  h before the onset. Feature window  $T_{FW} = [24, 48]$  is the number of hours before AKI onset, and we collect data with in this time frame. The collected data during this  $T_{FW}$  time is being used to predict if the patient will have AKI onset.

Our main objective was to study the relationship between AKI onset prediction with different time gaps and feature windows using the vital sign feature engineering framework we developed. Specifically, we considered the following two time ranges:

- 1. Time gap  $T_{TP}$  variation: The time gap from AKI onset to 48 h  $T_{TP} = [0, 48]$  before onset, using a feature window  $T_{FW} = 24$  of 24 h. In another word, we conducted a research where the time gap  $T_{TP}$  rolling from 0 h before onset to 48 h before onset. Coupled with.
- 2. Feature window  $T_{FW}$  variation: The feature window size from 24 h to 48 h before onset, with a time gap of 24 h.

Figure 2 illustrates the AKI prediction task, with a sequential EHRs problem set-up, using a feature window of 24 h versus 48 h with a time gap of 24 h. Non-AKI patients with a feature window of 24 h were randomly selected [17]. Table 1 shows the symbols used in this study.



**Figure 2.** Representation of the EHRs and AKI prediction task. AKI onset prediction with feature window variation of 24 h and 48 h with a time gap of 24 h is illustrated. Non-AKI patients with a feature window of 24 h were randomly selected.

## 2.4. Input Feature

Input features were broken down into two parts: Vital signs and administration information. Let F = (I, J, T, V) be a set of features defined by a set of *n* patients  $I = \{i_1, ..., i_n\}$ , a set of *m* admission IDs  $J = \{j_1, ..., j_m\}$ , a set of time stamps *T*, and a set of *k* vital signs  $V_{k,t} \in V$ at time *t*.

Vital Signs were recorded every 2 h, including systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure, oximetry, respiratory rate, pulse rate, and body temperature. Thus, we had 7 vital sign input features in total. Each feature included 3 generated features, as detailed in Section 2.5.2, and so k was 21 in our study.

Table 1. Notation.

Notations	Definition
y = f(F)	Target function
F = (I, J, T, V, L)	Set of features
Ι	Set of patients $I = \{i_1,, i_n\}$
J	Set of admission ID $J = \{j_1,, j_m\}$
Т	Set of time stamps
$T_{TP}$	Time gap where $T_{TP} = [0, 48]$ h before AKI onset
$T_{ST}$	Setting time
$T_{FW}$	Feature window where $T_F W = [24, 48]$ h before AKI onset
V	Set of vital signs $V_{k,t} \in V$
n	Total number of patients
т	Total number of admission IDs
k	Total number of vital signs
t	Time stamp
$H_q(V)$	Entropy for Vital sign V at the $q^{\text{th}}$ time
H	Target entropy
h	Total number of possible states
D0	AKI onset
D1	24 h before AKI onset
D2	48 h before AKI onset
D3	72 h before AKI onset
$V_{general}$	Mean and variance of vital signs
Ventropy	Entropy of vital signs
V <sub>merge</sub>	The combination of both $V_{general}$ and $V_{entropy}$
V <sub>mean</sub>	Mean of vital signs
$V_{variance}$	Variance of vital signs
fw	Feature window length
9	q is an iterative counter denoting the $q^{ ext{th}}$ time, where $q \in [1, 150]$
Ν	Number of computation times

#### 2.5. Entropy-Based Feature Engineering Framework

Model performance strongly correlates with data quality, where missing data may result in poor performance. Hence, we proposed a novel entropy-based feature engineering framework which is suitable for vital signs, based on their frequency of occurrence considering clinical availability. This framework was constructed through the following steps:

- 1. Step 1: We evaluated the setting time  $T_{ST}$ , which is the time interval between the patient's first admission time and their first data entry. This step is important, as the setting time dramatically affects the portion of missing data. In another word, we only consider data collected after setting time  $T_{ST}$ .
- 2. Step 2: The Shannon entropy was used to evaluate all vital signs *V*. This measures the probability distribution that characterizes the amount of missing information and data quality.
- 3. Step 3: We conducted missing value imputation on *V*. Both Steps 1 and 2 are critical for the data quality, as it is not measured on a frequent and consistent basis; yet, vital signs are crucial for AKI evaluation and indication. An overall workflow is shown in Figure 3.



Figure 3. Overall framework.

