# RESEARCH

**BMC Emergency Medicine** 



# Improving prediction accuracy of hospital arrival vital signs using a multi-output machine learning model: a retrospective study of JSAS-registry data

Yasuyuki Kawai<sup>1\*</sup>, Koji Yamamoto<sup>1</sup>, Keisuke Tsuruta<sup>1</sup>, Keita Miyazaki<sup>1</sup>, Hideki Asai<sup>1</sup> and Hidetada Fukushima<sup>1</sup>

# Abstract

**Background** Critically ill patients can deteriorate rapidly; therefore, prompt prehospital interventions and seamless transition to in-hospital care upon arrival are crucial for improving survival. In Japan, helicopter emergency medical services (HEMS) complement general emergency medical services (GEMS) by providing on-site care, reducing transport times, and aiding facility selection. Vital signs at hospital arrival determine initial management, but existing models are poor at predicting them, especially in patients receiving continuous interventions from both GEMS and HEMS. Therefore, we developed a machine-learning model to accurately predict the actual values of vital signs at hospital arrival using limited patient characteristic data and prehospital vital signs.

**Methods** Using data from the Japanese Society for Aeromedical Services registry, we retrospectively analyzed data from patients aged  $\geq$ 18 years transported by HEMS between April 2020 and March 2022. Patients with cardiac arrest during transport, missing vital signs, and data inconsistencies were excluded. The predictive model used prehospital vital signs from GEMS and HEMS contact times, demographic characteristics, and intervention information. The primary outcome was the actual values of vital signs measured at hospital arrival. After data preprocessing, we constructed a deep neural network multi-output regression model using Bayesian optimization. Model performance was assessed by comparing the predicted values with the actual hospital arrival measurements using mean absolute error,  $R^2$  score, residual standard deviation, and Spearman's correlation coefficient. Additionally, the NN model's performance was compared with alternative methods, namely HEMS contact values and change-based predictions derived solely from prehospital data.

**Results** The study included 10,478 patients (median age 70 years; 69% male). The model achieved mean absolute errors of 7.1 bpm for heart rate, 15.7 mmHg for systolic blood pressure, 10.8 mmHg for diastolic blood pressure, 2.9 breaths/min for respiratory rate, and 0.62 points for Glasgow Coma Scale score. The Spearman's correlation coefficients ranged from 0.54 to 0.86. The model outperformed other methods, especially for R<sup>2</sup> scores and residual standard deviations, demonstrating its superior ability to predict actual vital signs values.

\*Correspondence: Yasuyuki Kawai k6k6k@naramed-u.ac.jp

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

**Conclusion** The multi-output regression model accurately predicted the actual values of vital signs measured at hospital arrival using limited prehospital information, demonstrating the effectiveness of advanced modeling techniques.

Keywords Helicopter emergency medical services, Vital sign prediction, Machine learning

# Background

Critically ill patients can deteriorate rapidly, necessitating prompt prehospital interventions and a seamless transition to in-hospital care for optimal outcomes [1-4]. In Japan, following the initial response by general emergency medical services (GEMS), helicopter emergency medical services (HEMS) have been implemented to provide continuous care by delivering on-site interventions, reducing transport times, and facilitating the rapid initiation of hospital treatment [5–9].

Predicting the condition at hospital arrival from prehospital data is crucial because it enables hospital teams to prepare treatment for severe cases ahead of patient arrival, thereby reducing delays in initiating critical care [10, 11]. Moreover, assessing individual vital signs provides granular clinical insight that may be obscured when using composite scores, allowing for the early detection of specific physiological changes [12]. Existing predictive models for predicting patient condition upon arrival have limited accuracy [13]. Although several models have been developed for either GEMS or HEMS, none has integrated interventions from both services to predict vital sign changes [14–16]. This is a notable gap given that nearly all patients in Japan receive sequential interventions from GEMS followed by HEMS.

Additionally, the frequent unavailability of prehospital patient characteristic data can lead to inefficient resource allocation and treatment delays [17]. AI-based models offer a promising solution by capturing complex, non-linear relationships within available data, thereby enhancing prediction accuracy and facilitating earlier, more targeted and individualized clinical interventions [18, 19].

