A Data-Driven Context-Aware Health Inference System for Children during School Closures

ZHIHAN JIANG, Department of Electrical and Electronic Engineering, The University of Hong Kong, China
LIN LIN, Department of Mechanical Engineering, The University of Hong Kong, China
XINCHEN ZHANG, Department of Electrical and Electronic Engineering, The University of Hong Kong, China
JIANDUO LUAN, Department of Statistics and Actuarial Science, The University of Hong Kong, China
RUNNING ZHAO, Department of Electrical and Electronic Engineering, The University of Hong Kong, China
LONGBIAO CHEN, School of Informatics, Xiamen University, China
JAMES LAM, Department of Mechanical Engineering, The University of Hong Kong, China
KA-MAN YIP, HUNG-KWAN SO, WILFRED H.S. WONG, and PATRICK IP*, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, China
EDITH C.H. NGAI*, Department of Electrical and Electronic Engineering, The University of Hong Kong, China

Many countries have implemented school closures due to the outbreak of the COVID-19 pandemic, which has inevitably affected children's physical and mental health. It is vital for parents to pay special attention to their children's health status during school closures. However, it is difficult for parents to recognize the changes in their children's health, especially without visible symptoms, such as psychosocial functioning in mental health. Moreover, healthcare resources and understanding of the health and societal impact of COVID-19 are quite limited during the pandemic. Against this background, we collected real-world datasets from 1,172 children in Hong Kong during four time periods under different pandemic and school closure conditions from September 2019 to January 2022. Based on these data, we first perform exploratory data analysis to explore the impact of school closures on six health indicators, including physical activity intensity, physical functioning, self-rated health, psychosocial functioning, resilience, and connectedness. We further study the correlation between children's

*Co-corresponding authors.

Authors' addresses: Zhihan Jiang, zhjiang@connect.hku.hk, Department of Electrical and Electronic Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China; Lin Lin, linlin00wa@gmail.com, Department of Mechanical Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China; Xinchen Zhang, u3008407@connect.hku.hk, Department of Electrical and Electronic Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China; Jianduo Luan, u3591503@connect.hku.hk, Department of Statistics and Actuarial Science, The University of Hong Kong, Faculty of Science, Hong Kong, China; Running Zhao, rnzhao@connect.hku.hk, Department of Electrical and Electronic Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China; Longbiao Chen, longbiaochen@xmu.edu.cn, School of Informatics, Xiamen University, Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen, China; James Lam, james.lam@hku.hk, Department of Mechanical Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China; Ka-Man Yip, kmanyip@hku.hk; Hung-Kwan So, hkso@hku.hk; Wilfred H.S. Wong, whswong@hku.hk; Patrick Ip, patricip@hku.hk, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, LKS Faculty of Medicine, Hong Kong, China; Edith C.H. Ngai, chngai@eee.hku.hk, Department of Electrical and Electronic Engineering, The University of Hong Kong, Faculty of Engineering, Hong Kong, China.

2474-9567/2023/3-ART18 $15.00
https://doi.org/10.1145/3580800

contextual characteristics (i.e., demographics, socioeconomic status, electronic device usage patterns, financial satisfaction, academic performance, sleep pattern, exercise habits, and dietary patterns) and the six health indicators. Subsequently, a health inference system is designed and developed to infer children's health status based on their contextual features to derive the risk factors of the six health indicators. The evaluation and case studies on real-world datasets show that this health inference system can help parents and authorities better understand key factors correlated with children's health status during school closures.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: data analysis, school closures, health inference, risk factor analysis

ACM Reference Format:
Zhihan Jiang, Lin Lin, Xinchen Zhang, Jianduo Luan, Running Zhao, Longbiao Chen, James Lam, Ka-Man Yip, Hung-Kwan So, Wilfred H.S. Wong, Patrick Ip, and Edith C.H. Ngai. 2023. A Data-Driven Context-Aware Health Inference System for Children during School Closures. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 18 (March 2023), 27 pages. https://doi.org/10.1145/3580800

1 INTRODUCTION

It is reported that, since March 2020, more than 90% of schools have experienced temporary closure to slow the transmission of disease and protect children (5-17 years of age) due to COVID-19 [4]. School closures would inevitably have an impact on children's physical and mental health because schools are critical settings for physical activity and essential sources of mental health services for children [74]. For example, school closures increase Body Mass Index (BMI) and childhood obesity prevalence [3] and make children irritable and rebellious [13]. Anxiety, depression, and stress are more prevalent during school closures [68]. Meanwhile, physical and mental health in childhood may track into the adult [55]. Therefore, it is important for parents to pay special attention to children during school closures and have a comprehensive understanding of the short-term and long-term effects of school closures on children's physical and mental health.

However, in practice, parents often fail to recognize the changes in their children's health, particularly their mental health, until visible symptoms appear [60]. Besides, regular professional counseling, which can help parents keep abreast of their children's health status, is unaffordable for many families from developing countries without sound healthcare systems [45]. Moreover, many countries have witnessed shortages of healthcare resources during the COVID-19 pandemic [13]. Although the development of wearable devices such as smartwatches makes it feasible to track children's physical activity, sleep patterns, and many other indices of physical health [62], they cannot track mental health and are still unaffordable for low-income families [42].

On the other hand, in the literature, researchers have proposed numerous works studying the relationships between children's contextual characteristics and their physical and mental health. For example, apart from school closures, children's demographics [34], school performance [56], electronic device usage patterns [36], and economic circumstances [67] have significant impacts on children's health. Notably, school closures may have different impacts on children with different contextual characteristics. As studied in [42], children from low economic groups are more vulnerable to school closures. These contexts of children are easily accessible for parents, providing us new opportunities to infer children's health status cost-effectively.

Motivated by the above analysis, this work devotes to building a context-aware health inference system for children during school closures. First, we collect large-scale acceleration data and children's contextual data (i.e., demographics, socioeconomic status, electronic device usage patterns, financial satisfaction, academic performance, sleep patterns, exercise habits, and dietary patterns) using accelerometers and questionnaires from the children in primary and secondary schools in Hong Kong. We also leverage scales that have been widely used in clinical trials and population studies worldwide to measure children's physical and mental health and select six health indicators. Second, we conduct exploratory data analysis to explore the impacts of school closures on the health indicators and the correlation between children's contextual characteristics and the health indicators. Third, we build a context-aware health inference system to infer children's health status and evaluate potential risk factors based on children's contextual features. To this end, we have considered the following issues.

- *Different school closure types.* Due to the different pandemic-control regimes, generally, there are three types of classes: full-day face-to-face (full-day school opening), half-day face-to-face (half-day school class along with half-day online class), and full-day online (full-day school closures). To evaluate the influence of different types of classes, we collect the data during three different learning modes: full-day face-to-face (L0), half-day face-to-face combined with half-day online (L1), and online (L2), and analyze the data based on the three learning modes to evaluate the influence of different types of school closures.
- *Short-term effects and long-term effects.* School closures may have different effects on children in the short and long terms. Analyzing the data collected just before or after school closures cannot reflect long-term effects. Therefore, we collected the data during four time periods, i.e., before the outbreak of COVID-19 (T0: September 2019 - January 2020), full-day school closures due to COVID-19 (T1: March 2020 - April 2020), half-day school reopening when COVID-19 was under control (T2: October 2020 - November 2020), and after long-term half-day face-to-face classes (T3: October 2021 - January 2022). By analyzing the differences among four time periods, we can evaluate the influence of school closures both in the short and long terms.
- *Health status measurement.* To derive indicators for children's physical and mental health, we need to select appropriate measurements. We use accelerometers to measure children's physical activity intensity and use PedsQL [72], EQ5D [29], CD-RISC [18], and RSCS [59], which are scales widely used in clinical trials and population studies worldwide, to measure children's physical functioning, psychosocial functioning, self-rated health, resilience, and connectedness. Based on data collected by accelerometers and scales, we derived six health indicators for children's physical and mental health, i.e., *MVPA*, *PHYF*, *VVAS*, *PSYF*, *RESI*, and *CONN*, and built the system to infer these six indicators based on children's contextual features. These six indicators are elaborated in Section 3.
- *Risk factors analysis.* The relationships between contextual features and health indicators are complicated, and the importance of different features needs to be further evaluated. To analyze the risk factors for the indicators, we first divide the indicators into the at-risk and normal statuses and build inference models to predict the probability of normal status. Then, we use a game-theoretic approach, named SHAP [41], to choose the most important features with the aim of achieving good inference performance. Finally, we use SHAP to explain the influence of the contextual features on the predicted probability of at-risk health based on the simplified optimal models to analyze the risk factors.

Based on the above discussion, the main contributions of this paper are summarized as follows:

- To the best of our knowledge, this is the first data-driven context-aware system to *comprehensively* infer children's health status from the perspectives of physical activity intensity, physical functioning, self-rated health, psychosocial functioning, resilience, and connectedness, and analyze their risk factors during school closures. Such a *low-cost* system can help parents track their children's health status *timely* and provide guidance for authorities to allocate resources *cost-effectively*.
- We evaluate the inference performance and conduct case studies based on *real-world* datasets from children in primary and secondary schools in Hong Kong during four time periods from September 2019 to January 2022. The results show that the system can achieve an average area under receiver operating characteristic curve of 0.7881 in the six indicators.
- Following the exploratory analysis, feature importance analysis, and risk factor analysis, the key findings are as follows:
  - If the pandemic situation mitigates, half-day school closures would be a better choice than full-day school closures. Compared with full-day school closures, half-day school closures have significantly

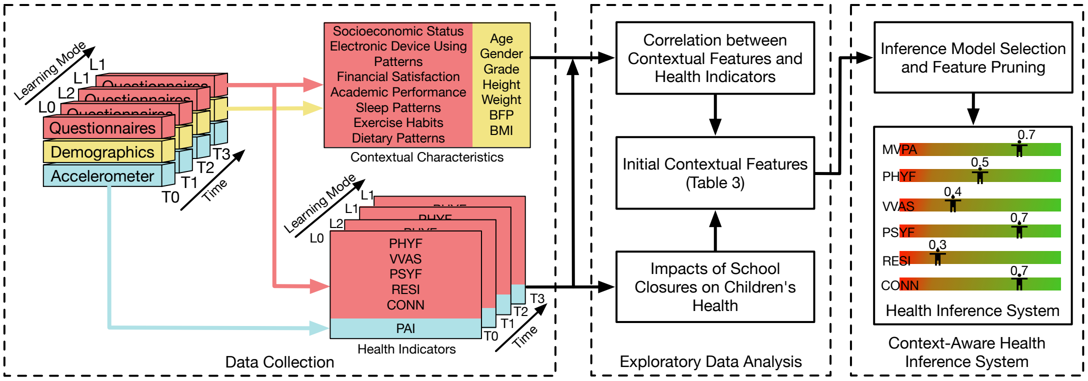


Fig. 1. Overview of the framework.

less impact on children's physical activity intensity and have no significant harm to other indicators, and even increase secondary school students' connectedness.

– The adverse impacts of full-day school closures on physical activity intensity are more likely to persist for female children, while for others, the half-day school reopening improves their physical activity intensity in the short term. In the long term, only secondary school children's physical activity intensity returns to the level before the outbreak of COVID-19.
– Full-day school closures worsen female children's physical functioning in the short term but improve their psychosocial functioning in the long term, and improve secondary school students' self-rated health in the short term.
– Although children from low-income families face more difficulties during school closures, they tend to have higher resilience. The negative influence of objective lower financial conditions can be partly offset by subjective higher financial satisfaction.
– Children's time spent on electronic devices, whether for gaming or learning, should be limited within a reasonable range, and parents should also reduce their own time spent on electronic devices. Higher addiction levels of parents to electronic devices may hinder their children's physical activity intensity, physical functioning, psychosocial functioning, and connectedness.
– Good sleep, frequent exercise, and healthy diets are panaceas. They are positively correlated with the probability of normal health indicators. Since there are parental and peer influences on children's exercise habits, parents should play an exemplary role and develop good exercise habits during school closures.

The remainder of this paper is organized as follows. In Section 2, related works are introduced to reveal the research gap. Section 3 introduces the processes of data collection and health indicator selection. Then, in Section 4, we conduct exploratory data analysis to understand the impacts of school closures on the indicators and obtain the correlation between contextual characteristics and health indicators. Subsequently, in Section 5, an inference system is established and several case studies are conducted. Finally, we discuss the insights, implications, and limitations of our work in Section 6 and conclude our work in Section 7. The framework of this work is shown in Fig. 1.

2 RELATED WORK

In this section, we discuss the existing works on mobile health inference, the health impact of school closures, and the machine learning in healthcare, to explain the research gap.