#### 2.5.1. Setting Time

The setting time,  $T_{ST}$ , is defined as the time interval between a patient *i*'s EHRs registered admission time with admission ID *j* and their first vital sign data entry. This evaluation is critical, with a direct contribution to the amount of missing data and, thus, affecting the model performance. The setting time analysis results are given in Section 3.2. In this step, we are emphasizing the Missing not at random (MNAR) type of missing data that was being introduced by the practical clinical administration setup.

#### 2.5.2. Vital Sign Entropy Feature Generation

Consider a discrete vital sign variable *V* for vital sign feature *k*, which occurs with probability  $P(V_{k,1}), ..., P(V_{k,h})$ , where *h* is the total number of possible states. According to the Shannon entropy, defined in [7], the entropy *H* of the discrete variable *V* at the *q*<sup>th</sup> time( $q \in [1, 150]$ ) can be written as follows:

$$H_q(V) = -\sum_{q=0}^h P(V_q) log P(V_q),$$
(1)

where fw is the feature window length. Then, h is:

$$h = \frac{|fw|}{4}.$$
 (2)

We propose our entropy-based vital sign feature, *H*, as follows:

$$H = \frac{1}{N} \sum_{q=0}^{N} H_q(V), \text{ where } N = 150.$$
(3)

In our setting, we set *h* equal to the feature window divided by 4, such that h = [6, 12]. As our feature window started at 24 h, *h* started at 6 and was incremented to 12 as the feature window became longer. Let *N* be the number of computation times. We obtained the numbers of 4 and *N* from multiple trials. Our proposed entropy feature, from Equation (3), takes *h* data points from the vital signs *V* to calculate *q* times of entropy,  $H_q$ , which are then accumulated and normalized *N* times to obtain our final entropy, *H*. The benefit of this is

that we obtain an entropy that objectively represents the overall data. The entropy can be regarded as another measure of the variation given feature windows.

The calculation of entropy on all vital signs V provides a quantified information evaluation of the average uncertainty regarding the outcomes of the vital sign features, as well as the vital sign feature qualities. It could also be interpreted as an evaluation of the information carried by the vital signs V.

For the vital signs stated in Section 2.4, the Shannon entropy was calculated. We also calculated the mean and variance for each vital sign *V*. Table 2 summarizes the feature generation for vital signs *V*. Let  $V_{general}$  denote the mean and variance of vital signs *V* and  $V_{entropy}$  denote the entropy of vital signs of *V*; then,  $V_{merge}$  is the combination of both  $V_{general}$  and  $V_{entropy}$ .

Table 2. Feature Generation.

Variable	Туре	Features			
Vital signs	Vital sign Mean	SBP, DBP, Pulse Pressure, Oximetry, Respiratory Rate, Pulse Rate, Body Temperature			
	Variance				

#### 2.5.3. Missing Value Imputation

Table 3 shows the average portion of missing data in our data set. While the entropy measures data quality based on probability, the amount of missing data results in missing data in the entropy features. In this step, we focus on the Missing at random (MAR) type of missing data as the biomedical data of patients may be documented on different frequency. As *h* is the frequency required to compute a single  $H_q(V)$ , if a patient has less data points for that given feature time, we will obtain a *miss* in the entropy features. Please note that we did not compute entropy-based features on imputed vital sign data, but only on the original data. This is how we made sure that we measured the data quality of the original data, not the data after imputation.

Table 3. Missing data analysis in vital signs.

Missing Proportion Mean(%)										
	SBP DBP Pulse Pressure Oximetry Respiratory Rate Pulse Rate Tem									
Entropy	0.23(0.07)	0.23(0.07)	0.24(0.07)	11.38(0.58)	0.83(0.11)	0.11(0.04)	7.76(0.31)			
Mean	0	0	0	0	0	0	0			
Variance	0	0	0	0	0	0	0			
Merge	0.08(0.12)	0.08(0.12)	0.08(0.12)	3.79(5.41)	0.28(0.4)	0.04(0.06)	2.59(3.69)			

For vital signs features, the median was used for the entropy  $V_{entropy}$  and variance. We defined a normal range for each vital signs feature, through the advice of clinicians. The missing value imputation method for the mean and other vital signs raw data were random imputations from the defined normal range, as shown in Table 4.