While no single indicator fully captures a patient's condition, vital signs remain important assessment parameters in the clinical setting [13, 20]. Vital signs include multiple interconnected items [21]; therefore, integrated predictions may improve prediction accuracy compared with single-item predictions. The present study constructed a machine learning model that integrates limited patient background information and prehospital vital signs to accurately predict vital signs upon patient arrival at a medical institution.

# Methods

# Study aim, design, and setting

This retrospective, observational study developed a machine-learning model that integrates limited patient background information with prehospital vital signs to accurately predict vital signs upon patient arrival at medical institutions.

# Data source

The study utilized data from the Japanese Society for Aeromedical Services Registry (JSAS-R), a nationwide database established in 2020 with the support from the Ministry of Health, Labour and Welfare Science Research Grant (Grant Number 202122064A). The JSAS-R prospectively records HEMS activities across Japan, covering 80.5% of all dispatches during the observation period from April 2020 to March 2022. The database is centrally maintained and collects prehospital data, including vital signs and activity times at the point of contact with GEMS and HEMS before hospital arrival. Interventions performed by GEMS and HEMS are also recorded, although specific implementation times and personnel are not provided. Patients receiving both GEMS and HEMS are transported to medical facilities by HEMS. While transport policies may vary between institutions, the data collection process is standardized across participating facilities.

#### Study population

We included patients aged  $\geq$  18 years who were transported by HEMS from the scene to a medical institution between April 2020 and March 2022. Patients who experienced cardiac arrest during transport were excluded due to significant differences in GEMS interventions and HEMS transport policies for such cases. We excluded cases with missing vital sign measurement times, calculated activity times that were negative or >480 minutes, and those with missing vital signs immediately before hospital arrival, which was the focus of this study.

# **Outcome definition**

The primary outcome of this study was defined as the actual values of vital signs measured at hospital arrival (Heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), and Glasgow Coma Scale (GCS) score).

# Model features

The input features for the model included:

1. **Vital signs:** HR, SBP, DBP, RR, and GCS score at the time of GEMS and HEMS contact, and their changes per unit time (Additional Fig. 1).

- 2. **Patient characteristics:** Age, sex, and etiology classification (whether caused by internal factors, such as diseases, or external factors, such as trauma).
- 3. Intervention: The presence or absence of tracheal intubation performed by GEMS and HEMS as our intervention parameter was due to its early use and significant impact. Although other important prehospital procedures, such as intravenous fluid administration and thoracostomy, are clinically relevant, they were not included, due to the lack of detailed and consistent timing, dosage, and contextual data in the JSAS-R database. Tracheal intubation was selected because, if sedation was administered at the time of intubation, the GCS is typically recorded as unchanged thereafter, while the respiratory rate reflects ideal ventilation based on assisted ventilation parameters. Therefore, to avoid introducing excessive complexity and uncertainty in our model, all HEMS interventions were treated uniformly, and tracheal intubation was used as a representative marker. We constructed our model under the assumption that all "HEMS interventions" were treated uniformly because the database lacked a record of the timing of interventions. In other words, the absence of detailed timing information meant that including each intervention's impact separately would introduce excessive complexity and uncertainty, making results interpretation overly difficult.

# Data preparation

Recording vital signs in emergency medicine can lead to inaccuracies [22–24]. To improve the reliability of the data, we performed the following preprocessing steps.

# 1. Detection and handling of outliers

We calculated the interquartile range (IQR) for each vital sign and defined outliers as data points below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR. Outliers were treated as missing values, and missing data processing was performed. However, cases with missing data on target vital signs immediately before hospital arrival were excluded.

# 2. Handling missing values after excluding outliers

For missing data, we used masking techniques and missing indicators to generate new features indicating the presence or absence of the missing values. We imputed the missing values using the iterative imputer method, which iteratively estimates missing values by leveraging correlations with other variables, thereby enabling accurate imputation.

# 3. Data standardization

To improve model convergence and computational efficiency, we standardized all continuous variables to have a mean of 0 and a standard deviation of 1. The data distributions before and after outlier processing and missingvalue imputation are shown in Additional Fig. 2.

# Feature preprocessing

We divided the entire dataset into external five-fold cross-validation splits to evaluate the generalization performance of the model. Within each external fold, we performed hyperparameter tuning using internal five-fold cross-validation. This nested cross-validation approach prevented overfitting and allowed for a more reliable performance evaluation. Figure 1 shows an overview of this process.

