Dosage Prediction in Pediatric Medication Leveraging Prescription Big Data

MIAOLIAN WU1, LINGHONG HONG2, YONGGEN ZHAO1, LONGBIAO CHEN3, AND JUE WANG1

1Children’s Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China
2Xiang’an Hospital of Xiamen University, Xiamen 361005, China
3Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China

Corresponding author: Miaolian Wu (chawml@zju.edu.cn)

This work was supported in part by the Natural Science Foundation of Zhejiang Province under Grant LSY19H300001, in part by the National Natural Science Foundation of China under Grant NSFC-61802325, in part by the Natural Science Foundation of Fujian Province under Grant 2018J01105, and in part by the China Fundamental Research Funds for the Central Universities under Grant 20720170040.

ABSTRACT Rational use of medicines is of great importance in pediatric clinical medication. As the pharmacokinetics and pharmacodynamics of the pediatric group is highly dynamic, it is a great challenge to determine the rational dosage for pediatric patients. Traditional clinical decision support systems for dosage guidance largely rely on manual collection of medication information, which usually suffers from incomplete and missing evidences for the pediatric group. In this paper, we propose a data-driven approach to accurately predict pediatric medication dosages by leveraging prescription big data. More specifically, we first identify two relevant factors of pediatric medication dosage, i.e., the physiology factors including patients’ body weight and age group, and the indication factors that affect clinical dosage patterns. We then extract the corresponding physiology and indication features, and propose a hybrid-learning-based method to adaptively integrate the two sets of heterogeneous features into a model for accurate pediatric dosage prediction. We evaluate our method on real-world prescription datasets from two tertiary children’s hospitals. Results show that our method predicts pediatric medication dosages with an accuracy above 81.3%, and consistently outperforms other baselines.

INDEX TERMS Dosage prediction, pediatric medication, prescription big data, rational use of medicines, data mining.

I. INTRODUCTION

Irrational use of medicines is a major problem worldwide. World Health Organization (WHO) estimates that more than 50% of all medicines are prescribed, dispensed or sold inappropriately [1]. The overuse, underuse or misuse of medicines not only waste medical resources, but also lead to significant patient harm in terms of medication errors (ME) and adverse drug events (ADE) [2]. Therefore, WHO is committed to promote rational use of medicines for clinical physicians and pharmacists, so as to ensure that “patients receive the appropriate medicines, in doses that meet their own individual requirements, for an adequate period of time, and at the lowest cost both to them and the community” [2]. Children are at higher risk of irrational use of medicines than adults are [3]. It has been recognized in clinical pediatrics that there are continuous biological and psychological development from birth through childhood and the adolescent years, which brings impact on the absorption, distribution, metabolism and excretion processes of medications, and consequently affects the pharmacology response in the pediatric groups [4].

One of the key challenges in pediatric medication is to calculate the rational dosage. Compared to adult medications, the rational dosage ranges for the pediatric group are more complicated and may vary significantly among individuals [3], [5], [6]. To address this issue, various Clinical Decision Support (CDS) systems have been developed and deployed in hospitals to facilitate the calculation and validation of rational medication dosages [7]. However, when applying CDS systems to pediatric hospitals, several issues arise. First, the dosage calculation of CDS systems are based on pre-compiled medical information databases, which contain the medicine description, usage, cautions, etc., collected from medicine instructions, pharmacopoeia,
clincial practice [8]. Due to the lack of dedicated medication instructions and standards for children, such pediatric medication information is usually incomplete or missing in these databases [5]. Second, the rules employed to guide rational medication dosage for adults are usually very simple, e.g., 1 capsule 3 times daily, which do not take into consideration the dosage adjustment based on children’s age, body weight, or indication information [9]. Consequently, the dosage suggestions of CDS systems deployed in pediatric hospitals tend to be error prone [9], [10], e.g., triggering clinically-insignificant false alarms or failing to detect irrational medication dosages.