Variable	Imputation Method	Variable
	Median	Shannon entropy, Variance
Vital		Mean
Signs	Random imputation	SBP, DBP, Pulse pressure,
U	from normal range	Oximetry, Respiratory rate,
		Pulse rate, Body temperature

Table 4. Missing Value Imputation.

#### 3. Results

This section presents the results and analysis of the cohort, setting time, time gap variation, and feature window variation with the proposed entropy-based framework.

#### 3.1. Cohort Analysis

The final cohort consisted of 4278 patients, of which 1631 had AKI onset during their ICU stay and labeled as AKI patients, while 2647 patients did not have AKI occurrence and were labeled as non-AKI patients. The training, validation, and testing data rates were set to 80%, 10%, and 10%, respectively. The training and validation cohort included 1492 AKI patients and 2382 non-AKI patients, while the testing cohort included 139 AKI patients and 625 non-AKI patients.

Table 5 shows a detailed comparison of the AKI and non-AKI patients cohort analysis, in terms of their demographics. It showed a statistical significant differences (p < 0.01) on time span in ICU in days, age and BMI between AKI and non-AKI group. AKI group tend to stay longer in the ICU, older in age, and had lower BMI value.

An aggregate analysis of vital sign and vasopressor medication between AKI and non-AKI cohort are shown in Table 6. In the comparison, the AKI group had a significant difference (p < 0.01) in vital signs, compared to the non-AKI group, except for pulse pressure on the training and validation cohort. In the testing cohort, only respiratory rate and pulse rate showed significant difference (p < 0.01). There were significant difference (p < 0.01) in the vasopressor medications of vasopressin, Norepinephrine, and Epinephrine between the AKI and non-AKI patients in training and validation sets. There were only significant difference (p < 0.01) in Norepinephrine and Epinephrine on the testing set.

In ventilatory support, statistic differences (p < 0.01) were shown in Table 7 in the fraction of inspired oxygen (FiO<sub>2</sub>), positive end-expiratory pressure and continuous positive airway pressure (PEEP/CPAP), mean arterial pressure (MAP), total respiratory rate (RR) between the AKI and non-AKI patients in training and validation sets. Mean airway pressure (Paw), exhaled VT and exhalations volume per time unit(MV) did not show statistic significant differences.

<b>/alue</b>
_
0 01 **
0.01
0.01 **
0.01 **
1.58
1.58
0

 Table 5. Cohort Analysis of Patient Population and Demographics.

\*: p < 0.05; \*\*: p < 0.01.

 Table 6. Cohort Analysis of Vital Signs and Medications.

Variable	Training and Validation				Testing			
Mean (STD)	All	AKI	Non-AKI	<i>p</i> -Value	AKI	Non-AKI	p-Value	
Vital signs, N	4278	1492	2382	_	139	265	_	
SBP(mmHg)	131.4 (23.68)	129.69 (25.04)	132.41 (22.68)	p < 0.01 **	130.75 (26.76)	132.23 (22.48)	0.56	
DBP(mmHg)	78 (16 77)	76.51	78.85	p < 0.01 **	76.15	79.59	0.05	
Pulse pressure	53.35	53.18	53.49	0.62	54.6	52.42	0.29	
Oximetry(%)	98.21	97.93	98.39 (2.37)	p < 0.01 **	98.09 (3.51)	98.3	0.53	
Respiratory rate	18.86	19.42	18.44	p < 0.01 **	20.54	18.6	p < 0.01 **	
Pulse rate(/min)	90.17 (19.98)	94.48	87.45 (18.68)	p < 0.01 **	95.22 (21.37)	87.68 (19.24)	p < 0.01 **	
Temperature(Celsius)	36.63	94.48 (21.12)	36.58 (0.87)	p < 0.01 **	36.76 (1.11)	36.57	0.05	
Medication, Vasopressors	()	( , , ,	()			()		
Vasopressin	24 (0.56%)	13 (0.87%)	7 (0.29%)	p < 0.05 *	1 (0.72%)	3 (1.13%)	0.69	
Norepinephrine	826 (19.31%)	398 (26.68%)	348 (14.61%)	p < 0.01 **	38 (27.34%)	42 (15.85%)	p < 0.01 **	
Dopamine	451 (10.54%)	156 (10.46%)	253 (10.62%)	0.87	14 (10.07%)	28 (10.57%)	0.88	
Epinephrine	278 (6.5%)	145 (9.72%)	102 (4.28%)	p < 0.01 **	16 (11.51%)	15 (5.66%)	p < 0.05 *	
Dobutamine	59 (1.38%)	23 (1.54%)	31 (1.3%)	0.54	(1.44%)	(1.13%)	0.79	

\*: p < 0.05; \*\*: p < 0.01.

