A Novel ED Triage Framework Using Conditional Imputation, Multi-Scale Semantic Learning, and Cross-Modal Fusion

Yi Xiao[1], Jun Zhang[1]*, Cheng Chi[2], and Chunyu Wang[2]

[1] State Key Laboratory of Digital Medical Engineering, School of Instrument Science and Engineering, Southeast University, 210096, China
{xiaoy, j.zhang}@seu.edu.cn
[2] Department of Emergency, Peking University People's Hospital, Beijing, 100044, China
chichengpku@bjmu.edu.cn, wdlym0226@163.com
* Corresponding author

Abstract. In emergency departments (ED), efficient triage is essential for timely patient care, but challenges like missing and sparse data often hinder the prediction performance of severity level and department. To address these issues, we propose a novel intelligent triage method that incorporates a Conditional Gaussian Mixture Imputation (CGMI) and a Feature Densification Module (FDM). The CGMI handles missing data through conditional probability modeling, while the FDM obtains correlations between variables by calculating the Manhattan distance between non-zero values in a one-hot coded feature. In addition, we design a multi-scale Feature Extraction Module (mFEM) to capture multi-level semantic information from patient complaints. Subsequently, two feature fusion strategies were introduced: early fusion and late fusion. The early fusion combines Principal Component Analysis (PCA)-processed features with another modality. The late fusion with enhancement introduces reverse features of another modality and applies an attention mechanism to obtain salient features. Experimental results show that our method outperforms existing approaches, achieving 84.83% sensitivity, 85.11% specificity, and 61.42% Cohen's Kappa for severity prediction and 90.89% sensitivity, 91.04% specificity, and 85.87% Cohen's Kappa for department prediction. Our method significantly improves the sensitivity, specificity, and robustness of ED triage, demonstrating superior performance and reliability in handling missing and sparse clinical data. The code is available at https://github.com/xiaoyiseu/CGMI.

Keywords: Emergency Department Triage, Missing Data, Feature Fusion.

1 Introduction

Triage is a critical emergency care step involving severity assessment and department allocation, which directly impacts patient outcomes. Manual triage methods rely heavily on subjective judgment or essential tools, making them highly dependent on com-

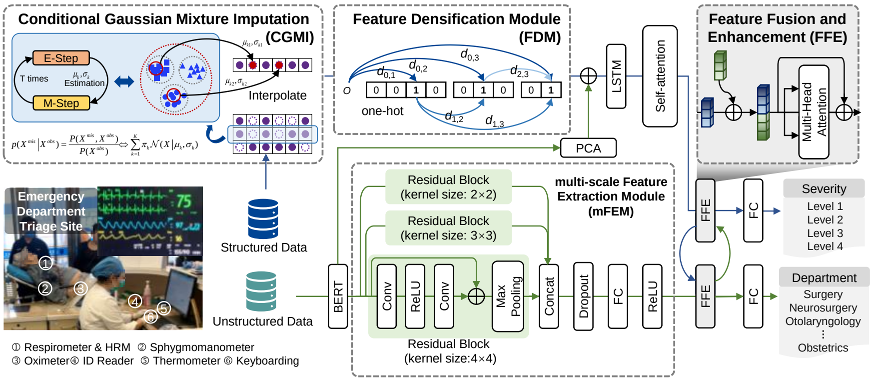


Fig. 1. Architecture of the proposed method, consisting of two branches: one for Severity prediction and one for Department prediction.

plete data [1]. While experienced physicians can make accurate decisions, less experienced practitioners struggle with incomplete or indistinct information, posing challenges in determining appropriate severity levels and department assignments.

The rapid development of artificial intelligence has made intelligent triage possible. Many existing methods have shown promising results in triage tasks, such as triage for specific conditions (e.g., intensive care or hospitalization [2]), specific diseases (e.g., sepsis [3], asthma and COPD [4], lung cancer [5], injury [6] [7], COVID-19 [8] [9]), and specific populations (e.g., elderly, adults, and children [1]). Since the emergency department (ED) serves a diverse patient population, accurate and efficient triage methods have become a key area of research [10] [11].

