Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 18 March 2026
Revised 15 April 2026
Accepted 10 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 69–81

KEYWORDS:
Artificial Intelligence; Digital Twin; Adolescent Health; Personalized Healthcare; Predictive Disease Management; Smart Hospitals

Artificial Intelligence-Assisted Digital Twin Architecture for Personalized Healthcare Monitoring and Predictive Disease Management in Smart Hospitals

[1]Dr. A. Mary Subashini, [2]Dr Arun Khatri, [3]Akanksha Singh Chandel, [4]R. Arulmozhi, [5]Dr. Srinivasa Galu K, [6]Dr. Sumit Kushwaha, [7]R. Naveenkumar, [8]Ramesh Mylapalli,

[1]Assistant Professor, Department of Artificial Intelligence & Machine Learning, Idhaya college for women, Kumbakonam, subasinghphd2020@gmail.comORCHID ID 0009-0005-0755-6583
[2]Professor, Department of Mittal School of Business, Lovely Professional University Punjab, Pincode144402 Orcid: 0000-0002-5895-7951 Email: khatriarun@yahoo.com
[3]Assistant Professor OBG Department Saveetha Medical College and Hospital,Chennai akankshadel08@gmail.com Orchid ID: 0009-0003-6096-7352
[4]Assistant Professor, Department of IT, Al-Ameen Engineering college (Autonomous), Erode, tamilnadu, India. Email id: anbuarulmozhi@gmail.com Orchid id: 0009-0003-1576-1128
[5]Professor, General Medicine, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu 631552
[6]Associate Professor, Department of Directorate of Research, Innovation, and Consultancy (DRIC), Chandigarh University Uttar Pradesh, Unnao - 209859, Uttar Pradesh, India. Email: sumit.kushwaha1@gmail.com ORCID: 0000-0002-3830-1736
[7]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India.Email: drnk1983@gmail.com 0000-0001-9033-9400
[8]Assistant Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India mylapalli2@gmail.com https://orcid.org/0009-0007-6919-4156

ABSTRACT: Continuous integration of physiological, clinical and contextual data is required in the personalized healthcare monitoring in smart hospitals. The conventional monitoring systems have disadvantages: fixed thresholds, limited personalization and they may miss detection of deterioration. In this work, artificial intelligence-enabled digital twin architecture for personalized monitoring of adolescent health and predictive disease management is proposed. Five-minute physiological observations were used to simulate a synthetic cohort of 5,000 adolescents ages 10-19 years for 30 days. Each patient-specific twin was continuously synchronized in a two-layer gated recurrent unit with attention to estimate the risk of the metabolic, respiratory, cardiovascular and overall deterioration. The proposed model successfully simulated the accuracy of 94.3%, F1-score of 92.8%, and an area under the receiver operating characteristic curve of 0.967. Prediction latency of mean was 42 ms, and early-warning time was 7.4 h. The architecture illustrates the scalability of decision support; however, validation with clinical data from adolescents is required to implement the architecture.

1. INTRODUCTION

Wearable sensors, bedside devices, electronic health records and connected medical technologies are increasingly emerging as options for smart hospitals to deploy and support in order to facilitate continuous patient monitoring. These systems create a ton of physiological, clinical, behavioral and contextual data that can enhance early disease detection. Traditional monitoring methods typically use a fixed threshold and population-level risk model, however. They are unable to properly capture the evolution of patient status or make meaningful predictions of a patient's status. A digital twin is a new solution that allows the creation of a virtual replica of a single patient, which is continuously updated. Together with AI, these twins can detect temporal health patterns, and foretell worsening before it happens.

Wearable, clinical and contextual information is often used separately in existing monitoring systems. This disintegration can result in incomplete patient-state evaluations, delayed warnings and too many false warnings. In addition, many healthcare DT systems fail to synchronise their patient-specific models and cannot adjust their parameters based on new observations. In adolescent health, these are important restrictions due to the physiological changes that take place during development, activity and the presence of underlying chronic conditions.