 Table 7. Cohort Analysis of Ventilatory Support.

Variable		Training an	d Validation	Testing			
Mean (STD)	All	AKI	Non-AKI	<i>p</i> -Value	AKI	Non-AKI	<i>p</i> -Value
Ventilatory Support	4287	1492	2382	-	139	265	-
FIO <sub>2</sub>	70.95 (25.87)	75.4 (26.07)	67.7 (25.27)	p < 0.01 **	78.62 (25.33)	65.89 (24.73)	p < 0.01 **
PEEPCPAP	4.87 (1.69)	5.06 (1.78)	4.71 (1.59)	p < 0.01 **	5.09 (1.72)	4.86 (1.71)	0.24
PAW	23.13 (6.73)	23.21 (6.72)	22.89 (6.63)	0.61	23.76 (8.1)	24.38 (6.71)	0.8
MAPS	11.62 (2.55)	11.93 (2.65)	11.34 (2.52)	p < 0.05 *	10.97 (1.47)	11.47 (2.06)	0.47
TOTRR	18.89 (5.36)	19.76 (5.8)	18.1 (4.81)	p < 0.01 **	19.01 (5.31)	17.71 (4.74)	0.44
VTEXH	0.52	0.52 (0.11)	0.53	0.28	0.51	0.49	0.63
MVEXH	9.57 (2.8)	9.79 (3.03)	9.38 (2.61)	0.13	9.72 (2.7)	9.08 (2.16)	0.44

 $\overline{*: p < 0.05}$ ; \*\*: p < 0.01.

#### 3.2. Setting Time and Missing Data Analysis Results

The setting time  $T_{ST}$  is the minimum time interval between a patient's admission time and their first vital sign data entry. Table 8 shows the mean setting times of Pulse pressure (SBP, DBP), Oximetry, Respiratory rate, Pulse rate, and Temperature. As the average setting time in most vital signs was around 3 h  $T_{ST} = 3$ , we set 3 h as our setting time in this research. Figure 4 visualizes the averaged setting time between patient's first vital sign data entry and ICU admission. In other words, if the time gap of the prediction task was less than 3 h, a lot of missing data may occur, thus affecting the model performance. In another word, we only consider data collected after  $T_{ST} = 3$  h of the patient's admission time.

Table 8. Mean setting time for vital signs.

Vital Signs	Mean Setting Time (h)
Pulse pressure (SBP, DBP)	2.93
Oximetry	8.39
Respiratory rate	3.00
Pulse rate	2.99
Temperature	3.05



Figure 4. Visualization plot for setting time between different vital signs.

#### 3.3. Classification and Evaluation Criteria

Random Forest (RF) machine learning algorithm [12] was employed to compare results between  $V_{entropy}$ ,  $V_{merge}$ , and  $V_{general}$ , where the time gap of AKI onset to 48 h before AKI onset and a feature window of 24 h was used. After finding the critical point in the time gap, we used it in the feature window variation task, studying feature windows of 24 h to 48 h with a time gap of 24 h.

The RF model can predict whether a patient will have AKI onset or not during ICU admission, by providing a probability. The probability is determined by the ratio of the decision trees that give positive results in the total number of trees. In our RF model, we set the number of multiple decision trees as 300 and the number of splits to 21. Ten-fold validation was used.

The performance was evaluated using the accuracy, area under the ROC curve (AU-ROC), and net reclassification improvement (NRI) metrics. The accuracy measures the percentage of correctly classified samples of the total samples under a given threshold. The AUROC evaluates the overall model performance. The NRI aims to provide an objective method to quantify improvements in categories in models [18]. It measures how well a new model correctly reclassifies subjects that were not correctly classified in the old model. In our case, the new model was the model with increment in time *t*, while the old model for comparison was the model with prior time (i.e., t - 1). F value from F test was conducted in order to compare the variance between two groups.

