Appendices

---- ✦ ----

APPENDIX A THE SUMMARY OF THE CONTEXT AND MOTION DATA

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. The context and motion features are summarized in TABLE 1. In tabular data, each feature has one numerical or categorical value [1]. Time series refers to sequences of values ordered by time [2]. The Body Mass Index (BMI) was calculated using height and weight. Specifically, the data are collected across the worldwide prevalence of the COVID-19 pandemic, which has greatly influenced children's daily activities since many schools in Hong Kong have experienced temporary closures and adopted different learning modes [3]. Therefore, we also incorporate learning modes as an important input feature. We also present the missing ratio of data. The highest ratio of missing values is 2.99% (socioeconomic status). It is suggested that a missing rate of 5% or less is inconsequential [4].

TABLE 1: Description of the context and motion features used in this study.

Category	Items	Data Type	Missing Ratio
Children's demographic	Age, gender, body mass index (BMI), body fat percentage (BFP)	Tabular	0.00%
Socioeconomic status	Parents' education levels, family income	Tabular	2.99%
Electronic device usage patterns	Children's time spent watching TV, playing games, online learning, and surfing the Internet on weekdays and weekends, parents' addiction levels to electronic devices from the perspectives of children and parents, and parents' restriction levels on children's use of electronic devices, communication between parents and children interrupted by electronic devices	Tabular	0.00%
Financial satisfaction	Financial satisfaction from the perspectives of parents, children, and the whole family	Tabular	1.71%
Academic performance	Children's interest and performance in the three major courses, Chinese, English, and Mathematics	Tabular	1.71%
Sleep patterns	Sleep quality, sleep problem, sleep regularity, sleep duration on weekdays and weekends, time spent on bed before falling asleep	Tabular	0.00%
Exercise habits	Proportions of friends and family with exercise habits, numbers of frequent and preferred exercise, numbers of frequent and preferred exercise locations, numbers of frequent and preferred exercise time, and reasons for more and less exercise	Tabular	0.00%
Dietary patterns	Unhealthy and healthy diet habits	Tabular	0.00%
Learning modes	Full-day face-to-face, half-day online and half-day face-to-face, and full-day online	Tabular	0.00%
Accelerometer Data	Tri-axial acceleration data	Time series	0.00%

As for the reliability of data, the highest ratio of missing values for context features is 2.99% (socioeconomic status). It is suggested that a missing rate of 5% or less is inconsequential [4]. As for the reliability of acceleration data, the devices used for acceleration data collection are ActiGraph wGT3X-BT accelerometers, which are very sensitive and accurate within +/- 0.5% of the data collected [3]. Children's demographics (e.g., height and weight) are measured by researchers, so we think they are highly reliable. For those rating scales used to derive the six indicators, i.e., the physical functioning (*PHYF*) and psychosocial functioning (*PSYF*) from PedsQL, health confidence (*VVAS*) from EQ5D, resilience (*RESI*) from CD-RISC, and connectedness (*CONN*) from RSCS, they have been widely used in clinics and research. We also use Cronbach's α [5] to measure the scale reliability. As shown in TABLE 2, Cronbach's $alpha$ of all scales is above 0.7, meaning that they have high reliability [6]. Specifically, to validate the *VVAS* score from EQ5D, we use the score of the five dimensions from EQ5D (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and the *VVAS* score to measure the criterion-related validity [7]. We conduct Spearman's rank correlation analysis between the five-dimension score and *VVAS* score, and the results of T0-T3 are -0.402, -0.466, -0.397, and -0.385, respectively, with p-value<0.001. Since a higher

TABLE 2: The Cronbach's α for the rating scales.

Scale	T0	T1	T2	T3
Physical Functioning	0.840	0.816	0.795	0.813
Psychosocial Functioning	0.890	0.910	0.875	0.895
Resilience	0.955	0.965	0.954	0.948
Connectedness	0.954	0.956	0.946	0.952

five-dimension score indicates worse health, we can find that the *VVAS* scores are significantly positively related to the participants' health status, showing the reliability of *VVAS*. Since the self-reported information from rating scales is highly reliable, we presume that other self-reported contextual characteristics (e.g., electronic device usage patterns and sleep patterns) are also reliable. Before the participants filled out the questionnaire, we emphasized that they must answer all questions truthfully. The contact information has also been collected for double-checking. The questionnaire collection strictly follows ethics. Informed written consent was obtained from the parents of the participants. Therefore, we consider the data collected in this work reliable.

