



Article

# In-Advance Prediction of Pressure Ulcers via Deep-Learning-Based Robust Missing Value Imputation on Real-Time Intensive Care Variables

Minkyu Kim <sup>1,†</sup> , Tae-Hoon Kim <sup>2,†</sup> , Dowon Kim <sup>1</sup>, Donghoon Lee <sup>1</sup> , Dohyun Kim <sup>1</sup>, Jeongwon Heo <sup>2</sup> , Seonguk Kang <sup>3</sup>, Taejun Ha <sup>4</sup>, Jinju Kim <sup>2</sup> , Da Hye Moon <sup>2,5</sup>, Yeonjeong Heo <sup>2,5</sup> , Woo Jin Kim <sup>2</sup> , Seung-Joon Lee <sup>2</sup> , Yoon Kim <sup>6</sup>, Sang Won Park <sup>7,8</sup> , Seon-Sook Han <sup>2,\*</sup> and Hyun-Soo Choi <sup>9,\*</sup>

- <sup>1</sup> Department of Research & Development, Ziovision Co., Ltd., Chuncheon 24341, Republic of Korea; minkyu.kim@ziovision.co.kr (M.K.); dakim1204@ziovision.co.kr (D.K.); dannylee@ziovision.co.kr (D.L.); dohyeon.kim@ziovision.co.kr (D.K.)
- <sup>2</sup> Department of Internal Medicine, Kangwon National University, Chuncheon 24341, Republic of Korea; blessing0104@naver.com (T.-H.K.); doctorhjw@naver.com (J.H.); kj451007@naver.com (J.K.); ansekgo@naver.com (D.H.M.); yonjong1954@naver.com (Y.H.); pulmo2@kangwon.ac.kr (W.J.K.); medfman@kangwon.ac.kr (S.-J.L.)
- <sup>3</sup> Department of Convergence Security, Kangwon National University, Chuncheon 24341, Republic of Korea; uk99@knuh.or.kr
- <sup>4</sup> Biomedical Research Institute, Kangwon National University Hospital, Chuncheon 24289, Republic of Korea; xvop11@naver.com
- <sup>5</sup> Department of Pulmonology, Kangwon National University Hospital, Chuncheon 24289, Republic of Korea
- <sup>6</sup> Department of Computer Science and Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea; yooni@kangwon.ac.kr
- <sup>7</sup> Department of Medical Informatics, School of Medicine, Kangwon National University, Chuncheon 24341, Republic of Korea; chicwon229@kangwon.ac.kr
- <sup>8</sup> Institute of Medical Science, School of Medicine, Kangwon National University, Chuncheon 24341, Republic of Korea
- <sup>9</sup> Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
- \* Correspondence: ssunimd@kangwon.ac.kr (S.-S.H.); choi.hyunsoo@seoultech.ac.kr (H.-S.C.)
- † These authors contributed equally to this work.



**Citation:** Kim, M.; Kim, T.-H.; Kim, D.; Lee, D.; Kim, D.; Heo, J.; Kang, S.; Ha, T.; Kim, J.; Moon, D.H.; et al.

In-Advance Prediction of Pressure Ulcers via Deep-Learning-Based Robust Missing Value Imputation on Real-Time Intensive Care Variables. *J. Clin. Med.* **2024**, *13*, 36. <https://doi.org/10.3390/jcm13010036>

Academic Editor: Alexander C. Katoulis

Received: 7 November 2023

Revised: 6 December 2023

Accepted: 15 December 2023

Published: 20 December 2023



**Copyright:** © 2023 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/>).

**Abstract:** Pressure ulcers (PUs) are a prevalent skin disease affecting patients with impaired mobility and in high-risk groups. These ulcers increase patients' suffering, medical expenses, and burden on medical staff. This study introduces a clinical decision support system and verifies it for predicting real-time PU occurrences within the intensive care unit (ICU) by using MIMIC-IV and in-house ICU data. We develop various machine learning (ML) and deep learning (DL) models for predicting PU occurrences in real time using the MIMIC-IV and validate using the MIMIC-IV and Kangwon National University Hospital (KNUH) dataset. To address the challenge of missing values in time series, we propose a novel recurrent neural network model, GRU-D++. This model outperformed other experimental models by achieving the area under the receiver operating characteristic curve (AUROC) of 0.945 for the on-time prediction and AUROC of 0.912 for 48h in-advance prediction. Furthermore, in the external validation with the KNUH dataset, the fine-tuned GRU-D++ model demonstrated superior performances, achieving an AUROC of 0.898 for on-time prediction and an AUROC of 0.897 for 48h in-advance prediction. The proposed GRU-D++, designed to consider temporal information and missing values, stands out for its predictive accuracy. Our findings suggest that this model can significantly alleviate the workload of medical staff and prevent the worsening of patient conditions by enabling timely interventions for PUs in the ICU.

**Keywords:** intensive care unit; clinical decision support systems; pressure ulcers; time series; missing values; machine learning; deep learning

## 1. Introduction

Pressure ulcers (PUs) are prevalent skin injuries in patients who remain immobile or cannot change positions for extended periods. PUs arise due to prolonged pressure on bony prominences, such as the back of the head, shoulders, elbows, and heels, or from blood circulation disorders [1–3]. In the early stages of PUs, simple interventions like wound dressings or patient repositioning are required. However, in advanced cases, the patient's condition may deteriorate significantly, and surgical procedures are required. Such interventions increase medical expenses [4,5] and the workload of the intensive care unit (ICU) staff [6].

In the ICU, various risk assessment tools, including the Braden scale, Gosnell scale, Norton scale, and Waterlow score, are used to gauge PU risks [7–12]. Recently, machine learning (ML)-based PU prediction systems have been developed [13–18] to effectively predict PU occurrences. These systems employ logistic regression, random forest, boosting machines, or multi-layer perceptron. However, they suffer two major problems: (1) they cannot handle time series data even though real-world data are time series and (2) they cannot appropriately handle missing values even though real-world data contain a massive number of missing values.

To overcome these problems, we develop recurrent neural network (RNN)-based PU prediction systems. RNN is a deep learning (DL) method for time series data. We employ five different RNNs. Simple RNN, gated recurrent unit (GRU) [19], and long-short-term memory (LSTM) [20] are representative RNN models. However, they are not suitable for missing-value handling. Therefore, we also employ GRU with decay (GRU-D) [21], which is specialized to impute missing values. Furthermore, we propose an enhanced version of GRU-D named GRU-D++, suitable for time series with a high missing-value rate.

In the empirical experiments, Simple RNN, GRU, and LSTM outperform traditional ML systems, which indicates the importance of time series information for PU prediction. GRU-D and GRU-D++ outperform RNN systems, demonstrating that their missing-value imputation mechanisms are very effective. In addition, we conduct additional experiments of GRU-D++ with different numbers of input variables, data rescaling, and model fine-tuning. These experiments can provide meaningful insights to researchers who want to employ GRU-D++ in other medical centers. For the reproducibility of our experiments, we have publicly opened the source code of GRU-D++ at <https://github.com/gim4855744/GRU-Dpp> accessed on 6 November 2023.

## 2. Methods

### 2.1. Study Population

We used the Medical Information Mart for Intensive Care IV (MIMIC-IV) [22], a large public database containing de-identified patient information admitted to the Beth Israel Deaconess Medical Center (BIDMC), as the internal dataset. Note that, although the MIMIC-IV is a public database, the use of full MIMIC-IV requires approval from PhysioNet. Only ICU patients were included in this study. We classified patients who had records of PUs or who had been assigned PU grades in the nursing records as the PU group. The remaining patients were classified as the non-PU group. Patients with PUs before ICU admissions were excluded from this study.

The external validation dataset was collected from the Kangwon National University Hospital (KNUH) in the Republic of Korea. We used adult patients (age  $\geq 18$ ) who were admitted to the ICU between January 2016 and August 2022 and had at least 48 h of records after admission. The dataset has been approved by the Institutional Review Board of KNUH (IRB, KNUH-2022-09-013-00).