With the emergence of medical big data [11], rich clinical information about pediatric medication can be accessed from Hospital Information Systems (HIS), providing us with new opportunities to address the above-mentioned issues. In this paper, we propose a data-driven approach to predict rational pediatric medication dosages by exploiting large-scale electronic prescription data. Instead of building pediatric medication databases and designing artificial dosage calculation rules, we exploit the clinical pediatric medication dosing experiences of physicians and pharmacists from historical prescriptions to train an adaptive model for rational dosage prediction. More specifically, the contributions of this work are:

- To the best of our knowledge, this is the first work on data-driven pediatric medication dosage prediction. By exploiting the knowledge in prescription big data, we are able to give accurate predictions of rational dosages, and reduce dosage errors in clinical medication practice.
- We propose a data-driven pediatric medication dosage prediction framework based on hybrid learning. First, we identify two relevant factors of pediatric medication dosages from prescription data: the physiology factors characterized by age group and body weight, and the indication factor denoted by diagnosis and indication. Then, we extract the corresponding physiology and indication features, and propose a hybrid learning model to adaptively integrate these heterogeneous factors. Finally, we exploit the rational dosages from historical prescriptions to train the model for accurate dosage prediction.
- We evaluate the proposed framework with real-world prescription big data collected from two children’s hospitals. Results show that our method predicts pediatric medication dosages with an accuracy above 81.3%, and consistently outperforms other baselines.

The rest of this paper is organized as follows. We first survey the related work in Section II, and then present the overview of the proposed framework in Section III. In Section IV and V, we detail the dosage relevant factor identification and feature extraction. In Section VI, we propose the hybrid learning model. We report the evaluation settings and results in Section VII, and conclude the work in Section VIII.

II. RELATED WORK
In this section, we briefly survey the existing literature from the following two perspectives: (1) rational use of medicine in pediatric groups, and (2) medical big data and the data mining technologies.

A. RATIONAL USE OF PEDIATRIC MEDICINE
Rational use of medicines has become an important topic in pediatric clinical medication practice [3]. Because of the complication in calculating right pediatric dosages, children are at higher risk of medication errors and adverse drug events than adults are [3]. In the United States, 5% to 27% of all pediatric medication orders result in a medication error, leading to 7,000 patient deaths annually from medication errors [12]. Specifically, [13] reported that the potential adverse drug event risks were three times higher in newborns than in other age groups, where dosage error was the most usual cause of ADEs.

In order to regulate rational use of pediatric medicines, various medication-related clinical decision support systems have been proposed by researchers and applied in pediatric hospitals [14]. A simple implementation of CDS system is to offer the physicians a complete list of dosing guidelines and parameters of each medication for reference [7]. A more intelligent approach for dosing is to provide specific recommendations for dose and frequency to physicians according to the patient’s specific conditions [7]. A complete evaluation of clinical decision support systems for medication administration can be found in [15].

In the kernel of these CDS systems are a variety of dosing rules, which are compiled from existing medicime information documentation and databases. However, due to the nature that pediatric medication information is constantly changing because of ongoing research and clinical experiences, it is difficult to ensure the correctness and timeliness of the dosing rules, leading to potential risks for pediatric medication safety [15], [16]. For example, it is reported in [17] that almost three-quarters of CDS alerts were overridden in clinical practice, and 40% of the overrides were not appropriate. To address these issues, in this paper, we propose to facilitate dosing recommendation directly from historical medication data instead of using medication information and rules.

B. MEDICAL BIG DATA ANALYTICS
The concept of big data is usually characterized by 4 Vs, i.e., volume, variety, velocity, and veracity [18]. Big data includes large-scale medical, environmental, financial, geographic, and social media information and continues to grow [19]. In the medical and healthcare field, large volumes of data are generated, which present a broad prospect in clinical applications [11], [20]. Medical big data come from various sources, such as hospital information systems, electronic medical record systems, laboratory information system, medical imaging and biomarker systems [11]. Medical big data analytics aims to exploit the knowledge in data to provide
predictive modeling and clinical decision support, disease or safety surveillance for public healthcare, and etc. [19]. The big data technologies employed in medical big data analytics involve various data mining methods, such as classification, clustering, and regression [18].

