Integrating Co-Clustering and Interpretable Machine Learning for the Prediction of Intravenous Immunoglobulin Resistance in Kawasaki Disease

HAOLIN WANG [ID] [1], **ZHILIN HUANG**[2], **DANFENG ZHANG**[2], **JOHAN ARIEF**[2], **TIEWEI LYU**[2], **AND JIE TIAN**[2]
[1]College of Medical Informatics, Chongqing Medical University, Chongqing 400016, China
[2]Department of Cardiology, Heart Centre, Children's Hospital of Chongqing Medical University, Ministry of Education Key Laboratory of Child Development and Disorders, National Center for Clinical Medicine Research in Children's Health and Disease, Chongqing 400014, China

Corresponding author: Jie Tian (jietian@cqmu.edu.cn)

ABSTRACT Identifying intravenous immunoglobulin-resistant patients is essential for the prompt and optimal treatment of Kawasaki disease, suggesting the need for effective risk assessment tools. Data-driven approaches have the potential to identify the high-risk individuals by capturing the complex patterns of real-world data. To enable clinically applicable prediction of intravenous immunoglobulin resistance addressing the incompleteness of clinical data and the lack of interpretability of machine learning models, a multi-stage method is developed by integrating data missing pattern mining and intelligible models. First, co-clustering is adopted to characterize the block-wise data missing patterns by simultaneously grouping the clinical features and patients to enable (a) group-based feature selection and missing data imputation and (b) patient subgroup-specific predictive models considering the availability of data. Second, feature selection is performed using the group Lasso to uncover group-specific risk factors. Third, the Explainable Boosting Machine, which is an interpretable learning method based on generalized additive models, is applied for the prediction of each patient subgroup. The experiments using real-world Electronic Health Records demonstrate the superior performance of the proposed framework for predictive modeling compared with a set of benchmark methods. This study highlights the integration of co-clustering and supervised learning methods for incomplete clinical data mining, and promotes data-driven approaches to investigate predictors and effective algorithms for decision making in healthcare.

INDEX TERMS Co-clustering, interpretable machine learning, medical informatics, predictive models, Kawasaki disease.

I. INTRODUCTION

Kawasaki disease (KD) is a nonspecific systemic vascular inflammation in infancy and early childhood. KD may lead to serious coronary complications and has become the leading cause of acquired cardiac disease in children worldwide with increased incidence in recent years. Many aspects of the etiology and pathophysiology of KD remain unknown. Approximately 25% of patients with KD develop

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal [ID].

coronary artery aneurysms without treatment, and coronary artery aneurysms from KD account for 5% of acute coronary syndromes in adults under 40 years of age [1]. Therefore, prompt diagnosis and treatment are essential for favorable clinical outcomes. In clinical practice, the timely initiation of treatment with intravenous immunoglobulin (IVIG) and aspirin can reduce the risk of coronary artery aneurysms significantly (to less than 5%) [2]. However, about 15%–20% of patients with KD are refractory to IVIG therapy and have an elevated risk of coronary artery lesion [3]. Existing studies show that risk-tailored initial therapy can improve

the outcome of patients with KD. Therefore, the early prediction of IVIG-resistance can have an important role in supporting the decision making of healthcare professionals, suggesting the need for effective risk assessment tools.

Mining Electronic Health Records (EHRs), which capture clinical data relating to all aspects of patient care (such as diagnosis, medication, and laboratory test), can facilitate cohort-wide investigations and discovery of clinical evidence on an unprecedented scale [4]. Learning from data provides potentially important opportunities for identifying underlying patterns to develop predictive models for improving healthcare outcomes with individualized diagnosis, prognosis, and administration of treatments [5]. Data-driven approaches attempt to utilize retrospectively collected EHRs data and computational methods to identify patients with IVIG-resistance, and enable prompt and optimal treatment for patients with KD. Recently, several clinical studies have investigated the clinical and laboratory factors to predict KD patients with IVIG-resistance who require further therapy. The utility of six models by using different risk factors for the prediction of IVIG-resistance are compared with a cohort of 504 patients with KD [6]. The presented model with the best performance has an AUC of 0.80, sensitivity of 0.72, and specificity of 0.75. Tan *et al.* have identified eight independent risk factors and developed a prediction model by using multivariate regression analysis [7]. The presented predictive model shows an AUC of 0.74, sensitivity of 0.76, and specificity of 0.59. Another study has performed multivariate logistic regression analyses to predict the lack of response to IVIG and concluded that several factors, which are associated with IVIG-resistance, are not good indicators for accurate prediction in terms of AUC score [8]. Meta-analysis was performed to identify predictive factors of resistance to IVIG to address the conflicting results reported in relevant publications [3], [9]. In addition to statistical techniques, machine learning-based approaches are developed for risk assessment tools. Takeuchi *et al.* have collected a dataset of 767 patients with KD, including 170 who are refractory to initial IVIG therapy and used the random forest classifier to identify IVIG-resistance in patients with KD [10].