Complete patient information is crucial for accurate triage, but missing data complicates decision-making. Existing approaches to handling missing data include deletion, modelling, and imputation, with the latter two being more commonly used. Modelling techniques, especially those leveraging machine learning, probabilistic, and generative models, have shown great promise in addressing these challenges. For example, an efficient alternating expectation-conditional maximization algorithm has been proposed for a generalized hyperbolic factor analyzer model to handle missing values [12]. A cross-modal generative adversarial network (GAN) has been developed, integrating cross-modal fusion and adversarial generation techniques to impute missing values in long-term time-series data [13]. In contrast, Imputation methods often fill in missing values by leveraging spatially or semantically nearby values. Such techniques encompass K-nearest neighbors [14], regression-based predictions (e.g., Multiple Imputation [15]), and random forest-based imputation [16]. Additionally, decentralized federated multiple imputation-based fuzzy clustering algorithms have been developed to handle complex longitudinal behavioral trial data across varying periods [17]. Approaches using weighted cross-entropy loss mitigate missing data impact by focusing on available labels [18]. However, challenges remain with data integrity, noise, bias, and the inability of one-hot encoding to capture the complex relationships in discrete triage data.

In this paper, we propose a novel intelligent triage method, as shown in Fig. 1, to address data missingness using Conditional Gaussian Mixture Imputation (CGMI), which estimates missing values based on the conditional distribution of observed data. We introduce a Feature Densification Module that extracts associations from sparse features to enhance feature representation. Additionally, we introduce two feature fusion strategies: early fusion incorporating principal component analysis (PCA)-derived global features and late fusion utilizing reverse-ordered features with cross-attention to highlight key information. The main contributions of this work are as follows:

- We propose CGMI, a structured data imputation method that estimates missing values using the conditional distribution of observed data.
- We introduce a cross-modal feature fusion module enhanced by multi-head attention, with early fusion (introducing another modal PCA dimension reduction feature) and late fusion (introducing another modal inverse feature).
- We design a multi-scale Feature Extraction Module (mFEM) to capture semantic information from patient complaints across scales.

2 Methods

2.1 Problem Definition

The triage data are divided into structured and unstructured categories. The structured data include vital signs (temperature: 0.29%, pulse rate: 45.67%, respiratory rate: 99.88%, blood pressure: 45.67%, oxygen saturation: 45.70%) and demographic details (sex: 0%, age: 0%, and mode of arrival: 4.53%). The percentages indicate the missing rates for each variable, with vital signs showing the highest levels of missingness, particularly the respiratory rate. The unstructured data consists of patient chief complaints.

Consider a set of unstructured data $U=\{U_i|i\in[1, N]\}$, $U_i=\varphi_u(t_i)$, and structured data $S=\{X_i|i\in[1,N]\}$, $X_i=\varphi_s(s_i)$, where t_i, s_i, N, and L represent a chief complaint, structured data, the sample size, and the feature dimension, respectively. $\varphi_u(\cdot)$ and $\varphi_s(\cdot)$ represent BERT [19] and the proposed Feature Densification Module (FDM), respectively.

2.2 Conditional Gaussian Mixture Imputation

The structured data can be represented as $X=\{X^{obs}, X^{mis}\}$, where X^{obs}, X^{mi} denote the observed and missing values, respectively. We first assume that the data distribution follows a Gaussian Mixture Model (GMM). To model the conditional distribution of the missing values, we use the conditional distribution of the Gaussian Mixture Model as follows:

$$p(X^{mis} \big| X^{obs}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(X^{mis} \big| \mathbb{E}[X^{mis} \big| X^{obs}, k], Cov[X^{mis} \big| X^{obs}, k]) \ , \tag{1}$$

where $\mathbb{E}[X^{mis} | X^{obs}, k]$ represents the conditional expectation of the missing data given the observed data and the k-th Gaussian component, this conditional expectation is computed as follows:

$$\mathbb{E}[X^{mis} \mid X^{obs}, k] = \mu_k + \sum\nolimits_{k, X^{mis}, X^{obs}} \sum\nolimits_{k, X^{obs}}^{-1} (X^{obs} - \mu_k) , \tag{2}$$

where $\Sigma_{k,\ Xmis,\ Xobs}$ represents the covariance between X^{mis} and X^{obs} in the k-th Gaussian component and $\Sigma_{k,\ Xobs}$ denotes the covariance of X^{obs} in the k-th component.