The study, therefore, suggests an Artificial Intelligence enabled digital twin framework for customized adolescent monitoring at smart hospitals. The architecture involves physiological measurements, clinical characteristics, and measurements of contextual variables and observations gathered using wearable devices. A continuous synchronized patient twin facilitates the prediction of metabolic and other physiological deterioration on 6, 12 and 24 hrs time horizon. It further develops customized risk scores and virtual monitoring suggestions.

The main contributions include a multimodal patient-state representation, an attention-based temporal risk-prediction model and a personalized synchronization mechanism. The framework is assessed by synthetic adolescent profiles under normal, moderate risk and deterioration scenarios. Its assessment considers its accuracy, fidelity of the digital twins, computational performance, and its prediction capability in early-warning. Robustness is also evaluated with respect to missing observations, sensor noise and temporary loss of observations. This simulation offers the proof of concept evidence although real clinical validation is still needed prior to practical deployment.

2. RELATED WORK

A digital twin has become a virtual patient that creates an environment that mimics physiological conditions and enables a continuous health assessment [1]. The healthcare digital twin is a system that integrates the collected patient clinical data along with the measured sensor data and simulation models to aid in disease progression and treatment effects simulation [2]. Physiological models have been used in the domain of cardiovascular monitoring, glucose regulation and respiratory assessment [6]. These systems can be used to aid clinical decision making, and to detect changes in patient state [12]. A lot of the models currently exist, however, follow a fixed model, limited data sources and infrequent synchronization [11] are common in many of these existing frameworks. As a result, their virtual states might not be representative of the quickly changing patient's state.

In addition, AI can enhance automated patient monitoring [10]. The traditional techniques have been used for disease classification and risk estimation, such as logistic regression, decision trees, random forests, support vector machines, etc. These methods are of limited ability for modelling long-term physiological dependencies, but computationally efficient. The recurrent neural network or the gated recurrent unit can utilize the sensor measurements at sequential points in time and account for changing temporal relations [4]. Attention mechanisms further boosts the prediction by emphasizing more on clinically relevant observations [5]. However, numerous models make predictions at the population level, without taking into account individual patient histories [8].

The Internet of Medical Things is the networking of wearable devices, bedside sensors and electronic health records in smart-hospital monitoring systems. Edge processing is suitable for processing signals quickly, and cloud platforms can be used to provide all-round storage and model training [7]. Such systems send out alerts if the measurement goes beyond a specific limit or the risk of something happening becomes severe. But the lack of consolidated data integration, communications and gaps in sensor readings can diminish the clinical value, along with excessive false alerts.

Predictive disease management studies include those targeting metabolic risk as well as asthma exacerbations, cardiovascular abnormalities and general physiological decline. Most studies focus either on a single condition or on a single prediction horizon. They tend not to combine a disease-specific prediction with continuously updated patient simulations. There thus exists a huge research gap between the AI based prediction and the operational digital-twin modelling [9]. The existing ones usually offer a virtual physiological representation or predictive analytics but do not offer multimodal data fusion, personalized multi-horizon prediction and robustness to missing or noisy observation together with continuous synchronization of the twin. This gap is tackled by the proposed architecture for personalized adolescent monitoring in smart hospitals [3] based on an integrated and simulation based framework.

3. PROPOSED METHODOLOGY

3.1 Overall AI-Assisted Digital Twin Architecture

The architecture involves wearable devices, bedside sensors, clinical records and context information to enable continuous personalized healthcare monitoring. Physiological data from wearable devices and bedside sensors are complemented by clinical, behavioural and environmental data as shown in Figure 1. These multimodal inputs are sent to the preprocessing module to validate, impute missing values and normalize them. This processed information is then fed into the patient-state modelling module where it is used to build an individualized virtual model of a patient (adolescent).