Most of the existing studies are insightful but suffer from small sample size and inconsistent potential risk factors. The lack of sensitive risk factors and effective methods limits the performance of these studies in accurately predicting potential patients with IVIG-resistance. Our paper addresses the two following challenges for clinical data mining to support medical decision making, and investigate effective data-driven approaches to improve prediction performance.

First, the incompleteness of clinical data presents a major challenge for predictive modeling with sufficient patient cohort and results in potential biases for estimation. A common problem in retrospective studies is the large amount of data missing from patient subsets. In clinical practice, physicians usually issue the minimum amount of laboratory testing and diagnostics to treat patients effectively.

Patient ID	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
1	13			4.14	321.6
2	19			4.75	204.5
3	23			3.76	118.5
4	19	0	2.1	4.12	135.5
5	29	0	2.5	4.71	165.8
6	20	0	2.5	3.56	157.4
7	24	0	2.5		
8	40	0	2.3		
9	21	2	2.5		
10	19	1	2.9		

FIGURE 1. Illustration of the clinical dataset with block-wise missing data.

Complex and individualized health conditions lead to a variety of personal health records. Tests that are not measured may suggest that the patient appears to be healthy. Thus, the missingness in EHRs can provide insights into the unobserved values that indicate the patient's health state. Common strategies to address the missing data such as patient selection reduce sample size, and simple data imputation may introduce potential biases. Most of the missing data imputation methods deal with data missing at random. However, missing values are usually associated with different clinical assessment measures, which are missing not at random. The groups of features from similar sources demonstrate similar block-wise gaps in data [11]. For example, an illustration of clinical dataset is shown in Figure 1. A clinical assessment measure consisting of feature 2 and 3 are missing for patients 1–3, and the results of a laboratory testing consisting of feature 4 and 5 are missing for patients 7–10. Commonly used missing data imputation and machine learning methods are inefficient in dealing with such block-wise missing data patterns, suggesting the need to develop methods considering the incompleteness of clinical data. For instance, assessing the model validity only on observed data entries is less sensitive to imputation uncertainty and effective in dealing with missingness for multi-view learning [12].

Second, advanced machine learning models have demonstrated huge success in various tasks by capturing complex patterns of real-world data and exhibited satisfactory performance in terms of accuracy. However, the predictions of these models are considered risky and not actionable for clinical applications due to the lack of interpretability or explainability. Understanding the reasons behind prediction is fundamental to provide applicable and actionable decision support. To address this limitation, there has been an increased interest in explainable machine learning, which aims to design inherently interpretable models to provide explanations of the risk factors for high descriptive accuracy without degrading the predictive accuracy. Considering the modeling stage of data science life cycle, interpretability can be defined as model-based and post hoc interpretability [13]. For instance, Letham *et al.* have developed a predictive model in the form

of sparse decision statements. This model is interpretable by human experts to estimate the risk of stroke in patients with atrial fibrillation [14]. Lundberg and Lee have presented a unified framework for interpretable prediction by assigning each feature an importance score for a certain prediction [15]. Both model-based and individualized interpretability can be addressed using interpretable learning models.

We present a multi-stage approach to identify homogeneous clusters considering the availability of clinical data and fit localized prediction models for each group. The proposed method performs co-clustering to identify data blocks to characterize the block-wise data missing patterns. Uncovering both feature and patient subgroups can support group-based feature selection using the group Lasso and the group-specific predictive models to take advantage of the patterns of individuals that may differ from the entire population groups. We integrate the Explainable Boosting Machine (EBM) to provide the contributing factors for the predictive model and explain why a certain prediction is made for a patient. We validate the proposed method using a cohort of 2796 patients extracted from the EHRs, and demonstrate the superior performance of the proposed method with the advantages of interpretability and flexibility.

The remainder of this paper is organized as follows. Section 2 briefly introduces related works for the feasible integration of co-clustering and interpretable machine learning. Section 3 presents the proposed framework to address incompleteness and interpretability for clinical data mining. Section 4 evaluates the performance of the proposed method by comparison with several benchmark models using a real-world EHRs dataset. The last section concludes this paper.

II. RELATED WORKS
For complex real-world data, a single model may not be sufficiently efficient in capturing the distributions and underlying structure in the data. Multiple classifier systems have demonstrated advantages for increasing accuracy over individual classifier models [16]. An effective approach to generate multiple classifiers is the divide-and-conquer strategy, which divides data samples into homogenous subgroups and learning localized models for each group [17]. For instance, clustering models can be performed to learn the underlying group structure, and classification models can utilize the obtained information to train the classification rules [18]. The interleaving of unsupervised clustering and supervised classification models can improve the performance of learning from complex data with better interpretability and reliability. This paper follows a sequential learning manner for clustering and classification. Different from previous studies, our partitioning based on co-clustering aims to characterize the data missing patterns of the dataset through grouping clinical features and patients simultaneously. In addition to identifying patient subgroups, the structural information obtained by co-clustering can enable localized feature selection for feature subsets [19].