$Cov[X^{mis} | X^{obs}, k]$ is the conditional covariance of the missing data given the observed data, which quantifies the uncertainty in imputing the missing values:

$$Cov[X^{mis} \mid X^{obs}, k] = \sum\nolimits_{k,X^{mis},X^{obs}} - \sum\nolimits_{k,X^{mis},X^{obs}} \sum\nolimits_{k,X^{obs}}^{-1} \sum\nolimits_{k,X^{obs},X^{mis}} , \quad (3)$$

The Expectation-Maximization (EM) algorithm is then iteratively applied to train the GMM to obtain the mean and variance of the target distribution. After training, missing values are sampled from the learned distribution. Once the GMM model converges, the global parameters of the target distribution for the training set are obtained, including the conditional expectation $\mathbb{E}[X^{mis} | X^{obs}, k]$ and conditional covariance $Cov[X^{mis} | X^{obs}, k]$. These values are then used to sample and impute the missing data:

$$X_i^{mis} \sim \mathcal{N}(\mathbb{E}[X^{mis} \mid X^{obs}, k], Cov[X^{mis} \mid X^{obs}, k]) , \quad (4)$$

where X_i^{mis} denotes the missing values obtained from the *i*-th sampling iteration.

2.3 Feature Densification of Structured Data

Given an input vector $S \in \mathbb{R}^{B \times F}$, where *B* denotes the batch size, and *F* represents the number of features, each sample *i* is first one-hot encoded. Then, the non-zero index set $I \in \mathbb{R}^{B \times N}$ within S_i is identified:

$$I_i = \{ j \mid S[i, j] \neq 0 \},\ i \in [1, B] , \quad (5)$$

Subsequently, for each sample, *i*, the distance matrix $D_i \in \mathbb{R}^{N \times N}$ between the non-zero feature indices is calculated. Each element $D_i[p, q]$ represents the Manhattan distance between the indices $I_i[p]$ and $I_i[q]$:

$$D_i[p, q] = \left| I_i[p] - I_i[q] \right|,\ p, q \in [1, N] , \quad (6)$$

To avoid feature redundancy, we concatenate its upper triangular portion into a feature vector, which can be expressed as:

$$\tilde{D} = \{ D_i^{up} \}_{i=1}^{B},\ D_i^{up} = \{ D_i[p, q] \mid p < q,\ i \in [1, B] \} , \quad (7)$$

2.4 Feature Extraction and Fusion

We incorporate components from the chief complaint data to address varying severity levels corresponding to the same structured data. PCA is applied to reduce the dimensionality of the chief complaints, which are then concatenated with the structured data. LSTM captures dependencies between features, creating a mixed data representation where structured data is dominant.

$$\hat{D} = LSTM(Concat(\tilde{D}, PCA(U))) , \quad (8)$$

To further enhance the causal relationships among variables of structured data, we employ Self-Attention:

$$\hat{S} = SA(\hat{D}) = softmax\left(\frac{QK^T}{\sqrt{d_k}} \right) V_\alpha = softmax\left(\frac{\hat{D} W_Q W_K^T \hat{D}^T}{\sqrt{d_k}} \right) \hat{D} W_V , \quad (9)$$

where W_Q, W_K, and W_V represent the learnable matrices for *Query*, *Key*, and *Value*, respectively. d_k is the dimensionality of *Key*.

To better capture semantic information from embeddings of unstructured data, we propose a multi-scale Feature Extraction Module (mFEM) as the backbone. The core component of this model consists of three parallel residual blocks designed to capture both local features and long-range dependencies using two layers of 1D dilated convolution. Each branch process can be expressed as:

$$F_i = max(ReLU(W_j \bullet U + b_j)),\ i = 1,2,3;\ j = 1,2\,, \tag{10}$$

where i represents the branch, W_j and b_j denote the weight matrix and bias vector of the convolutional layers in the residual block, respectively. Moreover, the final output is denoted as:

$$\tilde{U} = ReLU(W_f \bullet concat(F_1, F_2, F_3) + b_f)\,, \tag{11}$$

where W_f and b_f denote the weight matrix and bias vector of the linear layer, respectively. The features are then processed in reverse order and concatenated with another feature. These are passed through a cross-attention mechanism, followed by a linear layer with a *softmax* activation function to produce the final predictions.

3 Experiments

3.1 Datasets

This study retrospectively collected 161,196 pathological records from ED of Peking University People's Hospital between June 2020 and June 2022. The severity levels were categorized into Level 4, Level 3, Level 2, and Level 1, with increasing criticality. The departments included Surgery, Internal Medicine, Neurology, Otolaryngology, Obstetrics, Ophthalmology, Gynecology, Orthopedics, Trauma Center, and Neurosurgery. The dataset exhibits significant imbalance, with severe long-tail issues in severity-level prediction and department recommendation.