2.1 Mobile Health Inference

In recent years, researchers have proposed many interesting works on inferring health status by mobile devices [33]. For example, Zhang et al. [79] used mobility patterns and demographic data to predict hospital visits. Tong et al. [69] predicted patients' fatigue and EQ5D index based on sensing data. Li et al. [37] proposed a deep neural network to predict self-reported mood, health, and stress based on sensor data. Zhang et al. [80] quantified the causal effect of urban mobility behaviors on individual health. However, these works require various sensors to collect data, while in many developing countries, wearable devices are unaffordable for low-income families. Particularly, during pandemic periods, the mobility patterns of children are quite different from the usual time. Kao et al. [33] designed a mobile health system to predict health conditions based on the health labels provided by users. However, their system is for adults rather than children. Furthermore, it is worth noticing that healthcare resources are quite limited during pandemic periods. Therefore, it is important to not only infer children's health status but also provide indications on how to improve health. In this work, we build a system to infer children's health based on their contextual features which are easily available. It can further assess the risk factors of six health indicators on an individual level, giving personalized advice on improving children's health.

2.2 Impacts of School Closures on Children's Health

School closures have been implemented in many countries to slow down the transmission of the COVID-19 pandemic [75]. However, schools are important facilities for educating children and supporting their physical and social activities, so school closures would inevitably affect children's physical and mental health [13]. There are numerous works exploring the impact of school closures on children's health. For example, in [3], a microsimulation model was given to simulate the trajectory of a kindergarten cohort's BMI and discovered that school closures increased BMI and childhood obesity prevalence [3]. Many researchers conducted online surveys and found that school closures increased anxiety, frustration, and loneliness in children [13]. Also, a cross-sectional study has been conducted to show preliminary evidence on Long COVID in Children [10]. Although these works revealed the potential impact of school closures on children's health, they failed to help parents directly judge what has changed in their children's health and had no suggestions on how to offset the negative impact. Additionally, most of the studies only considered full-day closures, whereas half-day closures could also be an important strategy during pandemic periods. In this paper, we provide our insights into the impact of full-day and half-day school closures on children's health.

2.3 Machine Learning in Healthcare

Machine learning has been widely employed in healthcare. For example, in [11], depression was detected from social media based on Random Forest. Kaur et al. [35] predicted diseases based on disease-related attributes through Support Vector Machine (SVM), Random Forest (RF), and Multilayer Perception. Ishaque et al. [31] employed various machine learning algorithms to classify stress and relaxation states based on physiological signals and found that Boosting method achieved good performance and generalization. Similarly, Guo et al. [27] compared various algorithms, including SVM, RF, Gradient Boosting Decision Tree [76], and XGBoost (XGB) [15] for classifying teenagers' physical fitness levels, and XGB outperforms other classifiers. XGB stands out from other Boosting algorithms due to its high efficiency, low computational cost, and robustness to overfitting [15, 51]. It requires fewer data for training than deep-learning-based classifiers, especially applicable to the case that healthcare data is hard to obtain. Furthermore, model interpretability is very important in healthcare [61]. In this work, we compare various algorithms to find the one with the best performance to build the inference system.

3 DATA COLLECTION AND DESCRIPTION

In this section, we introduce the participants and the procedures for collecting data, processing data, and selecting health indicators. This study was approved by The Institutional Review Board of the University of Hong Kong / Hospital Authority Hong Kong West Cluster (IRB number: UW 19-516). Informed written consent was obtained from the parents of the participants.

3.1 Participants and Data Collection

The participants are students from seven primary schools and nine secondary schools in Hong Kong. Every participant was given a wristband with accelerometer and advised to wear it during the data collection periods except in a bath or shower. The brand of the wristband is ActiGraph wGT3X-BT[1] which can capture and record continuous, high-resolution physical activity and raw acceleration data. The accelerometer is very sensitive and is accurate within +/- 0.5% of the data collected [1].

There are four data collection periods in this study. The first period (T0) is from September 2019 to January 2020, i.e., before COVID-19, when students attended normal full-day classes. The second period (T1) is from March 2020 to April 2020, when COVID-19 cases increased rapidly in Hong Kong, and the government announced full-day school closures. The third period (T2) is from October 2020 to November 2020 and from April 2021 to June 2021, when students returned to schools on a half-day basis. The fourth period is from October 2021 to January 2022 (T3), when students had attended half-day classes for more than one year. In total, we have received 1,875 sets of accelerometer data. The mean duration is 13.85 days (standard deviation: 6.94). More detailed information can be found in Appendix 1.

Participants' age, gender, height, weight, and body fat percentage were measured and recorded by the researchers when distributing the wristbands. The participants were also asked to fill out a questionnaire, which has two sections: one for students and one for parents. It is composed of totally 16 parts, including questions concerning children's contextual characteristics and widely-used clinical scales measuring children's physical and mental health during their data collection periods. The questionnaire can be found in the supplementary files. We finally received 1,779 valid questionnaires.

3.2 Data Processing

The collected data was sorted into the following three aspects:

- *Accelerometer Data.* We used ActiLife Software 6.13.4 [2] to process the 1,875 sets of acceleration data and generate acceleration counts [3]. Note that although we asked participants to wear the wristband during the data collection periods except in a bath or shower, they might not strictly follow this requirement. Therefore, we further select the records of children who wore the wristbands for more than six hours in at least three days, excluding sleep time [77] out of the data collection periods. Finally, we obtain 1,172 valid sets of accelerometer data for further analysis.
- *Children's Demographics.* We recorded the participants' grades and measured their age, gender, height, weight, and body fat percentage (BFP) when distributing the wristbands, and we calculated the BMI using height and weight.
- *Questionnaire Data.* The subjective data are collected by the questionnaire (see Supplementary files). Since some questions in the questionnaire have relatively lower response rates, we finally determine seven types

[1]https://actigraphcorp.com/actigraph-wgt3x-bt/
[2]https://actigraphcorp.com/support/software/actilife/
[3]https://actigraphcorp.my.site.com/support/s/article/What-are-counts

Table 1. Description of participants' ages, family income, and parents' education levels across four data collection periods.

Time	**All (N=1172)**		**T0 (N=431)**		**T1 (N=107)**		**T2 (N=327)**		**T3 (N=307)**	
Group	**F (N=733)**	**M (N=439)**	**F (N=259)**	**M (N=172)**	**F (N=70)**	**M (N=37)**	**F (N=255)**	**M (N=72)**	**F (N=149)**	**M (N=158)**
Age	12.40±2.22	12.61±2.19	12.38±2.69	12.29±2.40	11.28±2.54	12.04±2.46	12.21±1.74	11.49±2.44	13.29±1.44	13.61±1.15
Income	6.38±2.14	6.23±2.05	6.17±2.26	5.88±1.96	6.87±1.20	6.22±0.94	6.24±2.20	5.36±1.77	6.75±2.11	7.00 ±2.20
Education	4.15±1.06	3.98± 1.03	3.98±1.12	3.83±1.02	4.37±0.57	3.89±0.35	4.13±1.07	3.64±0.83	4.31±1.07	4.31±1.13
Group	**P (N=521)**	**S (N=651)**	**P (N=243)**	**S (N=188)**	**P (N=75)**	**S (N=32)**	**P (N=152)**	**S (N=175)**	**P (N=51)**	**S (N=256)**
Age	10.48±1.30	14.08±1.30	10.32±1.20	14.96±1.09	10.00±0.94	15.16±0.67	10.65±1.58	13.27±1.28	11.45±0.36	13.86±1.03
Income	6.35±2.17	6.30±2.06	6.30±2.33	5.75±1.85	6.76±1.29	6.37±0.69	6.02±2.16	6.07±2.14	6.97±2.32	6.87±2.13
Education	4.18±1.10	3.99±1.00	4.13±1.21	3.64±0.83	4.30±0.58	3.97±0.41	4.15±1.13	3.91±0.95	4.27±1.07	4.31±1.10

The results are reported in the format of $M \pm SD\ (N)$, where M and SD are the mean and standard deviation of age. N is the number of participants. The family income is described by ten levels (1: < 4000, 2: [4000,7999], 3: [8000,11999], 4: [12000,15999], 5: [16000,19999], 6: [20000,29999], 7: [30000,39999], 8: [40000,49999], 9: [50000,79999], 10: ≥ 80000, in HKD). Education represents the parents' education level (the average of father and mother for each family), described by seven levels (1: no formal education, 2: primary school, 3: junior high school, 4: senior high school, 5: higher certificate/diploma, associate, 6: bachelor, 7: postgraduate).

of contextual characteristics with response rates greater than 70% from the questionnaires, namely, socioeconomic status[4], electronic device usage patterns, financial satisfaction, academic performance, sleep patterns, exercise habits, and dietary patterns. Besides, some widely used clinical scales are also included to measure children's health status (detailed in Section 3.3).

Finally, we select 1,172 participants whose acceleration data, demographics, and questionnaire data are all valid. Their ages range from 6 to 18 ($M = 12.48$, $SD = 2.21$). These participants are divided into different independent groups, respectively according to the three learning modes: full-day face-to-face classes (L0), half-day face-to-face classes (L1), and full-day online classes (L2), and four data collection periods: before the outbreak of COVID-19 (T0), full-day school closures due to COVID-19 (T1), half-day school reopening when COVID-19 was under control (T2), and after long-term half-day classes (T3). The descriptions of the ages, family income, and parents' education levels of participants regarding different data collection periods are shown in Table 1.

3.3 Health Indicators Selection

To profile children's physical and mental health, we derived six indicators based on the accelerometer data and the scales widely used in clinical trials and population studies worldwide, detailed as follows. The Cronbach's α [8] is used to measure the reliability of the scales.

- *Physical activity intensity (*PAI*).* Physical activity can be classified into three categories according to the intensity, i.e., light intensity activity (e.g., light walking), moderate intensity activity (e.g., biking on level ground), and vigorous intensity activity (e.g., jogging). Specifically, frequent moderate-to-vigorous intensity physical activity (*MVPA*) has great benefits on children's health [5], and it is also an important indication for children's health [28]. For example, the World Health Organization (WHO) recommended that children aged 5-17 years should do at least an average of 60 minutes per day of *MVPA* [52]. Besides, sedentary behavior (*SB*), which is a sitting or reclining posture requiring very low energy expenditure [22], impairs children's health. In this work, *PAI* is described by the daily average time for *SB* and *MVPA*, which are estimated by acceleration counts. By comparing the results using different cut-points with the recorded physical activity intensity in PE classes, we finally choose the most consistent ones, which are proposed by Mattocks et al. [44] (detailed in Appendix 2). It should be noted that although we use the wristbands and Mattocks et al. [44]'s cut points are hip-worn based, the ActiGraph automatically scales

[4]Socioeconomic status describes the individual position on a social-economic scale that measures factors such as education, income, and type of occupation [40]. In this work, we measure the family income levels, education levels, employment types, and occupation types to describe the family socioeconomic status of children.

the wrist counts [2]. These cut-points are utilized to infer the participants' average daily time of *SB* and *MVPA*.

- *Physical functioning (*PHYF*).* Physical functioning describes the ability to perform basic and instrumental activities in daily life [26]. In this paper, children's physical functioning is measured by the mean score of the Physical Functioning Scale in PedsQL measurement model [72], whose range is [0,4]. A higher score means worse physical functioning. PedsQL provides a modular approach to measure health-related quality of life (HRQOL). It consists of four scales, i.e., physical functioning, emotional functioning, social functioning, and school functioning, answered by children. The Cronbach's α of the four data collection periods (T0-T3) is 0.840, 0.816, 0.795, and 0.813, respectively.
- *Self-rated health (*VVAS*).* The self-rated health is a visual analogue scale from EQ5D [29]. EQ5D is an instrument that evaluates the generic quality of life. It has been widely used to rate respondents' health from five dimensions, i.e., mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and self-rated health, answered by children (a higher score indicates worse health status). *VVAS* is recorded on a vertical visual analogue scale, where the endpoints are labelled 'The best health you can imagine' and 'The worst health you can imagine' [5]. It is an important index of future morbidity and mortality [32]. It is used as a quantitative measure of the health status from children's subjective point of view, whose range is [0,100]. A higher score means better self-evaluated health status. To validate the reliability of the VVAS score from EQ5D, we first get the score in five dimensions from EQ5D (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and then measure the criterion-related validity [12] between the five-dimension score and the VVAS score. The results of Spearman's rank correlation analysis between the five-dimension score and VVAS score of T0-T3 are -0.402, -0.466, -0.397, and -0.385, respectively, with p-value < 0.001. Since a higher five-dimension score indicates worse health, we can find that the VVAS scores are positively correlated with the participants' health status, showing the reliability of VVAS.
- *Psychosocial functioning (*PSYF*).* Psychosocial functioning is the ability to perform daily activities and engage in relationships with others while gratifying others [46]. In this paper, children's psychosocial functioning is measured by emotional functioning, social functioning, and school functioning in PedsQL measurement model, whose range is [0,4]. A higher score means a worse *PSYF*. The Cronbach's α of the four data collection periods (T0-T3) are 0.890, 0.910, 0.875, and 0.895, respectively.
- *Resilience (*RESI*).* Resilience is the ability to withstand adversity and overcome difficulties in life [43]. In this paper, we use the Connor-Davidson Resilience Scale (CD-RISC) [18], which is a 25-item scale and has often been employed to measure children's resilience. The range is [25,125]. A higher score means a higher *RESI*. The Cronbach's α of the four data collection periods (T0-T3) are 0.955, 0.965, 0.954, and 0.948, respectively.
- *Connectedness (*CONN*).* Connectedness measures the children's connection and relationship with others, especially their families and teachers. In this paper, we use the Resnick Social Connectedness Scale (RSCS) [59] to evaluate children's connectedness. RSCS is often tested to explore the connectedness between children and their parents/teachers. *CONN* score is the mean score of the 17 items from RSCS. The range of the score is [1,5]. A higher score means a higher *CONN*. The Cronbach's α of the four data collection periods (T0-T3) are 0.954, 0.956, 0.946, and 0.952, respectively.