The experiences of clinicians and the advantages of big data should be integrated to provide reliable diagnosis and conclusions [19]. For example, Armstrong et al. [21] proposed a neural network-based data mining model to analyze the characteristic indexes of microcalcifications in mammograms. The model was able to accurately predicted microcalcifications in early images of patients. In rational use of medicines, the National Library of Medicine has launched Pillbox, an initiative to provide data and images for prescription, over-the-counter, homeopathic, and veterinary oral solid dosage medications (pills) marketed in the United States. Pillbox contains information of a variety of medicines including color, shape, dosage, and interactions with other medicines. A similar dataset is VigiBase by the World Health Organization, which aims to provide knowledge about safe use of medicines and assist therapeutic decisions in clinical practice. Based on the dataset, Suzuki et al. [22] discovered that concomitant medicines can change the occurrence frequency of drug-induced liver damage.

However, to the best of our knowledge, there are still few research work in prescription big data analytics for pediatric diagnosis and medication. One of the relevant ideas to our work was illustrated in [23], where the authors leverage prescription data analytics and artificial intelligence technologies to enable the accurate diagnoses of pediatric diseases. In this work, we leverage large-scale prescription datasets to exploit the knowledge of medication in clinical practices, and to provide rational medication dosage for pediatric groups.

III. FRAMEWORK OVERVIEW

We propose a pediatric dosage prediction framework, as illustrated in Fig. 1. At the learning stage, by conducting empirical studies on the historical prescription big data, we first identify two relevant factors of pediatric medication dosages, i.e., the physiology factors characterized by age group and body weight, and the indication factor denoted by diagnosis and indication. We then extract the corresponding sets of features from the prescription dataset, and propose an hybrid learning model to adaptively integrate the two sets of heterogeneous features. Finally, we exploit the rational dosages from historical prescription data to train a model for each medicine in the dataset. At the prediction stage, given a new prescription, we employ the trained hybrid model to predict the rational dosage for each medicine listed in the prescription, and provide dosage suggestions to physicians and pharmacists. We elaborate the details of the key components in the following sections.

IV. DOSAGE RELEVANT FACTOR IDENTIFICATION

Pediatric medication dosage calculation is more complicated than that of adults [5]. In clinical practice, most pediatric medicines are dosed according to the patient’s body weight (mg/kg) or body surface area (mg/m$^2$) [5]. Moreover, dosages also vary by the patient’s symptom and indication, therefore diagnostic information is helpful when calculating dosages. Based upon this prior knowledge, we first identify two relevant factors of pediatric medication dosages, and then extract the corresponding features from prescription data via correlation analysis.

More specifically, we exploit an anonymized dataset with one month of pediatric prescriptions from a children’s hospital in Zhejiang Province, China as an example. Based on a series of correlation analysis, the following two factors are identified: the physiology factors characterized by age group and body weight, and the indication factor denoted by diagnosis and indication. We elaborate the details as follows.

A. PHYSIOLOGY FACTORS

Physiology metrics of pediatric patients, such as age group, body weight, and body surface area, are usually the most important considerations in clinical pediatric dosage calculation [5]. Although body surface area provides the greatest accuracy in calculating pediatric dosages [3], it requires specialized nomograms and is usually time consuming and troublesome to calculate [5]. Therefore, we propose to estimate the baseline dosage from the age group and body weight indices, which are widely available in prescription datasets.

To this end, we conduct a correlation analysis of the rational dosage against the age group and body weight, respectively. More specifically, for a medicine $m$ used in the dataset, we extract its dosages vector $y^{(m)}$ and the corresponding age group vector $x_{ag}^{(m)}$ and body weight vector $x_{bw}^{(m)}$ in the prescription. We then calculate the Pearson correlation coefficient [24] of dosage and age group

$$
\rho_{y^{(m)}, x_{ag}^{(m)}} = \frac{\text{cov}(y^{(m)}, x_{ag}^{(m)})}{\sigma_{y^{(m)}} \sigma_{x_{ag}^{(m)}}}
$$

as well as the correlation of dosage and body weight

$$
\rho_{y^{(m)}, x_{bw}^{(m)}} = \frac{\text{cov}(y^{(m)}, x_{bw}^{(m)})}{\sigma_{y^{(m)}} \sigma_{x_{bw}^{(m)}}}
$$

where $\text{cov}$ is the covariance and $\sigma$ is the standard deviation of the vectors.