3.2 Implementation

All experiments were conducted on a single NVIDIA RTX 2080Ti GPU. The models were trained using the AdamW optimizer with a batch size 512, an initial learning rate of 0.001, and a weight decay 5e-5. Each model was trained for 100 epochs, with early stopping implemented after 10 epochs of no improvement.

The dataset was randomly split into training, validation, and test sets with a fixed random seed, using an 8:1:1 ratio. Performance was evaluated using several metrics, including Sensitivity (SE), Specificity (SP), F1-Score (F1), and the Cohen's Kappa coefficient (K), to assess the effectiveness of the proposed method.

3.3 Ablation Study

The ablation study in Table 1 highlights the performance improvements by introducing different components. Initially, when no modules are used, the severity prediction has a sensitivity (SE) of 80.28% and a specificity (SP) of 77.60%, while the department prediction achieves an SP of 89.79% and a K of 84.16%. Adding the Feature Extraction and Fusion (FEF) module increases SE to 82.76% for severity and improves department

Table 1. Ablation study of component contributions to triage prediction performance.

Component				Severity				Department			
CGMI	PCA	FDM	FEF	SE (%)	SP (%)	F1 (%)	K (%)	SE (%)	SP (%)	F1 (%)	K (%)
×	×	×	×	80.28	77.60	78.33	41.25	89.79	89.77	89.65	84.16
×	×	×	√	82.76	82.68	82.52	55.74	90.86	91.02	90.76	85.85
×	×	√	√	82.95	84.07	83.02	58.25	90.65	90.87	90.52	85.48
√	×	×	√	84.60	85.39	84.67	61.37	90.83	**91.23**	90.70	85.77
√	√	×	√	84.57	**85.41**	84.74	**61.67**	90.89	90.89	90.76	85.86
√	×	√	√	84.13	83.48	83.64	57.37	90.61	90.80	90.41	85.40
√	√	√	√	**84.83**	85.11	**84.77**	61.42	**90.89**	91.04	**90.76**	**85.87**

Bold for max, underlined for second max.

Table 2. Comparison of imputation methods for triage prediction performance.

Imputation Method	Severity				Department			
	SE (%)	SP (%)	F1 (%)	K (%)	SE (%)	SP (%)	F1 (%)	K (%)
MICE	**84.86**	84.81	84.59	60.28	90.66	91.02	90.54	85.49
GAN	84.14	**85.38**	84.34	61.08	90.65	90.88	90.57	85.55
VAE	79.58	78.03	78.31	41.79	90.78	90.90	90.65	85.67
RF	84.74	84.42	84.51	60.11	90.68	90.78	90.50	85.52
CGMI (*ours*)	84.83	85.11	**84.77**	**61.42**	**90.89**	**91.04**	**90.76**	**85.87**

Table 3. Triage performance comparison across different backbones.

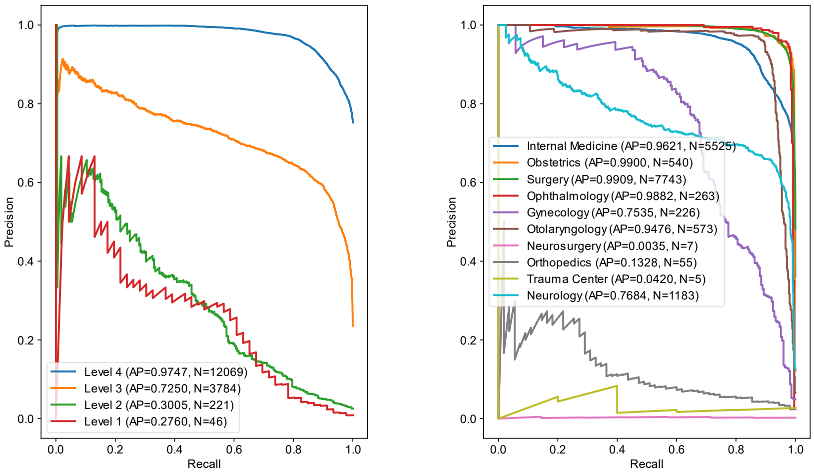
backbone	Severity				Department			
	SE (%)	SP (%)	F1 (%)	K (%)	SE (%)	SP (%)	F1 (%)	K (%)
Transformer	84.74	**85.49**	84.76	**61.66**	90.53	90.54	90.40	85.30
ResNet	84.40	83.72	84.00	58.21	90.56	90.58	90.40	85.31
TextCNN	84.41	83.99	84.08	58.66	90.62	90.51	90.47	85.40
NomFEM	84.73	85.43	84.72	61.48	90.40	90.52	90.27	85.09
ours	**84.83**	85.11	**84.77**	61.42	**90.89**	**91.04**	**90.76**	**85.87**