4 EXPLORATORY DATA ANALYSIS

After collecting heterogeneous data and deriving the six essential health indicators, in this section, we perform exploratory data analysis to answer two questions:

[5] https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/

(1) How do the six health indicators change across different learning modes (L0-L2) and different data collection periods (T0-T3)? In other words, what are the impacts of school closures on the six indicators?
(2) What is the correlation between contextual features (including demographics, socioeconomic status, electronic device usage patterns, financial satisfaction, academic performance, sleep pattern, exercise habits, and dietary patterns) and the six indicators?

To answer these questions, we leverage the following statistical analysis methods in this section:

- *Kruskal-Wallis H test* is a non-parametric method to test the statistically significant differences on a continuous or ordinal dependent variable between more than two independent groups [54]. In this work, we conduct Kruskal-Wallis H tests to find the significant differences on the health indicators between different learning modes, data collection periods, parents' occupation types, employment types, and marital status.
- *Spearman's rank correlation coefficient (ρ)* measures the strength and direction of association between two variables measured on an ordinal, interval, or ratio scale [78]. Since we have a large sample size, we follow the guideline of [64] that $|\rho| < 0.1$ indicates a negligible correlation, and $0.1 \leq |\rho| < 0.4$ indicates a weak correlation.
- *Mann-Whitney U test* is used to test the statistically significant differences on a continuous or ordinal dependent variable between two independent groups [48]. In this work, we conduct Mann-Whitney U tests to find the significant differences on the health indicators between female and male children and between primary and secondary school students.

All tests were two-sided, where p-value < 0.05 indicates the statistical significance. Specifically, for Kruskal-Wallis H tests, the reported p-values have been adjusted by the Bonferroni correction for multiple comparisons.

4.1 Impacts of School Closures on Health Indicators

We conduct the Kruskal-Wallis H test to determine whether there are statistically significant differences in the time for different *PAI* (*SB* and *MVPA*), the scores of *PHYF*, *VVAS*, *PSYF*, *RESI*, and *CONN* between different learning modes (L0-L2) and different data collection periods (T0-T3). The participants' average time of *SB* and *MVPA* (minutes/day), scores of *PHYF*, *VVAS*, *PSYF*, *RESI*, and *CONN* in regard to different learning modes and data collection periods can be found in Appendix 3.

The results are shown in Table 2, where p-value < 0.05 indicates the statistical significance, and all statistical tests were two-sided. All reported p-values have been adjusted by the Bonferroni correction for multiple tests. The pairwise comparisons between learning modes and between data collection periods and their detailed analysis can be found in Appendix 4. Generally, we find that school closures have no significant influence on children's *RESI* but significantly influence children's *PAI*, *PHYF*, *VVAS*, *PSYF*, and *CONN*.

More specifically, *MVPA* is the most influenced indicator. Considering learning modes, school closures, no matter half-day (L1) or full-day (L2), significantly increase children's sedentary behavior ($p < 0.001$) and decrease their *MVPA* ($p < 0.001$) compared with full-day school opening (L0), but the influence of half-day school closures is significantly less than the full-day ($p < 0.001, = 0.003$). Considering data collection periods, children's *MVPA* significantly decreases while *SB* significantly increases after the full-day school closures (T0→T1, $p < 0.001$). When schools reopen on a half-day basis (T1→T2), the *MVPA* and *SB* of children except females return to the level before COVID-19. However, in the long term, the *MVPA* and *SB* of children, except for secondary school students, fail to return to the level before COVID-19. Based on the above analysis, we can find that the trends of *MVPA* and *SB* are generally opposite. Therefore, we use *MVPA* to represent the *PAI* in the following analysis.

As for other indicators, full-day school closures significantly worsen female children's *PHYF* ($p = 0.002$), but it can be improved after the school reopening (T1→T2, $p = 0.035$). The *VVAS* of the secondary school students

Table 2. Results of Kruskal-Wallis H tests for different groups. All significant results are marked in bold.

Group	Indicator	Mean Rank Score L0	Mean Rank Score L1	Mean Rank Score L2	$\mathcal{X}^2(2)$	p-value	Mean Rank Score T0	Mean Rank Score T1	Mean Rank Score T2	Mean Rank Score T3	$\mathcal{X}^2(3)$	p-value
All	***SB***	**490.19**	**620.19**	**774.80**	**74.298**	**<0.001**	**490.19**	**774.80**	**558.72**	**685.67**	**96.575**	**<0.001**
	MVPA	**683.39**	**547.09**	**429.76**	**66.862**	**<0.001**	**683.39**	**429.76**	**591.12**	**500.19**	**78.289**	**<0.001**
	PHYF	**591.00**	**558.12**	**500.81**	**6.517**	**0.038**	591.00	500.81	554.89	561.56	6.583	0.086
	VVAS	537.14	527.13	552.42	0.578	0.749	537.14	552.42	543.08	510.97	2.239	0.524
	PSYF	**586.82**	**563.06**	**480.22**	**8.103**	**0.017**	**586.82**	**480.22**	**568.07**	**557.71**	**8.261**	**0.041**
	RESI	487.93	535.23	508.65	5.758	0.056	487.93	508.65	521.35	548.51	6.985	0.072
	CONN	533.32	571.15	538.64	3.673	0.159	533.32	538.64	579.04	562.89	4.060	0.255
Female	***SB***	**294.76**	**398.96**	**449.8**	**50.055**	**<0.001**	**294.76**	**449.80**	**363.03**	**460.46**	**69.964**	**<0.001**
	MVPA	**422.21**	**340.35**	**316.52**	**27.988**	**<0.001**	**422.21**	**316.52**	**359.09**	**308.28**	**33.403**	**<0.001**
	PHYF	**372.57**	**358.54**	**266.65**	**12.771**	**0.002**	**372.57**	**266.65**	**348.82**	**375.22**	**14.328**	**0.002**
	VVAS	338.56	331.91	352.69	0.507	0.776	338.56	352.69	342.15	314.91	2.342	0.505
	PSYF	**376.01**	**353.47**	**281.13**	**10.180**	**0.006**	**376.01**	**281.13**	**355.81**	**349.47**	**10.27**	**0.016**
	RESI	314.42	335.71	337.00	1.893	0.388	314.42	337.00	334.68	337.38	1.912	0.591
	CONN	340.51	362.03	312.56	3.803	0.149	340.51	312.56	364.07	358.64	3.870	0.276
Male	***SB***	**196.06**	**220.51**	**328.11**	**32.994**	**<0.001**	**196.06**	**328.11**	**165.47**	**245.60**	**52.725**	**<0.001**
	MVPA	**259.30**	**207.66**	**114.01**	**44.496**	**<0.001**	**259.30**	**114.01**	**258.53**	**184.48**	**61.342**	**<0.001**
	PHYF	219.59	198.98	236.00	4.474	0.107	219.59	236.00	192.82	201.70	2.731	0.193
	VVAS	199.30	195.83	200.02	0.100	0.951	199.30	200.02	207.71	191.22	1.003	0.801
	PSYF	212.03	210.02	200.10	0.273	0.872	212.03	200.10	211.32	209.45	0.285	0.963
	RESI	174.55	200.16	173.43	5.250	0.072	174.55	173.43	185.35	205.17	6.543	0.088
	CONN	193.48	209.68	224.21	2.771	0.250	193.48	224.21	215.09	207.36	2.964	0.397
Primary	***SB***	**222.76**	**270.89**	**358.13**	**47.772**	**<0.001**	**222.76**	**358.13**	**242.04**	**356.88**	**69.996**	**<0.001**
	MVPA	**298.66**	**243.17**	**187.22**	**36.069**	**<0.001**	**298.66**	**187.22**	**269.14**	**165.78**	**54.069**	**<0.001**
	PHYF	262.92	237.72	231.46	4.454	0.108	262.92	231.46	246.96	210.93	6.877	0.076
	VVAS	228.42	216.29	191.77	3.480	0.175	228.42	191.77	212.22	226.56	3.957	0.266
	PSYF	251.44	248.90	233.82	0.809	0.667	**251.44**	**233.82**	**266.30**	**198.41**	**9.369**	**0.025**
	RESI	194.56	208.80	190.76	1.699	0.428	194.56	190.76	194.46	241.15	7.492	0.058
	CONN	246.95	239.26	212.85	2.981	0.225	246.95	212.85	231.43	260.27	4.606	0.203
Secondary	***SB***	**299.48**	**320.68**	**553.50**	**50.907**	**<0.001**	**299.48**	**553.50**	**322.77**	**319.24**	**50.944**	**<0.001**
	MVPA	**355.99**	**329.35**	**104.63**	**49.255**	**<0.001**	**355.99**	**104.63**	**318.51**	**336.77**	**50.235**	**<0.001**
	PHYF	340.70	309.71	276.63	4.970	0.083	340.70	276.63	308.78	310.35	4.978	0.173
	VVAS	**280.39**	**323.62**	**383.90**	**11.105**	**0.004**	**280.39**	**383.90**	**329.25**	**319.67**	**11.394**	**0.010**
	PSYF	**358.37**	**303.35**	**252.98**	**14.752**	**0.001**	**358.37**	**252.98**	**303.27**	**303.41**	**14.752**	**0.002**
	RESI	289.47	327.87	337.63	5.987	0.050	289.47	337.63	328.77	327.24	5.995	0.112
	CONN	**247.79**	**346.85**	**288.31**	**38.556**	**<0.001**	**247.79**	**288.31**	**347.69**	**346.25**	**38.563**	**<0.001**

significantly improved after the full-day school closures. Female children's *PSYF* significantly improved after the outbreak of COVID-19 ($p = 0.049$). School closures have no significant influence on children's *RESI*. Full-day school closures have no significant influence on children's *CONN* (T0→T1). In contrast, when it turns to a half-day basis (T2), the secondary school students' *CONN* is improved (T0→T2, $p < 0.001$). This effect lasts after long-term half-day face-to-face classes.

4.2 Correlation between Contextual Features and Health Indicators

We conduct statistical analysis to reveal the correlation between contextual features and health indicators. Detailed analysis can be found in Appendix 5. The key findings are elaborated as follows:

- *MVPA*. Female children have significantly less *MVPA* than male children ($U = 145688, p = 0.007$). Besides, secondary school students have significantly less *MVPA* than primary ones ($U = 98028$, $p < 0.001$). Older children tend to have less *MVPA* (Spearman's $|\rho| > 0.3$, $p < 0.05$). Besides, children with higher BMI, higher BFP, lower educated parents, more time spent on electronic devices, worse sleep patterns and exercise patterns tend to have less *MVPA* ($|\rho| > 0.1, 0.2$, $p \leq 0.001$).