The AI prediction engine utilizes a gated recurrent unit (GRU) with an attention mechanism to understand the changing patient state and predict the risks of diseases over multiple horizons. The digital-twin synchronization module compares the state predicted from the virtual patient model against any new measurements, and continually adjusts the virtual patient model. The "feedback" connection in Fig. 1 means the digital twin can react to changes in the patient's condition over time. The customized risk-management module then categorizes the estimated risk as low, moderate or high. It then provides simulated alerts and monitoring suggestions on the clinical dashboard. The proposed architecture, therefore, creates a closed monitoring loop, with multimodal data acquisition, patient-state modelling, prediction of temporal risk and the synchronization of the digital twin with the patient's state, finally leading to personalised decision support.

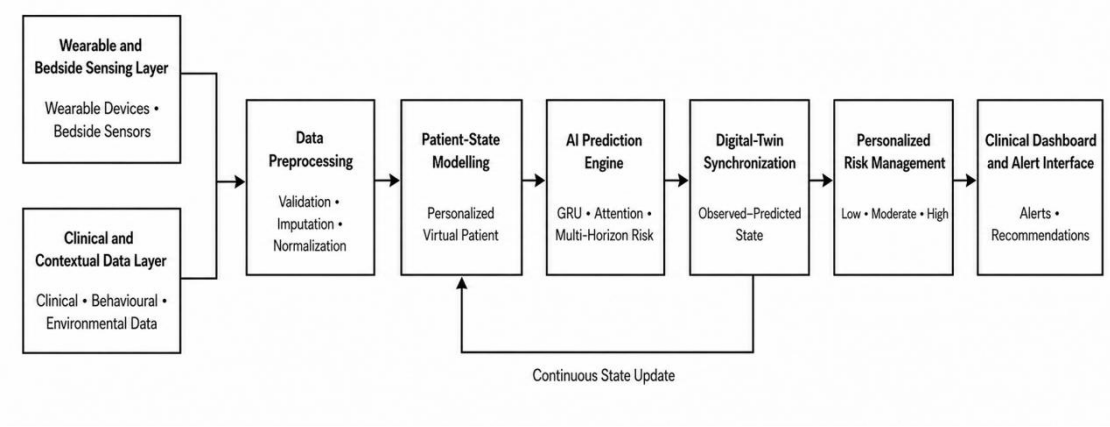

Figure 1. Overall Architecture of the Proposed AI-Assisted Healthcare Digital Twin

3.2 Multimodal Data and Synthetic Patient Population

The architecture proposed integrates the physiological, clinical, behavioral and contextual variables in a complete synthetic patient representation of each of the patients. Physiological inputs comprise heart rate, heart-rate variability, oxygen saturation, respiratory rate, body temperature, systolic and diastolic blood pressure, blood glucose and peak expiratory flow. Clinical data includes demographic data such as age, sex, BMI, medical history, family disease history, information about medications and preexisting medical conditions. Behavioural and contextual inputs include physical activity, sleep duration, stress level, medication adherence, air quality index (AQI), meal events and medication events.

There are 5,000 synthetic adolescent patient profiles (10-19 years). Physiological measurements are made every five minutes and each patient is followed for 30 days. This setup yields 288 observations/patient/day and 8,640 observations/patient for

the entire monitoring period. This gives a simulation with 43.2 million patient observations that have all been given time stamps. Table 1 summarizes the generation of all continuous variables, whose range is clinically plausible. To avoid the same health trajectories, temporal variation, measurement noise, and patient variability are introduced.

The simulated observation windows are 70% of the normal conditions. The remaining 10% correspond to deterioration events, and 20% of the states are moderate-risk. The deterioration category includes metabolic deterioration, asthma exacerbation and cardiovascular abnormality. These classes are defined by changes in a set of physiological variables, rather than having been labeled descriptively by the programmer.

The patient profiles used in the synthetic patients are split on a patient level into training, validation and testing sets. There are 3,500 patients in the training group, 750 patients in the validation group and 750 patients in the test group. Patient level separation means that each patient's record is in one group only and that information is not leaked between training and evaluation.