#### 3.4. Classification Performance with Time Gap Variation

In this section, we reveal the time gap variation  $T_{TP}$  results. We studied the time gap from AKI onset to 48 h  $T_{TP} = [0, 48]$  before onset using a feature window size of 24 h  $T_{FW} = 24$ . This is critical for identifying the most important point of time between slight AKI signs and emerging AKI oscillation. We compared the accuracy, AUROC, and NRI performance of  $V_{entropy}$ ,  $V_{mean}$ ,  $V_{variance}$ , and  $V_{merge}$ . The Model performance results are shown in Figure 5. It was shown in AUROC that the  $V_{merge}$ ,  $V_{mean}$  and  $V_{variance}$ , performed steadily with different time gap variation.  $V_{entropy}$  performed better when time gap is longer.

Figure 6 shows the F value from F test of AUROC along with the time gap. The F value on AUROC showed that the peak in F value was at 30 h; namely 30 h was the point that contributed the most information and was the turning point from slight AKI signs to emerging AKI oscillation. Therefore, clinicians may consider the patient has higher risk of having AKI onset. The overall model performance increased drastically, when starting at this point in time. The trend was most obvious in NRI for both  $V_{entropy}$  and  $V_{merge}$ . As the NRI seeks to quantify whether a new marker provides clinically relevant improvements in model prediction, this result indicates that when passing the critical point of 30 h, the shorter the time gap and the more information available, the better the models performed.



Figure 5. Visualization plot for time gap variance on AUROC, Accuracy and NRI performance.



Figure 6. Visualization plot for time gap variance F value from F test on AUROC.

As 30 h was the critical point of time between slight AKI signs and rapid AKI oscillation, as shown in in Tables 9 and 10, we conducted a comparison between using a time gap of less than 30 h (after critical point, rapid AKI oscillation) and a time gap of more than 30 h (before critical point, slight AKI signs). Statistically significant differences (p < 0.01) were observed in and  $V_{entropy}$  in AUROC;  $V_{entropy}$  and  $V_{merge}$  inaccuracy, NRI and precision.  $V_{entropy}$  and  $V_{mean}$  in Recall

Feature AUROC Mean (STD) Accuracy(%) Mean (STD) NRI(%) Mean (STD) Types Time Before After Before After Before After p Value p Value p Value Gap Critical Critical Critical Critical Critical Critical Point Point Point point Point Point p < 0.01p < 0.01p < 0.0169.09 67.36 65.66 64 02 -0.9-4.19Entropy (1.86)-0.22(2.27) -1.29 (0.99)(0.82)(0.80)(1.08)70.81 70.35 66.01 65.48 0.07 0.09 Mean 0.06 (0.55) (1.10)(0.31)(0.53)(0.42)(1.09)67.68 67.67 64.13 64.18 0.19 0.17 Variance 0.96 0.79 0.95 (0.49)(0.56) (0.31)(0.60)(0.85)(1.23) p < 0.05p < 0.0575.87 75.58 70.3 69.89 1.3 0.2 0.09 Merge (0.95) (0.30)(0.34)(0.31) (0.30) (0.46)

**Table 9.** Comparison between using a time gap of less than 30 h (rapid AKI oscillation) and a time gap greater than 30 h (slight AKI signs) on AUROC, Accuracy and NRI.

\*: p < 0.05; \*\*: p < 0.01.

**Table 10.** Comparison between using a time gap of less than 30 h (rapid AKI oscillation) and a time gap greater than 30 h (slight AKI signs) on Recall, and Precision.

Feature Types	Α	UROC Mean (ST	D)	Accuracy(%) Mean (STD)			
Time Gap	pp Before After p Value Critical Critical p Value Point Point		Before Critical Point	After Critical Point	p Value		
Shannon entropy	55.74 (0.89)	53.88 (1.44)	p < 0.01 **	60.57 (1.04)	58.42 (1.44)	<i>p</i> < 0.01 **	
Mean	60.82 (1.37)	59.49 (0.99)	p < 0.05 *	59.98 (0.46)	59.51 (0.69)	0.19	
Variance	55.72 (1.24)	55.1 (0.94)	0.25	58.31 (0.35)	58.45 (0.82)	0.65	
Merge	66.76 (0.67)	66.33 (0.99)	0.4	64.7 (0.44)	64.18 (0.58)	p < 0.05 *	

\*: p < 0.05; \*\*: p < 0.01.