### 2.2. Data Collection

We used 48 variables. Our input variables included patient demographics, vital signs, laboratory findings, medication and treatment information, underlying diseases, the Braden scales, and the sedation scale. In the MIMIC-IV, some patients have the Richmond-

Agitation Sedation Scale (RASS), and the others have the Riker Sedation-Agitation Scale (SAS). Therefore, we converted RASS to SAS. Figure 1 shows the list of our input variables. In this study, we use real-world datasets collected hourly. Therefore, some of the variables have large amounts of missing values; for instance, the laboratory events pH and lactate are missing 72% and 80%, respectively, whereas vital signs such SBP, DBP, and MBP are 0.6% missing.

Demographics	Vital Sign	Laboratory Finding	Braden Scale
Gender	Systolic BP	pH	Perception
Age	Diastolic BP	PaCO <sub>2</sub>	Moisture
Height	Mean BP	PaO <sub>2</sub>	Activity
Weight	Heart Rate	Bicarbonate	Mobility
BMI	Respiratory Rate	Lactate	Nutrition
	Temperature	BUN	Friction / Shear
	SpO <sub>2</sub>	Creatinine	
Underlying Disease	Medication	Glucose	Invasive Procedure
Diabetes	Relaxant	Cholesterol	Invasive Ventilation
Cerebrovascular Disease	Sedation	Albumin	CRRT
Hypertension	Vasopressor	Total Bilirubin	ECMO
Arterial Disease		AST	
Skin Disease		ALT	Others
	Sedation Scale	WBC	Repositioning
	SAS	Hemoglobin	Restraint
			Diarrhea

**Figure 1.** List of input variables.

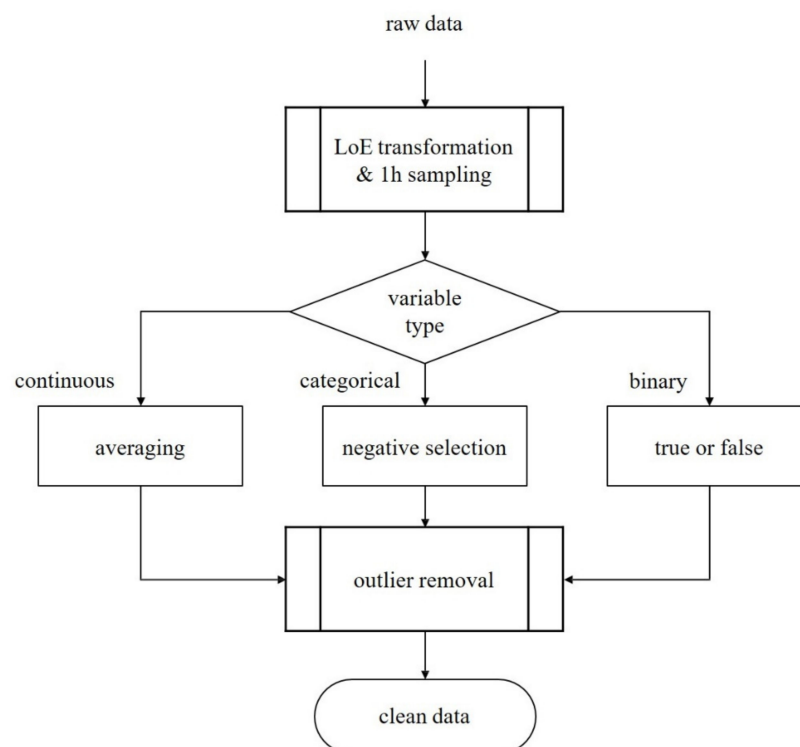
### 2.3. Data Preprocessing

We performed the following preprocessing steps to prepare patients' time series:

- We sampled patient data hourly from ICU admission to discharge;
- If multiple sampling data exist in an hour, we selected the average values for continuous variables, the most negative values for categorical variables, and binarized medication information;
- We applied the interquartile range method to discard outliers.

Figure 2 displays a flow chart of the preprocessing steps.

The input variables of a dataset have distinct scales, which is problematic for effectively training ML models. Thus, we performed min–max scaling with the range  $[-1, 1]$ . We performed forward filling and then mean filling to impute missing values. This imputation was not performed when using GRU-D and GRU-D++ because they automatically impute missing values internally.



**Figure 2.** Flow chart of the preprocessing steps.

#### 2.4. Prediction Models

In this study, we compare nine ML and DL models. Logistic regression (LR), decision tree (DT), random forest (RF), and extreme gradient boosting machine (XGBoost) are representative machine learning models that have been widely used in previous studies. However, they can only take 1 h of data as input and cannot capture time-varying information. Therefore, they are not suitable to handle real-time data.

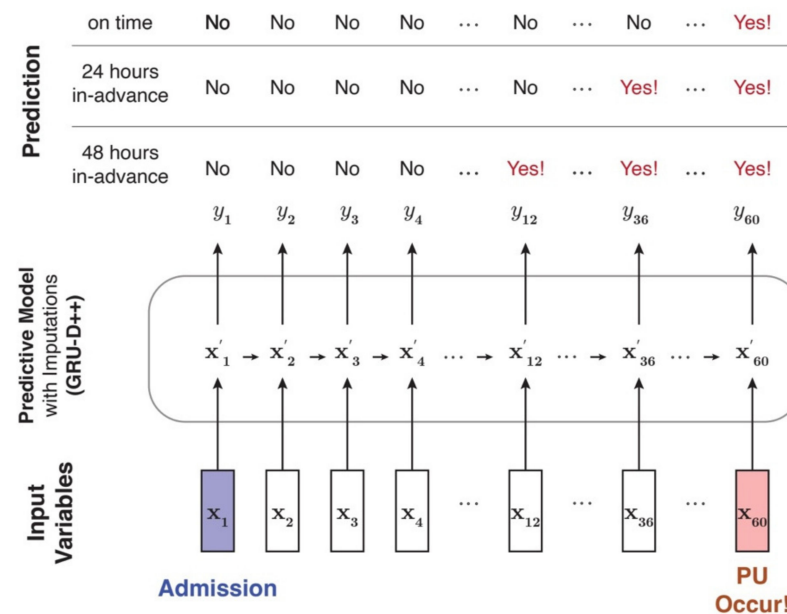
RNN, GRU, and LSTM are deep learning models for time series data. Since they take information from the previous time step as input again, they can capture time-varying information. Due to this advantage, they are widely used in real-time applications. However, the missing value problem is another challenge in the medical domain. Real-world electronic health records (EHR) contain many missing values, and missing values must be imputed appropriately before being input into a machine learning model.

GRU-D automatically imputes missing values internally and has demonstrated good performance in EHR datasets compared to other ML and DL models with traditional imputation methods such as mean fill. This result indicates that the imputation mechanism of GRU-D is effective. However, we found that GRU-D still has a problem in high missing-value rate datasets. GRU-D requires the last observation of a variable to generate an imputation value. However, the last observation is unavailable when the first value of a variable is missing.

In recent years, various missing value imputation methods for multivariate time series have been proposed [23–27]. Notably, many of these methods adopt a two-phase learning process [23–27], wherein the entire dataset is first imputed using an imputation model, followed by the training of a classification model. However, this two-phase approach is time-consuming for both training and inference. Some recent methods utilize generative adversarial networks (GANs) [24–26], but it is well known that training GANs is notoriously difficult.

To overcome this problem, we developed a novel deep learning model named GRU-D++ in this study. Our GRU-D++ is an end-to-end (training and inference of imputation and classification are simultaneously performed) and RNN-based model, practical in real-world scenarios. In addition, we focused on making GRU-D++ able to handle high missing-rate datasets effectively. To the best of our knowledge, GRU-D++ is the first model that explicitly

considers high missing-rate data. Details of GRU-D++ can be found in Appendix A. Figure 3 depicts the overview of our proposed real-time in-advance PU prediction system.