# Model construction

As multiple vital signs are interrelated, we constructed a multi-output regression model using a deep neural network (DNN) to capture these interactions. The details of the model are as follows:

- 1. Input layer: Preprocessed features were input.
- 2. **Hidden layers:** We introduced multiple residual blocks to increase the network depth while mitigating the vanishing gradient problem. Batch normalization and dropout were applied to improve the training stability and generalization performance.
- 3. **Output layer:** Output nodes corresponding to each target vital sign.
- 4. **Hyperparameter tuning:** We performed Bayesian optimization using Optuna to optimize hyperparameters, such as the optimizer type, learning rate, number of units in each layer, dropout rate, number of residual blocks, and loss function weights. By weighting the loss function, we adjusted the impact of the prediction errors for each vital sign. The search ranges are listed in Additional Table 1.
- 5. **Model training and evaluation:** The training was halted when the performance of the validation data stopped improving. We introduced a learning rate scheduler to decay the learning rate as the training progressed.

The evaluation metrics included the mean absolute error (MAE), mean squared error (MSE),  $R^2$  score, Spearman's correlation coefficient, and standard deviation of residuals to evaluate the variance of the prediction errors. The purpose of each evaluation metric is detailed in Additional File 1.



Fig. 1 Model architecture. NNPM: neural network-based prediction model; HCV: helicopter contact-based prediction values; CRPM: change rate-based prediction model; MAE: mean absolute error; Spearman's rank correlation coefficient

# **Comparison of prediction models**

To evaluate prediction accuracy and reliability, we compared the machine learning-based multi-output regression model with the following two alternatives:

- 1. **HEMS contact values:** Reflecting the state at the start of the medical interventions.
- 2. **Change-based prediction:** Linear prediction using the rate of change per unit time of vital signs between the GEMS and HEMS interventions. Calculated as follows:

$$VS_{pred} = VS_{HEMS} + \left(\frac{VS_{HEMS} - VS_{GEMS}}{\Delta t_{GEMS-HEMS}}\right) \\ \times \Delta t_{HEMS-Hospital}$$

VS<sub>pred</sub>: Predicted vital signs upon hospital arrival

VS<sub>HEMS</sub>: Vital signs at HEMS contact

VS<sub>GEMS</sub>: Vital signs at GEMS contact

 $\Delta t_{GEMS-HEMS}$ : Time interval from GEMS contact to HEMS contact (minutes)

 $\Delta t_{\text{HEMS-Hospital}}$ : Time interval from HEMS contact to hospital arrival (minutes)

This change-based prediction equation was derived from our previous study [25], which focused on the rate of change in vital signs during the GEMS phase. Our analysis revealed that the changes observed during GEMS interventions alone did not fully account for the additional impact of HEMS interventions. Therefore, this linear model, which relies exclusively on prehospital data, serves as a conventional benchmark for the NN model's performance evaluation.

However, please note the following points: In the main analysis, the NN model was trained on the full dataset using imputation to handle missing prehospital data, along with the actual vital sign measurements at hospital arrival. In contrast, the HEMS contact values and change-based predictions rely solely on the available prehospital information. Thus, these reference methods serve as conventional benchmarks rather than direct competitors.

# Statistical analysis

Continuous variables with many outliers are expressed as medians (interquartile ranges), while categorical variables are presented as counts and percentages.

To evaluate the prediction accuracy of the models, we compared the predicted vital sign values with the actual values measured at hospital arrival using scatter plots for visual assessment and calculated the MAE, Spearman's correlation coefficient,  $R^2$  score, and standard deviation of the residuals as quantitative metrics (Additional Fig. 3).

# Sensitivity analysis

We conducted a sensitivity analysis on the data with missing values by building a model using only cases with no missing values.

To ensure reproducibility, we set random seeds to maintain consistency in the results.

#### **Programming environment**

Data analysis and machine learning were performed using Google Colaboratory with Python 3.10.12, scikitlearn 1.3.2, TensorFlow 2.17.0, and Optuna 4.0.0.

#### Results

Among 25,815 initial patients, 10,478 were included after applying the exclusion criteria. Of the excluded patients, 60% were excluded because of missing prehospital arrival vital signs (Fig. 2).

