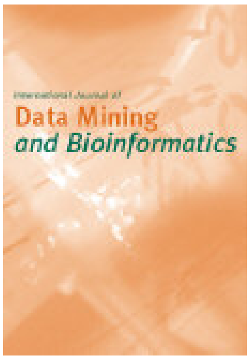


International Journal of Data Mining and Bioinformatics

ISSN online: 1748-5681 - ISSN print: 1748-5673
https://www.inderscience.com/ijdmb

Data mining-based integration method of infant emergency and critical information in modern hospital

Juan Xiao, Jina Zhang, Xiaoli Liu

DOI: 10.1504/IJDMB.2023.10058132

Article History:

Received:	24 February 2023
Last revised:	10 April 2023
Accepted:	19 June 2023
Published online:	17 October 2023

Data mining-based integration method of infant emergency and critical information in modern hospital

Juan Xiao* and Jina Zhang

College of Nursing,
HeNan Medical College,
ZhengZhou, 450000, China
Email: xiaojuan102886@163.com
Email: 1043648539@qq.com
*Corresponding author

Xiaoli Liu

Department of Nursing,
The Second People's Hospital of Henan Province,
ZhengZhou, 450000, China
Email: 1287195659@qq.com

Abstract: In this paper, a modern hospital infant emergency and critical information integration method based on data mining is designed. First of all, the data types of children's critical information in modern hospitals are analysed. Then, metadata is extracted through mapping relationship. Finally, the data missing value is filled in by the mean filling method, and the support and correlation of the data are calculated by the association rule algorithm, and the information integration model is constructed to realise the information data integration. The test results show that the error of the proposed method for the integration of children's emergency and critical information in modern hospitals is always lower than 0.3%, the throughput is always above 75 Mbps, and the maximum integration time is only 2.12 s, which has good practical application performance.

Keywords: data mining; modern hospitals; children are in critical condition; information integration; metadata; variable linear method.

Reference to this paper should be made as follows: Xiao, J., Zhang, J. and Liu, X. (2023) 'Data mining-based integration method of infant emergency and critical information in modern hospital', *Int. J. Data Mining and Bioinformatics*, Vol. 27, No. 4, pp.312–325.

Biographical notes: Juan Xiao received his Master's degree in Xinxiang Medical University Pathology and Pathophysiology in 2021. He is currently an Associate Professor at the Vocational School College of Nursing. His research fields are paediatric nursing, critical care and clinical nursing education.

Jina Zhang received his Master's in Epidemiology and Health Statistics from the Zhengzhou University in 2013. He is currently an Associate Professor at the Vocational School College of Nursing. His research directions are clinical nursing, nursing quality and safety management and nursing education.

Xiaoli Liu obtained her Bachelor's degree in the Zhengzhou University Nursing in 2011. He is currently the Deputy Chief Nurse of the Nursing Department of the Second People's Hospital of Henan Province. Her research direction is medical nursing.

1 Introduction

With the rapid rise of the electronic information technology industry, the informatisation level of the medical industry has also increased. Most hospitals have established their own key medical information systems. A large number of key medical data are stored and integrated in the medical information system (Karolina et al., 2022). Among them, the medical information data of children in the modern hospital is the key component of the system data. These critical and critical medical information data of children cover a variety of key information about children's diseases, which mainly includes text, numbers, images and other multimedia data (Ramgopal et al., 2022). Because all emergencies of young children are critical diseases, the integration of known disease information can help other young children to quickly diagnose critical diseases, and is conducive to the formulation of treatment plans for children's critical diseases. However, there are various forms and a large number of medical information data for children with acute and critical diseases, which makes it difficult to manage them (Shivani et al., 2022). Information integration is a dynamic process that takes information resources as the object, organically integrates and optimises multiple information. Through information integration, data can be more standardised and consistent, data conflicts and contradictions between different data sources can be avoided, and the quality and reliability of data can be improved. This allows users to get the information they need more quickly, eliminating the need to frequently switch between different systems or data sources. In addition, the entry and storage of medical information data are only limited to common management methods, which cannot effectively achieve management. To this end, the effective integration of the information data of infant emergency and critical care in modern hospitals is the key link in its management. The information integration of infant emergency and critical care in modern hospitals can effectively manage data in all aspects (Ceylan et al., 2021). However, with the diversity of information data forms and the non-change of attributes, there are also major problems in the integration process summary. To this end, relevant researchers targeted. The information integration method of infant emergency and critical illness in modern hospital has been designed and improved.