As a result, we obtain a mean $\rho_{y^{(m)}, x_{ag}^{(m)}} = 0.902$ (std = 0.057) and a mean $\rho_{y^{(m)}, x_{bw}^{(m)}} = 0.884$ (std = 0.066), respectively. These high correlation indices indicate that the rational pediatric dosages are highly correlated to the patients’ age group and body weight. For example, Fig. 2 shows the correlation matrix of Ibuprofen Suspension, one commonly used pediatric medicine for treating pain, fever, and inflammation. We can observe that the major pediatric age group
M. Wu et al.: Dosage Prediction in Pediatric Medication Leveraging Prescription Big Data

**FIGURE 1.** An overview of the proposed framework.

**FIGURE 2.** The correlation matrix between the dosage of Ibuprofen Suspension and the physiology features (age group and body weight) in the sample pediatric prescription dataset. The color bar on the right denotes the rational dosage measured in ml. The specification of Ibuprofen Suspension is 30ml:0.6g.

is 0–5 years with body weight range of 5–15 kg. The dosage amount increases quasi-linearly as age and weight grows.

**B. INDICATION FACTOR**

Indication is another important relevant factor of pediatric medication dosages [25]. In clinical practice, physicians may choose different dosage levels according to different indications for a same medication. For example, when Intravenous immunoglobulin (IVIg) is used to prevent and treat immunoglobulin deficiency, its regular dosage range is usually 200–400 mg/kg. However, when IVIg is used in the treatment of mucocutaneous lymphnode syndrome (i.e., Kawasaki disease), its regular dosage can be up to 2,000 mg/kg. Without considering the factor of indications, incorrect dosing of medicine might introduce risk for pediatric group and incur medication errors or adverse drug events.

Therefore, we conduct a correlation analysis on the dataset to find out whether indications are in fact associated with dosage patterns by calculating their Cramér’s V [26]. More specifically, for a medicine $m$ used in the dataset, we enumerate its indications in a categorical variable $I(m)$, and its dosage scales in a categorical variable $D(m)$. For example, we obtain 8 unique indication categories and 14 unique dosage categories for Ibuprofen Suspension. We then construct an indication-dosage contingency table $T(m)$ by counting the occurrences of each indication and dosage category combination in the prescription dataset. For example, Fig. 3 shows the visualization of the contingency table of Ibuprofen Suspension, where an indication category corresponds to a specific dosage pattern (denoted by a row).
Based on the contingency table $T^{(m)}$, we calculate its Pearson’s chi-squared test statistic $\chi^2$ [26], and then calculate its Cramér’s $V$ by

$$\phi_i = \sqrt{\frac{\chi^2}{N(k-1)}}$$

(3)

where $N$ is the number of indication-dosage category combinations, and $k$ is the lesser number of categories of indication and dosage variable. It is proven that $\phi_i \in [0, 1]$ and a large $\phi_i$ indicates a strong association between indication and dosage. As a result, we obtain a mean $\phi_i = 0.372$ (std $= 0.183$) for all the medicines, indicating that indications and dosages are associated, and therefore could be used to predict dosage patterns.

V. FEATURE EXTRACTION

With the above-identified dosage relevant factors, we extract the corresponding features from the historical training dataset to build a prediction model. More specifically, for each medicine $m$, the following two categories of features are extracted.

Physiology Features: for each medication record containing medicine $m$, we extract two key physiology variables: the patient’s age group $x_a^{(m)}$ measured in year, and the patient’s body weight $x_w^{(m)}$ measured in kg, and build its physiology feature vector $\mathbf{x}_p^{(m)} = [x_a^{(m)}, x_w^{(m)}]$.

Indication Features: for medicine $m$, we first build its indication category $I^{(m)}$ by enumerating its corresponding indications from the historical training dataset. Then, for each medication record containing medicine $m$, we extract its indication feature $\mathbf{x}_i^{(m)}$ using the one-hot encoding of the primary and secondary indications based on $I^{(m)}$.

VI. THE HYBRID LEARNING MODEL

In this step, our objective is to predict the rational pediatric medication dosage based on the extracted features. One of the intuitive methods is to concatenate the physiology and indication features into a vector, and build a regression or classification model to predict the rational dosage. However, due to the considerable variety of the two categories of features, such a direct concatenation of the two heterogeneous features does not perform well, especially when some features play a dominate role in specific medication conditions [27].