Co-clustering is an extension of cluster analysis to detect coherent patterns in multidimensional data [20]. Co-clustering is known as bi-clustering for matrix, in which group row and column clusters simultaneously identify homogeneous data blocks via appropriate permutations [21]. Widely used algorithms for co-clustering consist of spectral, model-based, factorization-based, and information-theoretic-based methods. Many of these algorithms seek to discover a diagonal structure. In our case, the number of row and column clusters can be different. Thus, nondiagonal co-clustering methods are suitable. The information-theoretic-based co-clustering method has become very popular due to its speed of convergence and scalability. Mutual information is suitable for co-clustering, which uses rows and column information symmetrically [22]. In general, the co-clustering algorithms can be implemented using alternating iterative procedures by fixing one of the partitions (such as row partition) and searching for an optimal partition of the other set (such as column partition) [23].

Although advanced machine learning methods, such as deep learning and ensemble models, can provide accurate predictions, the inherent complexity obfuscates the interpretation and limits the use of these models for clinical decision support. Understanding why a model makes a specific prediction provides insight and engenders appropriate trust in predictions. Interpretable machine learning has emerged as a key aspect of data-driven medical decision making. For instance, a model is developed to represent the feature importance of the risk prediction of hypoxaemia during surgery [24]. Two case studies are presented for pneumonia risk and hospital readmission predictions by using Generalized Additive Models (GAMs) with pairwise interactions [25]. GAMs can be interpreted easily with single variables due to its additive property and can provide flexible functions to model the linear and the nonlinear relationships among individual predictors to uncover latent patterns. GAMs usually provide more accurate predictions than commonly used machine learning models by capturing the nonlinear patterns in data. EBM is an implementation of the GA^2M algorithm, which has high accuracy and intelligibility [26]–[28]. EBM is adopted for the prediction of group-specific IVIG-resistance to capture nonlinear patterns and pairwise interactions enhanced by ensemble learning.

III. METHODS
For the EHRs datasets, the clinical features can be naturally organized into groups associated with different clinical assessment measures. The available set of diagnostic and laboratory testing records issued by medical experts can reveal inherent similarities and health conditions of patient subpopulation. Co-clustering is a well-suited approach to group clinical features and patients into subgroups and make the complex and incomplete data easy to handle and interpret. An indicator matrix of $n \times d$, which represented d-dimensional features of n patients, was introduced in the clinical dataset to characterize the data missing patterns.
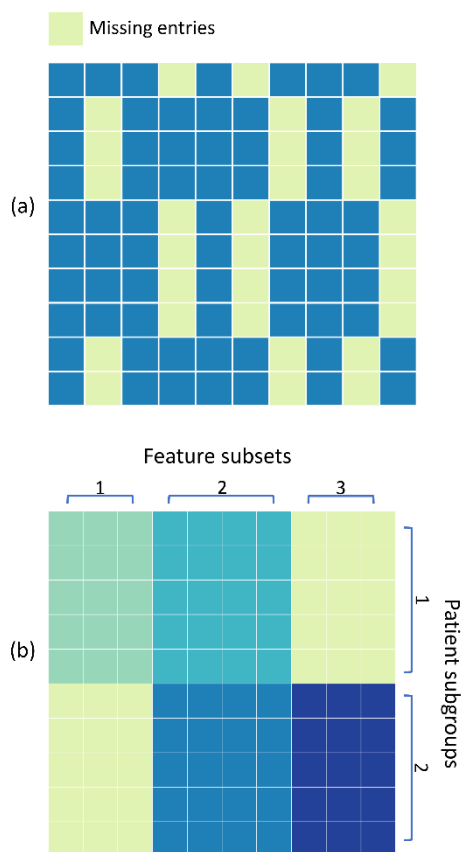
FIGURE 2. (a) Indicator matrix of the clinical dataset. (b) Reorganized matrix with row and column clusters.

FIGURE 3. Proposed prediction workflow.

The entry M_{ij} in the indicator matrix was set to 0 or 1 for data missing or not, respectively. Constant co-clusters can be identified for blocks via co-clustering methods. An illustrative example is shown in Figure 2. The x-axis represented the clinical features, whereas the y-axis represented the patients. The observed and missing elements were labeled using distinct colors. Based on the availability of clinical features, the co-clustering of the indicator matrix of patients with KD dataset can organize features and patients simultaneously into clusters. Figure 2 (b) shows three feature (column) clusters and two patient (row) subgroups.

This paper followed a sequential learning manner in integrating co-clustering for prediction. The proposed diagram is presented in Figure 3. The applied co-clustering enabled group-based feature selection and patient subgroup-specific predictive models. For instance, the illustrative example in Figure 2 shows that feature groups 1 and 2 should be selected for patient subgroup 1, and feature groups 2 and 3 should be selected for patient subgroup 2. Two predictive models can be trained for patient subgroups considering the availability of clinical data.