Some researchers put forward a method of integrating hospital infant emergency and critical care information multi-group information data based on generating confrontation network. This method introduces the generation of antagonism network model, which can realise the integration of two group data and their interaction network. Acquire information from the interactive network and the two sets of omics data through the confrontation network model, and fuse them to generate synthetic data with better prediction signals (Ahmed et al., 2022). This method has a good effect on the integration of children's emergency and critical information in modern hospitals, but the integration takes a long time and reduces the work efficiency in practical application. Other scholars have combined the best performance internal standard correction and support vector regression algorithm to study the data integration and standardisation methods of large-scale metabolomics. The best performance internal standard correction is combined with support vector regression normalisation to completely eliminate systematic and random errors and matrix effects. The adoption of norm ISWSVR reduces the relative standard deviation of median cross-validation of data, increases the correlation between data, improves the classification accuracy of biomarkers, and is well compatible with quantitative data (Ding et al., 2022). However, the metadata in this method is relatively complex, resulting in some errors in the data integration results. Some scholars in this field have put forward a method of integrating the information data of infant emergency and critical care in modern hospitals by using the parameters of the simplified external model to estimate the ratio. This method makes limited assumptions about the similarity of the distribution in the two research populations, obtains information about the coefficient ratio from the external model through orthogonal variables, and updates the correlation results between the parameters in the generalised linear model and the parameters of the generalised linear model with omitted covariant, and obtains the final integration results (Taylor et al., 2022). However, with the increase of data volume, the network throughput in integration is small, which cannot meet the actual application requirements of massive data integration.

In order to solve the shortcomings of the above methods, reduce and improve the error and time consumption of the integration of children's emergency and critical information in modern hospitals, and improve the network throughput in the integration process, a method of integrating children's emergency and critical information in modern hospitals based on data mining technology is designed.

2 Analysis and extraction of data types of children's emergency and critical information in modern hospitals

2.1 Analysis of data types of children's emergency and critical information in modern hospitals

In order to realise the integration of children's emergency and critical information in modern hospitals, it is necessary to first determine the metadata of children's emergency and critical information in modern hospitals. The information data of children's emergency illness in modern hospitals is a kind of medical data for children's health. The data needs to be divided in detail according to the composition type (Lipovetsky, 2022). The classification of the types of children's emergency and critical information data in modern hospitals is the basis of its metadata. Therefore, in the face of children's patients, doctors, medical devices, etc., these objects affect the key of the whole children's emergency and critical information (Wu et al., 2022). The data types of children's emergency and critical information in modern hospitals are mainly divided into children's clinical data, children's health big data, biological big data and operation big data. Table 1 shows the types and details of emergency and critical information of children in modern hospitals.

Among the data types of children's acute and critical information in modern hospitals, it is determined according to a variety of data types (Zhao and Li, 2022). Among these data, it is also necessary to determine the types of data in different databases. The modern hospital infant emergency and critical information database is shown in Table 2.

Table 1 Data types of children's emergency and critical information in modern hospitals

Data type	*Detailed information*
Clinical data of young children	Electronic archives, biomedical effects, signals and some key clinical data
Big data of infant health	The lifestyle, environment and behaviour of children's individual health
Biological big data	Critical data related to children obtained from biomedical laboratories, clinical trials and public neighbourhoods
Operation big data	Various medical institutions, child care centres, etc.

Table 2 Types of information database of children's acute and critical diseases in modern hospitals

Database name	*Database type*	*Details*
Entrez Gene	Gene	Gene sequence annotation
Ensembl	Gene	Recent gene sequence data
GO	Gene	Gene ontology data
db	Genetic variation	Polymorphic data
HPO	Phenotype and disease	Phenotypic characteristics
SMPDB	Metabolic pathway	Metabolic pathway data
HMDB	Metabolism and disease	Provide metabolic disease data

In the analysis of data types of children's emergency and critical information in modern hospitals, the data types are determined according to the data objects and sources of children's emergency and critical information in modern hospitals, and the data types existing in different databases are determined to realise the analysis of data types of children's emergency and critical information in modern hospitals. Based on this type of data, it provides a data basis for the subsequent metadata extraction of infant critical care information in modern hospitals.