- *PHYF*. Female children have significantly worse *PHYF* than males ($U = 136556, p = 0.019$), and secondary school students have significantly worse *PHYF* than primary school students ($U = 179852, p < 0.001$). Older children tend to have worse *PHYF* ($|\rho| > 0.1, p < 0.001$). Besides, children with higher BFP, worse sleep patterns, exercise patterns, and dietary habits tend to have worse *PHYF* ($|\rho| > 0.1, 0.2, p < 0.001$).
- *VVAS*. Secondary school students have significantly worse *VVAS* than primary school students ($U = 99527.5$, $p < 0.001$). Older children tend to have worse *VVAS* ($|\rho| > 0.2$, $p < 0.001$). Besides, children with lower BMI and BFP, less time on electronic games, better sleep patterns, exercise patterns, and dietary habits tend to have better *VVAS* ($|\rho| > 0.1, 0.2$, $p < 0.001$). Besides, the *VVAS* of children is also significantly associated with mothers' occupation types ($p = 0.047$)
- *PSYF*. Secondary school students have significantly worse *PSYF* than primary school students ($U = 189605.5$, $p < 0.001$). Older children tend to have worse *PSYF* ($|\rho| > 0.1$, $p < 0.001$). Besides, children with lower BFP, less time spent playing electronic games, better academic performance, better sleep patterns, and better dietary habits tend to have better *PSYF* ($|\rho| > 0.1, 0.3, p < 0.001$). Specifically, children whose parents are married and living together have significantly better *PSYF* ($p = 0.035$).
- *RESI*. Children with higher educated parents, more time spent watching TV on weekdays, higher financial satisfaction, better academic performance and interest, better sleep patterns, better exercise patterns, and better dietary habits tend to have better *RESI* ($|\rho| > 0.1, 0.2, p < 0.05$). Children whose fathers are full-time employed and whose parents are married and living together have significantly better *RESI* ($p = 0.026$).
- *CONN*. Secondary school students have significantly worse *CONN* than primary ones ($U = 105339.5, p < 0.001$). Besides, children with lower BMI, higher educated parents, higher family income levels, less time spent playing electronic games, higher financial satisfaction and interests in major courses, better sleep patterns, better exercise patterns, and better dietary habits tend to have better *CONN* ($|\rho| > 0.1, 0.2, 0.3$, $p < 0.001$). Specifically, children whose fathers are full-time employed and whose parents are married and living together have significantly better *CONN* ($p = 0.001$). Specifically, the *CONN* of children is also significantly associated with mothers' and fathers' occupation types ($p = 0.015, < 0.001$).

5 CONTEXT-AWARE HEALTH INFERENCE SYSTEM

Motivated by the exploratory data analysis, we build a context-aware health inference system to bridge the contextual features (including school closures) and the health indicators with machine learning models to infer children's physical and mental health. By explaining the model, we analyze the feature importance to reduce the model complexity and assess the risk factors. We first introduce the input features and health profiling. Second, we compare different machine learning algorithms and select the one with the best performance to build the system. Then, we analyze the feature importance and prune the model by removing the redundant features. Finally, we present the risk factor analysis and case studies.

5.1 Contextual Feature Description

Based on the above analysis, we derive the contextual features from the perspectives of children's demographic, socioeconomic status, electronic device usage patterns, financial satisfaction, academic performance, sleep patterns, exercise habits, and dietary patterns. Specifically, we use one-hot encoding to process the categorical features, such as gender, school type, and parents' employment and occupation types. Finally, we obtain the initial features, detailed in Table 3.

5.2 Health Profiling

We use the six indicators to profile children's physical and mental health. Each indicator is divided into two statuses, i.e., at-risk and normal. For *MVPA*, according to the WHO guideline [52], if the average time of *MVPA* per

Table 3. Description of the initial input contextual features and corresponding labels.

Category	Contextual Feature Labels and Description
Children's demographic	***Age***: the child's age, ***Gender: F***: female, ***Gender: M***: male, ***BMI***: body mass index, ***BFP***: body fat percentage, ***Grade year***: the child's grade year, ***Grade: P***: primary school, ***Grade: S***: secondary school
Socioeconomic status Father: * is '*Fa*' Mother: * is '*Mo*'	***Edu. (*)***: parents' education levels, ***Income***: family income level, ***Empl. (*): PT/FT/U***: parents' employment types (part-time, full-time, or unemployed), ***Occ. (*): ME/P/AP/CSW/SSW/SAFW/CRW/PMOA/EO/SAH/Others***: parents' occupation types† ***MS: married and living together, MS: divorced/separated, MS: cohabit***: parents' marital status
Electronic device usage patterns Weekdays: ** is '*WD*' Weekends: ** is '*WE*'	***T (TV/G/L/S, **)***: children's time spent watching TV (TV), playing game (G), online learning (L), and surfing the Internet (S) in weekdays and weekends, ***Addi. (C/P)***: parents' addiction levels to electronic devices from the perspectives of children and parents, ***Restriction***: parents' restriction levels on children's use of electronic devices, ***Comm. (ED)***: communication between parents and children interrupted by electronic devices
Financial satisfaction	***Fin. Sat. (P/C/F)***: financial satisfaction from the perspectives of parents, children, and the whole family
Academic performance	***Int. (C/E/M)***: the child's interest in the three major courses, Chinese, English, and Mathematics ***Per. (C/E/M)***: the child's performance in the three major courses, Chinese, English, and Mathematics
Sleep patterns	***SQ***: sleep quality, ***SP***: sleep problem, ***SR***: sleep regularity, ***Sat. (sleep)***: sleep satisfaction, ***SD (WD/WE)***: sleep duration on weekdays and weekends, ***T (bed, awake)***: time spent on bed before falling asleep, ***Obs. (sleep)***: obstruction levels caused by sleep problem, ***Nap duration***: nap duration
Exercise habits Frequent: † is '*F*' Preferred: † is '*P*'	***Exer. Habit (F/R)***: proportions of friends and family with exercise habits, ***Exer. (†)***: numbers of frequent and preferred exercise, ***Loc. (†, Exer.)***: numbers of frequent and preferred exercise locations, ***T (†, Exer.)***: numbers of frequent and preferred exercise time, ***R (M/F, Exer.)***: reasons for more and less exercise
Dietary patterns	***Unhealthy/Healthy Diet***: Unhealthy and healthy diet habits
Learning mode and time	***L0/L1/L2***: different learning modes, ***T0/T1/T2/T3***: different data collection periods

†: There are 11 occupation types, detailed in the supplementary questionnaires.

day is less than 60 minutes, it is at-risk status; otherwise, it is at normal status. Besides, physical and psychosocial functioning is at-risk if the scores are higher than 1 SD above the mean of the total participants [71], while *VVAS*, *RESI*, and *CONN* are at-risk if their scores are lower than 1 SD below the mean of the total participants.

5.3 Inference Model Selection

The pipeline of the inference model selection is shown in Fig. 2. We train classifiers to infer probabilities of at-risk status for the six indicators based on the contextual features. To select the models with best performance, 1,172 participants are divided into a training set (80%) and a test set (20%). The following algorithms are used to train the classifiers.

- *Logistic Regression (LR) [47]* uses a logistic function to model the probabilities of two possible classes.
- *Random Forest (RF) [30]* is a machine learning algorithm for classification and regression by constructing plenty of decision trees at training time.
- *AdaBoost (AB) [63]*, short for *Adaptive Boosting*, refers to a specific approach of training a boosted classifier, using an ensemble of decision trees to predict the target label.
- *Support Vector Machine (SVM) [19]* is a supervised learning algorithm, which builds a model and make it be a non-probabilistic binary linear classifier by assigning new examples to one category.

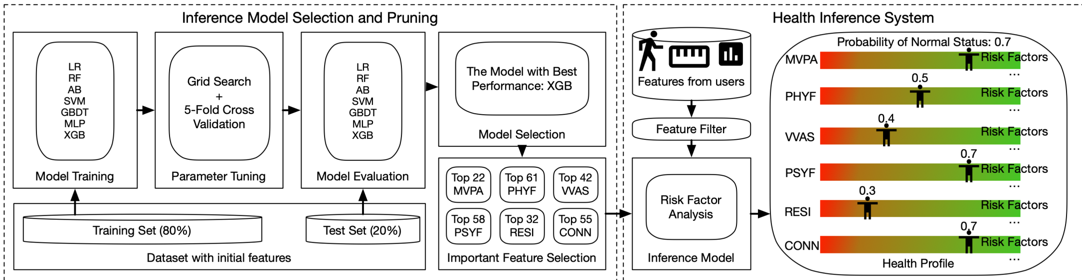


Fig. 2. Inference model selection pipeline and the health inference system.

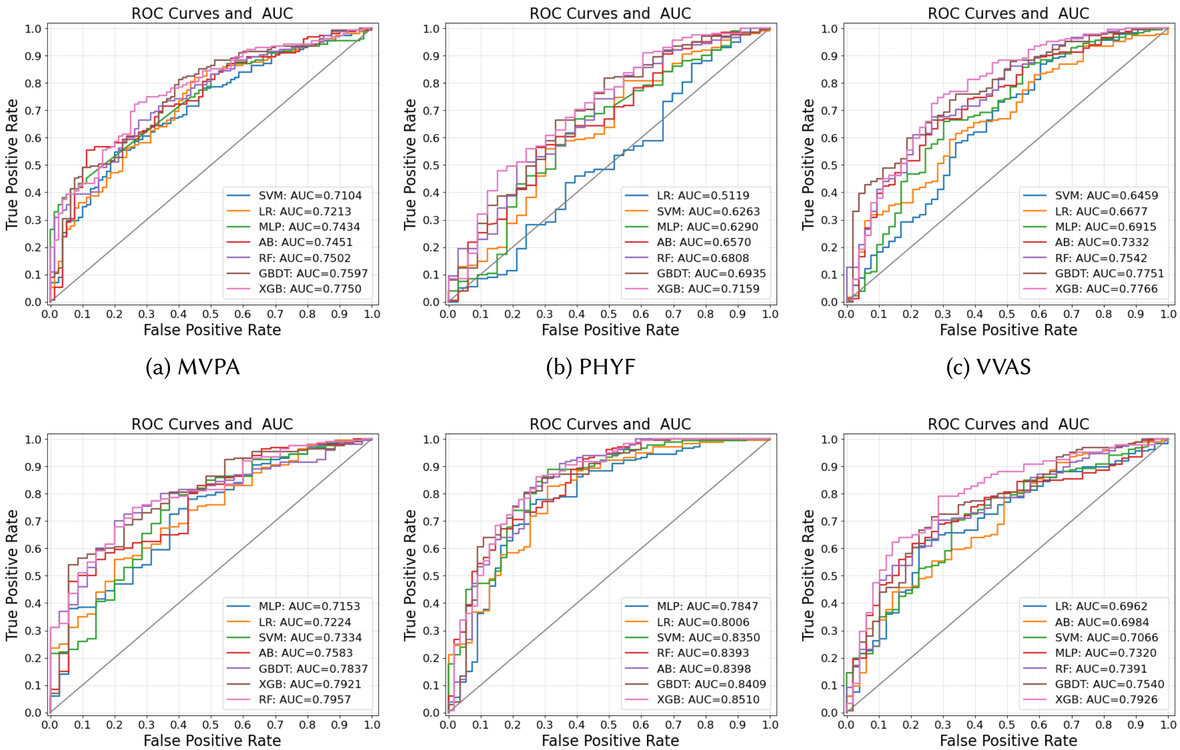
- *Gradient Boosting Decision Tree (GBDT) [76]* is an iterative decision tree algorithm, by combining several weak learners, that is, individual decision trees, to come up with one strong learner.
- *Multilayer Perception (MLP) [50]* is an artificial neural network consisting of an input layer, an output layer, and one or more hidden layers. Backpropagation is used for training.
- *XGBoost (XGB) [15]* is an ensemble machine learning algorithm based on gradient boosted decision trees, where decision trees are established in sequential form.

Since some initial features are correlated with each other. We conducted principal components analysis (PCA) on the features for those methods that would be influenced by multicollinearity [14, 58], such as LR. Specifically, although MLP does not suffer from multicollinearity, we also conducted PCA to speed up its convergence. The correlated features would influence RF when interpreting the data since it would randomly use the correlated features, so we also conducted PCA for RF. The explained variance in PCA is set to 95%. For other decision-tree-based (including boosted trees) methods, such as AB, GBDT, and XGB, they are very robust to the correlated features [14, 58], so we directly input the initial features.

We select the parameters achieving the best performance for each model. The parameters are determined by grid search and validated through five-fold cross-validation [6] on the training set. Then, we compare the performance of the models on the test set. We use the area under receiver operating characteristics (ROC) curves (AUC) [9] as the evaluation metrics. The results are shown in Fig. 3.