NomFEM indicates that the proposed model does not include the mFEM module.

prediction with an SP of 90.86% and a K of 85.85%. Including the Feature Densification Module (FDM) further improves performance, with severity prediction SE rising to 82.95% and K reaching 58.25%. Introducing Conditional Gaussian Mixture Imputation (CGMI) results in the highest improvement for severity prediction, with SE reaching 84.60% and K at 61.37%. When added to the CGMI and FEF combination, PCA provides only slight additional gains. The best overall performance is achieved when all modules are combined, with severity prediction reaching a SE of 84.83% and K of 61.42%, and department prediction showing a K of 85.87%. These results confirm the critical role of each module in improving performance, particularly in handling missing data and enhancing model robustness in ED triage.

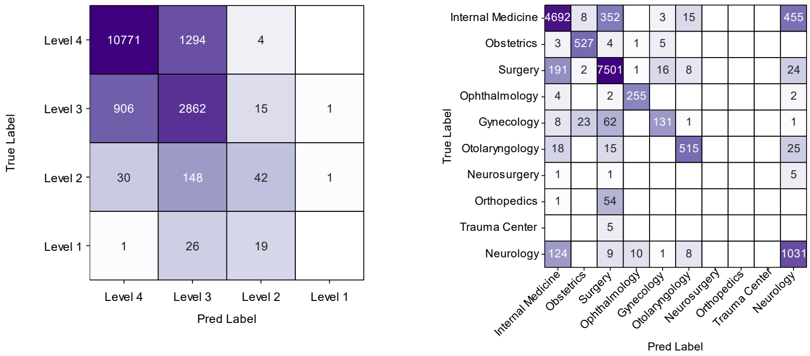
3.4 Comparisons with the Previous Method

Table 2 compares the performance of different imputation methods, such as GAN [20],



(a) Severity (b) Department

Fig. 2. Precision-Recall (PR) curves for severity-level prediction and department recommendation using the proposed method in testing set.



(a) Severity (b) Department

Fig. 3. Performance of our method for severity-level and department prediction.

Variational Auto-Encoder (VAE) [21] [22], Multiple Imputation by Chain Equations (MICE) [15], Random Forest-based Imputation (RF) [16] for severity and department prediction. The proposed CGMI outperforms all other methods, achieving the highest F1 of 84.77% and K of 61.42% for severity prediction. It also leads in department prediction with SE of 90.89%, SP of 91.04%, and F1 of 90.76%. MICE and RF show similar performance but fall short of CGMI overall, while GAN shows competitive results, particularly in K. VAE, though effective in department prediction, performs poorly in severity prediction. These results demonstrate CGMI's superior performance in handling missing data.

To validate the effectiveness of our mFEM, this section further analyzes the department prediction task by replacing the backbone with Transformer [23], ResNet [24], and TextCNN [25]. Table 3 compares triage performance across different backbone modules under the proposed method. The proposed model outperforms all other backbones, achieving the highest SE of 84.83% and K of 61.42% for severity prediction and the best performance in department prediction with all metrics. Among the other backbones, Transformer shows competitive results with SE of 84.74% and K of 61.66%, while ResNet and TextCNN perform slightly worse in sensitivity and specificity. NomFEM, although practical, shows lower performance than the proposed method. These results highlight the superior performance of the proposed model, demonstrating its effectiveness in both tasks.