Feature selection plays an essential role to increase computational efficiency, prevent overfitting and improve learning performance. Given the existence of the group structure of
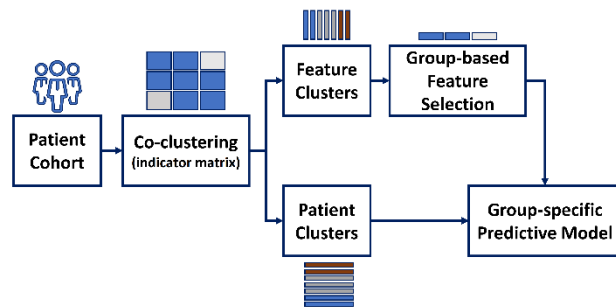
features associated with various clinical assessment measures, we tend to select or not select features in the same feature group simultaneously. Commonly used methods such as filter methods based on mutual information and knockoff framework for feature selection usually ignore the group feature structures. In contrast, group Lasso which assumed that the features formed k disjoint groups is adopted in this study. It performed regularization on the model parameters corresponding to the groups of coefficients [29].

We train a localized predictive model for each patient subgroup to increase prediction accuracy and take advantage of the patient group structure. The adopted EBM is designed to be explainable and intelligible, and demonstrates high accuracy comparable with state-of-the-art machine learning methods. The improvements of EBM over traditional GAMs include the integration of ensemble learning techniques and the automatic detection of pairwise feature interactions in the form of [28]:

$$g(E[y]) = \beta_0 + \sum f_j(x_j) + \sum f_{ij}(x_i, x_j)$$

This model can be effectively learned using gradient boosting with shallow tree-like ensembles [26], [27].

In summary, the proposed method performed co-clustering to characterize the availability of clinical data, group Lasso for group-based feature selection, and EBM for group-specific prediction in a sequential manner.

IV. EXPERIMENTS

The dataset comprised 2796 patients admitted from 2007 to 2016 in the Children's Hospital of Chongqing Medical University in a de-identified format. The main diagnosis of the enrolled patients in their discharge records was KD. The patients who received IVIG treatment before admission and those who did not receive IVIG treatment during hospitalization were excluded. The collected structured EHRs record of the patients consisted of 227 features, including their demographic information (i.e., age and gender), symptom (e.g., days of fever), image diagnosis (e.g., degree of coronary artery lesions), and laboratory test (e.g., platelet count). However, around 41% of the values were missing in this dataset. Unlike conventional approaches discarding features and samples with high missing rate to identify several significant risk factors for prediction, this paper aimed to develop methods
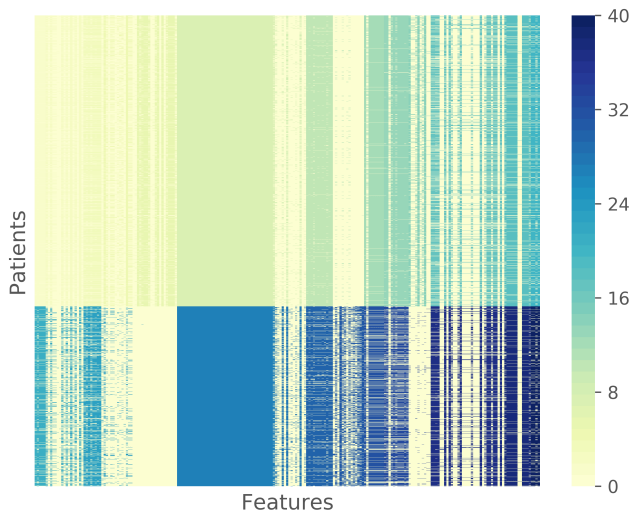
FIGURE 4. Reorganization of patients with KD dataset by using co-clustering to identify feature groups and patient subgroups.

FIGURE 5. ROC curves for the prediction of IVIG-resistance enhanced by the proposed framework.

FIGURE 6. (a) Importance (contribution) score of risk factors for the prediction model. (b) Risk score contributed by factors for individualized prediction.

to take advantage of the incomplete clinical data, obtain an informative and accurate information to improve prediction accuracy. IVIG-resistance is defined as persistent or recurrence of fever at any time during 48 hours to 2 weeks after the initial IVIG treatment. A total of 184 patients presented IVIG-resistance, whereas the rest were IVIG-responsive, resulting in imbalanced data distribution, which is a well-posed problem in the study of medical or healthcare events. Given this issue, the classifier usually leans to be partial toward the major class, leading to poor accuracy for the minority class [30]. Over-sampling was performed using SMOTE [31] to increase the prevalence of IVIG-resistance to 15%. Data balancing aims to obtain the typical incidence rate reported in existing studies [3] and improve the capability of the predictive models. It is worth noting that data balancing is feasible for validating predictive models in our paper but may introduce biases in causality studies [32]. The final dataset consisted of 3017 patients with 459 IVIG-resistance samples.