We can find that XGBoost achieves the best performance on *MVPA* ($AUC = 0.7750$), *PHYF* ($AUC = 0.7159$), *VVAS* ($AUC = 0.7766$), *RESI* ($AUC = 0.8510$), and *CONN* ($AUC = 0.7926$). RF outperforms XGBoost on *PSYF* ($AUC_{XGB} = 0.7921$, $AUC_{RF} = 0.7957$), but there is no big difference. XGboost requires fewer data for training and has good interpretability compared with deep-learning-based methods [39]. It stands out from other algorithms due to its high efficiency, low computational cost, good performance, and generalization [15, 51].

Considering that the positive (normal) and negative (at-risk) samples are quite imbalanced (most of the children are with normal health status), we select the optimal thresholds based on the G-mean according to the ROC curves. G-mean is the geometric mean of sensitivity and specificity, which is one of the unbiased evaluation metrics for imbalanced classification. $G - mean = \sqrt{(\text{True Positive Rate}) \times (1 - \text{False Positive Rate})}$. The thresholds that maximize the G-mean value are selected, which are 0.601, 0.716, 0.766, 0.847, 0.790, and 0.775 for *MVPA*, *PHYF*, *VVAS*, *PSYF*, *RESI*, and *CONN*, respectively. Also, instead of directly presenting whether the children are in normal health status or at-risk status, the system presents the probabilities. A lower probability of normal status (a higher probability of at-risk status) indicates a higher probability of the child suffering from the health issues. The probability value is for reference rather than a certain diagnosis. Therefore, the higher the probability of at-risk status, the higher confidence we would have in the child suffering the issue.



(a) MVPA (b) PHYF (c) VVAS

(d) PSYF (e) RESI (f) CONN

Fig. 3. ROC curves and AUC of the seven algorithms on the six indicators.

5.4 Feature Importance Analysis

We further prune the model by removing the less important features for each indicator. First, we use SHAP, which is a game theoretic approach to obtain the importance values for each feature [41]. Then, we rank the features according to their importance values. Initially, there were 91 features. For each indicator, we iteratively select the top 1, top 2, ..., and top 91 most important features and evaluate the inference performance until achieving close or even better performance compared with using the original 91 features.

Fig. 4 shows the summary of SHAP values of the top 25 most important features for the six indicators, and the summary of SHAP values for all 91 features can be found in Appendix 6. The selected important features for each indicator are detailed in Appendix 7. More specifically, for *MVPA*, its top 22 features achieve an AUC of 0.7826 ($AUC_{all} = 0.7750$. For *PHYF*, its top 61 features achieve an AUC of 0.7238 ($AUC_{all} = 0.7159$). For *VVAS*, its top 42 features achieve an AUC of 0.7787 ($AUC_{all} = 0.7766$). For *PSYF*, its top 58 features achieve an AUC of 0.7909 ($AUC_{all} = 0.7921$). For *RESI*, its top 32 features achieve an AUC of 0.8603 ($AUC_{all} = 0.8510$). For *CONN*, its top 55 features achieve an AUC of 0.7926 ($AUC_{all} = 0.7926$).

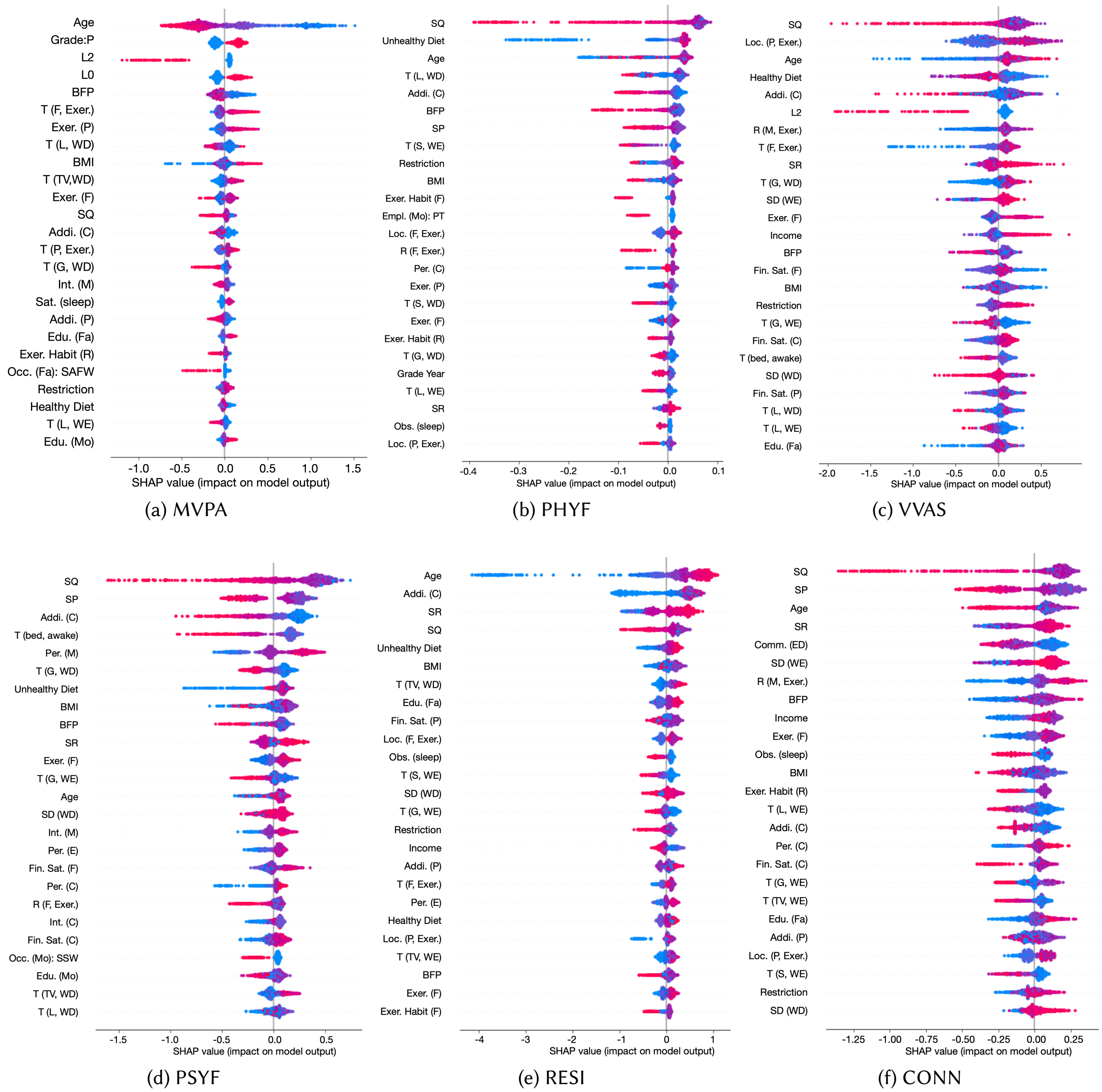


Fig. 4. Top 25 most important features for each indicator. Red color means higher feature values and blue means lower feature values. Lower SHAP values mean higher risks. The description of labels can be found in Table 3

5.5 Risk Factor Analysis

Based on the summary plot of SHAP values, we can identify the general correlation between contextual features and the predicted probability of at-risk status. By figuring out the factors correlated to the high predicted probability of at-risk health status, potential risk factors of the health indicators can be identified. We elaborate on the risk factor analysis for the health indicators as follows.

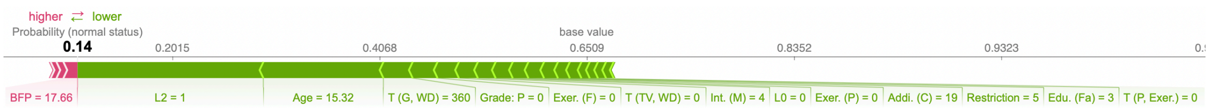
5.5.1 MVPA. As shown in Fig. 4a, generally, an increase in age, BFP, addiction levels to electronic devices of parents, children's time on electronic devices for learning and gaming on weekdays, interest in Math, and worse sleep quality (a higher score means worse sleep quality) increases the predicted probability of at-risk status. A decrease in BMI, time on watching TV on weekdays, number of frequent and preferred time periods for exercise, number of preferred exercises, and sleeping satisfaction also increases the predicted probability of at-risk status. When it is online learning (L2), and the father is a skilled agricultural and fishery worker, the predicted probability of at-risk status increases. When the child is a primary school student, the learning mode is full-day face-to-face (L0), and the father has higher education levels, the predicted probability of normal status increases.

5.5.2 PHYF. As shown in Fig. 4b, the predicted probability of at-risk status generally increases with worsening sleep quality and sleep problem (higher scores), higher levels of obstruction caused by sleep problem, more unhealthy diet habits (lower scores), increasing body fat percentage, increasing parents' addiction levels to electronic devices, and more time spent on electronic devices. Lower proportions of friends with exercise habits would also increase the predicted probability of at-risk status. When the parent is part-time employed, the learning mode is full-day online or full-day face-to-face, and the parent's occupation is service and sales worker or craft and related worker, the predicted probability of at-risk status usually would increase.

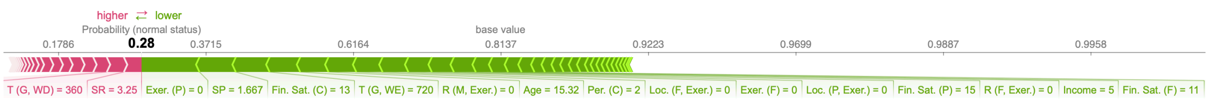
5.5.3 VVAS. As shown in Fig. 4c, the monotonic relationships between features and the predicted probability of at-risk *VVAS* are not obvious. But younger children tend to increase the predicted probability of at-risk *VVAS*, and more time spent on electronic devices for gaming but less time on weekends decreases the predicted probability of at-risk status. Healthier diet patterns (lower healthy diet scores and higher unhealthy diet scores), longer sleep duration on weekends, more exercise, higher family income levels, and higher interest in Math, English, and Chinese increase the predicted probability of normal status. Besides, when the learning mode is online learning (L2), and the mother is part-time employed, the possibility of at-risk status increases. When the mother is unemployed, and the time is T3 (after long-term half-day face-to-face classes), the predicted probability of at-risk status decreases.

5.5.4 PSYF. As shown in Fig. 4d, generally, worse sleep patterns (worse sleep quality and problem, more time spent on bed before falling asleep, and less sleep regularity), higher addiction levels to electronic devices of parents from the perspective of children, more communication between parents and children interrupted by electronic devices, more time spent playing electronic games, lower performance and interest in Math, Chinese, and English, more unhealthy diet habits, less healthy diet habits, higher body fat percentage, lower BMI, and younger age increase the predicted probability of at-risk status. Higher income levels increase the probability of normal status. Also, if the mother is part-time employed and the occupation type is service and sales worker, or the father is stay-at-home, and the child's gender is female, the predicted probability of at-risk status would increase. If parents are married and living together, the occupation type of the mother is stay-at-home, the time is T3, and the learning mode is L1, the predicted probability of normal status would increase.

5.5.5 RESI. As shown in Fig. 4e, children with older age, higher BMI, and lower BFP tend to decrease the predicted probability of at-risk *RESI*. The predicted probability of at-risk status decreases with increasing addiction levels of parents from the perspective of children, increasing time spent watching TV, decreasing time spent



(a) *MVPA* (Probability of at-risk status = 0.86)



(b) *CONN* (Probability of at-risk status = 0.72)

Fig. 5. Risk factor analysis for a 15-year-old boy whose *MVPA* and *CONN* are at-risk status.

playing electronic games and surfing the Internet on weekends, decreasing communication between parents and children interrupted by electronic devices, increasing sleep quality, lower obstruction levels caused by sleep problem, fewer unhealthy diet habits, more frequent and preferred time periods for exercise, more frequent locations for exercise, lower family income levels, better academic performance, and higher education levels of fathers. Besides, if the father is employed part-time, the predicted probability of at-risk status usually increases.

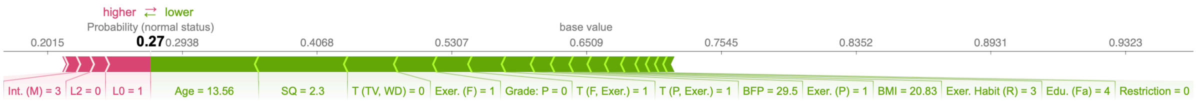
5.5.6 CONN. As shown in Fig. 4f, the predicted probability of at-risk *CONN* increases with worse sleep quality and problem (higher scores), less sleep regularity (lower scores), higher levels of obstruction caused by sleep problem, older age, lower BFP, more communication between parents and children interrupted by electronic devices, more time spent on electronic devices for gaming, learning, surfing, and watching TV on weekdays, higher parents' addiction levels, lower restriction levels on children's using electronic devices, lower family income, lower education levels of fathers, lower financial satisfaction, and worse performance in Chinese and English. Besides, when the mother's occupation type is skilled agricultural and fishery workers or professionals, the father's occupation type is skilled agricultural and fishery workers or craft and related workers, and the parents are not married or living together, the predicted probability of at-risk status increases. When the time is T2 or T3 (half-day school reopening after COVID-19), the predicted probability of at-risk status decreases.