**Figure 3.** Overview of the proposed in-advance PU prediction system. The blue cell indicates the patient’s admission time, and the red cell indicates the PU occurrence time. Here,  $x_t$  represents the input variables updated every hour, and  $x'_t$  represents imputed complete variables. Furthermore,  $y_t$  is the output of whether PU occurs or not, which was used for training and evaluating our model. The definition of  $y_t$  can be changed according to the in-advance prediction setting. For example, in 24 h of in-advance prediction, if PU occurs 60 h after admission, the first “yes” should be  $y_{36}$ .

### 3. Results

#### 3.1. Baseline Characteristics

The internal cohort, MIMIC-IV, consists of 67,175 patients, whereas the external cohort, KNUH, consists of 6876 patients. In the internal cohort, we used 53,740 patients for training and 13,435 for validation. Table 1 displays the baseline characteristics of both the internal and external cohorts.

**Table 1.** Baseline characteristics.

Variables	MIMIC-IV (Internal)						KNUH (External)		
	Development Cohort			Validation Cohort			Validation Cohort		
	Control (n = 51,600)	Pressure Ulcer (n = 2140)	p-Value	Control (n = 12,876)	Pressure Ulcer (n = 559)	p-Value	Control (n = 6317)	Pressure Ulcer (n = 559)	p-Value
<b>Demographics</b>									
Male, n (%)	28,600 (55.43%)	1265 (59.11%)	<0.001	7125 (55.34%)	352 (62.97%)	<0.001	1575 (44.82%)	164 (65.86%)	<0.001
Age *	63.74 (±17.10)	65.44 (±15.22)	<0.001	63.76 (±16.89)	64.98 (±16.25)	0.096	61.32 (±20.90)	72.60 (±13.08)	<0.001
Height (cm) *	168.56 (±9.08)	169.14 (±10.34)	0.004	168.49 (±9.07)	169.52 (±10.52)	0.009	161.75 (±8.57)	162.32 (±9.39)	0.314
Body Weight (kg) *	81.92 (±35.39)	85.49 (±28.56)	<0.001	81.98 (±27.48)	83.94 (±24.21)	0.097	61.89 (±14.38)	56.78 (±14.33)	<0.001
BMI *	28.79 (±13.01)	29.79 (±9.35)	<0.001	28.84 (±9.44)	29.12 (±7.59)	0.481	23.61 (±4.88)	21.46 (±4.77)	<0.001

Table 1. Cont.

Variables	MIMIC-IV (Internal)						KNUH (External)		
	Development Cohort			Validation Cohort			Validation Cohort		
	Control (n = 51,600)	Pressure Ulcer (n = 2140)	p-Value	Control (n = 12,876)	Pressure Ulcer (n = 559)	p-Value	Control (n = 6317)	Pressure Ulcer (n = 559)	p-Value
<b>Underlying Disease</b>									
DM, n (%)	15,243 (29.54%)	728 (34.02%)	<0.001	3661 (28.44%)	192 (34.35%)	0.003	221 (3.50%)	38 (6.80%)	<0.001
CVD, n (%)	8236 (15.96%)	465 (21.730%)	<0.001	2019 (15.68%)	118 (21.11%)	<0.001	793 (12.55%)	111 (19.86%)	<0.001
Hypertension, n (%)	32,541 (63.06%)	1385 (64.72%)	0.125	8081 (62.76%)	374 (66.91%)	0.052	141 (2.23%)	10 (1.79%)	0.593
Arterial Disease, n (%)	5958 (11.55%)	385 (17.99%)	<0.001	1496 (11.62%)	100 (17.89%)	<0.001	71 (1.12%)	10 (1.79%)	0.233
Skin Disease, n (%)	4852 (9.40%)	510 (23.83%)	<0.001	1235 (9.59%)	136 (24.33%)	<0.001	101 (1.60%)	55 (9.84%)	<0.001

\*: mean and standard deviation, DM: diabetes mellitus, CVD: cerebrovascular disease, BMI: body mass index.

### 3.2. Predictive Performances

We used the area under the receiver operating characteristics curve (AUROC) and the area under the precision–recall curve (AUPRC) to evaluate the experimental models. AUROC is a common evaluation metric for binary classification. AUPRC is a similar metric to AUROC, but it considers class imbalance. We not only evaluated the performances at the PU occurrence time but also assessed early prediction performances (i.e., 12, 24, and 48 h in-advance predictions).

Table 2 presents the performances of the experimental models on the internal validation set. The results indicated that DT performed poorly with an AUROC of 0.569 at PU occurrence time, whereas the other ML models, LR, RF, and XGBoost exhibited higher performances with AUROCs of 0.818, 0.818, and 0.814 at the PU occurrence time. All RNN models exhibited higher performances than the ML models. In particular, GRU exhibited an AUROC of 0.918 at the PU occurrence time. GRU-D outperformed other RNNs. Furthermore, GRU-D++ exhibited the best performances in all experimental settings, indicating the superiority of its imputation mechanism.

We evaluated the performance of the Braden scale, which is widely used to measure PU risks in the ICU. The AUROC of the Braden score showed lower results at all times compared to the AI model (Table 3).

Table 2. Performance comparison of various models on the internal validation set.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.818 (±0.000)	0.433 (±0.001)	0.814 (±0.000)	0.438 (±0.001)	0.810 (±0.000)	0.442 (±0.001)	0.804 (±0.000)	0.450 (±0.001)
DT	0.569 (±0.007)	0.145 (±0.006)	0.567 (±0.007)	0.151 (±0.005)	0.566 (±0.005)	0.156 (±0.004)	0.569 (±0.004)	0.172 (±0.003)
RF	0.818 (±0.000)	0.407 (±0.005)	0.813 (±0.001)	0.412 (±0.004)	0.808 (±0.001)	0.414 (±0.004)	0.801 (±0.001)	0.424 (±0.004)
XGBoost	0.814 (±0.002)	0.404 (±0.006)	0.807 (±0.003)	0.404 (±0.006)	0.805 (±0.006)	0.410 (±0.012)	0.797 (±0.003)	0.424 (±0.011)
Simple RNN	0.860 (±0.021)	0.489 (±0.037)	0.864 (±0.003)	0.506 (±0.015)	0.859 (±0.010)	0.511 (±0.022)	0.837 (±0.003)	0.507 (±0.015)
GRU	0.918 (±0.003)	0.657 (±0.011)	0.913 (±0.002)	0.657 (±0.009)	0.905 (±0.001)	0.649 (±0.008)	0.885 (±0.003)	0.630 (±0.007)
LSTM	0.909 (±0.005)	0.625 (±0.020)	0.903 (±0.004)	0.619 (±0.019)	0.900 (±0.007)	0.620 (±0.020)	0.880 (±0.005)	0.605 (±0.013)



Table 2. Cont.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
GRU-D	0.944 ( $\pm 0.003$ )	0.737 ( $\pm 0.008$ )	0.938 ( $\pm 0.002$ )	0.723 ( $\pm 0.009$ )	0.930 ( $\pm 0.003$ )	0.712 ( $\pm 0.005$ )	0.909 ( $\pm 0.002$ )	0.686 ( $\pm 0.007$ )
GRU-D++	<b>0.945</b> ( $\pm 0.003$ )	<b>0.742</b> ( $\pm 0.005$ )	<b>0.940</b> ( $\pm 0.002$ )	<b>0.730</b> ( $\pm 0.004$ )	<b>0.933</b> ( $\pm 0.001$ )	<b>0.722</b> ( $\pm 0.005$ )	<b>0.912</b> ( $\pm 0.003$ )	<b>0.699</b> ( $\pm 0.007$ )

LR: logistic regression, DT: decision tree, RF: random forest, XGBoost: extreme gradient boosting machine, RNN: recurrent neural network, GRU: gated recurrent unit, LSTM: long short-term memory, GRU-D: gated recurrent unit with a decay.