2.2 Metadata extraction of children's critical information in modern hospitals

After the determination of the data type of children's emergency and critical information in modern hospitals, in order to improve the accuracy of the integration of children's emergency and critical information in modern hospitals, it is necessary to integrate on the basis of the key metadata of information. Therefore, this paper extracts these different types of metadata (Wang et al., 2022).

Because the text and other types of data in the information data of children's critical illness in modern hospitals are unstructured data, it is very difficult to integrate and store them, and it is difficult to obtain key data information directly from their text. Therefore, in the metadata extraction of the data information, the unstructured data is converted into structured general data for the convenience of subsequent integration operations (Oh et al., 2022). In the process of transforming from unstructured to structured, it is realised by means of data ontology mapping. Suppose that the metadata set in the data table is determined in the modern hospital infant critical and critical information data, and the unstructured dataset is set to exist:

$$a_i \in A \tag{1}$$

In formula (1), a_i represents the data table of the first metadata in the unstructured dataset, which also exists in the structured dataset. Set a mapping of the corresponding relationship between the ontology class in the ontology data and the emergency and critical information data of children in modern hospitals (Li et al., 2021), which can be described as:

$$\operatorname{map}(a_i) = \operatorname{map}(a_i) + \{y_i\} \tag{2}$$

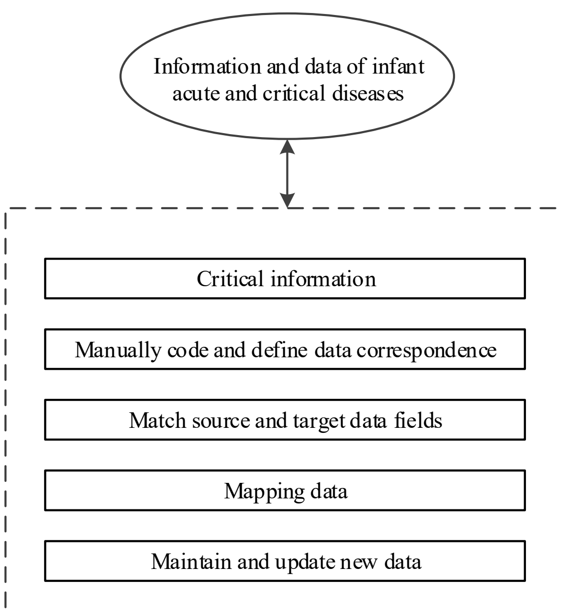
In formula (2), map(a_i) represents the mapping description of the corresponding relationship, and y_i represents the ontology class of the i metadata of the modern hospital infant critical care information.

After the transformation of structured data, that is, the metadata of children's emergency and critical information data in modern hospitals should be marked with a certain special marking principle. In this way, the structural transformation of children's emergency and critical information data in modern hospitals can be realised. The design criteria are expressed as follows:

$$O_i = \sum a_i \int \operatorname{map}(a_i) \nabla A \tag{3}$$

In formula (3), O_i represents the i metadata.

Figure 1 Data mapping process of children's emergency and critical information in modern hospitals



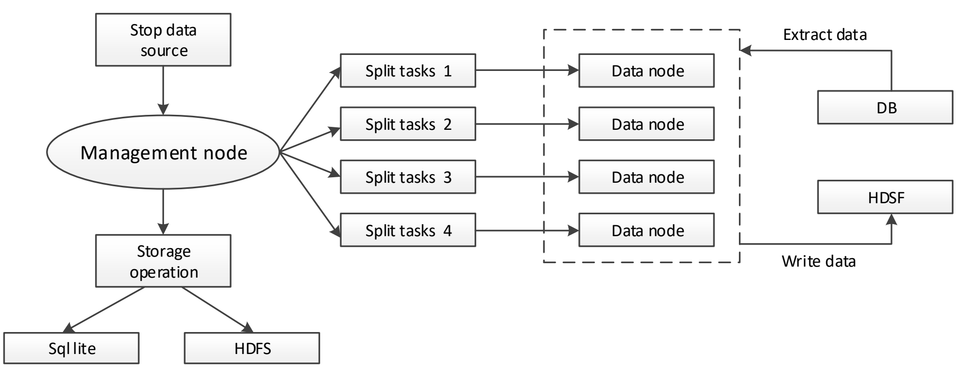
On this basis, we extract the metadata from the structured information data of children's acute and critical illness in modern hospitals. Through the mapped tag data, establish a data extraction path (Jackson and Vogel, 2022), extract metadata from these structured data (Sun et al., 2022), and uniformly describe it with ontology data, that is, map the modern hospital infant emergency and critical information data into multiple ontology metadata, and the implementation process is shown in Figure 1.