We utilized the Precision-Recall curves to evaluate the impact of imbalanced data on the proposed method. Fig. 2 (a) and (b) show the high Performance for categories with large data volumes and poor detection results for the few samples with tiny data volumes. For example, in the severity prediction task, the proposed method achieved an average precision (AP) of 0.7250 for level 4 (N=37,84) and 0.9747 for level 4 (N=12,069), indicating its superior performance in identifying severity level. The proposed method performs well in the department prediction task on larger sample categories, such as Internal Medicine, Surgery, and Ophthalmology. However, performance deteriorates for long-tail categories with fewer samples, as limited training data leads to insufficient feature learning and significant bias. Analysis of the confusion matrix (Fig. 3) and the original data reveals that feature overlap, particularly in long-tail categories, impairs the model's ability to distinguish between these classes. For example, in structured data, identical features may correspond to different severity levels due to the influence of chief complaints. Common terms like "dizziness" appear in the chief complaints across multiple departments, causing the model to favor more populous categories. Similar issues arise in the Trauma Center, where terms like "fall injury" and "fracture" overlap with Surgery and Orthopedics, leading to similar biases. Despite some distinctions in structured data, the model's ability to mitigate this bias is limited, highlighting the challenge of improving long-tail sample recognition in imbalanced datasets.

4 Conclusion

In this paper, we propose a novel approach to enhance severity and department prediction in ED triage by addressing the challenges of missing and sparse data. Our method incorporates Conditional Gaussian Mixture Imputation (CGMI) to mitigate the impact of missing structured data, a Feature Densification Module (FDM) to capture relationships among sparse variables, and a multi-scale Feature Extraction Module (mFEM) to improve the semantic representation of unstructured data. Experimental results demonstrate that the proposed method significantly outperforms existing approaches, achieving higher sensitivity, specificity, F1 score and Cohen's Kappa value across both prediction tasks. Future work will explore further optimization and broader clinical applications of the model.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China under Grant 61873066 and Grant 62173090, and the Fundamental Research Funds for the Central Universities 2242024K40039. This work used retrospectively collected pathological records from the ED of Peking Union Medical College Hospital. The data collection was granted by the Independent Ethics Committee for Clinical Research of Peking Union Medical College Hospital under approval No. I-23PJ130.

Disclosure of Interests. The authors have no competing interests.

References

1. Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo, C. A., Hasegawa, K.: Emergency Department Triage Prediction of Clinical Outcomes Using Machine Learning Models. Critical Care **23**(1), 64 (2019).
2. El-Bouri, R., Eyre, D. W., Watkinson, P., Zhu, T., Clifton, D. A.: Hospital Admission Location Prediction via Deep Interpretable Networks for the Year-Round Improvement of Emergency Patient Care. IEEE Journal of Biomedical and Health Informatics **25**(1), 289–300 (2021).
3. Liu, R., Rizzo, S., Whipple, S., Pal, N., Pineda, A. L., Lu, M., Arnieri, B., Lu, Y., Capra, W., Copping, R., Zou, J.: Evaluating Eligibility Criteria of Oncology Trials Using Real-World Data and AI. Nature **592**(7855), 629–633 (2021).
4. Goto, T., Camargo, C. A., Faridi, M. K., Yun, B. J., Hasegawa, K.: Machine Learning Approaches for Predicting Disposition of Asthma and COPD Exacerbations in the ED. The American Journal of Emergency Medicine **36**(9), 1650–1654 (2018).
5. Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L.-C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S.-C., Kuo, P.-C., Lungren, M. P., Palmer, L. J., Price, B. J., Purkayastha, S., Pyrros, A. T., Oakden-Rayner, L., Okechukwu, C., Seyyed-Kalantari, L., Trivedi, H., Wang, R., Zaiman, Z., Zhang, H.: AI Recognition of Patient Race in Medical Imaging: A Modelling Study. The Lancet Digital Health **4**(6), e406–e414 (2022).
6. Yu, J. Y., Heo, S., Xie, F., Liu, N., Yoon, S. Y., Chang, H. S., Kim, T., Lee, S. U., Hock Ong, M. E., Ng, Y. Y., Do Shin, S., Kajino, K., Cha, W. C.: Development and Asian-Wide Validation of the Grade for Interpretable Field Triage (GIFT) for Predicting Mortality in Pre-Hospital Patients Using the Pan-Asian Trauma Outcomes Study (PATOS). The Lancet Regional Health - Western Pacific **34**, 100733 (2023).
7. Kong, J. S., Lee, K. H., Kim, O. H., Lee, H. Y., Kang, C. Y., Choi, D., Kim, S. C., Jeong, H., Kang, D. R., Sung, T.-E.: Machine Learning-Based Injury Severity Prediction of Level 1 Trauma Center Enrolled Patients Associated with Car-to-Car Crashes in Korea. Computers in Biology and Medicine **153**, 106393 (2023).
8. Liang, W., Yao, J., Chen, A., Lv, Q., Zanin, M., Liu, J., Wong, S., Li, Y., Lu, J., Liang, H., Chen, G., Guo, H., Guo, J., Zhou, R., Ou, L., Zhou, N., Chen, H., Yang, F., Han, X., Huan, W., Tang, W., Guan, W., Chen, Z., Zhao, Y., Sang, L., Xu, Y., Wang, W., Li, S., Lu, L., Zhang, N., Zhong, N., Huang, J., He, J.: Early Triage of Critically Ill COVID-19 Patients Using Deep Learning. Nature Communications **11**(1), 3543 (2020).
9. Rohanian, O., Kouchaki, S., Soltan, A., Yang, J., Rohanian, M., Yang, Y., Clifton, D.: Privacy-Aware Early Detection of COVID-19 Through Adversarial Training. IEEE Journal of Biomedical and Health Informatics **27**(3), 1249–1258 (2023).