We investigate the data-driven approaches for the prediction of IVIG-resistance of patients with KD by using the collected dataset. First, we implemented several benchmark models, including regression models (logistic regression, Lasso, and Ridge regression), popular machine learning models (K-nearest neighbors (KNN), multinomial naive Bayes (MNB), and multiple layer perceptron (MLP)) and ensemble methods (random forest (RF), lightGBM (GBM) [33], XGboost (XGB) [34], and EBM). Multivariate imputation was performed for missing value imputation [35]. Second, we presented the study of those baseline methods enhanced by the proposed co-clustering-based framework.

For the proposed method, the CROINFO algorithm was performed for the co-clustering of the indicator matrix of the collected dataset [23]. The co-clusters of the collected dataset are shown in Figure 4. The x-axis represented the 227 clinical features, whereas the y-axis represented the
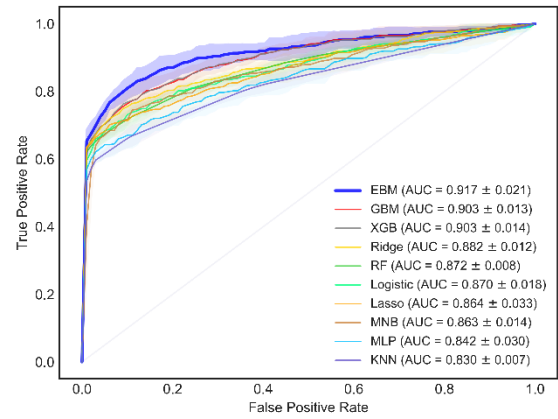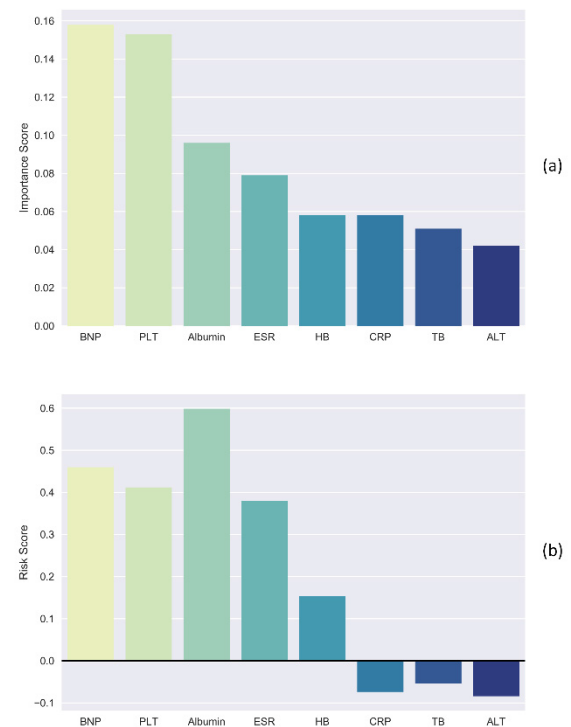
3017 patients. The number of row and column clusters were determined by cross-validation of the training set to optimize the overall classification performance in terms of the AUC score. The presented study identified 20 feature (column) clusters and 2 patient (row) clusters, resulting in 40 co-clusters (blocks labeled as 1-40) with missing entries (labeled as 0). These co-clusters characterized the different data missing patterns of the dataset. Considering the group-based data missing patterns associated with clinical assessment measures, the group Lasso was performed for feature selection for each patient subgroup. EBM-based prediction

TABLE 1. Experimental results (mean ± standard deviation) for the prediction of IVIG-resistance.

	METHOD	AUC	AP	PRECISION	RECALL	F1
BASELINE METHODS	Lasso	0.848±0.027	0.763+0.024	0.888+0.049	0.658+0.037	0.754+0.016
	Logistic	0.853±0.029	0.759+0.033	0.930+0.052	0.643+0.034	0.758+0.012
	Ridge	0.860±0.029	0.778+0.028	0.955+0.019	0.638+0.029	0.765+0.017
	KNN	0.820±0.012	0.673+0.014	0.944+0.015	0.551+0.012	0.696+0.009
	MNB	0.849±0.017	0.703+0.019	0.817+0.051	0.621+0.020	0.704+0.013
	MLP	0.836±0.024	0.680+0.039	0.813+0.040	0.601+0.044	0.689+0.021
	RF	0.855±0.009	0.739+0.007	0.986+0.019	0.606+0.010	0.750+0.006
	GBM	0.884±0.013	0.787+0.007	0.938+0.035	0.634+0.014	0.756+0.008
	XGB	0.882±0.011	0.789+0.011	0.936+0.048	0.647+0.042	0.763+0.018
	EBM	0.878±0.006	0.786+0.009	0.948+0.051	0.640+0.041	0.762+0.013
METHODS ENHANCED BY THE PROPOSED FRAMEWORK	Lasso	0.864±0.033	0.763+0.020	0.915+0.038	0.651+0.036	0.759+0.012
	Logistic	0.870±0.018	0.757+0.017	0.906+0.063	0.632+0.046	0.741+0.016
	Ridge	0.882±0.012	0.780+0.023	0.945+0.013	0.634+0.020	0.759+0.016
	KNN	0.830±0.007	0.677+0.012	0.845+0.055	0.580+0.022	0.686+0.006
	MNB	0.863±0.014	0.690+0.023	0.845+0.053	0.645+0.017	0.730+0.010
	MLP	0.842±0.030	0.723+0.017	0.914+0.061	0.599+0.034	0.722+0.015
	RF	0.872±0.008	0.756+0.013	0.983+0.018	0.610+0.013	0.753+0.005
	GBM	0.903±0.013	0.808+0.015	0.916+0.082	0.653+0.055	0.757+0.007
	XGB	0.903±0.014	0.809+0.014	0.932+0.068	0.647+0.041	0.760+0.009
	EBM	**0.917**±0.021	**0.835**+0.022	0.928+0.075	0.669+0.051	**0.773**+0.021