#### 3.5. Classification Performance with Feature Window Variation

For feature window variation, we varied the feature window size from 24 h to 48 h before onset  $T_{FW} = [24, 48]$ , with a time gap of 24 h  $T_{TP} = 24$ . The feature window variation provided insights, in terms of clinical practice, about how much data is required for the best prediction performance and how different features react differently to different time windows. The model performance in the feature variation task is shown in Figure 7. A comparison of 24 h and 48 h feature windows is shown in Tables 11 and 12.

The performance of  $V_{merge}$  was the highest among all the other features, such as  $V_{mean}$ ,  $V_{variance}$ , and  $V_{entropy}$ . Both the AUROC accuracy and NRI of  $V_{merge}$  consistently improved with increasing feature data availability. The longer the feature window, the more data was available for the model to learn from. Model performance with  $V_{entropy}$  outperformed  $V_{mean}$  when the feature window size was bigger than 28 h. The overall model performance of  $V_{entropy}$  improved drastically with more data availability, as shown in Figure 7.

In terms of accuracy, we see a merging trend in performance of the proposed  $V_{entropy}$  feature. We can reasonably conclude that with fewer features available, our proposed entropy-based feature  $V_{entropy}$  performed better. While there the accuracy was steady in  $V_{mean}$ , and there is no evidence that more data in  $V_{variance}$  would make the prediction better.



**Figure 7.** Visualization plot for feature window variation performance, with the feature window size varying from 24 h to 48 h before onset, with a time gap of 24 h.

In terms of NRI,  $V_{entropy}$  was the highest among other features after 32 h. This implied that the increase in feature window had a high impact on  $V_{entropy}$  model performance, such that the  $V_{variance}$  performance was more sensitive to the length of feature window and data availability. The model performance of  $V_{entropy}$  was sensitive to the feature window size of 24 h to 36 and 38 to 48 h of feature window size.  $V_{mean}$  was not sensitive to feature window variance. Similar behavior was found in  $V_{variance}$ , the amount of available data was critical to  $V_{variance}$  between 24 to 30 h of feature window size In comparison, despite the fact that  $V_{entropy}$  performed well, in terms of accuracy, we drew the same conclusion as for AUROC; that is,  $V_{entropy}$  improved drastically with more data and a longer feature window. The NIR showed that  $V_{entropy}$  is steadily sensitive to data availability.

In Tables 11 and 12 we showed an improvement comparison between feature window size of 24 h  $T_{FW}$  = 24 and 48 h  $T_{FW}$  = 48. The  $V_{entropy}$  feature had the highest performance improvement in AUROC, NRI, Recall and Precision. At the same time, more data may not improve the performance of  $V_{mean}$ . Compared to  $V_{variance}$ , more data available improves  $V_{merge}$ 's performance across all metrics.

 Table 11. Comparison of 24 h and 48 h feature windows on AUROC, Accuracy, NRI.

Feature Types	AUROC (%)			Accuracy (%)				NRI (%)	
Feature Window	24 h	48 h	Improvement 24 h		48 h	Improvement 24 h		48 h	Improvement
Shannon entropy	70.01	76.8	6.79	65.82	69.33	3.5	-	8.32	8.32
Mean	72.48	71.31	-1.16	67.91	67.42	-0.48	-	-0.88	-0.88
Variance	67.08	69.51	2.42	64.7	65.96	1.26	-	3	3
Merge	77.64	81.24	3.59	71.62	73.67	2.05	-	4.4	4.4

Table 12. Comparison of 24 h and 48 h feature windows on Precision and Recall.

Feature Types	Recall(%)		F1-Score			NRI (%)			
Feature Window	24 h	48 h	Improvem	ent 24 h	48 h	Improvem	ent 24 h	48 h	Improvement
Shannon entropy	54.55	64.59	10.03	55.37	59.32	3.94	54.96	61.84	6.88
Mean	60.56	59.28	-1.27	57.56	57.48	-0.08	59.02	58.37	-0.65
Variance	52.2	54.97	2.76	53.97	55.9	1.92	53.07	55.43	2.35
Merge	67.43	70.88	3.45	61.48	64.37	2.52	64.51	67.47	2.95

#### 4. Discussion and Limitation

In this study, we proposed a novel entropy-based feature engineering framework for vital signs, based on their frequency of records and clinical availability. Both quantitative analysis of the features and data quality are taken into analysis. In addition, we conducted feature window and time gap experiment in order to determine the best time window to maximize the accuracy of AKI prediction in ICU.