APPENDIX B THE ALGORITHM FOR FINDING THE MAXIMAL AVERAGE IMPORTANCE WITH CUSTOMIZED TIME WINDOW

The motion pattern m_i with a shape of 3×10080 is input into Algorithm 1 as matrix M to find the maximal average importance with time window W. The number of channels C is three, and the overall time length T is 10080. In the visualization system, users can select the time window (denoted by W) to customize the measurement of feature importance for the motion pattern. In the algorithm, the three sequences of m_i are first combined into one (with size 1×10080) using the Root Mean Square (RMS). After that, an algorithm with $O(T)$ time complexity is used to find the time slot with the length of W that has the maximal average importance. The above steps will be repeated to find the second, third, ... most important time slot. The previous most important time slots are removed from the sequence and conduct the algorithm iteratively. After that, we can get the importance of the input motion pattern during specific time points with a customized time window.

Algorithm 1: The algorithm for finding the maximal average importance with time window W for motion pattern.

Input: M: The $C \times T$ raw feature importance for the motion pattern of an individual, where C is the number of channels.
$\quad\quad\quad$ W: The time window for importance assessment selected by users.
$\quad\quad\quad$ T: The overall time length.
Output: v: The max average importance with time window W.
$\quad\quad\quad\quad$ max_i, max_j: The start and end points of the time window.

`// Combine the multi-channel sequences using RMS.`

1 **for** $0 \leq t \leq T - 1$ **do**

2 $\quad$ $A[t] = \sqrt{\sum_{i=0}^{C-1} M[i,t]^2 / C}$

3 **end for**

4 $i \leftarrow 0, j \leftarrow 0, s \leftarrow 0, v \leftarrow 0, max_i \leftarrow 0, max_j \leftarrow 0$

`// Find the max sum subarray with length W.`

5 **while** $j < T$ **do**

6 $\quad$ $s \leftarrow s + A[j]$

7 $\quad$ **if** $(j - i + 1 < W)$ **do**

8 $\quad\quad$ $j \leftarrow j + 1$

9 $\quad$ **end if**

10 $\quad$ **else if** $(j - i + 1 == W)$ **do**

11 $\quad\quad$ $v \leftarrow max(s, v), max_i \leftarrow i, max_j \leftarrow j$

12 $\quad\quad$ $s \leftarrow s - A[i], i \leftarrow i + 1, j \leftarrow j + 1$

13 $\quad$ **end else if**

14 **end while**

15 $v \leftarrow v/W$

16 **return** v, max_i, max_j

APPENDIX C EXPERT INTERVIEW PROCEDURES AND QUESTION LIST

The procedures and question list for expert interviews are detailed in TABLE 3.

TABLE 3: Procedures and question list for expert interviews.