models were then learned for two patient subgroups for the prediction of IVIG-resistance.

The stratified five-fold cross-validation was performed to evaluate the performance of the benchmark models and the proposed method. The AUC that considered all possible thresholds for binary classification was used as a standard metric to quantify the performance. In terms of the AUC scores, all methods presented satisfactory performance for prediction (AUC > 0.82) as shown in Table 1. For those baseline methods, the ensemble learning methods including GBM, XGB and EBM outperformed other methods in providing accurate prediction for IVIG- resistance as expected since they demonstrate state-of-the-art performance in many prediction tasks [28]. Despite the difference in performance is not significant, this study focuses on EBM for its interpretability and effectiveness for prediction. The integration of the proposed framework can improve the performance for all those methods consistently under the same settings.

The mean receiver operating characteristic (ROC) curves with one standard deviation of the evaluated methods enhanced by the proposed framework are shown in Figure 5. The proposed method integrating co-clustering and EBM demonstrated superior performance by presenting the ROC curve at the top left corner with the highest AUC score (AUC = 0.917), followed by GBM and XGB.

The performance was further evaluated using average precision score (AP), precision, recall, and F1 score. The average score with standard deviation is presented in Table 1. The AP (micro) summarized the precisions achieved at each threshold. In terms of the AP and F1 score, the proposed method demonstrated best performance. The computation of precision and recall depends on the threshold for classification such that they are not suitable for comparison separately. The reported precision and recall used threshold that maximized the F1 score for each model. Unlike the AUC score, the integration of the proposed framework may not improve the precision/recall/F1 score of each base classifier. This is partly because the performance is essentially improved by training multiple localized classifiers supported by co-clustering, but the single threshold used for binary classification to evaluate the overall performance may not optimal for each classifier.

In addition to prediction accuracy, interpretability is an important aspect for the data-driven models to support clinical decision making. The trained model for prediction is highly interpretable by keeping the contribution of each feature additive. For illustration, the contribution of several clinical features for the prediction of IVIG-resistance can be sorted in accordance with their importance, as shown in Figure 6 (a). The top-ranked risk factors identified by the EBM model in our experiments included brain natriuretic peptide (BNP), platelet count (PLT), albumin, erythrocyte

sedimentation rate (ESR), hemoglobin (HB), C-reactive protein (CRP), total bilirubin (TB), and alanine aminotransferase (ALT), which were consistent with the risk factors reported in published studies [3], [7].

Figure 6 (b) illustrates a correct prediction for a patient with IVIG-resistance to support individual level risk prediction with a list of clinical features that are most salient to each prediction. Five risk factors supported IVIG-resistance, whereas three other factors lowered the risk for this patient. The interpretation and understanding of the prediction model and individualized factors may be helpful for the decision making of healthcare professionals. Supplementary materials related to this study are available online (https://github.com/wanghaolin/coclustering_clinical_prediction).

V. CONCLUSION

We investigate data-driven learning approaches for the prediction of IVIG-resistance by using retrospectively collected EHRs, and present a multi-stage approach integrating co-clustering and EBM. The proposed method results in interpretable models considering the incompleteness of EHRs data with improved predictive performance. We demonstrate that co-clustering is feasible for exploiting the block-wise missing patterns of clinical data by grouping patients and features simultaneously, and can be integrated with supervised learning algorithms to enhance the capability to capture the underlying structure of complex real-world data. Despite the improved performance for predictive modeling, there are several limitations can be addressed in future studies. First, the proposed method should be further validated using multi-center datasets and evaluated prospectively. Second, generating multiple classifiers using different subsets of the dataset can increase the flexibility of the method, leading to the reduced information used for the training of each classifier. Thus, a sufficient amount of data is required to ensure the effectiveness of the prediction model. Third, joint optimization of the co-clustering and classification have the potential to further improve the performance. The adoption of interpretable machine learning models can offer opportunities for identifying high-risk patients at an early stage to enable prompt treatment and obtaining novel insights from EHRs.