10. Sax, D. R., Warton, E. M., Kene, M. V., Ballard, D. W., Vitale, T. J., Timm, J. A., Adams, E. S., McGauhey, K. R., Pines, J. M., Reed, M. E., Kaiser Permanente Clinical Research on Emergency Services & Treatments (CREST) Network, Rauchwerger, A. S., Di Lena, D. D.: Emergency Severity Index Version 4 and Triage of Pediatric Emergency Department Patients. JAMA Pediatrics **178**(10), 1027 (2024).
11. Xiao, Y., Zhang, J., Chi, C., Ma, Y., Song, A.: Criticality and Clinical Department Prediction of ED Patients Using Machine Learning Based on Heterogeneous Medical Data. Computers in Biology and Medicine **165**, 107390 (2023).
12. Wei, Y., Tang, Y., McNicholas, P. D.: Flexible High-Dimensional Unsupervised Learning with Missing Data. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(3), 610–621 (2020).
13. Kang, M., Zhu, R., Chen, D., Liu, X., Yu, W.: CM-GAN: A Cross-Modal Generative Adversarial Network for Imputing Completely Missing Data in Digital Industry. IEEE Transactions on Neural Networks and Learning Systems **35**(3), 2917–2926 (2024).
14. Marchang, N., Tripathi, R.: KNN-ST: Exploiting Spatio-Temporal Correlation for Missing Data Inference in Environmental Crowd Sensing. IEEE Sensors Journal **21**(3), 3429–3436 (2021).
15. Azur, M. J., Stuart, E. A., Frangakis, C., Leaf, P. J.: Multiple Imputation by Chained Equations: What Is It and How Does It Work? International Journal of Methods in Psychiatric Research **20**(1), 40–49 (2011).
16. Stekhoven, D. J., Bühlmann, P.: MissForest - Nonparametric Missing Value Imputation for Mixed-Type Data. Bioinformatics **28**(1), 112–118 (2012).
17. Ngo, H., Fang, H., Rumbut, J., Wang, H.: Federated Fuzzy Clustering for Decentralized Incomplete Longitudinal Behavioral Data. IEEE Internet of Things Journal **11**(8), 14657–14670 (2024).
18. Chen, B., Oliva, J., Niethammer, M.: A Unified Model for Longitudinal Multi-Modal Multi-View Prediction with Missingness. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, Linguraru, M. G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J. A., Eds., Lecture Notes in Computer Science, Springer Nature Switzerland: Cham, 15012, pp 410–420 (2024).
19. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019*, Minneapolis, Minnesota, Vol.1, p.2, pp 4171–4186 (2019).
20. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. arXiv (2014).
21. Kingma, D. P., Welling, M.: Auto-Encoding Variational Bayes. arXiv (2022).
22. Barrejón, D., Olmos, P. M., Artés-Rodríguez, A.: Medical Data Wrangling With Sequential Variational Autoencoders. IEEE Journal of Biomedical and Health Informatics **26**(6), 2737–2745 (2022).
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need. In *Proc. Adv. Neural Inf. Process. Syst.*, pp 5998–6008 (2017).
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv (2015).
25. Kim, Y.: Convolutional Neural Networks for Sentence Classification. arXiv (2014).