Table 3. Performance comparison of different input variables on the internal validation set.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
Braden scale	0.730	0.201	0.730	0.212	0.731	0.223	0.732	0.245
GRU-D++	<b>0.945</b> ( $\pm 0.003$ )	<b>0.742</b> ( $\pm 0.005$ )	<b>0.940</b> ( $\pm 0.002$ )	<b>0.730</b> ( $\pm 0.004$ )	<b>0.933</b> ( $\pm 0.001$ )	<b>0.722</b> ( $\pm 0.005$ )	<b>0.912</b> ( $\pm 0.003$ )	<b>0.699</b> ( $\pm 0.007$ )
GRU-D++ <sup>10</sup>	0.923 ( $\pm 0.004$ )	0.670 ( $\pm 0.012$ )	0.918 ( $\pm 0.003$ )	0.665 ( $\pm 0.004$ )	0.911 ( $\pm 0.005$ )	0.658 ( $\pm 0.003$ )	0.888 ( $\pm 0.007$ )	0.637 ( $\pm 0.008$ )
GRU-D++ (w/o Braden)	0.934 ( $\pm 0.002$ )	0.671 ( $\pm 0.009$ )	0.928 ( $\pm 0.002$ )	0.660 ( $\pm 0.010$ )	0.920 ( $\pm 0.001$ )	0.648 ( $\pm 0.010$ )	0.901 ( $\pm 0.002$ )	0.631 ( $\pm 0.011$ )

GRU-D++<sup>10</sup>: GRU-D++ trained with top ten important variables, GRU-D++ (w/o braden): GRU-D++ trained without the Braden scales.

We used 48 variables in the experiment, but these variables may not be available in other medical centers. Since the utilizability of the prediction system is essential, we trained and evaluated GRU-D++ again using the top ten variables, which have the highest SHAP values. The top ten variables are displayed in Figure 4. In Table 3, GRU-D++<sup>10</sup> indicates GRU-D++ with the top ten variables. GRU-D++<sup>10</sup> outperformed other RNN models trained with 48 variables, even though it only uses ten variables. Additionally, we conducted the GRU-D++ with 42 variables out of 48 variables, excluding the six variables corresponding to the Braden scales, and the results were presented on GRU-D++ (w/o Braden) in Table 3.

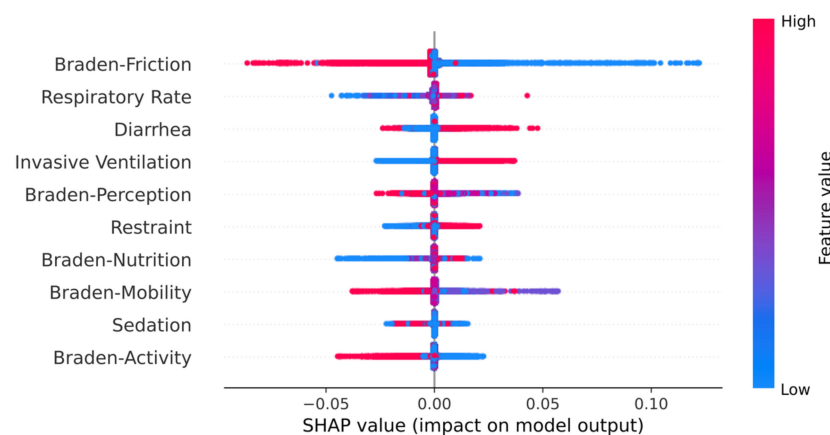


Figure 4. SHAP values of top ten influential variables.

Table 4 shows the performances of GRU-D++ on the external validation set. GRU-D++ achieved a promising performance (an AUROC of 0.807 at the PU occurrence time). To improve the performance of GRU-D++ on the external validation set, we rescale the external set to have a similar distribution to the internal set. Consequently, GRU-D++ (rescale) exhibited a significantly higher performance than GRU-D++ (AUROC 0.895 vs. 0.807 at the PU occurrence time). In addition, we evaluated the performance of GRU-D++ after fine-tuning. Specifically, we randomly split the external set into training (10%), model selection (10%), and validation (80%) sets. As a result, GRU-D++ (fine-tune) exhibited the

highest performances on the external set. A comparison of all experimental models on the external set is presented in Appendix A.

**Table 4.** Performances of GRU-D++ on the external validation set.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
GRU-D++	0.807 ( $\pm 0.015$ )	0.350 ( $\pm 0.026$ )	0.827 ( $\pm 0.018$ )	0.380 ( $\pm 0.031$ )	0.826 ( $\pm 0.035$ )	0.398 ( $\pm 0.054$ )	0.848 ( $\pm 0.016$ )	0.445 ( $\pm 0.030$ )
GRU-D++ (rescale)	0.895 ( $\pm 0.032$ )	0.534 ( $\pm 0.078$ )	0.892 ( $\pm 0.022$ )	0.538 ( $\pm 0.044$ )	0.889 ( $\pm 0.017$ )	0.539 ( $\pm 0.049$ )	0.864 ( $\pm 0.017$ )	0.514 ( $\pm 0.023$ )
GRU-D++ (fine-tune)	0.898 ( $\pm 0.018$ )	0.523 ( $\pm 0.051$ )	0.901 ( $\pm 0.016$ )	0.542 ( $\pm 0.045$ )	0.895 ( $\pm 0.019$ )	0.543 ( $\pm 0.050$ )	0.897 ( $\pm 0.008$ )	0.575 ( $\pm 0.027$ )

## 4. Discussion

### 4.1. Analysis of Predictive Performances

We conducted extensive experiments using various ML and DL models to predict PU occurrences. DT exhibited poor predictive performance because the model is susceptible to overfitting, that is, it cannot accurately predict unseen data. By contrast, other baseline models, LR, RF, and XGBoost, exhibited good performances because of their strong regularization abilities. This is a well-known phenomenon, and previous studies have demonstrated that XGBoost has good predictive power.

The RNN models, Simple RNN, GRU, and LSTM, demonstrated considerably higher performances than the ML models, highlighting the importance of considering time-varying information. Many researchers still widely use boosting machine for their studies, but we showed that RNN models are very helpful for time series data. In addition, GRU and LSTM significantly outperformed Simple RNN, which is unsuitable for capturing long-term temporal information. This result suggests that patients' long-term information is important in predicting PU occurrences. GRU outperformed LSTM, which is a well-known phenomenon. LSTM requires more data samples than GRU and is susceptible to the overfitting problem.

GRU-D and GRU-D++ significantly outperformed the other ML and RNN models, demonstrating that conventional imputation techniques are insufficient to handle missing values and that the trainable imputation methods of GRU-D and GRU-D++ are effective. In addition, GRU-D++ outperformed GRU-D, which indicates that GRU-D++ is more effective in handling high missing-value rate datasets than GRU-D. We also evaluated GRU-D++ with fewer variables (i.e., ten of the highest SHAP scores) and found that it outperformed the other RNNs with 48 variables. This is an important finding because other medical centers may be unable to collect all 48 variables.

We evaluated GRU-D++ on the external validation set and determined that it exhibited good predictive performances (AUROC > 0.8). An issue with directly applying GRU-D++ trained with the internal validation set to the external validation set is that the imputation mechanism of GRU-D++ was trained to predict the internal validation set accurately. However, the distributions of the internal and external validation sets may differ. Thus, the imputation mechanism may fail to generate accurate imputation values on the external validation set, which can result in a deterioration in the predictive performance of GRU-D++. To address this problem, we rescaled the external validation set to have a distribution similar to the internal validation set as much as possible. GRU-D++ with rescaling exhibited a considerably improved performance over basic GRU-D++, demonstrating that GRU-D++ can achieve predictive performances close to those of the internal validation set on any other external sets (e.g., datasets from other medical centers) via rescaling. Furthermore, we retrained (fine-tuned) GRU-D++ with a small part of the external validation set. It demonstrated superior predictive performance with a few data samples and is effective for other medical centers.