5.6 Case Studies

We present three cases to illustrate the effectiveness of the system. For each indicator, the predicted probability of normal status is presented, and a lower probability of normal status means a higher probability of at-risk status. The risk factors increasing the predicted probability of at-risk status are in green, while the factors contributing to the predicted probability of normal status are in red. The factors closer to the central point (the probability value) are more important.

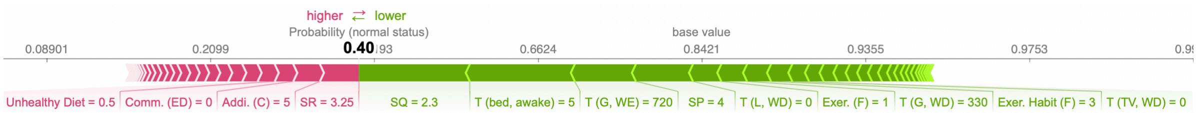
5.6.1 Case 1: A 15-Year-Old Boy with At-Risk MVPA *and* CONN. Fig. 5 shows the risk factors of a child (15.32-year-old boy) whose *MVPA* and *CONN* are at-risk status.

Fig. 5a shows that the probability of normal *MVPA* is 0.14, meaning that the child has a high probability of at-risk *MVPA*. The most important risk factor is L2 = 1 (online learning mode), meaning that full-day online learning mode significantly increases the probability of at-risk *MVPA*. Also, the child is a secondary school student; more academic pressure may decrease the time for physical activity. As for the exercise habits, the child has no frequent or preferred exercise and no preferred time period for exercise. Another important risk factor is his electronic usage patterns. The child spent 360 minutes per day playing electronic games on weekdays, and parents have fewer restrictions on the child's use of electronic devices (Restriction=5, in the least quarter). Addiction (C) = 19 means that the interaction between parents and the child is frequently interrupted by parents'

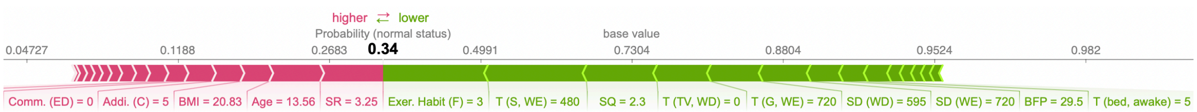


(a) *MVPA* (Probability of at-risk status = 0.73)

(b) *PHYF* (Probability of at-risk status = 0.41)



(c) *PSYF* (Probability of at-risk status = 0.60)



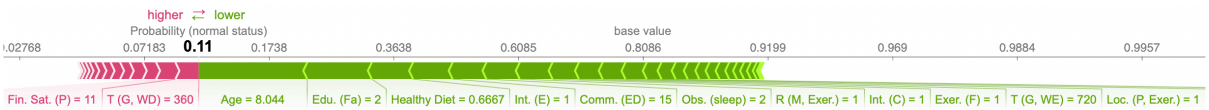
(d) *RESI* (Probability of at-risk status = 0.66)

Fig. 6. Risk factor analysis for a 13-year-old girl whose *MVPA*, *PHYF*, *PSYF*, and *RESI* are at-risk status.

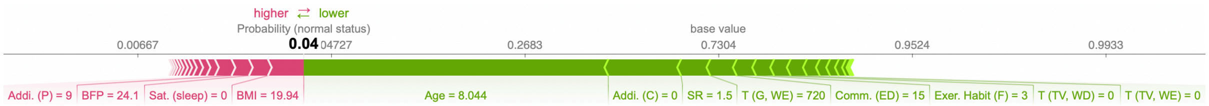
using electronic devices (the average value is 9.89, and a higher value means more frequent), which increases the probability of at-risk *MVPA*. Besides, the father's education level is junior high school, which is relatively low and increases the probability of at-risk *MVPA*. Therefore, to help the boy reduce the probability of at-risk *MVPA*, he may consider doing more exercise. Both the child and parents should spend less time on electronic devices, which is also consistent with our common sense.

Fig. 5b shows that the probability of normal *CONN* is 0.28, which means the child has a relatively high probability of at-risk *CONN*. Bad exercise habits are also important risk factors for *CONN*. Besides, serious sleep problem ($SP = 1.667$, in the worst quarter) increases the probability of at-risk *CONN*, but the child has high sleep regularity ($SR = 3.25$), decreasing the probability of at-risk *CONN*. Financial satisfaction scores from the perspective of both children and parents are high (higher scores mean lower satisfaction), and the income level of 5 is below the average, which increases the probability of at-risk *CONN*. Specifically, the child's time spent playing electronic games increases the probability of at-risk *MVPA* but decreases the probability of at-risk *CONN*. Therefore, to help the boy reduce the probability of at-risk *CONN*, the family may try to increase the financial satisfaction, and the boy should do more exercise and improve sleep quality.

5.6.2 Case 2: A 13-Year-Old Girl with At-Risk MVPA*,* PHYF*,* PSYF*, and* RESI*.* Fig. 6 shows the risk factors of a child (13.56-year-old girl) whose *MVPA*, *PHYF*, *PSYF*, and *RESI* are at-risk status. Fig. 6a shows the risk factors of *MVPA*. Similar to the previous case, the child is a secondary school student, and more academic pressure may increase the probability of at-risk status. Another important risk factor is poor sleep quality ($SQ = 2.3$). The child has a BFP of 29.5, which is a little high for a 13-year-old girl, increasing the probability of at-risk *MVPA*. According to the feature importance analysis of *MVPA*, we can find that the probability of at-risk *MVPA*



(a) *VVAS* (Probability of at-risk status = 0.89)



(b) *RESI* (Probability of at-risk status = 0.96)

Fig. 7. Risk factor analysis for a 8-year-old boy whose *VVAS* and *RESI* are at-risk status.

decreases with increasing time spent watching TV on weekdays, and the child spent no time watching TV on weekdays, increasing the probability of at-risk *MVPA*. Besides, the child has only one frequent and preferred exercise and one frequent time period for exercise. None of her family has exercise habits (Exer. Habit (R) = 3), and the father's education level is senior high school, increasing the probability of at-risk *MVPA*. It is before the outbreak of COVID-19 ($T0 = 1$), which decreases the probability of at-risk *MVPA*. Specifically, her interest in Math is 3, which belongs to the least interest quarter. Therefore, to improve her *MVPA*, the child may consider improving her sleep quality, and both the child and parents should improve their exercise habits.

Fig. 6b shows that the probability of normal *PHYF* is 0.56. Note that also it is over 0.5, the optimal threshold according to the ROC curve is 0.716, so it is classified into at-risk status. The most important risk factor is also poor sleep quality, and severe sleep problem ($SP = 4$) increases the probability of at-risk *PHYF*. Bad exercise habits are also important risk factors. Besides, the child spent too much time surfing the Internet and playing electronic games on weekdays. The parents' addiction level to electronic devices from the perspective of children is low, decreasing the probability of at-risk *PHYF*. Therefore, to improve the child's *PHYF*, it may be important to improve the sleep quality and reduce the time spent playing electronic games or surfing the Internet on Weekdays.

Poor sleep quality and exercise habits are also the most important risk factors for *PSYF* and *RESI*. As shown in Fig. 6c, the child spent too much time playing electronic games, increasing the probability of at-risk *PSYF*, but the high sleep regularity, few unhealthy diet habits, and lower parents' addiction levels to electronic devices decrease the probability of risk. Fig. 6d shows that the high BFP also increases the probability of at-risk *RESI*, but no communication between parents and the child is interrupted by parents' use of electronic devices, and high sleep regularity decreases the risk. Therefore, to improve the child's *PSYF* and *RESI*, it may be important to improve sleep patterns and exercise habits.

5.6.3 Case 3: A 8-Year-Old Boy with At-Risk VVAS *and* RESI*.* As shown in Fig. 7, an 8-year-old boy has at-risk *VVAS* and *RESI*. The education level of the father is primary school. The child has less healthy diet habits (0.667, higher scores means fewer, in the least quarter), and the obstruction level caused by sleep problem is higher than the average, increasing the probability of at-risk *VVAS*. Also, the communication between parents and the child is frequently interrupted by parents' use of electronic devices (Comm. (ED) = 15), and the child has bad exercise habits, increasing the probabilities of at-risk *VVAS* and at-risk *RESI*. The risk factors of *VVAS* also include the child's no interest in Chinese (Int. (C) = 1, the worst level). Nevertheless, he spent too much time playing electronic games and online learning on weekends, increasing the risk probability. Specifically, the financial satisfaction from the perspective of parents is low (Fin. Sat (P) = 11), but it decreases the probability of at-risk. After looking into the information of the child, we find that the child's financial satisfaction is relatively high (Fin.

Sat. (C) = 7, lower scores mean higher satisfaction), but the family income level is very low. Maybe the parents spent more of their income on the child, which decreases the probability of at-risk *VVAS*. The high financial satisfaction from the perspective of children also decreases the probability of at-risk *RESI*. The risk factors of *RESI* also include low sleep regularity, too much time spent playing electronic games on weekends, and none of his friends with exercise habits. Therefore, to improve the child's *VVAS* and *RESI*, the child may consider improving his diet habits, sleep quality, and exercise habits, and spending less time on electronic devices.

6 DISCUSSION

This paper explores the impacts of learning modes and contextual characteristics on children's physical and mental health, based on which we build a context-aware health inference system to infer children's health status from six health indicators: *MVPA*, *PHYS*, *VVAS*, *PSYF*, *CONN*, and *RESI*. In what follows, we conclude and discuss the key insights into the impacts of school closures, the implications, and the limitations of our work.

6.1 Insights

6.1.1 Short-term and Long-Term Impacts of Full-Day School Closures and Half-Day School Closures. Full-day school closures would influence children's physical and mental health [74]. As shown in our analysis, the most immediate impact is the decrease in physical activity and the increase in sedentary behaviors. Compared with male children, school closures have greater effects on female children. As shown in our analysis, although COVID-19 was under control and schools reopened, the time of *MVPA* of many female children still failed to return to the level before the outbreak of COVID-19. However, in the long term, the *MVPA* of children, except for secondary school students, fail to return to the level before COVID-19. It is possibly because children have to wear a mask in school during pandemic periods so that they cannot do many high-intensity activities, and the *MVPA* of secondary school students had been at lower levels before the outbreak of COVID-19. Besides, the physical functioning of female children was also significantly worsened due to full-day school closures, which is consistent with previous work [7, 49] that the likelihood of poor physical functioning is higher among females. However, full-day school closures improved female children's psychosocial functioning, which is not in line with the previous work [16] that girls are more likely to experience depression, anxiety, and stress symptoms than boys. It is possibly because the interaction with peer groups has a greater impact on female children's mental health [21]. During full-day school closures, their interaction with peers decreased, and thereby they were judged less by others, which increased their psychosocial functioning. This effect continued after the school reopening, possibly because of wearing masks. Secondary school students' self-rated health was also improved during school closures. A short-term closure of schools might relieve the stress and thereby increase their self-rated health. Although half-day school closures also decrease *MVPA* and increase *SB*, compared with full-day school closures, half-day school closures have significantly less negative impact. Besides, Half-day school closures have no significant harm to other indicators compared with full-day face-to-face classes and even increase secondary school students' connectedness with others. Therefore, when the pandemic is under control to some extent, half-day school closures may be a better choice than full-day school closures.

6.1.2 High-Income Families versus Low-Income Families. Children from low-income families tend to face greater health risks during school closures than those from high-income families, which is consistent with previous research indicating that children from lower-income families face greater difficulties as a result of the COVID-19 pandemic [24, 70]. As shown in our analysis, children from lower-income families tend to have worse connectedness with others and worse self-rated health. However, low-income families still have some strengths. We find that children from lower-income families tend to have higher resilience, which is consistent with previous work that children from low-income families are more independent and show surprising resilience and creativity in overcoming difficulties [17, 53]. Besides, financial satisfaction, which is evaluated based on both the levels of income (objective) and how adequate and stable the income is to satisfy the needs [73] (subjective), is also an important factor. Although some children are from low-income families, their parents sacrifice their own financial satisfaction to meet the needs of their children, which partly offsets the impact of low family income. For example, we find that children with high financial satisfaction tend to have higher resilience and connectedness with others.