The patient demographics are shown in Table 1. The median patient age was 70 years, and 69% of the patients were male. Trauma was the most common etiology (41%). Missing vital signs were particularly prevalent at the time of GEMS contact, with 36% and 20% missing GCS scores and RR measurements, respectively. At the time of HEMS contact, missing data were less frequent, but the RR still had a missing rate of 8%.

# Prediction accuracy of the multi-output regression model

The MAEs for predicting each vital sign in the multi-output regression model were as follows: HR 7.1 bpm, SBP 15.7 mmHg, DBP 10.8 mmHg, RR 2.9 breaths/min, and GCS score 0.62 points. The Spearman's correlation coefficients between the true values and the predicted values were high for HR (0.83) and the GCS (0.86) and low for SBP (0.68), DBP (0.55), and RR (0.54). These MAEs and correlation coefficients did not notably differ compared with those from the HEMS contact observations; however, the neural network (NN) model showed more accurate predictions in terms of  $R^2$  scores and standard deviations of residuals. Specifically, the  $R^2$  scores for the NN model's predictions were 0.42 for SBP and 0.29 for DBP, substantially higher than the 0.10 and -0.18 obtained using HEMS contact observations (Table 2).

These results indicated high prediction accuracy for HR and GCS scores, with the NN model consistently outperforming the other methods in terms of  $R^2$  and variance reproduction. Although the RR and BP predictions had relatively low correlation coefficients, the clinical predictions were within 10% of the observed values. The scatterplots in Fig. 3 visually depict these relationships and prediction performances across the different methods.

#### Sensitivity analysis

The sensitivity analysis, including only 5,112 cases without missing values showed slight improvements in accuracy and similar trends as the results after data imputation (Additional Table 2, Additional Fig. 3).

# Discussion

This study developed a multi-output regression model that integrates intervention data from both GEMS and HEMS to predict patient vital signs upon hospital arrival. Our NN model achieved high accuracy, particularly in predicting HR and GCS scores with low MAEs. In comparison to conventional methods used as benchmarks (that is, simple HEMS contact measurements and linear change-based predictions relying solely on prehospital data), the NN model demonstrated superior performance, as evidenced by higher R<sup>2</sup> scores and reduced residual variance. These results demonstrate that, even with limited prehospital data, a model that accounts for



Fig. 2 Flowchart of patient selection. JSAS-R, Japanese Society for Aeromedical Service Registry; HEMS, helicopter emergency medical service; CPR, cardiopulmonary resuscitation

# Table 1 Eligible patient characteristics

Variables	Overall		
		n=10478	
Patient demographics			
Age (years), median (IQR)		70 (55, 80)	
Sex (male), n (%)		7266 (69)	
Underlying condition, n (%)			
Trauma		4553 (43)	
Cerebrovascular diseases		2144 (20)	
Other internal causes		1632 (16)	
Cardiovascular diseases		990 (9)	
Other external causes		565 (5)	
Unknown		594 (6)	
Vital sign (n (%), median (IQR))			
GEMS at contact	HR (n=9515)	82 (70, 97)	
	SBP (n=9330)	141 (118, 165)	
	DBP (n=8945)	84 (70, 99)	
	RR (n=8383)	20 (18, 24)	
	GCS (n=6756)	15 (13, 15)	
HEMS at contact	HR (n=10191)	81 (70, 94)	
	SBP (n = 10027)	141 (120, 163)	
	DBP (n=9852)	83 (70, 97)	
	RR (n=9647)	20 (18, 24)	
	GCS (n = 10364)	15 (13, 15)	
HEMS final measurement	HR	80 (69, 94)	
	SBP	139 (120, 158)	
	DBP	82 (70, 93)	
	RR	20 (17, 23)	
	GCS	15 (13, 15)	
Duration of intervention			
GEMS (min), median (IQR)		22 (17, 28)	
HEMS (min), median (IQR)		26 (21, 32)	

IQR, interquartile range; GEMS, ground emergency medical service; HEMS, helicopter emergency medical service; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; RR, respiratory rate; GCS, Glasgow coma scale

interactions among factors can more accurately predict the actual vital sign values at hospital arrival.