REFERENCES

[1] B. W. McCrindle, A. H. Rowley, J. W. Newburger, J. C. Burns, A. F. Bolger, M. Gewitz, A. L. Baker, M. A. Jackson, M. Takahashi, P. B. Shah, T. Kobayashi, M.-H. Wu, T. T. Saji, and E. Pahl, "Diagnosis, treatment, and long-term management of Kawasaki disease: A scientific statement for health professionals from the American Heart Association," *Circulation*, vol. 135, no. 17, pp. e927-e999, 2017, doi: 10.1161/cir.0000000000000484.

[2] J. W. Newburger, M. Takahashi, and J. C. Burns, "Kawasaki disease," *J. Amer. College Cardiol.*, vol. 67, no. 14, pp. 1738–1749, 2016, doi: 10.1016/j.jacc.2015.12.073.

[3] X. Li, Y. Chen, Y. Tang, Y. Ding, Q. Xu, L. Sun, W. Qian, G. Qian, L. Qin, and H. Lv, "Predictors of intravenous immunoglobulin-resistant kawasaki disease in children: A meta-analysis of 4442 cases," *Eur. J. Pediatrics*, vol. 177, no. 8, pp. 1279–1292, Aug. 2018, doi: 10.1007/s00431-018-3182-2.

[4] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, pp. 395–402, May 2012, doi: 10.1038/nrg3208.

[5] G. S. Collins and K. G. M. Moons, "Reporting of artificial intelligence prediction models," *Lancet*, vol. 393, no. 10181, pp. 1577–1579, 2019, doi: 10.1016/S0140-6736(19)30037-6.

[6] W. Qian, Y. Tang, W. Yan, L. Sun, and H. Lv, "A comparison of efficacy of six prediction models for intravenous immunoglobulin resistance in kawasaki disease," *Italian J. Pediatrics*, vol. 44, no. 1, p. 33, Dec. 2018, doi: 10.1186/s13052-018-0475-z.

[7] X.-H. Tan, X.-W. Zhang, X.-Y. Wang, X.-Q. He, C. Fan, T.-W. Lyu, and J. Tian, "A new model for predicting intravenous immunoglobin-resistant kawasaki disease in chongqing: A retrospective study on 5277 patients," *Sci. Rep.*, vol. 9, no. 1, p. 1722, Dec. 2019, doi: 10.1038/s41598-019-39330-y.

[8] M. K. Kim, M. S. Song, and G. B. Kim, "Factors predicting resistance to intravenous immunoglobulin treatment and coronary artery lesion in patients with Kawasaki disease: Analysis of the Korean nationwide multicenter survey from 2012 to 2014," *Korean Circulat. J.*, vol. 48, no. 1, pp. 71–79, 2018.

[9] J.-Y. Baek and M. S. Song, "Meta-analysis of factors predicting resistance to intravenous immunoglobulin treatment in patients with kawasaki disease," *Korean J. Pediatrics*, vol. 59, no. 2, p. 80, 2016.

[10] M. Takeuchi, R. Inuzuka, T. Hayashi, T. Shindo, Y. Hirata, N. Shimizu, J. Inatomi, Y. Yokoyama, Y. Namai, Y. Oda, M. Takamizawa, J. Kagawa, Y. Harita, and A. Oka, "Novel risk assessment tool for immunoglobulin resistance in kawasaki disease: Application using a random forest classifier," *Pediatric Infectious Disease J.*, vol. 36, no. 9, pp. 821–826, Sep. 2017, doi: 10.1097/inf.0000000000001621.

[11] K. Srinivasan, C. Foe-Parker, N. Goebel, R. Herzl, M. R. Mehl, B. Gilligan, J. Heerwagen, K. Kampschroer, K. Canada, F. Currim, S. Ram, C. Lindberg, E. Sternberg, P. Skeath, B. Najafi, J. Razjouyan, and H.-K. Lee, "Feature importance and predictive modeling for multi-source healthcare data with missing values," in *Proc. 6th Int. Conf. Digit. Health Conf. (DH)*, 2016, pp. 47–54.

[12] G. Chao, J. Sun, J. Lu, A.-L. Wang, D. D. Langleben, C.-S. Li, and J. Bi, "Multi-view cluster analysis with incomplete data to understand treatment effects," *Inf. Sci.*, vol. 494, pp. 278–293, Aug. 2019, doi: 10.1016/j.ins.2019.04.039.

[13] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.

[14] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model," *Ann. Appl. Statist.*, vol. 9, no. 3, pp. 1350–1371, Sep. 2015.

[15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.

[16] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "Dynamic classifier selection: Recent advances and perspectives," *Inf. Fusion*, vol. 41, pp. 195–216, May 2018.