#### 4.2. Clinical Findings

Similar to previous studies, we developed PU prediction systems using various ML and DL models, classifying data into PU and non-PU groups. However, although previous studies [3,8] have predicted PU occurrences using average values of each variable as input, we used time series data to predict PU occurrences. This study overcomes the performance degradation problem of existing studies using time series information and a novel missing value imputation mechanism. Our system is very effective in real-time in-advance prediction.

The MIMIC-IV dataset used for model development has been widely used in various studies because of its diverse variables. However, the dataset has a disproportionate representation, with Whites and Blacks accounting for approximately 78.56% and Asians making up only approximately 2.94% of the total dataset. Due to this racial imbalance, applying the results of studies only performed with the MIMIC-IV to Asians is inappropriate. One advantage of this study is presenting a prediction system suitable for diverse ethnicities by validating the system developed using the MIMIC-IV with Asian data (KNUH).

An analysis of the SHAP values revealed that, as expected, the Braden scores were among the top ten essential variables for predicting PU occurrences. Interestingly, PU prediction using only 42 variables without the Braden score demonstrated only 1.2–1.4% degradation in AUROC to that with 48 variables (Table 3). This finding indicates that the 42 variables contain valuable information about PU occurrences. Furthermore, using GRU-D++ allows for accurate predictions without the Braden scales, given that these scales inherently have been measured based on the patient's condition related to the 42 variables. Clinically, this could potentially reduce the workload of nurses who measure and record the Braden score. We anticipate that this model could be a valuable tool for future clinical use.

#### 5. Conclusions

PU adversely affects patient health and increases the workload of medical staff in the ICU. Therefore, early prediction of PU is crucial. This study compared various ML and DL models to develop accurate PU prediction systems. As a result, we developed the GRU-D++ model, which covers a high missing-value rate. This model demonstrated better performance than the Braden scale, and it even showed high performance when implemented without the Braden scale. It is expected to contribute significantly to reducing the workload of medical staff in the future.

GRU-D++ can be helpful to other researchers aiming to predict PU occurrences accurately. Furthermore, in future work, we plan to develop a system that integrates the precise prediction of PU occurrence region and grade predictions for PU.

**Author Contributions:** Conceptualization, J.H., S.-S.H. and H.-S.C.; data curation, S.K., T.H., J.K. and S.W.P.; formal analysis, M.K. and D.K. (Dowon Kim); methodology, M.K.; supervision, D.H.M., Y.H., W.J.K., S.-J.L. and Y.K.; validation, D.L. and D.K. (Dohyun Kim); writing—original draft, M.K. and T.-H.K.; writing—review and editing, S.-S.H. and H.-S.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C1074).

**Institutional Review Board Statement:** This study was approved by the Institutional Review Board of Kangwon National University Hospital (KNUH-2022-09-013-001, 25 October 2022).

**Informed Consent Statement:** This study was a retrospective chart review. Therefore, we obtained approval for consent waiver from our institution's Institutional Review Board (IRB).

**Data Availability Statement:** Data sharing is possible via the approval process of the Data Review Committee of Kangwon National University Hospital.

**Conflicts of Interest:** Minkyu Kim, Dowon Kim, Donghoon Lee and Dohyun Kim have been involved as Ziovision Co., Ltd. Yoon Kim and Hyun-Soo Choi have been involved as a membership in Ziovision Co., Ltd.

## Appendix A

### Appendix A.1. Details of GRU-D++

Having missing values in real-world multivariate time series causes poor prediction results. To address this problem, many studies have been conducted to impute missing values [25–27]. GRU-D [23] can directly handle missing values. GRU-D implicitly learns the weighted sum between the means and last observations of input variables. It outperforms RNNs with conventional imputation methods on the MIMIC-III and PhysioNet datasets, which are real-world electronic health records. However, GRU-D is still not suitable for high missing-value rate datasets. To overcome this limitation, we proposed a novel GRU-based model for missing values called GRU-D++ in this study.

Let  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) \in \mathbb{R}^{T \times M}$  be a multivariate time series, where  $T$  is the number of time steps and  $M$  is the number of input variables. Given a multivariate time series  $\mathbf{X}$ , we define a mask  $\mathbf{M} \in \{0, 1\}^{T \times M}$  for  $\mathbf{X}$  as follows:

$$\mu_{t,m} = \begin{cases} 0, & \text{if } x_{t,m} \text{ is N/A} \\ 1, & \text{otherwise} \end{cases} \in \mathbf{M} \quad (\text{A1})$$

where N/A indicates a missing value. Then, the last observation  $\tilde{\mathbf{X}} \in \mathbb{R}^{T \times M}$  for  $\mathbf{X}$  and the corresponding mask  $\tilde{\mathbf{M}} \in \{0, 1\}^{T \times M}$  are defined as follows:

$$\tilde{x}_{t,m} = \begin{cases} x_{t,m}, & \text{if } \mu_{t,m} = 1 \\ \tilde{x}_{t-1,m}, & \text{if } \mu_{t,m} = 0 \text{ and } t > 1 \\ \text{N/A}, & \text{otherwise} \end{cases} \in \tilde{\mathbf{X}} \quad (\text{A2})$$

$$\tilde{\mu}_{t,m} = \begin{cases} 0, & \text{if } \tilde{x}_{t,m} \text{ is N/A} \\ 1, & \text{otherwise} \end{cases} \in \tilde{\mathbf{M}} \quad (\text{A3})$$

We then record time intervals  $\Delta \in \mathbb{R}^{T \times M}$  as follows:

$$\delta_{t,m} = \begin{cases} x_{t,m} - s_{t-1} + \delta_{t-1,m}, & \text{if } t > 1 \text{ and } \mu_{t-1,m} = 0 \\ x_{t,m} - s_{t-1}, & \text{if } t > 1 \text{ and } \mu_{t-1,m} = 1 \\ 0, & \text{otherwise} \end{cases} \in \Delta \quad (\text{A4})$$

where  $\mathbf{s} = (s_1, s_2, \dots, s_T) \in \mathbb{R}^T$  indicates the timestamps for  $\mathbf{X}$ . (A4) is the same as the method detailed in [23]. Figure A1 illustrates an example of  $\mathbf{X}$ ,  $\tilde{\mathbf{X}}$ ,  $\mathbf{M}$ ,  $\tilde{\mathbf{M}}$ ,  $\mathbf{s}$ , and  $\Delta$ .

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} 7 & 2 & \text{N/A} \\ 4 & \text{N/A} & \text{N/A} \\ \text{N/A} & 5 & 4 \\ \text{N/A} & \text{N/A} & \text{N/A} \\ 9 & \text{N/A} & 9 \\ 3 & \text{N/A} & 1 \end{bmatrix} & \tilde{\mathbf{X}} &= \begin{bmatrix} 7 & 2 & \text{N/A} \\ 4 & 2 & \text{N/A} \\ 4 & 5 & 4 \\ 4 & 5 & 4 \\ 9 & 5 & 9 \\ 3 & 5 & 1 \end{bmatrix} & \mathbf{s} &= \begin{bmatrix} 0 \\ 0.3 \\ 0.7 \\ 2.1 \\ 2.5 \\ 3.4 \end{bmatrix} \\ \mathbf{M} &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} & \tilde{\mathbf{M}} &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & \Delta &= \begin{bmatrix} 0 & 0 & 0 \\ 0.3 & 0.3 & 0.3 \\ 0.4 & 0.7 & 0.7 \\ 1.8 & 1.4 & 1.4 \\ 2.2 & 1.8 & 1.8 \\ 0.9 & 2.7 & 0.9 \end{bmatrix} \end{aligned}$$

**Figure A1.** An example of  $\mathbf{X}$ ,  $\tilde{\mathbf{X}}$ ,  $\mathbf{M}$ ,  $\tilde{\mathbf{M}}$ ,  $\mathbf{s}$ , and  $\Delta$ . This time series has three variables and six time steps. N/A = not applicable.