To address these challenges, we propose an hybrid learning model to adaptively aggregate these features for accurate dosage prediction. More specifically, we first build an individual predictor for each category of feature, and then train an aggregator to adaptively aggregate the outputs of the predictors for accurate dosage predictions. We elaborate the details as follows.

A. THE PREDICTORS

First, for each medicine $m$, we extract its dosage category $D^{(m)}$ by enumerating its corresponding dosages from the historical training dataset. Then, we build the following predictors for the two categories of features, respectively.

Physiology Predictor: since the physiology features demonstrate quasi-linear correlation with the dosages, we train a decision tree-based linear classifier $\psi_p^{(m)}$, which accepts the physiology feature $\mathbf{x}_p^{(m)}$ as input and outputs the probability estimation of the dosage categories $\mathbf{p}_p^{(m)}$.

$$\mathbf{p}_p^{(m)} = \psi_p^{(m)}(\mathbf{x}_p^{(m)})$$

(4)

Indication Predictor: we exploit a linear mapping function to predict the probabilities of dosage categories given specific indications. More specifically, for each medication record containing medicine $m$, we estimate the dosage category probability distribution $\mathbf{p}_i^{(m)}$ based on the indication-dosage contingency table $T^{(m)}$,

$$\mathbf{p}_i^{(m)} = \sigma(T^{(m)} \mathbf{x}_i^{(m)})$$

(5)

where $\sigma$ is the softmax function [28] to convert the contingency counts to probability estimations in $[0, 1]$.

B. THE AGGREGATOR

Finally, we need to aggregate the probability estimation of the above-mentioned predictors to an accurate dosage prediction. To this end, we first combine the probability estimations into a hybrid feature vector, i.e.,

$$\mathbf{P}^{(m)} = [\mathbf{p}_p^{(m)}, \mathbf{p}_i^{(m)}]$$

(6)

We then train an Artificial Neural Network (ANN) classification model $M^{(m)}$ to aggregate the probability feature, i.e.,

$$\hat{y}^{(m)} = M^{(m)}(\mathbf{P}^{(m)})$$

(7)

The objective of this hybrid learning architecture is to learn an optimal mapping between the hybrid probability estimation feature and the ground-truth dosage category. In this way, the two categories of factors relevant to medication dosages are adaptively integrated for an accurate dosage prediction.

VII. EVALUATION

We evaluate our method with real-world, anonymized prescription big data collected from two pediatric hospitals. We first introduce the basic information about the datasets used, and then elaborate the experiment settings. Finally, we present the results of pediatric dosage prediction.

A. DATASET DESCRIPTION

We collected two anonymized datasets from two tertiary children’s hospitals in Zhejiang Province, China and Fujian Province, China, respectively. We deployed the datasets on Apache Spark [29] for distributed data storage and processing. We use the Python machine learning library scikit-learn [30] for data analytics and model training. The frameworks and algorithms were run on a cluster consisting of three high-performance servers.

We note that it is crucial to ensure the validity of the dataset, otherwise the wrong data samples in the training set may
lead to wrong learning models and bring risks to medication dosage decisions. Therefore, we conducted a training data cleansing process to remove prescriptions with wrong medication dosages from the following two aspects. First, we deliberately removed the wrong prescriptions identified by existing clinical decision support systems (already deployed in the two tertiary hospitals) and confirmed by pharmacists from the datasets. Second, we have removed the wrong samples concerning dosage errors marked by the prescription review program, which are operated in the two tertiary hospitals on a monthly basis, respectively. After data cleansing, we obtain 144,603 prescriptions in dataset ZHEJIANG and 1,665,420 prescriptions in FUJIAN, respectively. The summary of these data are shown in Table 1.

### B. EXPERIMENT SETTINGS

**Evaluation Plan:** we first group prescriptions by medicines, and extract the corresponding features and dosage categories from these prescriptions. For each medicine, we build a separate hybrid model to predict the rational dosage category. We randomly select 70% of the prescriptions for training, and the left 30% for evaluation.