According to the mapped information data relationship of acute and critical children in modern hospitals, the concept of data ontology and the related concept are linked together, and the metadata (Lu and Welch, 2022) of acute and critical information data of children in modern hospitals is extracted. Assuming that there are n acute and critical information data for children in modern hospitals, the attributes of the i metadata are set to t_i and the ontology properties to c_i. Then, the results of the metadata for modern hospital children extracted during the process are:

$$R(x) = \sum c_i / t_i \int v' \tag{4}$$

In formula (4), $R(x)$ represents the metadata extraction collection of acute and critical childhood information in modern hospitals, and v' represents the principal components of the medium metadata extraction. According to the method of extracting metadata, the schematic diagram of metadata extraction of acute and critical children in modern hospitals is drawn, as shown in Figure 2.

Figure 2 Schematic diagram of metadata extraction of children's critical information in modern hospitals



By transforming the unstructured modern hospital infant critical and critical information metadata into structured general data, we can determine the attributes between different metadata, set transformation constraints, and extract information metadata through mapping relationships to provide a data basis for the next step of information integration.

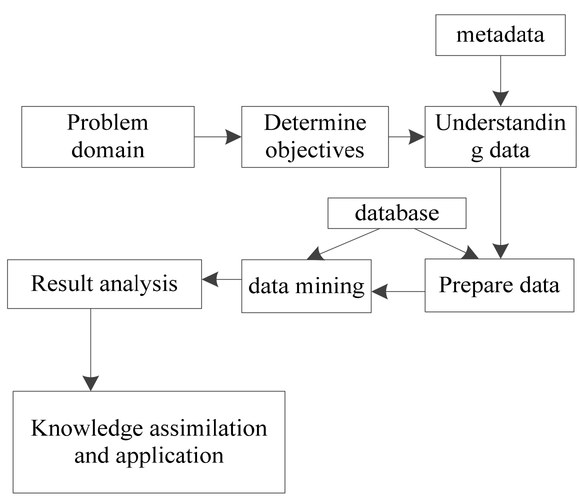
3 Design of information integration method of infant emergency and critical illness in modern hospital based on data mining technology

After the extraction of the metadata of children's critical information in modern hospitals, this paper introduces data mining technology to integrate the information data of children's critical information in modern hospitals. By setting the attribute missing value, mining the missing value of children's emergency and critical information data in modern hospitals, filling the missing value with the mean filling method, calculating the support and relevance of the data with the help of the association rule algorithm, determining the key characteristics of the data and classifying it through the variable linear method, and constructing the information integration model to achieve the final information integration.

Data mining technology is to extract the data hidden in the database that is not easy to be found by finding completely different, fuzzy and other random data in the database. The technology originated from and was influenced by many disciplines. This technology includes many algorithms such as database, artificial intelligence and machine learning (Lu and Welch, 2022). With the help of the key algorithms in this technology, the integration of research objects is realised, which is characterised by fast speed and accurate integration objectives. Therefore, this paper introduces this technology for integrated design.

Modern hospital infant emergency and critical information data has its particularity. During the data integration process, the data can be effectively queried and the potential relationship between the data can be analysed. These processes are the key steps in the data mining process. In this integration, we will understand relevant knowledge and solve key data problems by mining the data targets of children's emergency and critical information in modern hospitals, and then understand the data after in-depth mining and prepare the integrated data (Rautenstrauch et al., 2022). The implementation process of data mining of children's emergency and critical information in modern hospitals is shown in Figure 3.

Figure 3 Implementation process of data mining for children's emergency and critical information in modern hospitals



The general data mining process in the data mining of children's emergency and critical information in modern hospitals is relatively simple, and there are many data missing due to storage problems. In this in-depth mining, the data mining technology is used to mine the missing data of children's emergency and critical information data in modern hospitals and improve the integration of subsequent data (Thiele et al., 2021). The incomplete set of information data of infant acute and critical illness in modern hospitals is as follows:

$$D = \{d_1, d_2, \ldots, d_m\} \tag{5}$$

In formula (5), D represents the dataset of acute and critical illness information of children in modern hospitals, d_1, d_2, …, d_m represents the data composition in the set, and m represents the amount of data.