Because time intervals between each time step can be different, we applied a trainable soft time decaying mechanism to input variables and hidden states of GRU-D++. The decaying weight vector  $\gamma_{x_t}$  and  $\gamma_{h_t}$  for the input  $x_t$  and hidden state  $h_t$  are defined as follows:

$$\gamma_{x_t} = \exp(-\max(\mathbf{0}, \delta_t \mathbf{W}_{x_t} + \mathbf{b}_{x_t})) \odot \tilde{\mu}_t, \quad (\text{A5})$$

$$\gamma_{h_t} = \exp(-\max(\mathbf{0}, \delta_t \mathbf{W}_{h_t} + \mathbf{b}_{h_t})), \quad (\text{A6})$$

where  $\mathbf{W}_{x_t} \in \mathbb{R}^{M \times M}$ ,  $\mathbf{W}_{h_t} \in \mathbb{R}^{M \times K}$ ,  $\mathbf{b}_{x_t} \in \mathbb{R}^M$ , and  $\mathbf{b}_{h_t} \in \mathbb{R}^K$  are trainable weights and biases. Unlike GRU-D,  $\mathbf{W}_{x_t}$  is a dense matrix, and  $\gamma_{x_t, m}$  is held to 0 if the last observation  $\tilde{\mu}_{t, m}$  is missing. These decaying mechanisms help GRU-D++ capture temporal dynamics in various time intervals. The trainable imputation method is defined as follows:

$$x'_{t, m} = m_{t, m} x_{t, m} + (1 - m_{t, m}) \times (\gamma_{x_t, m} \tilde{x}_{t, m} + (1 - \gamma_{x_t, m}) \bar{x}_m) \quad (\text{A7})$$

where  $\bar{x}_m$  indicates the arithmetic mean of the  $m^{\text{th}}$  variable. Finally, GRU-D++ is defined as follows:

$$h'_{t-1} = \gamma_{h_t} h_{t-1}, \quad (\text{A8})$$

$$r_t = \sigma(x'_t \mathbf{W}_r + h'_{t-1} \mathbf{U}_r + m_t \mathbf{V}_r + \tilde{m}_t \mathbf{Q}_r + \mathbf{b}_r), \quad (\text{A9})$$

$$z_t = \sigma(x'_t \mathbf{W}_z + h'_{t-1} \mathbf{U}_z + m_t \mathbf{V}_z + \tilde{m}_t \mathbf{Q}_z + \mathbf{b}_z), \quad (\text{A10})$$

$$\tilde{h}_t = \tanh(x'_t \mathbf{W} + (r_t \odot h'_{t-1}) \mathbf{U} + m_t \mathbf{V} + \tilde{m}_t \mathbf{Q} + \mathbf{b}), \quad (\text{A11})$$

$$h_t = (\mathbf{1} - z_t) \odot h'_{t-1} + z_t \odot \tilde{h}_t. \quad (\text{A12})$$

#### Appendix A.2. Further Experiments

In this section, we present a comparison of all experimental models on the external validation set. Table A1 displays the performances of the experimental models on the external validation set, where GRU outperforms GRU-D++. The imputation mechanisms of GRU-D and GRU-D++ were trained to predict the internal set accurately. Therefore, they may fail to generate precise imputation values on the external set because the distributions of the internal and external sets may differ. Consequently, GRU-D and GRU-D++ exhibit lower performances than GRU. To address this problem, we rescaled the external validation set to have a similar distribution to the internal set. Table A2 presents the performances of the experimental models after rescaling, and a considerable improvement in the GRU-D++ performance. Furthermore, we evaluated the performances of RNNs on the external validation set after fine-tuning. Thus, we retrained the RNNs with a small subset of the external set and evaluated the performances on the remaining set. Table A3 demonstrates the performances of the RNNs on the external set after fine-tuning, where GRU-D++ consistently outperforms the other experimental models in all settings.

**Table A1.** Performance comparison on the external validation set.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.653 ( $\pm 0.008$ )	0.212 ( $\pm 0.002$ )	0.655 ( $\pm 0.008$ )	0.220 ( $\pm 0.002$ )	0.656 ( $\pm 0.008$ )	0.229 ( $\pm 0.003$ )	0.667 ( $\pm 0.008$ )	0.249 ( $\pm 0.003$ )
DT	0.504 ( $\pm 0.015$ )	0.125 ( $\pm 0.005$ )	0.502 ( $\pm 0.037$ )	0.131 ( $\pm 0.010$ )	0.519 ( $\pm 0.012$ )	0.141 ( $\pm 0.006$ )	0.520 ( $\pm 0.014$ )	0.153 ( $\pm 0.009$ )
RF	0.652 ( $\pm 0.011$ )	0.195 ( $\pm 0.010$ )	0.631 ( $\pm 0.038$ )	0.203 ( $\pm 0.016$ )	0.645 ( $\pm 0.029$ )	0.215 ( $\pm 0.012$ )	0.662 ( $\pm 0.033$ )	0.237 ( $\pm 0.016$ )
XGBoost	0.603 ( $\pm 0.025$ )	0.189 ( $\pm 0.010$ )	0.605 ( $\pm 0.022$ )	0.199 ( $\pm 0.015$ )	0.627 ( $\pm 0.026$ )	0.221 ( $\pm 0.018$ )	0.635 ( $\pm 0.016$ )	0.233 ( $\pm 0.015$ )

Table A1. Cont.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
RNN	0.768 (±0.033)	0.290 (±0.027)	0.753 (±0.028)	0.295 (±0.024)	0.759 (±0.023)	0.309 (±0.025)	0.751 (±0.008)	0.315 (±0.006)
GRU	<b>0.880</b> <b>(±0.015)</b>	<b>0.469</b> <b>(±0.035)</b>	<b>0.875</b> <b>(±0.006)</b>	<b>0.471</b> <b>(±0.025)</b>	<b>0.873</b> <b>(±0.010)</b>	<b>0.480</b> <b>(±0.031)</b>	<b>0.867</b> <b>(±0.008)</b>	<b>0.490</b> <b>(±0.024)</b>
LSTM	0.816 (±0.042)	0.370 (±0.054)	0.833 (±0.013)	0.390 (±0.027)	0.829 (±0.017)	0.405 (±0.029)	0.815 (±0.027)	0.409 (±0.041)
GRU-D	0.762 (±0.019)	0.267 (±0.015)	0.768 (±0.015)	0.285 (±0.009)	0.777 (±0.022)	0.307 (±0.021)	0.759 (±0.026)	0.291 (±0.029)
GRU-D++	0.807 (±0.015)	0.350 (±0.026)	0.827 (±0.018)	0.380 (±0.031)	0.826 (±0.035)	0.398 (±0.054)	0.848 (±0.016)	0.445 (±0.030)