Vital sign-based models present statistical challenges owing to repeated measurements and multicollinearity [26]. The multi-output regression model used in this study improved the model's performance by incorporating deep learning techniques such as residual blocks [27] and batch normalization [28]. The use of residual blocks mitigates the vanishing gradient problem associated with deep networks, allowing the model to learn complex nonlinear relationships. Batch normalization and dropout improved training stability and reduced the risk of overfitting. In addition, Bayesian optimization using Optuna enabled efficient hyperparameter tuning to maximize the model's performance [29]. The optimization of the loss function weights allowed the model to appropriately reflect the importance of each vital sign. These modeling techniques and tuning strategies significantly contributed to improving the prediction accuracy. By considering the effects of continuous interventions from the GEMS and HEMS on prediction accuracy, the model more accurately reflected the complex realities of emergency medicine. This approach demonstrated the practical and effective application of predictive models in emergency settings.

The lower predictive accuracy of RR and GCS scores compared with that for the other vital signs may be due, in part, to the fact that these variables were treated as continuous values in the regression model despite their discrete nature. Although respiratory rate is theoretically a continuous variable, in clinical practice, it is typically measured by counting breaths over a short interval (for example, 10 or 15 seconds) and then extrapolating to a per-minute value, resulting in discrete numbers (see Additional Fig. 2 for the actual distribution).

Table 2 Comparison of predicted vital sign values at nospital arrival with actual measurements across different i
---

	Variable	HR	SBP	DBP	RR	GCS
HEMS contact values	MAE	7.3	18.8	13.5	2.9	0.38
	R <sup>2</sup>	0.63	0.1	-0.18	0.1	0.84
	SD	11	25.9	18.4	4.3	1.3
	Spearman correlation	0.83	0.65	0.52	0.54	0.93
Change-based predicted values	MAE	13.9	29.8	22.8	5.9	0.99
	R <sup>2</sup>	-0.68	-1.5	-2.7	-2.7	0.42
	SD	23.5	43	32.5	8.7	2.4
	Spearman correlation	0.68	0.51	0.37	0.42	0.77
NN predicted values	MAE	7.1	15.7	10.8	2.9	0.62
	R <sup>2</sup>	0.69	0.42	0.29	0.32	0.84
	SD	10	20.8	14	3.9	1.4
	Spearman correlation	0.83	0.68	0.55	0.54	0.86

Table 2 compares three methods for predicting hospital arrival vital signs: (1) HEMS contact values (initial HEMS measurements), (2) change-based predictions (extrapolated from GEMS and HEMS data), and (3) NN predictions from our multi-output regression model. Performance metrics (MAE, R<sup>2</sup>, SD of residuals, and Spearman's correlation) were calculated by comparing each method's predictions with the actual hospital arrival values. HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; RR, respiratory rate; GCS, Glasgow coma scale; MAE, mean absolute error; R<sup>2</sup>, R-squared; SD, standard deviation; Spearman's rank correlation coefficient



Fig. 3 Scatterplots comparing three prediction model outputs with true vital sign values. Scatterplots comparing the helicopter contact values, rate of change predictions, and neural network (NN) predictions with the true values for heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), and Glasgow Coma Scale (GCS) scores. The NN model shows higher accuracy, with values closer to the diagonal line

Consequently, the model may have struggled to capture their characteristics accurately, leading to decreased prediction accuracy. In addition, high missing rates and potential measurement errors owing to manual assessments may have also affected the results. Moreover, the RR is often measured over 15 s and multiplied by four, thus introducing inter-observer variability [30]. Additionally, the GCS includes subjective evaluation components that may lead to inconsistent data collection [31]. These issues highlight the need for improved data collection methods and standardization of measurement techniques. Future models that account for the discrete nature of the RR and GCS scores, along with improved data reliability, are expected to improve predictive accuracy.

While previous studies have attempted to predict vital signs by focusing on single parameters or isolated interventions [32, 33], our study extends this work by examining the sequential effects of interventions from both GEMS and HEMS using a multi-output regression model. This approach offers additional insight into the complex interplay of factors in emergency care. However, given the limitations of our data– including a high proportion of missing values and the lack of detailed prehospital intervention timing– these findings should

be interpreted with caution. Although our model demonstrated improved prediction accuracy of the dataset, further validation is necessary before the results can be applied in real-world clinical practice.