[17] M. Deodhar and J. Ghosh, "SCOAL: A framework for simultaneous coclustering and learning from complex data," *ACM Trans. Knowl. Discovery Data*, vol. 4, no. 3, pp. 1–31, Oct. 2010.

[18] W. Cai, S. Chen, and D. Zhang, "A simultaneous learning framework for clustering and classification," *Pattern Recognit.*, vol. 42, no. 7, pp. 1248–1259, Jul. 2009, doi: 10.1016/j.patcog.2008.11.029.

[19] Y. Li, M. Dong, and J. Hua, "Simultaneous localized feature selection and model detection for Gaussian mixtures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 953–960, May 2009.

[20] H. Yan, "Coclustering of multidimensional big data: A useful tool for genomic, financial, and other data analysis," *IEEE Syst., Man, Cybern. Mag.*, vol. 3, no. 2, pp. 23–30, Apr. 2017, doi: 10.1109/MSMC.2017.2664214.

[21] F. Role, S. Morbieu, and M. Nadif, "CoClust: A Python package for co-clustering," *J. Stat. Softw.*, vol. 88, no. 7, p. 29, 2019, doi: 10.18637/jss.v088.i07.

[22] G. Govaert and M. Nadif, "Mutual information, phi-squared and model-based co-clustering for contingency tables," *Adv. Data Anal. Classification*, vol. 12, no. 3, pp. 455–488, Sep. 2018.

[23] G. Govaert and M. Nadif, *Co-Clustering: Models, Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 2013.

[24] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, and S.-I. Lee, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomed. Eng.*, vol. 2, no. 10, pp. 749–760, Oct. 2018, doi: 10.1038/s41551-018-0304-0.

[25] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for HealthCare: Predicting pneumonia risk and hospital 30-day readmission," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2015, pp. 1721–1730.

[26] Y. Lou, R. Caruana, and J. Gehrke, "Intelligible models for classification and regression," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2012, pp. 150–158.

[27] Y. Lou, R. Caruana, J. Gehrke, and G. Hooker, "Accurate intelligible models with pairwise interactions," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2013, pp. 623–631.

[28] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A unified framework for machine learning interpretability," 2019, *arXiv:1909.09223*. [Online]. Available: http://arxiv.org/abs/1909.09223

[29] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," in *Data Classification: Algorithms and Applications*. New York, NY, USA: Chapman & Hall, 2014, pp. 37–64, doi: 10.1201/b17320.

[30] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surveys*, vol. 52, no. 4, pp. 1–36, Sep. 2019, doi: 10.1145/3343440.

[31] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[32] J.-J. Beunza, E. Puertas, E. García-Ovejero, G. Villalba, E. Condes, G. Koleva, C. Hurtado, and M. F. Landecho, "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," *J. Biomed. Informat.*, vol. 97, Sep. 2019, Art. no. 103257, doi: 10.1016/j.jbi.2019.103257.

[33] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.

[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[35] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: What is it and how does it work?" *Int. J. Methods Psychiatric Res.*, vol. 20, no. 1, pp. 40–49, Mar. 2011, doi: 10.1002/mpr.329.
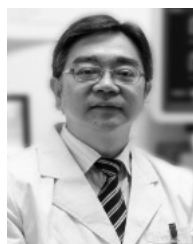
DANFENG ZHANG is currently pursuing the Ph.D. degree with the Children's Hospital of Chongqing Medical University, Chongqing, China. Her research interests include coronary artery lesion and intravenous immunoglobulin resistance of Kawasaki disease.

JOHAN ARIEF received the M.B.B.S. degree in clinical medicine from Chongqing Medical University, in 2017. He is currently pursuing the master's degree in pediatrics with the Children's Hospital of Chongqing Medical University, China. His research interests include congenital heart disease and childhood autoimmune disorders.

TIEWEI LYU received the M.D. degree from Chongqing Medical University. He is currently a Professor with Chongqing Medical University. His research interests include basic and clinical research of cardiovascular diseases in children.

HAOLIN WANG received the B.S. degree in electronics and information engineering from Beihang University, in 2010, the M.S. degree in software engineering from the University of Science and Technology of China, in 2013, the Ph.D. degree in computer applied technology from the University of Chinese Academy of Sciences, and the joint Ph.D. degree from the City University of Hong Kong, in 2018. He is currently a Postdoctoral Researcher with the College of Medical Informatics, Chongqing Medical University. His research interests include medical data mining and intelligent systems.

ZHILIN HUANG is currently pursuing the Ph.D. degree with the Children's Hospital of Chongqing Medical University, Chongqing, China. His research interests include congenital heart disease and diagnosis of Kawasaki disease.

JIE TIAN received the M.D. degree from Chongqing Medical University. From 1994 to 1996, he has worked with the Affiliated Research Centers and Hospitals, Cornell University. He was with Oxford University, in 2001. He is currently a Professor with Chongqing Medical University. He has published over 190 research articles. His research interests include cardiovascular and pediatric diseases. He serves on the editorial board of several medical journals.

• • •