The etiology of AKI is usually multi-factorial. Dehydration, infection, renal toxic medications, and contrast medium injection during computed tomography scanning are common risk factors for AKI [19]. Clinicians are usually aware of AKI only when patients' urine output decrease or elevated serum creatinine occur. In 2014, the FDA approved a commercial urine stress biomarker, Nephrocheck<sup>®</sup>, to be used for ICU patients for early prediction of AKI with 12-h time gap windows in advance [20]. However, the commercial kit is very expensive and has not been approved by other countries other than the United States. On the other hand, we sought to integrate an AKI prediction algorithm into electronic health information system in our hospital to provide a real time AKI risk probability for ICU patients without spending additional cost and nursing labors. To achieve this goal, we mathematically define the EHR data availability according to different time gaps and feature windows to apply state-of-the-art machine learning techniques.

Compared to recent deep learning-based AKI prediction studies [6,21,22], we not only conducted AKI prediction, but also proposed an overall framework that covers feature pre-processing, missing value imputation, and proposed a novel entropy-based vital sign feature engineering. Moreover, we conducted a fine-grained analysis considering both time gap variation and feature window variation. In feature window examination, we tested the feature window length from 24 h to 48 h before AKI onset, with a fixed time gap of 24 h. Our results demonstrated that early AKI can be predicted, based on our proposed vital sign entropy-based features, with a feature window of 24 h and a time gap of 24 h. The best-performing model was obtained when using all available features with the longest feature window. The result implies that both the frequency of clinical records and data quality are important for AKI prediction model.

Our study has several limitations. First, our cohort is derived from a single hospital, and the prediction algorithm has not been externally validated with other cohorts which may limit its generalization. Second, the best feature window length and time gap length may vary depending on the availability of the clinical records. However, our work could still provide a practical data mining process for developing a prediction model in ICU.

## 5. Conclusions

We introduced an end-to-end practical framework from missing data handling, entropybased feature engineering, to different time gap and feature window length analysis in an ICU data set for AKI prediction. In the missing data handling, our proposed framework is able to address both the missing at random (MAR) and missing not at random (MNAR) types of missing data in clinical practice. As RF algorithms are widely used in various settings, our proposed method provides the practical missing data handling that will improve model performance. We studied the relationships among different time gap variations and feature window variations with the proposed vital sign entropy feature for AKI prediction. This work could provide a guidance for feature windows selection and missing data processing during the development of a prediction model in ICU.

Author Contributions: Conceptualization: C.-T.H. (Chun-Te Huang), C.-H.C. and C.-M.L.; Methodology: T.-J.W., L.-C.C. and R.-K.S.; Software: Y.-L.T.; Data Curation: C.-C.H. and M.-S.W.; Writing: R.-C.C., C.-T.H. (Chun-Te Huang) and C.-M.L.; Investigation: C.-L.W.; Validation: C.-T.H. (Chia-Tien Hsu) and K.-C.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Ministry of Science and Technology in Taiwan under grant number MOST 109-2321-B-075A-001-1.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Institutional Review Board of Taichung Veterans General Hospital under case number SE20249B, approved on 7 May 2020.

**Informed Consent Statement:** Patient consent was waived, due to retrospective study and all data had been deidentified under review by the Institutional Review Review Board of Taichung Veterans General Hospital.

**Data Availability Statement:** The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available, due to the constraint of Taiwanese Personal Information Protection Act.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