This technology has the potential to significantly improve the quality of emergency medical care. Specifically, this model can provide valuable information for preparing hospital reception systems and allocating appropriate medical resources before patient arrival. By accurately predicting a patient's condition upon arrival, medical teams can formulate effective initial responses and optimize resource utilization, potentially reducing treatment delays and improving patient outcomes [34].

# Limitations

This study has some limitations. First, a major limitation of this study is the substantial amount of missing prehospital data, which resulted in the exclusion of 12,158 cases. This exclusion may have introduced selection bias by omitting patients with more severe or atypical clinical profiles, particularly since vital sign recording is often insufficient in highly urgent cases [22–24]. This limitation may affect the model's overall performance and restrict the generalizability of our findings. Although we employed an iterative imputation method with missing indicators to mitigate the impact of missing data and preserve inter-variable correlations, this approach may not fully capture the underlying variability or all critical clinical nuances. Importantly, sensitivity analyses using only complete cases revealed similar trends in model performance, supporting the robustness of our imputation strategy despite the high missing rate. Future studies should explore alternative or additional methods for handling missing data to further validate these results. Moreover, recent studies have demonstrated that advances in digital technology, such as automated vital sign capture and real-time data transmission, can significantly reduce data loss and transcription errors in prehospital settings [35]. These innovations hold promise for enhancing data completeness and, ultimately, the reliability of predictive models in emergency care. Second, treating RR and GCS scores as continuous variables, despite their inherently discrete nature, may have reduced prediction accuracy. Because these scores are typically recorded as whole numbers with specific clinical thresholds, alternative modeling approaches (for example, ordinal regression or categorical analysis) might better capture their true distribution and clinical significance, thereby improving predictive performance [36]. Third, our intervention data were limited to the presence or absence of tracheal intubation. All HEMS interventions were treated uniformly because the database lacked precise timing for prehospital interventions. Tracheal intubation was selected as the sole intervention parameter since, following intubation (often with sedation), the GCS remains unchanged, and the respiratory rate reflects ventilatory support. Therefore, we were unable to fully evaluate the impact of other interventions, such as intravenous fluid administration or thoracostomy. The outcomes in intubated cases should be interpreted with caution. Fourth, advanced NNs have a low interpretability owing to their black-box nature. Interpretability is crucial for model adoption in medical fields; therefore, future work should consider incorporating interpretable methods such as attention mechanisms or feature importance analyses. Fifth, the complexity of the model poses challenges for immediate clinical applications in terms of computational resources and inference time. Real-time responses are required in emergency settings, necessitating model simplification and inference speed optimization for practical implementation. Finally, because the study targeted cases with specific interventions in a limited region, external validation in other regions or with different intervention conditions is required to assess the generalizability.

# Conclusions

The multi-output regression model developed in this study demonstrated high accuracy in predicting the vital signs upon hospital arrival in patients who underwent interventions from both GEMS and HEMS. Improving the prediction accuracy will provide the foundation for rapid emergency responses by optimizing medical resources and preparing care plans in advance.

#### Abbreviations

HEMS	Helicopter Emergency Medical Services
GEMS	General Emergency Medical Services
ISAS-R	Japanese Society for Aeromedical Services Registry
HR	Heart rate
SBP	Systolic blood pressure
OBP	Diastolic blood pressure
R	Respiratory rate
GCS	Glasgow coma scale
MAE	Mean absolute error
MSE	Mean squared error
₹ <sup>2</sup>	R-squared
NN	Neural network

#### Supplementary information

The online version contains supplementary material available at https://doi.or g/10.1186/s12873-025-01233-9.

Supplementary Material 1	
Supplementary Material 2	
Supplementary Material 3	
Supplementary Material 4	
Supplementary Material 5	
Supplementary Material 6	

#### Acknowledgements

Not applicable.

#### Author contributions

YK conceived and designed the study; KY collected the data; YK, KY, KT, and KM analyzed and interpreted the data; YK performed machine learning analyses; YK wrote the draft manuscript. All authors have read and approved the manuscript.

#### Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

#### Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

#### Declarations

#### Ethics approval and consent to participate

This study was approved by the Ethics Committee of Nara Medical University (Approval No. 3684). The need for written informed consent was waived by the Ethics Committee because the data were anonymized. The study was conducted in accordance with the principles of the Declaration of Helsinki.

#### **Consent for publication**

Not applicable.

#### **Clinical trial number** Not applicable.

# **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Emergency and Critical Care Medicine, Nara Medical University, Shijo-cho, 840, Kashihara City, Nara 634-8522, Japan

Received: 27 October 2024 / Accepted: 2 May 2025 Published online: 13 May 2025

#### References

- 1. Wilson MH, Habig K, Wright C, Hughes A, Davies G, Imray CHE. Pre-hospital emergency medicine. Lancet. 2015;386:2526–34.
- Noc M, Fajadet J, Lassen JF, Kala P, MacCarthy P, Olivecrona GK, et al. Invasive coronary treatment strategies for out-of-hospital cardiac arrest: a consensus statement from the European association for percutaneous cardiovascular interventions (EAPCI)/stent for life (SFL) groups. EuroIntervention. 2014;10:31–37.
- 3. Mitra B, Bade-Boon J, Fitzgerald MC, Beck B, Cameron PA. Timely completion of multiple life-saving interventions for traumatic haemorrhagic shock: a retrospective cohort study. Burns Trauma. 2019;7:22.
- Latif RK, Clifford SP, Baker JA, Lenhardt R, Haq MZ, Huang J, et al. Traumatic hemorrhage and chain of survival. Scand J Trauma Resusc Emerg Med. 2023;31:25.
- Ohsaka H, Yanagawa Y, Nagasawa H, Takeuchi I, Jitsuiki K, Madokoro S, et al. A report concerning collaboration between a physician-staffed helicopter (doctor helicopter) and firefighting/rescue helicopter. Air Med J. 2018;37:325–28.
- Endo A, Kojima M, Uchiyama S, Shiraishi A, Otomo Y. Physician-led prehospital management is associated with reduced mortality in severe blunt trauma patients: a retrospective analysis of the Japanese nationwide trauma registry. Scand J Trauma Resusc Emerg Med. 2021;29:9.
- Abe T, Takahashi O, Saitoh D, Tokuda Y. Association between helicopter with physician versus ground emergency medical services and survival of adults with major trauma in Japan. Crit Care. 2014;18:R146.
- Hosomi S, Kitamura T, Sobue T, Nakagawa Y, Ogura H, Shimazu T. Association of pre-hospital helicopter transport with reduced mortality in traumatic brain injury in Japan: a nationwide retrospective cohort study. J Neurotrauma. 2022;39:76–85.
- Muramatsu KI, Omori K, Kushida Y, Nagasawa H, Takeuchi I, Jitsuiki K, et al. An analysis of patients with acute aortic dissection who were transported by physician-staffed helicopter. Am J Emerg Med. 2021;44:330–32.
- Kang DY, Cho KJ, Kwon O, Kwon JM, Jeon KH, Park H, et al. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scand J Trauma Resusc Emerg Med. 2020;28:1–8.
- Dehli T, Monsen SA, Fredriksen K, Bartnes K. Evaluation of a trauma team activation protocol revision: a prospective cohort study. Scand J Trauma Resusc Emerg Med. 2016;24:1–7.
- Riccalton V, Threlfall L, Ananthakrishnan A, Cong C, Milne-Ives M, Le Roux P, et al. Modifications to the national early warning score 2: a scoping review. BMC Med. 2025;23:154.
- Williams T, Ho K, Tohira H, Fatovich D, Bailey P, Brink D, et al. 14 Initial prehospital vital signs to predict subsequent adverse hospital outcomes. BMJ Open. 2017;7:A5.3–A6.
- Bourke-Matas E, Doan T, Bowles KA, Bosley E. A prediction model for prehospital clinical deterioration: the use of early warning scores. Acad Emerg Med. 2024;31:1139–49.
- Schellenberg M, Biswas S, Bardes JM, Trust MD, Grabo D, Wilson A, et al. Prehospital time decreases reliability of vital signs in the field: a dual center study. Am Surg. 2021;87:943–48.
- Björkman J, Raatiniemi L, Setälä P, Nurmi J. Shock index as a predictor for short-term mortality in helicopter emergency medical services: a registry study. Acta Anaesthesiol Scand. 2021;65:816–23.
- Ovidiu Popa T, Carmen Cimpoesu D, Lucian Nedelea P. Prehospital emergency care in acute trauma conditions. Emergency medicine and trauma. IntechOpen. 2019. doi:https://doi.org/10.5772/intechopen.86776. Available from.