Table A2. Performance comparison on the external validation set with rescaling.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.643 (±0.010)	0.217 (±0.004)	0.644 (±0.009)	0.226 (±0.004)	0.646 (±0.010)	0.235 (±0.004)	0.657 (±0.007)	0.255 (±0.004)
DT	0.542 (±0.022)	0.139 (±0.011)	0.503 (±0.060)	0.133 (±0.013)	0.539 (±0.024)	0.151 (±0.011)	0.541 (±0.009)	0.161 (±0.005)
RF	0.513 (±0.051)	0.162 (±0.026)	0.590 (±0.044)	0.197 (±0.023)	0.581 (±0.067)	0.202 (±0.029)	0.555 (±0.073)	0.202 (±0.029)
XGBoost	0.565 (±0.033)	0.179 (±0.016)	0.569 (±0.059)	0.187 (±0.022)	0.555 (±0.039)	0.192 (±0.019)	0.587 (±0.048)	0.218 (±0.016)
RNN	0.755 (±0.043)	0.278 (±0.028)	0.774 (±0.029)	0.321 (±0.025)	0.779 (±0.031)	0.336 (±0.029)	0.749 (±0.011)	0.324 (±0.010)
GRU	0.890 (±0.007)	0.513 (±0.024)	0.882 (±0.003)	0.503 (±0.017)	0.880 (±0.006)	0.515 (±0.015)	<b>0.874</b> <b>(±0.005)</b>	<b>0.521</b> <b>(±0.006)</b>
LSTM	0.834 (±0.014)	0.393 (±0.027)	0.838 (±0.031)	0.394 (±0.048)	0.819 (±0.017)	0.390 (±0.026)	0.811 (±0.028)	0.398 (±0.045)
GRU-D	0.824 (±0.045)	0.370 (±0.088)	0.840 (±0.039)	0.407 (±0.070)	0.810 (±0.035)	0.378 (±0.050)	0.726 (±0.083)	0.289 (±0.103)
GRU-D++	<b>0.895</b> <b>(±0.032)</b>	<b>0.534</b> <b>(±0.078)</b>	<b>0.892</b> <b>(±0.022)</b>	<b>0.538</b> <b>(±0.044)</b>	<b>0.889</b> <b>(±0.017)</b>	<b>0.539</b> <b>(±0.049)</b>	0.864 (±0.017)	0.514 (±0.023)

Table A3. Performance comparison on the external validation set with finetuning.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
RNN	0.786 (±0.014)	0.307 (±0.017)	0.773 (±0.032)	0.318 (±0.019)	0.766 (±0.012)	0.324 (±0.021)	0.771 (±0.012)	0.338 (±0.005)
GRU	0.873 (±0.011)	0.458 (±0.029)	0.872 (±0.008)	0.466 (±0.024)	0.868 (±0.007)	0.473 (±0.026)	0.864 (±0.007)	0.485 (±0.023)
LSTM	0.826 (±0.032)	0.375 (±0.047)	0.832 (±0.026)	0.398 (±0.025)	0.838 (±0.020)	0.414 (±0.036)	0.811 (±0.022)	0.399 (±0.036)
GRU-D	0.858 (±0.015)	0.426 (±0.046)	0.853 (±0.023)	0.424 (±0.044)	0.856 (±0.017)	0.442 (±0.035)	0.845 (±0.018)	0.438 (±0.029)
GRU-D++	<b>0.898</b> <b>(±0.018)</b>	<b>0.523</b> <b>(±0.051)</b>	<b>0.901</b> <b>(±0.016)</b>	<b>0.542</b> <b>(±0.045)</b>	<b>0.895</b> <b>(±0.019)</b>	<b>0.543</b> <b>(±0.050)</b>	<b>0.897</b> <b>(±0.008)</b>	<b>0.575</b> <b>(±0.027)</b>

**Table A4.** Performance comparison on the external validation set without Braden scales.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.707 (±0.002)	0.234 (±0.001)	0.709 (±0.002)	0.244 (±0.001)	0.711 (±0.002)	0.254 (±0.002)	0.716 (±0.002)	0.274 (±0.002)
DT	0.516 (±0.008)	0.129 (±0.004)	0.513 (±0.008)	0.133 (±0.003)	0.525 (±0.012)	0.143 (±0.006)	0.532 (±0.023)	0.157 (±0.013)
RF	0.683 (±0.004)	0.211 (±0.006)	0.680 (±0.011)	0.217 (±0.007)	0.694 (±0.008)	0.230 (±0.009)	0.693 (±0.004)	0.251 (±0.003)
XGBoost	0.676 (±0.010)	0.220 (±0.009)	0.675 (±0.008)	0.227 (±0.005)	0.683 (±0.013)	0.239 (±0.005)	0.692 (±0.008)	0.264 (±0.005)
RNN	0.761 (±0.014)	0.277 (±0.010)	0.777 (±0.015)	0.290 (±0.009)	0.775 (±0.022)	0.304 (±0.015)	0.772 (±0.006)	0.323 (±0.006)
GRU	<b>0.856</b> <b>(±0.010)</b>	<b>0.413</b> <b>(±0.021)</b>	<b>0.843</b> <b>(±0.008)</b>	<b>0.406</b> <b>(±0.015)</b>	<b>0.848</b> <b>(±0.016)</b>	<b>0.426</b> <b>(±0.026)</b>	<b>0.837</b> <b>(±0.027)</b>	<b>0.434</b> <b>(±0.058)</b>
LSTM	0.844 (±0.020)	0.389 (±0.026)	0.833 (±0.020)	0.388 (±0.043)	0.827 (±0.016)	0.385 (±0.038)	0.813 (±0.010)	0.390 (±0.022)
GRU-D	0.749 (±0.028)	0.270 (±0.029)	0.750 (±0.025)	0.286 (±0.032)	0.761 (±0.025)	0.302 (±0.033)	0.778 (±0.018)	0.344 (±0.030)
GRU-D++	0.783 (±0.033)	0.328 (±0.051)	0.795 (±0.034)	0.337 (±0.046)	0.803 (±0.030)	0.382 (±0.031)	0.818 (±0.034)	0.405 (±0.039)

**Table A5.** Performance comparison on the external validation set without Braden scales after rescaling.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
LR	0.697 (±0.003)	0.236 (±0.003)	0.697 (±0.003)	0.245 (±0.003)	0.699 (±0.003)	0.254 (±0.004)	0.702 (±0.003)	0.273 (±0.004)
DT	0.515 (±0.046)	0.132 (±0.019)	0.546 (±0.030)	0.148 (±0.016)	0.537 (±0.032)	0.148 (±0.013)	0.453 (±0.097)	0.138 (±0.016)
RF	0.538 (±0.050)	0.164 (±0.020)	0.612 (±0.042)	0.198 (±0.010)	0.552 (±0.058)	0.185 (±0.016)	0.525 (±0.056)	0.180 (±0.018)
XGBoost	0.537 (±0.057)	0.176 (±0.028)	0.592 (±0.046)	0.197 (±0.024)	0.560 (±0.062)	0.193 (±0.025)	0.568 (±0.050)	0.209 (±0.020)
RNN	0.771 (±0.026)	0.284 (±0.013)	0.761 (±0.016)	0.287 (±0.007)	0.751 (±0.009)	0.301 (±0.004)	0.764 (±0.014)	0.323 (±0.009)
GRU	0.866 (±0.008)	0.446 (±0.022)	0.844 (±0.019)	0.426 (±0.032)	0.850 (±0.021)	0.447 (±0.036)	0.826 (±0.040)	0.439 (±0.062)
LSTM	0.840 (±0.008)	0.380 (±0.010)	0.821 (±0.033)	0.381 (±0.029)	0.798 (±0.039)	0.368 (±0.043)	0.791 (±0.023)	0.376 (±0.021)
GRU-D	0.811 (±0.020)	0.408 (±0.036)	0.853 (±0.013)	0.469 (±0.016)	0.845 (±0.035)	0.468 (±0.048)	0.810 (±0.059)	0.431 (±0.115)
GRU-D++	<b>0.868</b> <b>(±0.037)</b>	<b>0.492</b> <b>(±0.078)</b>	<b>0.888</b> <b>(±0.012)</b>	<b>0.530</b> <b>(±0.034)</b>	<b>0.889</b> <b>(±0.014)</b>	<b>0.546</b> <b>(±0.040)</b>	<b>0.874</b> <b>(±0.032)</b>	<b>0.543</b> <b>(±0.062)</b>