Set an attribute missing value for each modern hospital infant critical and critical information data in formula (5), and express the missing attribute set of each data as:

$$S(d_i) = \{x_j^e \mid \forall x_j^e \in S \cap x_j^e \cup d_e\} \tag{6}$$

In formula (6), $S(d_e)$ represents the missing attribute set of acute and critical childhood information data in modern hospitals, x_j^e represents the j attribute in the missing attribute e state, and S represents the missing pattern value.

Fill in the missing data of emergency and critical care information of children in the mining hospital effectively to make it more complete data. In order to improve the quality of integrated data, this paper uses the mean filling method to achieve filling (Lu and Zhou, 2021), whose filling formula is expressed as:

$$k_j^\alpha = \sum\nolimits_{k \in S} \frac{w_j^\alpha}{f_e} \tag{7}$$

In formula (7), w_j^α represents the index set of all missing data with inaccurate attributes in the α dimension, f_e represents the missing dataset, and k_j^α represents the mean filling result.

Based on the above mined information data of critical children in modern hospitals, the correlation between these data is further determined, so as to reduce the tedious degree of similarity data integration. The correlation between these data is determined according to the association rule algorithm in data mining techniques. Between the missing information of children in modern hospitals are two, in which the support in the things data includes F and G things, and the support between the data of these two things is expressed as:

$$\text{support}(F \to G) = H(F \cup G) \tag{8}$$

In formula (8), → represents the association rule, support($F \to G$) represents the support degree between the data of two things, and H represents the support probability.

According to the calculation results of the support degree, we can determine the support degree of the thing data between the missing data of children's emergency and critical information in modern hospitals, and then we can determine the correlation between the data in the two thing datasets. The calculation formula is:

$$con(F, G) = \frac{1}{m} \sum\nolimits_{i=1}^{n} v_w \int H \tag{9}$$

In formula (9), *con*(F, G) represents the correlation between the data in the two datasets of things, and v_w represents the coefficient of correlation between the data.

According to the correlation between the determined information of children in modern hospitals, the key characteristics of these data are determined, and the key characteristics of these data are determined by variable linear method. The results are as follows:

$$P(X,Y)=\frac{\sum_i (x_u-y_u)(x_u-y_u)^2}{\sqrt{\sum (x_u-y_u)^2}} \tag{10}$$

In formula (10), $P(X, Y)$ represents the dataset of key features, and x_u and y_u represent two linear data.

Determine the key characteristics of these data according to the variable linear method, estimate the distance between these data to determine the final similar data, and integrate them to a certain extent. The integrated result is:

$$g(x)=\frac{1}{n}\sum_{x=1}^{n} K\left[\frac{(x_u-y_u)}{h}\right] \tag{11}$$

In formula (11), $g(x)$ represents similar data results after integration, K represents interference coefficient in integration, and h represents high nuclear density in any data integration.

On this basis, the above modern hospital infant emergency and critical information data will be classified, and the final integration will be based on different databases after classification. The classification is realised with support vector machine classification algorithm, and the classification formula is:

$$\mu=\sum_{i=1}^{n} x_u\varepsilon_i+b \tag{12}$$

In formula (12), μ represents the result after classification, ε_i represents the optimal function of quadratic classification, and b represents the bias value in the classification.

On this basis, a modern hospital infant emergency and critical care information integration model based on data mining is constructed, and the above classified data is input into it to achieve the final integration. The constructed model is:

$$\gamma(x)=sign(\mu(i))\,\omega\sum_{i=1}^{n} x_u\varepsilon_i+b \tag{13}$$

In formula (13), $\gamma(x)$ represents the integration result, ω represents the data weight value, and *sign* represents the integrator.

In the integration of children's emergency and critical information in modern hospitals, by setting attribute missing values, mining the missing values of children's emergency and critical information data in modern hospitals, filling in the missing values through the mean filling method, calculating the support and relevance of the data with the help of association rule algorithm, determining the key characteristics of the data through the variable linear method and classifying, building the information integration model, and achieving the final information integration.

4 Experimental analysis

4.1 Experimental scheme design

In the test, the internal information storage platform of children's hospital was selected as the research object for integration research. The operating system of the platform is Windows 10, and the operating memory of the system is 32 GB. The paediatric emergency department of the hospital has been running for 15 years. Therefore, there are many critical and critical information data of children in the internal information storage platform of the children's hospital. This time, the critical and critical information of children under three years of age in the platform is selected as the research object, and the specific research parameters are shown in Table 3.