**Evaluation Metrics:** since the dosage category prediction is a multi-class classification problem, we evaluate the performance of our models with prediction accuracy scores. More specifically, for each medicine $m$, we organized the classification results into a confusion matrix [31] $C_{i,j}^{(m)}$, where each row of the matrix represents the instances in a predicted dosage category and each column represents the instances in a ground truth dosage category. Each element $C_{i,j}^{(m)}$ counts the number of medication record that are predicted as in dosage category $i$ while the actual dosage is in category $j$. With the confusion matrix, we define the accuracy score $S^{(m)}$ as the fraction of correct predictions over all the samples, i.e.

$$S^{(m)} = \frac{1}{\sum \sum C_{i,j}^{(m)}} \sum C_{i,j}^{(m)}$$ (8)

where $N_d^{(m)}$ denotes the number of dosage categories.

**Baseline Methods:** we design the following baselines to compare with our method on the same dataset.

- **PHY:** this baseline only exploits the physiology features to predict the dosage category using a decision tree-based linear classification model.
- **IND:** this baseline only exploits the indication features to predict the dosage category with a linear mapping function similar to the proposed method.

### C. RESULTS

Table 2 shows the average accuracy scores of dosage prediction accuracy using the proposed method as well as the baselines. We can see that in both datasets, the proposed HYBRID method achieves best performance with regard to prediction accuracy scores. More specifically, the baseline method PHY only considers age and weight features, which ignores the dosage adjustment based on indications, and thus does not achieve significant prediction accuracy. The IND baseline attempts to predict dosage categories with only indication features while ignoring the physiology differences of patients, and thus fails to achieve consistent prediction results. The ANN baseline directly combines the two categories of features for dosage modeling, and the results show that the prediction accuracy is improved. In the proposed HYBRID method, we take one step forward by integrating the probability estimation instead of directly combining the heterogeneous features, and therefore significantly improve the prediction accuracy in both datasets.

### D. CASE STUDY

We conduct a case study on Ibuprofen Suspension, one of the most popular pediatric medication in both ZHEJIANG and FUJIAN datasets. In the ZHEJIANG dataset, the age and weight distributions of the corresponding patients are shown in Fig. 2. We can see different dosage categories (denoted by different colors) under different combinations of age groups and body weights. More specifically, we present the detailed prediction accuracy scores with regard to different age groups using the proposed HYBRID method in Fig. 4. The age groups

### TABLE 1. Summary of datasets from two hospitals.

|                      | ZHEJIANG  | FUJIAN    |
|----------------------|-----------|-----------|
| Data collection period | 2016/06–2016/07 | 2010/01–2010/07 |
| # Prescriptions      | 144,603   | 1,665,420 |
| # Medicines          | 376       | 460       |
| # Patients           | 69,199    | 263,935   |

### TABLE 2. Accuracy scores of dosage prediction with different methods.

| Methods | ZHEJIANG | FUJIAN |
|---------|----------|--------|
| PHY     | 0.535    | 0.508  |
| IND     | 0.237    | 0.217  |
| ANN     | 0.648    | 0.634  |
| HYBRID (proposed) | 0.847 | 0.813 |
are defined as new born (birth to 1 month of age), infant (1 month to 2 years of age), child (2 to 12 years of age), and adolescent (12-18 years of age). We can see that the prediction accuracy of the new born group is relatively higher than other age groups. The possible reason is that the age range of the new born group is quite narrow (only one month), and this group tends to be more sensitive to overdosaging risks. Therefore, the dosing patterns are usually quite moderate and easy to predict (normally 3ml). We also observe that in other age groups, the prediction accuracy increases as the patients’ age grow, which indicates that the dosage patterns are more clear among elder pediatric groups.

The indication-dosage mapping scheme of Ibuprofen Suspension is illustrated in Fig. 3, where specific indication shows specific dosage pattern. More specifically, we present the detailed indications of Ibuprofen Suspension as well as their corresponding ICD-10 codes [33] in Table 3. We can see that Ibuprofen Suspension is used in the treatment of various inflammation indications, since it can ease pain, swelling, and fever. By considering different dosage patterns of different indications, our method is able to give more accurate dosage prediction results.