# References

- Ostermann, M.; Bellomo, R.; Burdmann, E.A.; Doi, K.; Endre, Z.H.; Goldstein, S.L.; Kane-Gill, S.L.; Liu, K.D.; Prowle, J.R.; Shaw, A.D.; et al. Controversies in acute kidney injury: Conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. *Kidney Int.* 2020, *98*, 294–309. [CrossRef]
- Hoste, E.A.; Bagshaw, S.M.; Bellomo, R.; Cely, C.M.; Colman, R.; Cruz, D.N.; Edipidis, K.; Forni, L.G.; Gomersall, C.D.; Govil, D.; et al. Epidemiology of acute kidney injury in critically ill patients: The multinational AKI-EPI study. *Intensive Care Med.* 2015, 41, 1411–1423. [CrossRef]
- 3. Zabala-Blanco, D.; Mora, M.; Barrientos, R.J.; Hernández-García, R.; Naranjo-Torres, J. Fingerprint Classification through Standard and Weighted Extreme Learning Machines. *Appl. Sci.* **2020**, *10*, 4125. [CrossRef]
- 4. Chen, J.; Zhao, H.; Cao, Z.; Guo, F.; Pang, L. A Customized Semantic Segmentation Network for the Fingerprint Singular Point Detection. *Appl. Sci.* **2020**, *10*, 3868. [CrossRef]
- 5. Fujita, H.; Cimr, D. Decision support system for arrhythmia prediction using convolutional neural network structure without preprocessing. *Appl. Intell.* **2019**, *49*, 3383–3391. [CrossRef]
- Tomašev, N.; Glorot, X.; Rae, J.W.; Zielinski, M.; Askham, H.; Saraiva, A.; Mottram, A.; Meyer, C.; Ravuri, S.; Protsyuk, I.; et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019, 572, 116–119. [CrossRef] [PubMed]
- 7. Shannon, C.E. A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. 2001, 5, 3–55. [CrossRef]
- 8. Lent, C.S. Information and entropy in physical systems. In *Energy Limits in Computation*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–63.
- 9. Li, P.; Karmakar, C.; Yearwood, J.; Venkatesh, S.; Palaniswami, M.; Liu, C. Detection of epileptic seizure based on entropy analysis of short-term EEG. *PLoS ONE* 2018, *13*, e0193691.
- 10. Chicote, B.; Irusta, U.; Aramendi, E.; Alcaraz, R.; Rieta, J.J.; Isasi, I.; Alonso, D.; Baqueriza, M.D.M.; Ibarguren, K. Fuzzy and sample entropies as predictors of patient survival using short ventricular fibrillation recordings during out of hospital cardiac arrest. *Entropy* **2018**, *20*, 591. [CrossRef]
- 11. Chen, C.H.; Huang, P.W.; Tang, S.C.; Shieh, J.S.; Lai, D.M.; Wu, A.Y.; Jeng, J.S. Complexity of heart rate variability can predict stroke-in-evolution in acute ischemic stroke patients. *Sci. Rep.* **2015**, *5*, 1–5.
- 12. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 13. Awad, A.; Bader-El-Den, M.; McNicholas, J.; Briggs, J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int. J. Med Inform.* **2017**, *108*, 185–195. [CrossRef] [PubMed]
- 14. Lin, K.; Hu, Y.; Kong, G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int. J. Med Inform.* **2019**, *125*, 55–61. [CrossRef] [PubMed]
- 15. Celi, L.A.G.; Tang, R.J.; Villarroel, M.C.; Davidzon, G.A.; Lester, W.T.; Chueh, H.C. A clinical database-driven approach to decision support: Predicting mortality among patients with acute kidney injury. *J. Healthc. Eng.* **2011**, *2*, 97–110. [CrossRef]
- 16. Khwaja, A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin. Pract.* **2012**, *120*, c179–c184. [CrossRef] [PubMed]
- 17. Moor, M.; Rieck, B.; Horn, M.; Jutzeler, C.; Borgwardt, K. Early Prediction of Sepsis in the ICU using Machine Learning: A Systematic Review. *medRxiv* 2021, *8*, 348.
- 18. Pencina, M.J.; D'Agostino, R.B., Sr.; D'Agostino, R.B., Jr.; Vasan, R.S. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat. Med.* **2008**, *27*, 157–172. [CrossRef]
- Makris, K.; Spanou, L. Acute kidney injury: Definition, pathophysiology and clinical phenotypes. *Clin. Biochem. Rev.* 2016, *37*, 85.
   Ilaria, G.; Kianoush, K.; Ruxandra, B.; Francesca, M.; Mariarosa, C.; Davide, G.; Claudio, R. Clinical adoption of Nephrocheck<sup>®</sup> in the early detection of acute kidney injury. *Ann. Clin. Biochem.* 2021, *58*, 6–15. [CrossRef]
- 21. Wang, Y.; Wei, Y.; Yang, H.; Li, J.; Zhou, Y.; Wu, Q. Utilizing imbalanced electronic health records to predict acute kidney injury by ensemble learning and time series model. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 1–13. [CrossRef]
- 22. Abdullah, S.S.; Rostamzadeh, N.; Sedig, K.; Garg, A.X.; McArthur, E. Predicting Acute Kidney Injury: A Machine Learning Approach Using Electronic Health Records. *Information* **2020**, *11*, 386. [CrossRef]