Table 3 Test parameters

Parameter	*Details*
Sample data volume / piece	20,000
Data type / class	4
Data noise / dB	< 2
Data integration interval / s	1.0
Test experiment error / %	< 1
Iterations / time	100

In order to ensure the consistent effect of integration during the test, all experimental sample data are processed in a unified environment and tested and studied under a unified condition. In order to ensure the quality and accuracy of the original data for subsequent data analysis and mining, it is necessary to clean the data, reduce noise, improve the accuracy and efficiency of data mining algorithms, and make the conclusions more scientific and reliable. In order to highlight the effectiveness of the proposed method, the comparison method is selected. The comparison method is: the proposed method, the information data integration method based on the generation of confrontation network, and the information data integration method based on the best performance internal standard correction and support vector regression algorithm; the error of the integration result of the sample infant critical and critical information, the system throughput in the integration and the time consuming result of the integration are used as test indicators to measure the final effect of the integration method. The smaller the integration error, the larger the system throughput and the shorter the integration time, the better the effect of the integration method and the better the actual application performance.

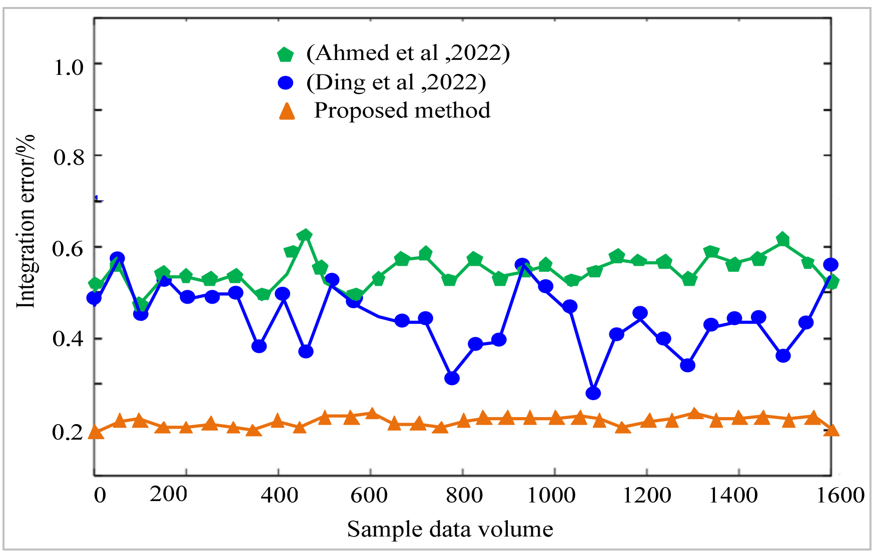
4.2 Analysis of experimental results

4.2.1 Integration error

Using the proposed method, the information data integration method based on the generation of confrontation network, and the information data integration method based on the best performance internal standard correction and support vector regression algorithm, the integration error results are analysed for 1,600 random samples of children's critical information data, and the error results are shown in Figure 4.

Analysis of the experimental results in Figure 4 shows that the three information data integration methods have certain differences in the integration error results of random 1,600 samples of acute and critical disease information data of young children. The error of the information data integration method based on the generative adversarial network fluctuates between 0.46%~0.62%. The error of the information data integration method based on standard correction and support vector regression algorithm in the best performance fluctuates between 0.26%~0.58%. The error of data integration of the proposed method fluctuates between 0.18%~0.22%, and the error is always less than 0.3%, the fluctuation range is small, and the error is far lower than that of the two methods compared, indicating that the accuracy of information data integration using the proposed method is high, and the proposed method has certain feasibility.