**TABLE 3. The relevant indications of Ibuprofen Suspension.**

| No | Indication       | ICD-10 Code |
|----|-----------------|-------------|
| 1  | herpangina      | B08         |
| 2  | acute pharyngitis | J02        |
| 3  | acute tonsillitis | J03        |
| 4  | suppurative tonsillitis | J03      |
| 5  | pneumonia        | J12         |
| 6  | acute tracheobronchitis | J20    |
| 7  | allergic rhinitis | J30         |
| 8  | acute nasal sinusitis | J32     |

**FIGURE 4.** The prediction accuracy scores of Ibuprofen Suspension with regard to different age groups in the ZHEJIANG dataset.

In the future, we plan to extend our work in the following directions. First, we plan to involve more data sources from other hospital information systems, especially data from clinical laboratories, to investigate more relevant factors of pediatric medication dosages. Second, we plan to investigate the reasons for wrong dosage predictions, including overdosing or underdosing by model overfitting, and then leverage the knowledge to improve our predictive models. Third, we plan to integrate our method with the existing clinical decision support systems to provide dosing recommendation for physicists and pharmacists in clinical practice.

**ACKNOWLEDGMENT**

We would like to thank the reviewers for their constructive suggestions.

**REFERENCES**

[1] The Pursuit of Responsible Use of Medicines: Sharing and Learning from Country Experiences, World Health Org., Geneva, Switzerland, 2012.

[2] Promoting Rational Use of Medicines: Core Components, WHO Regional Office for South-East Asia, 2011.

[3] R. Kelishadi and F. Mousavinasab, “Rational use of medicine in the pediatric age group: A summary on the role of clinical pharmacists,” J. Res. Pharmacy Pract., vol. 1, no. 1, pp. 10–13, 2012.

[4] K. Williams, D. Thomson, I. Seto, D. G. Contopoulou-Ioannidis, J. P. Ioannidis, S. Curtis, E. Constantin, G. Batamanbane, L. Hartling, and T. Klassen, “Standard 6: Age groups for pediatric trials,” Pediatrics, vol. 129, pp. S153–S160, Jun. 2012.

[5] A. M. M. Frey, “Pediatric dosage calculations,” J. Infusion Nursing, vol. 8, no. 5, pp. 373–379, 1985.

[6] S. S. Ranganathan and M. Gazarian, “Rational use of medicines (RUM) for children in the developing world: Current status, key challenges and potential solutions,” in Optimizing Treatment for Children in the Developing World, Cham, Switzerland: Adis, 2015, pp. 225–238.
[7] G. J. Kuperman, A. Bobb, T. H. Payne, A. J. Avery, T. K. Gandhi, G. Burns, D. C. Classen, and D. W. Bates, “Medication-related clinical decision support in computerized provider order entry systems: A review,” J. Amer. Med. Inform. Assoc., vol. 14, no. 1, pp. 29–40, 2007.

[8] C. Yang, L. Yang, X. Xiang, Y. Tang, H. Wang, N. Bobai, and X. Zhang, “Interventions assessment of prescription automatic screening system in chinese hospitals: A systematic review,” Drug Inf. J., vol. 46, no. 6, pp. 669–676, Nov. 2012.

[9] A. L. Potts, F. E. Barr, D. F. Gregory, L. Wright, and N. R. Patel, “Computerized physician order entry and medication errors in a pediatric critical care unit,” Pediatics, vol. 113, no. 1, pp. 59–63, Jan. 2004.

[10] C. J. Mullett, R. S. Evans, J. C. Christenson, and J. M. Dean, “Development and impact of a computerized pediatric antifibrotic decision support program,” Pediatics, vol. 108, no. 4, e75, Oct. 2001.

[11] C. H. Lee and H.-J. Yoon, “Medical big data: Promise and challenges,” Kidney Res. Clin. Pract., vol. 36, no. 1, pp. 3–11, Mar. 2017.

[12] M. L. Rinke, D. G. Bundy, C. A. Velasquez, S. Rao, Y. Zerhouni, K. Lobner, C. H. Lee and H.-J. Yoon, “Medical big data: Promise and challenges,” in World Health Organization. (2016). [Online]. Available: http://www.who.int/classifications/icd/en/