6.1.3 Electronic Devices are Double-Edged Sword. Attitudes toward children's use of electronic devices are always a controversial topic [65]. Electronic devices have many benefits, especially during school closures. They are required to take online classes. However, as shown in our analysis, spending too much time on electronic devices increases the probability of at-risk status of children's physical and mental health, which is consistent with previous work that increased time on electronic devices for both gaming and learning increased psychosocial problems in children [70]. Interestingly, we find that the time spent watching TV is positively correlated with children's resilience. It is possibly because children usually watch TV with the family, and it enhances family communication [25]. Therefore, parents should appropriately limit the time children spend on electronic devices no matter for gaming or learning. Besides, our analysis shows that parents' addiction levels to electronic devices are also significantly correlated with children's health. To help children develop proper electronic usage patterns, it is important for parents to play an exemplary role.

6.1.4 Good Sleep, Frequent Exercise, and Healthy Diet are Panaceas. Based on the analysis, sleep patterns, exercise habits, and dietary patterns correlated with the six indicators. Better sleep patterns, including better sleep quality, higher sleep regularity and satisfaction, less sleep problem, and reasonable sleep time, lower the probability of health risks from the six perspectives. Frequent exercise also reduces the probability of health risks from all six perspectives. Besides the direct benefits of exercise on physical health, exercise can keep the BFP and BMI within a healthy range, which increases children's body satisfaction and thus increases their connectedness with others [20] and promotes their psychosocial health. Furthermore, we find that higher proportions of friends and family with exercise habits also positively correlated with children's own health, which is in line with previous work that there are parental and peer influences on children's exercise habits [66]. Therefore, parents should also play an exemplary role and develop good exercise habits.

6.2 Implications

The proposed system can help families comprehensively understand the contexts correlated to their children's health. It presents the importance and influence of various contextual features at an individual level, which provides intuitive indications for individuals on potential factors hindering their health. Furthermore, the system facilitates a comprehensive understanding of the influence of school closures on different child populations. The correlation between the contextual features and the health indicators revealed by the system could help the authorities allocate the resources more flexibly and equally. With the knowledge of which groups are suffering more from the pandemic control measures enforced by the government, priority could be given to these groups when allocating resources.

Compared with traditional clinical scales, each of which usually only evaluates one perspective of the health status, the proposed system is more comprehensive and ubiquitous. To use the clinic scales, parents should have priori knowledge of those scales. More specifically, they should select the scales corresponding to the health issues of their children and learn how to interpret the results. While in this system, we have predefined the six essential health indicators, and the users only need to input their contextual features to estimate their health status from various perspectives. Also, no device is required to be worn by children.

We suggest that the proposed system can be an auxiliary tool for the self-monitoring of health, especially when healthcare resources are quite insufficient (e.g., during pandemics). We believe that the proposed system could help parents sense the changes in their children's health without visible symptoms, especially mental health, such as psychosocial functioning. Furthermore, it can also contribute to the screening of specific populations for social work. More attention should be paid to families with specific characteristics that are more influenced by the pandemic control measures. For the health indicators that have been incorporated into the system, the proposed system can be used independently or combined with the scales (e.g., PedsQL and EQ5D) since it can provide more information than the scales (e.g., the important features and risk factors.) Moreover, the system can be easily extended to incorporate more health indicators.

In this work, we employed G-mean to select the thresholds for the system, while in practice, thresholds can be dynamically selected based on the application scenarios. If we want to increase the system's sensitivity to at-risk status (negative), thresholds lower the false positive rate could be selected. However, it should be noted that a lower false positive rate usually leads to a lower true positive rate, meaning that a lot of participants with normal health status would be classified as at-risk status. The thresholds can be selected according to the balanced need of the sensitivity and specificity of the system. If we want the system to be more sensitive to at-risk status, we could choose the corresponding thresholds lowering the false positive rate. If we want the system to be more sensitive to normal status, the thresholds increasing the true positive rate can be selected.

6.3 Limitations and Future Work

6.3.1 Demographic Diversity. All participants are Chinese children, which may lack demographic diversity in terms of race and affect the generality of our findings and system. Also, we only consider binary gender following existing literature in this domain. As a result, our analysis based on gender groups may fail to reveal the influence of the complexities of gender. In the future, we can deploy the system online and improve the diversity based on more data and feedback.

6.3.2 Study Design. We explore the impacts of school closures on the health indicators in Section 4.1. Limited by the realistic condition, we compared the indicators with different learning modes at different periods rather than simultaneously, which is not able to completely eliminate the influence of the closure of other infrastructure. Since schools are critical settings for physical activity and essential sources of mental health services for children [13, 74] and children spend most of their time in schools, we think the school closures would be the most significant factor leading to the differences, especially for Hong Kong which is a high-density metropolitan area with limited public space for physical activities. Also, the results from the questionnaire show that schools are the most frequent places (60.32%) for exercise. Parks (always accessible) are the second (40.19%), and courts are the third (24.66%). Only 14.93% of participants selected gyms. Therefore, we think it is reasonable to claim that the differences are mainly caused by school closures. However, other potential confounders and unmeasured factors, such as the closure of courts and gyms, may also influence some children's behaviors. In addition, a territory study ranging from kindergarten to university students can provide a comprehensive picture of the impact of school closure on physical and mental health across various age groups.

6.3.3 Contextual Diversity. We try to cover as many aspects as possible when selecting contextual features based on the literature review and empirical study. Also, the experiment results show that the included features can achieve satisfactory performance. However, since the real-world settings are really complex, there may exist other factors which can improve the inference performance, such as children's living conditions, the housing built environment, and the infection rate. In the future, we plan to incorporate more features from various perspectives.

6.3.4 Health Indicator Diversity. Currently, the health inference system includes six essential health indicators. However, since there are numerous health issues, there are still many health indicators that can be considered.

Therefore, the system is designed to be extendable. In the future, we plan to incorporate more indicators for various health issues, such as ADHD.

6.3.5 Privacy Concerns. Children's detailed contextual features are used to train the inference model, which raises privacy concerns. To address this issue, we plan to leverage the federated learning scheme [38]. For example, users can download a global model and train local models to improve the global model by uploading trained parameters while keeping users' private data on devices.

6.3.6 Causality Analysis. When interpreting machine learning models using interpretability tools such as SHAP, it might be impractical to generalize causal effects due to the unmeasured confounding features and correlated input features, especially in complex real-world settings [23, 57]. However, by explaining the general correlation between contextual features and health indicators, it is still feasible to figure out how these features influenced the predicted probability of at-risk health status and thus provide indications of potential risk factors. In the future, we plan to employ causal models to obtain more solid casual insights.

7 CONCLUSION

Many schools have experienced temporary closures due to the COVID-19 pandemic, which has a great influence on children's physical and mental health globally. However, it is difficult for parents to track and get advice on improving their children's health status timely and comprehensively. In this work, we conduct exploratory data analysis based on real-world datasets collected from children in primary and secondary schools in Hong Kong during four time periods (i.e., before the outbreak of COVID-19, full-day school closures due to COVID-19, half-day school reopening when COVID-19 was under control, and after long-term half-day face-to-face classes) from September 2019 to January 2022 to explore the impacts of school closures on children's health and the correlation between contextual characteristics and children's health. We further build a context-aware system that can infer children's physical and mental health with six indicators (physical activity intensity, physical functioning, self-rated health, psychosocial functioning, resilience, and connectedness) based on children's contextual features (demographics, socioeconomic status, electronic device usage patterns, financial satisfaction, academic performance, sleep pattern, exercise habits, and dietary patterns) during school closures. By explaining the influence of the contextual features on the inference results, the system can identify risk factors for individuals and provide guidance for authorities to allocate resources *cost-effectively*. Findings from our analysis and case studies also provide insights into the impacts of school closures and contextual characteristics on children's health.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers and editors for their constructive suggestions. This study was supported by the UGC CRF fund (Reference No. : C7149-20GF, Project Title: Impact of COVID-19 pandemic on Hong Kong children and support to families during crisis), Chow Tai Fook Enterprises, InspiringHK Sports Foundation, HKU Start-up Grant, the COVID-19 Action Seed Fund from the Faculty of Engineering at The University of Hong Kong, and the Committee on Research and Conference Grants (CRCG) under Grant 202011159116.

REFERENCES

[1] ActiGraph. 2014. wGT3X-BT User's Manual. (2014). https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2018/03/11084614/wGT3X-BT_UsersManual_RevD_Q3_2014.pdf.

[2] ActiGraph. 2020. What does the "Worn on Wrist" option do in the Scoring tab? https://actigraphcorp.my.site.com/support/s/article/What-does-the-Worn-on-Wrist-option-do-in-the-Scoring-tab. (2020). Accessed: 2022-04-18.

[3] Ruopeng An. 2020. Projecting the impact of the coronavirus disease-2019 pandemic on childhood obesity in the United States: A microsimulation model. *Journal of sport and health science* 9, 4 (2020), 302–312.

[4] Richard Armitage and Laura B Nellums. 2020. Considering inequalities in the school closure response to COVID-19. *The Lancet Global Health* 8, 5 (2020), e644.

[5] Thomas Astell-Burt, Xiaoqi Feng, and Gregory S Kolt. 2014. Green space is associated with walking and moderate-to-vigorous physical activity (MVPA) in middle-to-older-aged adults: findings from 203 883 Australians in the 45 and Up Study. *British Journal of Sports Medicine* 48, 5 (2014), 404–406.
[6] Safial Islam Ayon and Md Milon Islam. 2019. Diabetes prediction: a deep learning approach. *International Journal of Information Engineering and Electronic Business* 12, 2 (2019), 21.
[7] Lisa C Barry, Julian D Ford, and Robert L Trestman. 2014. Comorbid mental illness and poor physical function among newly admitted inmates in Connecticut's jails. *Journal of Correctional Health Care* 20, 2 (2014), 135–144.
[8] J Martin Bland and Douglas G Altman. 1997. Statistics notes: Cronbach's alpha. *Bmj* 314, 7080 (1997), 572.
[9] Christopher D Brown and Herbert T Davis. 2006. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems* 80, 1 (2006), 24–38.
[10] Danilo Buonsenso, Daniel Munblit, Cristina De Rose, Dario Sinatti, Antonia Ricchiuto, Angelo Carfi, and Piero Valentini. 2021. Preliminary evidence on long COVID in children. *Acta Paediatrica (Oslo, Norway: 1992)* 110, 7 (2021), 2208.
[11] Fidel Cacheda, Diego Fernandez, Francisco J Novoa, Victor Carneiro, et al. 2019. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research* 21, 6 (2019), e12554.
[12] Jose Castro-Piñero, Enrique G Artero, Vanesa España-Romero, Francisco B Ortega, Michael Sjöström, Jaana Suni, and Jonatan R Ruiz. 2010. Criterion-related validity of field-based fitness tests in youth: a systematic review. *British journal of sports medicine* 44, 13 (2010), 934–943.
[13] Sonia Chaabane, Sathyanarayanan Doraiswamy, Karima Chaabna, Ravinder Mamtani, and Sohaila Cheema. 2021. The impact of COVID-19 school closure on child and adolescent health: A rapid systematic review. *Children* 8, 5 (2021), 415.
[14] Tianqi Chen, Michaël Benesty, Tong He, and Y Tang. 2018. Understand your dataset with xgboost. (2018). R Document.
[15] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, San Francisco, CA, USA, 785–794.
[16] Alicia Tamarit Chulia, Usue de la Barrera, Estefanía Mónaco, Konstanze Schoeps, and Inmaculada Montoya Castilla. 2020. Psychological impact of COVID-19 pandemic in Spanish adolescents: risk and protective factors of emotional symptoms. *Revista de psicología clínica con niños y adolescentes* 7, 3 (2020), 73–80.
[17] Steven J Condly. 2006. Resilience in children: A review of literature with implications for education. *Urban education* 41, 3 (2006), 211–236.
[18] Kathryn M Connor and Jonathan RT Davidson. 2003. Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and anxiety* 18, 2 (2003), 76–82.
[19] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.
[20] Carla Crespo, Magdalena Kielpikowski, Paul E Jose, and Jan Pryor. 2010. Relationships between family connectedness and body satisfaction: A longitudinal study of adolescent girls and boys. *Journal of youth and adolescence* 39, 12 (2010), 1392–1401.
[21] Tanya E Davison and Marita P McCabe. 2006. Adolescent body image and psychosocial functioning. *The Journal of social psychology* 146, 1 (2006), 15–30.
[22] Leandro Fornias Machado de Rezende, Juan Pablo Rey-López, Victor Keihan Rodrigues Matsudo, and Olinda do Carmo Luiz. 2014. Sedentary behavior and health outcomes among older adults: a systematic review. *BMC public health* 14, 1 (2014), 1–9.
[23] Eleanor Dillon, Jacob LaRiviere, Scott Lundberg, Jonathan Roth, and Vasilis Syrgkanis. 2022. Be careful when interpreting predictive models in search of causal insights. (2022). https://shap.readthedocs.io/en/latest/overviews.html.
[24] Danielle G Dooley, Asad Bandealy, and Megan M Tschudy. 2020. Low-income children and coronavirus disease 2019 (COVID-19) in the US. *JAMA pediatrics* 174, 10 (2020), 922–923.
[25] Brent M Foster and Mark Smith. 2001. Television and Its Impact on Family Communication: A Literature Review for the Family Communication Educator. (2001). ERIC.
[26] Carol Ewing Garber, Mary L Greaney, Deborah Riebe, Claudio R Nigg, Patricia A Burbank, and Phillip G Clark. 2010. Physical and mental health-related correlates of physical function in community dwelling older adults: a cross sectional study. *BMC geriatrics* 10, 1 (2010), 1–10.
[27] Junqi Guo, Lan Yang, Rongfang Bie, Jiguo Yu, Yuan Gao, Yan Shen, and Anton Kos. 2019. An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Computer Networks* 151 (2019), 166–180.
[28] William L Haskell, I-Min Lee, Russell R Pate, Kenneth E Powell, Steven N Blair, Barry A Franklin, Caroline A Macera, Gregory W Heath, Paul D Thompson, and Adrian Bauman. 2007. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation* 116, 9 (2007), 1081.
[29] Michael Herdman, Claire Gudex, Andrew Lloyd, MF Janssen, Paul Kind, David Parkin, Gouke Bonsel, and Xavier Badia. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research* 20, 10 (2011), 1727–1736.
[30] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1. IEEE, Montreal, QC, Canada, 278–282.