Figure 4 Analysis of the error results of the integration of the information data of children's critical illness (see online version for colours)

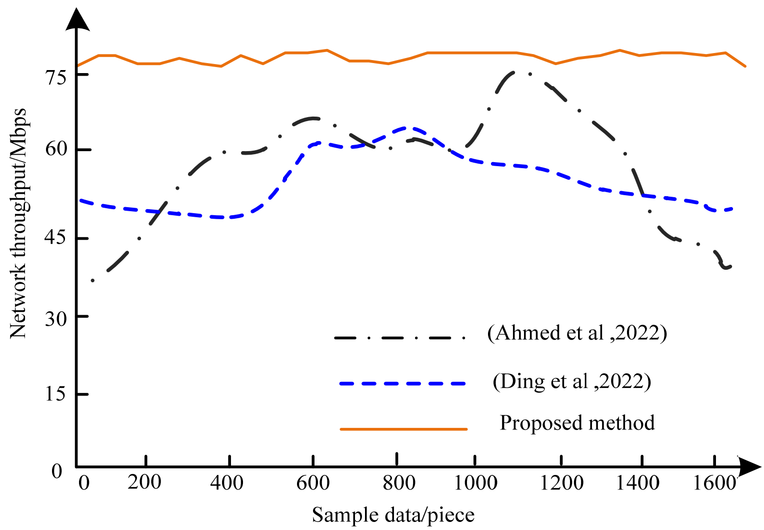


4.2.2 *Network throughput*

Using the proposed method, the information data integration method based on the generation of confrontation network, and the information data integration method based on the best performance internal standard correction and support vector regression algorithm, the throughput of the network during the integration of 1,600 random samples of children's critical information data is analysed, and the results are shown in Figure 5.

Analysis of the experimental results in Figure 5 shows that with the increase of data volume, the throughput results of the network of the three information data integration methods for random 1,600 samples of acute and critical diseases of young children also fluctuate to a certain extent. When integrated by the information data integration method based on the generative adversarial network, the throughput of the network fluctuates between 38 Mbps~75 Mbps. The throughput of the network is always maintained above 75 Mbps during the information data integration based on the standard correction and support vector regression algorithm in the best performance, which is much higher than that of the two methods compared, indicating that the throughput of the network in the integration process of the proposed method is high, which proves that the data integration effect is better and has good applicability.

Figure 5 Analysis of throughput results of the network during the integration of children's emergency and critical information data (see online version for colours)



4.2.3 Integration time consumed

Using the proposed method, the information data integration method based on the generation of confrontation network, and the information data integration method based on the best performance internal standard correction and support vector regression algorithm, the integration time consuming analysis was carried out for 1,600 random samples of children's critical information data, and the results are shown in Table 4.

Table 4 Time consumption analysis of data integration of children's emergency and critical information(s)

Sample data volume / piece	*Information data integration method based on generating countermeasure network*	*Information data integration method based on best performance internal standard correction and support vector regression algorithm*	*Proposed method*
200	2.54	1.21	1.12
400	3.65	2.38	1.24
600	4.98	3.51	1.45
800	5.02	4.97	1.65
1,000	6.12	5.65	1.76
1,200	6.56	6.43	1.81
1,400	7.21	6.89	1.98
1,600	7.43	7.21	2.12

Analysis of the experimental results in Table 4 shows that with the increase of data volume, the three information data integration methods have different time-consuming results for the integration of 1,600 random samples of acute and critical disease information data of young children. When the maximum data volume reaches 1,600, the integration time of the information data integration method based on generative adversarial network is 7.43 s, the integration time of the information data integration method based on standard correction and support vector regression algorithm in the best performance is 7.21 s, and the integration time of the proposed method is 2.12 s, which is much lower than that of the two comparative methods and takes a short time. It shows that the proposed method can effectively shorten the data integration time, improve work efficiency, and have better practical application performance.

5 Conclusions

The integration of children's emergency and critical information in modern hospitals is conducive to the development of the whole medical field. Facing the problems of large integration error, small network throughput and long integration time, this paper designs a method of integrating children's emergency and critical information in modern hospitals based on data mining technology:

1 Determine the data type and extract metadata through mapping relationship; mining the missing value of the information data of children's acute and critical illness in modern hospitals, filling it with the mean filling method, using the association rule algorithm to calculate the support and relevance of the data, determining the key features of the data and classifying it through the variable linear method, and constructing the information integration model to achieve information integration.

2 The test results show that the proposed method can effectively reduce the data integration error and integration time, improve the network throughput during the integration process, and have good practical application performance and high feasibility.

Acknowledgements

This work was supported by the Department of Education of Henan Province. The research and practice of innovative nursing talents training model with post competence as the core, project number: Jiao Gao [2022]111.

References

Ahmed, K.T., Sun, J., Cheng, S., Yong, J. and Zhang, W. (2022) 'Multi-omics data integration by generative adversarial network', *Bioinformatics*, Vol. 38, No. 1, pp.179–186.