**Table A6.** Performance comparison on the external validation set without Braden scales after finetuning.

	On Time		12 h In Advance		24 h In Advance		48 h In Advance	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
RNN	0.771 (±0.015)	0.296 (±0.009)	0.782 (±0.013)	0.315 (±0.007)	0.782 (±0.014)	0.329 (±0.015)	0.784 (±0.001)	0.346 (±0.004)
GRU	0.856 (±0.010)	0.426 (±0.018)	0.844 (±0.011)	0.406 (±0.020)	0.852 (±0.018)	0.436 (±0.029)	0.847 (±0.020)	0.447 (±0.046)
LSTM	0.853 (±0.009)	0.406 (±0.022)	0.851 (±0.012)	0.408 (±0.028)	0.837 (±0.025)	0.405 (±0.025)	0.835 (±0.001)	0.418 (±0.009)
GRU-D	0.852 (±0.010)	0.424 (±0.021)	0.85 (±0.012)	0.439 (±0.031)	0.857 (±0.016)	0.449 (±0.037)	0.845 (±0.006)	0.459 (±0.024)
GRU-D++	<b>0.881</b> <b>(±0.025)</b>	<b>0.500</b> <b>(±0.073)</b>	<b>0.867</b> <b>(±0.020)</b>	<b>0.475</b> <b>(±0.059)</b>	<b>0.878</b> <b>(±0.016)</b>	<b>0.504</b> <b>(±0.048)</b>	<b>0.875</b> <b>(±0.011)</b>	<b>0.518</b> <b>(±0.039)</b>

## References

1. Araujo, S.M.; Sousa, P.; Dutra, I. Clinical decision support systems for pressure ulcer management: Systematic review. *JMIR Med. Inform.* **2020**, *8*, e21621. [\[CrossRef\]](#)
2. PAN Pacific. Prevention and treatment of pressure ulcers: Quick reference guide. *Natl. Press. Ulcer Advis. Panel* **2014**, *75*, 1–72.
3. Šín, P.; Hokynková, A.; Marie, N.; Andrea, P.; Krč, R.; Podroužek, J. Machine learning-based pressure ulcer prediction in modular critical care data. *Diagnostics* **2022**, *12*, 850. [\[CrossRef\]](#)
4. Kim, E.; Ham, S.; Jung, B.K.; Park, J.-W.; Kim, J.; Lee, J.H. Effect of baicalin on wound healing in a mouse model of pressure ulcers. *Int. J. Mol. Sci.* **2023**, *24*, 329. [\[CrossRef\]](#)
5. Park, S.-H.; Lee, Y.-S.; Kwon, Y.-M. Predictive validity of pressure ulcer risk assessment tools for elderly: A meta-analysis. *West. J. Nurs. Res.* **2016**, *38*, 459–483. [\[CrossRef\]](#)
6. Weststrate, J.T.; Bruining, H.A. Pressure sores in an intensive care unit and related variables: A descriptive study. *Intensive Crit. Care Nurs.* **1996**, *12*, 280–284. [\[CrossRef\]](#)
7. Cho, I.S.; Yoon, H.Y.; Im, P.S.; Sook, L.H. Availability of nursing data in an electronic nursing record system for a development of a risk assessment tool for pressure ulcers. *J. Korean Soc. Med. Inform.* **2008**, *14*, 161–168. [\[CrossRef\]](#)
8. Cramer, E.M.; Seneviratne, M.G.; Sharifi, H.; Ozturk, A.; Hernandez-Boussard, T. Predicting the incidence of pressure ulcers in the intensive care unit using machine learning. *eGEMs* **2019**, *7*, 49. [\[CrossRef\]](#)
9. Cicceri, G.; De Vita, F.; Bruneo, D.; Merlino, G.; Puliafito, A. A deep learning approach for pressure ulcer prevention using wearable computing. *Hum. Centric Comput. Inf. Sci.* **2020**, *10*, 5. [\[CrossRef\]](#)
10. Walther, F.; Heinrich, L.; Schmitt, J.; Eberlein-Gonska, M.; Roessler, M. Prediction of inpatient pressure ulcers based on routine healthcare data using machine learning methodology. *Sci. Rep.* **2022**, *12*, 5044. [\[CrossRef\]](#)
11. Choi, J.E.; Hwang, S.-K. Predictive validity of pressure ulcer risk assessment scales among patients in a trauma intensive care unit. *J. Korean Crit. Care Nurs.* **2019**, *12*, 26–38. [\[CrossRef\]](#)
12. Qaseem, A.; Mir, T.P.; Starkey, M.; Denberg, T.D.; Clinical Guidelines Committee of the American College of Physicians. Risk assessment and prevention of pressure ulcers: A clinical practice guideline from the american college of physicians. *Ann. Intern. Med.* **2015**, *162*, 359–369. [\[CrossRef\]](#)
13. Dweekat, O.Y.; Lam, S.S.; McGrath, L. An integrated system of multifaceted machine learning models to predict if and when hospital-acquired pressure injuries (bedsores) occur. *Int. J. Environ. Res. Public Health* **2023**, *20*, 828. [\[CrossRef\]](#)
14. Silva, A.; Metrôlho, J.; Ribeiro, F.; Fidalgo, F.; Santos, O.; Dionisio, R. A review of intelligent sensor-based systems for pressure ulcer prevention. *Computers* **2022**, *11*, 6. [\[CrossRef\]](#)
15. Cox, J.; Schallom, M.; Jung, C. Identifying risk factors for pressure injury in adult critical care patients. *Am. J. Crit. Care* **2020**, *29*, 204–213. [\[CrossRef\]](#)
16. Haleem, A.; Javaid, M.; Khan, I.H. Current status and applications of artificial intelligence (ai) in medical field: An overview. *Curr. Med. Res. Pract.* **2019**, *9*, 231–237. [\[CrossRef\]](#)
17. Ribeiro, F.; Fidalgo, F.; Silva, A.; Metrôlho, J.; Santos, O.; Dionisio, R. Literature review of machine-learning algorithms for pressure ulcer prevention: Challenges and opportunities. *Informatics* **2021**, *8*, 76. [\[CrossRef\]](#)
18. Song, J.; Gao, Y.; Yin, P.; Li, Y.; Li, Y.; Zhang, J.; Su, Q.; Fu, X.; Pi, H. The random forest model has the best accuracy among the four pressure ulcer prediction models using machine learning algorithms. *Risk Manag. Healthc. Policy* **2021**, *14*, 1175–1187. [\[CrossRef\]](#)
19. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv* **2014**, arXiv:1409.1259. [\[CrossRef\]](#)
20. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
21. Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* **2018**, *8*, 6085. [\[CrossRef\]](#)
22. Johnson, A.; Bulgarelli, L.; Pollard, T.; Horng, S.; Celi, L.A.; Mark, R. “MIMIC-IV” (Version 2.0). *PhysioNet* **2022**. [\[CrossRef\]](#)
23. Cihan, P.; Ozger, Z.B. A New Heuristic Approach for Treating Missing Value: ABCimp. *Elektron. Ir Elektrotehnika* **2019**, *25*, 48–54. [\[CrossRef\]](#)
24. Luo, Y.; Zhang, Y.; Cai, X.; Yuan, X. E<sup>2</sup>GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Palo Alto, CA, USA, 10–16 August 2019.
25. Luo, Y.; Cai, X.; Zhang, Y.; Xu, J. Multivariate Time Series Imputation with Generative Adversarial Networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–12.
26. Zhang, Y.; Zhou, B.; Cai, X.; Guo, W.; Ding, X.; Yuan, X. Missing value imputation in multivariate time series with end-to-end generative adversarial networks. *Inf. Sci.* **2021**, *551*, 67–82. [\[CrossRef\]](#)
27. Fortuin, V.; Baranchuk, D.; Rätsch, G.; Mandt, S. GP-VAE: Deep Probabilistic Time Series Imputation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Online, 26–28 August 2020; PMLR: London, UK, 2020.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.