[31] Syem Ishaque, Alice Rueda, Binh Nguyen, Naimul Khan, and Sridhar Krishnan. 2020. Physiological Signal Analysis and Classification of Stress from Virtual Reality Video Game. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. IEEE, Montreal, QC, Canada, 867–870.

[32] Marja Jylhä. 2009. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social science & medicine* 69, 3 (2009), 307–316.

[33] Hsien-Te Kao, Shen Yan, Homa Hosseinmardi, Shrikanth Narayanan, Kristina Lerman, and Emilio Ferrara. 2020. User-Based Collaborative Filtering Mobile Health System. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–17.

[34] Mark S Kaplan, Jason T Newsom, Bentson H McFarland, and Leining Lu. 2001. Demographic and psychosocial correlates of physical activity in late life. *American journal of preventive medicine* 21, 4 (2001), 306–312.

[35] Pavleen Kaur, Ravinder Kumar, and Munish Kumar. 2019. A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications* 78, 14 (2019), 19905–19916.

[36] Jung Jae Lee, Man Ping Wang, Tzu Tsun Luk, Ningyuan Guo, Sophia Siu-Chee Chan, and Tai Hing Lam. 2020. Associations of electronic device use before and after sleep with psychological distress among chinese adults in Hong Kong: cross-sectional study. *JMIR Mental Health* 7, 6 (2020), e15403.

[37] Boning Li and Akane Sano. 2020. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–26.

[38] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. 2020. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine* 37, 3 (2020), 50–60.

[39] Yawen Li, Liu Yang, Bohan Yang, Ning Wang, and Tian Wu. 2019. Application of interpretable machine learning models for the intelligent decision. *Neurocomputing* 333 (2019), 273–283.

[40] Chu Lim and Omar Thanoon. 2013. 2 - Social and Ethnic Determinants of Obesity. In *Obesity*, Tahir Mahmood and Sabaratnam Arulkumaran (Eds.). Elsevier, Oxford, 13–21.

[41] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., Long Beach, CA, USA, 4765–4774.

[42] Abbey R Masonbrink and Emily Hurley. 2020. Advocating for children during the COVID-19 school closures. *Pediatrics* 146, 3 (2020), 1–10.

[43] Ann S Masten and Marie-Gabrielle J Reed. 2002. Resilience in development. *Handbook of positive psychology* 74 (2002), 88.

[44] Calum Mattocks, Sam Leary, Andy Ness, Kevin Deere, Joanne Saunders, Kate Tilling, Joanne Kirkby, Steven N Blair, and Chris Riddoch. 2007. Calibration of an accelerometer during free-living activities in children. *International Journal of Pediatric Obesity* 2, 4 (2007), 218–226.

[45] Korapat Mayurasakorn, Bonggochpass Pinsawas, Pichanun Mongkolsucharitkul, Kitti Sranacharoenpong, and Sa-nga Damapong. 2020. School closure, COVID-19 and lunch programme: Unprecedented undernutrition crisis in low-middle income countries. *Journal of paediatrics and child health* 56, 7 (2020), 1013–1017.

[46] Shubham Mehta, Pankaj Kumar Mittal, and Mukesh Kumar Swami. 2014. Psychosocial functioning in depressive patients: a comparative study between major depressive disorder and bipolar affective disorder. *Depression research and treatment* 2014 (2014), 1–7.

[47] Scott Menard. 2002. *Applied logistic regression analysis*. Vol. 106. Sage, University of Colorado.

[48] Nadim Nachar et al. 2008. The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology* 4, 1 (2008), 13–20.

[49] Hadi Nobari, Mohamad Fashi, Arezoo Eskandari, Santos Villafaina, Álvaro Murillo-Garcia, and Jorge Pérez-Gómez. 2021. Effect of COVID-19 on health-related quality of life in adolescents and children: a systematic review. *International Journal of Environmental Research and Public Health* 18, 9 (2021), 4563.

[50] Leonardo Noriega. 2005. *Multilayer perceptron tutorial*. Technical Report. School of Computing. Staffordshire University.

[51] Adeola Ogunleye and Qing-Guo Wang. 2020. XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 17, 6 (2020), 2131–2140.

[52] World Health Organization et al. 2008. Pacific physical activity guidelines for adults: Framework for accelerating the communication of physical activity guidelines. (2008).

[53] Dennis K Orthner, Hinckley Jones-Sanpei, and Sabrina Williamson. 2004. The resilience and strengths of low-income families. *Family relations* 53, 2 (2004), 159–167.

[54] Eva Ostertagová, Oskar Ostertag, and Jozef Kováč. 2014. Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials* 611 (10 2014), 115–120.

[55] Christiane Otto, Franziska Reiss, Catharina Voss, Anne Wüstner, Ann-Katrin Meyrose, Heike Hölling, and Ulrike Ravens-Sieberer. 2021. Mental health and well-being from childhood to adulthood: design, methods and results of the 11-year follow-up of the BELLA study.

European child & adolescent psychiatry 30, 10 (2021), 1559–1577.

[56] Joshua R Polanin, Dorothy L Espelage, Jennifer K Grotpeter, Elizabeth Spinney, Katherine M Ingram, Alberto Valido, America El Sheikh, Cagil Torgal, and Luz Robinson. 2021. A meta-analysis of longitudinal partial correlations between school violence and mental health, school performance, and criminal or delinquent acts. *Psychological bulletin* 147, 2 (2021), 115.

[57] Mattia Prosperi, Yi Guo, Matt Sperrin, James S Koopman, Jae S Min, Xing He, Shannan Rich, Mo Wang, Iain E Buchan, and Jiang Bian. 2020. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence* 2, 7 (2020), 369–375.

[58] Sri Rahayu, Teguh Sugiarto, MADU Ludiro, Ahmad Subagyo, et al. 2017. Application of Principal Component Analysis (PCA) to Reduce Multicollinearity Exchange Rate Currency of Some Countries in Asia Period 2004-2014. *International Journal of Educational Methodology* 3, 2 (2017), 75–83.

[59] Lynn Rew, Margaret Taylor-Seehafer, Nancy Y Thomas, and Ronald D Yockey. 2001. Correlates of resilience in homeless adolescents. *Journal of nursing scholarship* 33, 1 (2001), 33–40.

[60] Eric Robinson and Angelina R Sutin. 2016. Parental perception of weight status and weight gain across childhood. *Pediatrics* 137, 5 (2016), e20153957.

[61] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.

[62] Sara E Schaefer, Marta Van Loan, and J Bruce German. 2014. A Feasibility Study of Wearable Activity Monitors for Pre-Adolescent School-Age Children. *Prev Chronic Dis* 11, 130262 (2014), 1–8.

[63] Robert E Schapire. 2003. *The boosting approach to machine learning: An overview.* Springer, Florham Park, NJ 07932 US. 149–171 pages.

[64] Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* 126, 5 (2018), 1763–1768.

[65] Kaveri Subrahmanyam, Patricia Greenfield, Robert Kraut, and Elisheva Gross. 2001. The impact of computer use on children's and adolescents' development. *Journal of Applied Developmental Psychology* 22, 1 (2001), 7–30.

[66] Saulius Sukys, Daiva Majauskienė, Vida J Cesnaitiene, and Diana Karanauskiene. 2014. Do parents' exercise habits predict 13–18-year-old adolescents' involvement in sport? *Journal of sports science & medicine* 13, 3 (2014), 522.

[67] Elizabeth Sweet, Arijit Nandi, Emma K Adam, and Thomas W McDade. 2013. The high price of debt: Household financial debt and its impact on mental and physical health. *Social science & medicine* 91 (2013), 94–100.

[68] Suqin Tang, Mi Xiang, Teris Cheung, and Yu-Tao Xiang. 2021. Mental health and its correlates among children and adolescents during COVID-19 school closure: The importance of parent-child discussion. *Journal of affective disorders* 279 (2021), 353–360.

[69] Catherine Tong, Matthew Craner, Matthieu Vegreville, and Nicholas D Lane. 2019. Tracking fatigue and health state in multiple sclerosis patients using connnected wellness devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–19.

[70] Winnie WY Tso, Rosa S Wong, Keith TS Tung, Nirmala Rao, King Wa Fu, Jason Yam, Gilbert T Chua, Eric YH Chen, Tatia Lee, Sherry KW Chan, et al. 2020. Vulnerability and resilience in children during the COVID-19 pandemic. *European child & adolescent psychiatry* 2022, 31 (2020), 1–16.

[71] James W Varni, Tasha M Burwinkle, Michael Seid, and Douglas Skarr. 2003. The PedsQL™* 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambulatory pediatrics* 3, 6 (2003), 329–341.

[72] James W Varni, Michael Seid, and Cheryl A Rode. 1999. The PedsQL™: measurement model for the pediatric quality of life inventory. *Medical care* 37, 2 (1999), 126–139.

[73] Esperanza Vera-Toscano, Victoria Ateca-Amestoy, and Rafael Serrano-Del-Rosal. 2006. Building financial satisfaction. *Social Indicators Research* 77, 2 (2006), 211–243.

[74] Russell M Viner, Simon Russell, Rosella Saulle, Helen Croker, Claire Stansfield, Jessica Packer, Dasha Nicholls, Anne-Lise Goddings, Chris Bonell, Lee Hudson, et al. 2021. Impacts of school closures on physical and mental health of children and young people: a systematic review. (2021). MedRxiv.

[75] Russell M Viner, Simon J Russell, Helen Croker, Jessica Packer, Joseph Ward, Claire Stansfield, Oliver Mytton, Chris Bonell, and Robert Booy. 2020. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health* 4, 5 (2020), 397–404.

[76] Ping Xuan, Chang Sun, Tiangang Zhang, Yilin Ye, Tonghui Shen, and Yihua Dong. 2019. Gradient boosting decision tree-based method for predicting interactions between target genes and drugs. *Frontiers in genetics* 10 (2019), 459.

[77] Jane Jie Yu, Catherine M Capio, Bruce Abernethy, and Cindy Hui Ping Sit. 2021. Moderate-to-vigorous physical activity and sedentary behavior in children with and without developmental coordination disorder: Associations with fundamental movement skills. *Research in developmental disabilities* 118 (2021), 104070.

[78] Jerrold H Zar. 2005. Spearman rank correlation. (2005). Encyclopedia of biostatistics.

[79] Yunke Zhang, Fengli Xu, Tong Li, Vassilis Kostakos, Pan Hui, and Yong Li. 2021. Passive health monitoring using large scale mobility data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–23.

[80] Yunke Zhang, Fengli Xu, Tong Xia, and Yong Li. 2021. Quantifying the Causal Effect of Individual Mobility on Health Status in Urban Space. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–30.