- Choi A, Lee K, Hyun H, Kim KJ, Ahn B, Lee KH, et al. A novel deep learning algorithm for real-time prediction of clinical deterioration in the emergency department for a multimodal clinical decision support system. Sci Rep. 2024;14:30116.
- Holtenius J, Mosfeldt M, Enocson A, Berg HE. Prediction of mortality among severely injured trauma patients A comparison between TRISS and machine learning-based predictive models. Injury. 2024;55:111702.
- Candel BG, Duijzer R, Gaakeer MI, Ter Avest E, Sir Ö, Lameijer H, et al. The association between vital signs and clinical outcomes in emergency department patients of different age categories. Emerg Med J. 2022;39:903–11.
- Forkan ARM, Khalil I. A clinical decision-making mechanism for context-aware and patient-specific remote monitoring systems using the correlations of multiple vital signs. Comput Methods Programs Biomed. 2017;139:1–16.
- Armstrong B, Walthall H, Clancy M, Mullee M, Simpson H. Recording of vital signs in a district general hospital emergency department. Emerg Med J. 2008;25:799–802.
- Glasin J, Henricson J, Lindberg LG, Wilhelms D. Wireless vitals—Proof of concept for wireless patient monitoring in an emergency department setting. J Biophotonics. 2019;12:e201800275.
- Kjær J, Milling L, Wittrock D, Nielsen LB, Mikkelsen S. The data quality and applicability of a Danish prehospital electronic health record: a mixed-methods study. PLOS ONE. 2023;18:e0293577.
- Kawai Y, Yamamoto K, Miyazaki K, Takano K, Asai H, Nakano K, et al. Comparison of changes in vital signs during ground and helicopter emergency medical services and hospital interventions. Air Med J. 2022;41:391–95.
- Guo Y, Logan HL, Glueck DH, Muller KE. Selecting a sample size for studies with repeated measures. BMC Med Res Methodol. 2013;13:100. https://doi.or g/10.1186/1471-2288-13-100.
- He K, Zhang X, Ren S, Sun J Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition 2016; pp. 770–78.
- Wu S, Li G, Deng L, Liu L, Wu D, Xie Y, et al. L1-norm batch normalization for efficient training of deep neural networks. IEEE Trans Neural Netw Learn Syst. 2019;30:2043–51.
- Akiba T, Sano S, Yanase T, Ohta T, Koyama M Optuna: a next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019; pp. 2623–31
- Brabrand M, Hallas P, Folkestad L, Lautrup-Larsen CH, Brodersen JB. Measurement of respiratory rate by multiple raters in a clinical setting is unreliable: a cross-sectional simulation study. J Crit Care. 2018;44:404–06.
- Reith FCM, Van den Brande R, Synnot A, Gruen R, Maas AIR. The reliability of the Glasgow Coma Scale: a systematic review. Intensive Care Med. 2016;42:3–15.
- Al Jalbout N, Balhara KS, Hamade B, Hsieh YH, Kelen GD, Bayram JD. Shock index as a predictor of hospital admission and inpatient mortality in a US national database of emergency departments. Emerg Med J. 2019;36:293–97.
- Wang IJ, Bae BK, Park SW, Cho YM, Lee DS, Min MK, et al. Pre-hospital modified shock index for prediction of massive transfusion and mortality in trauma patients. Am J Emerg Med. 2020;38:187–90.
- Coslovsky M, Takala J, Exadaktylos AK, Martinolli L, Merz TM. A clinical prediction model to identify patients at high risk of death in the emergency department. Intensive Care Med. 2015;41:1029–36.
- Kim Y, Groombridge C, Romero L, Clare S, Fitzgerald MC. Decision support capabilities of telemedicine in emergency prehospital care: systematic review. J Med Internet Res. 2020;22:e18959..
- Verhulst B, Neale MC. Best practices for binary and ordinal data analyses. Behav Genet. 2021;51:204–14.

# **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.