Step	Items		
Step 1	Introduce the project background and workflows (10 minutes).		
Step 2	Ask interviewees to freely explore the system and Q&A (15 minutes).		
Step 3	Task-driven exploration (30 minutes).		
	Task list	Task 1: Describe the basic statistics of the participants involved. (**R1**)	
		Task 2: Find significant correlated features. (**R2**)	
		Task 3: Describe the patterns of motion features. (**R1**)	
		Task 4: Find the most important features of different health indicators and compare their difference. (**R5**)	
		Task 5: Describe the influence of the most important features on each health indicator. (**R5**)	
		Task 6: Summarize the characteristics of salient clusters in the graph in Group View. (**R3, R4**)	
		Task 7: Compare the context and motion features and their importance between different groups. (**R4**)	
		Task 8: Select the interested groups and health indicators, and then identify several participants with worse health status. (**R3, R4**)	
		Task 9: Select two individuals of interest and compare their health profiles, motion features, and context features. (**R1, R3**)	
		Task 10: For the above selected two individuals, explore their feature importance and influence. (**R5**)	
Step 4	Freely explore the system and find interesting points (15 minutes).		
Step 5	Conduct a semi-structured interview (30 minutes).		
	Question List	System Workflow	Q1: (Context and Motion Feature Exploration) Is it more efficient and effective to use our system to explore the context and motion features from the perspective of all participants, groups, and individuals?
			Q2: (Feature Importance and Influence Analysis) Is it more accurate and useful to use our system to identify important features and analyze their influence than the methods (e.g., hypothesis–based) you used previously?
			Q3: (Group and Individual Comparison) Is it intuitive and clear to use our system to compare the features and their importance and influence between different groups or individuals?
			Q3: (Health Profiling) Is it useful and clear to use our system to profile the children's physical and mental health?
			Q4: Is the system workflow clear that follows your analytical logic?
			Q5: Is the system easy to learn and easy to use?
			Q6: Is the system helpful to reveal insights into the multimodal data and health profiles of children?
			-Can you verify the results you already know? Please give an example.
			-Can you get any new insights from the system? Please give an example.
		Visualization and Interaction	Q7: Do these visualization and interactions meet the design requirements well?
			Q8: Is it easy to learn and use the system? Why?
			Q9: Is it easy to learn and read different views? Why?
			Q10: Is it easy to find a group/person of interest for further exploration? Why?
			Q11: Is it easy to learn and read the charts in the system? Why?
			Q12: Is it easy to learn and read the interactions in the system? Why?
			Q13: Which views are intuitive and useful to you? Why?
			Q14: Which top 3 components impress you most? Why?
			Q15: Are the interactions intuitive? Why?
			Q16: Which top 3 components impress you most? Why?
			Q17: Is anything confusing, not intuitive, or too complicated to understand?
			Q18: Which part of the system do you think can be further improved? How?

APPENDIX D THE CONFIGURATION OF THE HEALTH PROFILING MODEL

The detailed configuration of the health profiling model is shown in TABLE 4.

TABLE 4: The configuration of the health profiling model.

Layer	Parameters Setting
Gate (Context)	CONV1D(1, 1, 1), Sigmoid
Gate (Motion)	CONV1D(3, 3, 1), Sigmoid
Context Encoder	(50×128, Dropout, ReLU), (128×64, Dropout, ReLU), (64×64, Dropout, ReLU)×2
Motion Encoder	GroupNorm(1, 3), CONV1D(3, 32, 10), MaxPool1D(3), (CONV1D(32, 32, 5), MaxPool1D(3))×2, Group-Norm(4, 32), GRU(32, 64)
Fusion Part	(128×128, Dropout, ReLU), (128×64, ReLU), (64×64, ReLU), (64×32, Dropout, ReLU), (32×32, Dropout, ReLU), (32×16, Dropout, ReLU), (16×16, Dropout, ReLU), (16×1, Dropout, ReLU)
Output	Sigmoid

REFERENCES

[1] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.
[2] J. D. Hamilton, *Time series analysis*. Princeton University Press, 2020.
[3] Z. Jiang, L. Lin, X. Zhang, J. Luan, R. Zhao, L. Chen, J. Lam, K.-M. Yip, H.-K. So, W. H. Wong, P. Ip, and E. C. Ngai, "A data-driven context-aware health inference system for children during school closures," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, 2023.
[4] J. L. Schafer, "Multiple imputation: A primer," *Statistical Methods in Medical Research*, vol. 8, no. 1, pp. 3–15, 1999.
[5] J. M. Bland and D. G. Altman, "Statistics notes: Cronbach's alpha," *BMJ*, vol. 314, no. 7080, p. 572, 1997.
[6] O. F. Morera and S. M. Stokes, "Coefficient α as a measure of test score reliability: Review of 3 popular misconceptions," *American Journal of Public Health*, vol. 106, no. 3, pp. 458–461, 2016.
[7] J. Castro-Piñero, E. G. Artero, V. España-Romero, F. B. Ortega, M. Sjöström, J. Suni, and J. R. Ruiz, "Criterion-related validity of field-based fitness tests in youth: A systematic review," *British Journal of Sports Medicine*, vol. 44, no. 13, pp. 934–943, 2010.