Ceylan, B., Yayla, C., Aykac, K., Ozsurekci, Y. and Ceyhan, M. (2021) 'Characteristics and management of children with COVID-19 in a tertiary care hospital in Turkey', *Clinical Pediatrics*, Vol. 60, No. 3, pp.170–177.

Ding, X., Yang, F., Chen, Y., Xu, J., He, J., Zhang, R. and Abliz, Z. (2022) 'Norm ISWSVR: a data integration and normalization approach for large-scale metabolomics', *Analytical Chemistry*, Vol. 94, No. 21, pp.7500–7509.

Jackson, C.A. and Vogel, C. (2022) 'New horizons in the stormy sea of multimodal single-cell data integration', *Molecular Cell*, Vol. 82, No. 2, pp.248–259.

Karolina, B., Gerda, G. and Milda, E. (2022) 'A retrospective analysis of tick-borne encephalitis in children treated in Kaunas Hospital during 2012 to 2019', *The Pediatric Infectious Disease Journal*, Vol. 41, No. 9, pp.702–705.

Li, H., Xiao, X., Wu, X., Ye, L. and Ji, G. (2021) 'scLINE: a multi-network integration framework based on network embedding for representation of single-cell RNA-seq data', *Journal of Biomedical Informatics*, Vol. 122, No. 23, p.103899.

Lipovetsky, S. (2022) 'Multivariate data integration using R: methods and applications with the mixOmics package', *Technometrics*, Vol. 64, No. 3, pp.429–431.

Lu, L. and Welch, J.D. (2022) 'PyLiger: scalable single-cell multi-omic data integration in Python', *Bioinformatics*, Vol. 11, No. 10, p.10.

Lu, L. and Zhou, J. (2021) 'Research on mining of applied mathematics educational resources based on edge computing and data stream classification', *Mobile Information Systems*, Vol. 2021, No. 7, pp.1–8.

Oh, T.H., Park, H.M., Kim, J.W. and Lee, J.M. (2022) 'Integration of reinforcement learning and model predictive control to optimize semi-batch bioreactor', *AIChE Journal*, Vol. 68, No. 6, pp.1–16.

Ramgopal, S., Sepanski, R.J. and Martin-Gill, C.C. (2022) 'Empirically derived age-based vital signs for children in the out-of-hospital setting', *Annals of Emergency Medicine*, Vol. 17, No. 1, pp.1065–1079.

Rautenstrauch, P., Vlot, A.H.C., Saran, S. and Ohler, U. (2022) 'Intricacies of single-cell multi-omics data integration', *Trends in Genetics*, Vol. 38, No. 2, pp.128–139.

Shivani, G., Josh, B. and Brendan, B. (2022) 'Equivalent adverse outcomes for home and hospital-based infusions of infliximab in children and adolescents with inflammatory bowel disease', *Inflammatory Bowel Diseases*, Vol. 162, No. 3, pp.97–106.

Sun, C., Zhou, Z., Zeng, X., Li, Z., Wang, Y. and Deng, F. (2022) 'A multi-model-integration-based prediction methodology for the spatiotemporal distribution of vulnerabilities in integrated energy systems under the multi-type, imbalanced, and dependent input data scenarios', *Applied Energy*, Vol. 320, No. 15, pp.1–18.

Taylor, J.M.G., Kyuseong, C. and Han, P. (2022) 'Data integration: exploiting ratios of parameter estimates from a reduced external model', *Biometrika*, Vol. 12, No. 3, pp.632–654.

Thiele, S.T., Lorenz, S. and Kirsch, M. (2021) 'Multi-scale, multi-sensor data integration for automated 3-D geological mapping', *Ore Geology Reviews*, Vol. 136, No. 12, p.104252.

Wang, H.Y., Zhao, J.P., Zheng, C.H. and Su, Y.S. (2022) 'scCNC: a method based on capsule network for clustering scRNA-seq data', *Bioinformatics*, Vol. 38, No. 15, pp.3703–3709.

Wu, L., Li, Z. and Abourizk, S. (2022) 'Automating common data integration for improved data-driven decision-support system in industrial construction', *Journal of Computing in Civil Engineering*, Vol. 36, No. 2, pp.1–17.

Zhao, Y. and Li, L. (2022) 'Multimodal data integration via mediation analysis with high-dimensional exposures and mediators', *Human Brain Mapping*, Vol. 43, No. 8, pp.2519–2533.