Table 1. Simulated Patient Characteristics and Input-Variable Ranges

Category	Variable	Simulated range or value
Population	Number of synthetic patient profiles	5,000
Population	Age	10–19 years
Population	Female patients	49%
Population	Male patients	51%
Monitoring	Monitoring duration	30 days
Monitoring	Sampling interval	5 minutes
Monitoring	Observations per patient	8,640
Physiological	Heart rate	45–180 beats/min
Physiological	Heart-rate variability	15–120 ms
Physiological	Oxygen saturation	85–100%
Physiological	Respiratory rate	10–35 breaths/min
Physiological	Body temperature	35.5–40.0°C
Physiological	Systolic blood pressure	85–160 mmHg
Physiological	Diastolic blood pressure	50–105 mmHg
Physiological	Blood glucose	55–250 mg/dL
Physiological	Peak expiratory flow	150–650 L/min
Clinical	Body mass index	14–38 kg/m²
Clinical	Family disease history	Binary: 0 or 1
Clinical	Existing chronic condition	Binary: 0 or 1
Behavioural	Daily physical activity	500–20,000 steps
Behavioural	Sleep duration	3–11 hours/day
Behavioural	Stress level	0–10
Behavioural	Medication adherence	0–100%

Contextual	Air-quality index	0–300
Contextual	Meal event	Binary: 0 or 1
Contextual	Medication event	Binary: 0 or 1
Scenario	Normal observations	70%
Scenario	Moderate-risk observations	20%
Scenario	Deterioration observations	10%
Data division	Training group	3,500 patients (70%)
Data division	Validation group	750 patients (15%)
Data division	Testing group	750 patients (15%)

3.3 Data Preprocessing and Personalized Patient-State Modelling

Collected physiological, clinical, behavioural and contextual data is range verified to identify implausible measurements. Changes out of the physiological ranges are identified and corrected or eliminated. Linear temporal interpolation (LTI) is used to estimate missing measurements up to 10%. Missing segments are imputed with feature-based median imputation for larger missing segments. There is a 5 sample window moving average filter applied to the time series to reduce random sensor noise while preserving clinically meaningful time trends. Normalizing all continuous variables by Min-Max scaling to equip features with the same measurement scales to similar influences on the training of the model. Normalized Value of feature (x) is equal to

$$x'_{i,t} = \frac{x_{i,t} - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

where $x_{i,t}$ represents the original value of a feature for patient i at time t. The terms $x_{min}$and x_{max} denote the minimum and maximum permitted values of that feature. Equation (1) transforms each continuous variable into the range $[0,1]$.

Different measurements from various sources are synchronized by using their timestamps. The synchronized observations are windowed to capture recent patient history, in a synchronized sliding window manner. If measurements are taken every 5 minutes, then there are 288 measurements in a 24 hour observation period. This window is moved by 30 minutes, to produce sequential input sequences. The class-weight adjustment is done during training to eliminate the bias against the normal class. The weight of a class k is given by:

$$w_k = \frac{N}{KN_k}, \quad (2)$$

where N is the total number of training samples, K is the number of risk classes and N_k is the number of samples belonging to class k Equation (2) assigns to greater deterioration events less importance.

After pre-processing, a customized digital twin is generated for each synthetic profile of an adolescent patient. A static clinical profile, a dynamic physiological state, a behavioural state, an environment state and a historical risk trajectory as well as patient specific parameters are present in each twin.Each twin contains a static clinical profile, dynamic physiological state, behavioural state, environmental context, historical risk trajectory and patient-specific parameters. The complete state of patient i at time t is represented as

$$S_{i,t} = [C_i; P_{i,t}; B_{i,t}; E_{i.t}; H_{i,t}], \quad (3)$$

where C_i represents the static clinical attributes of patient i. *The* vectors $P_{i,t}$, $B_{i,t}$and $E_{i.t}$ represent the current physiological, behavioural and environmental states. The vector $H_{i,t}$contains historical health measurements and previous risk estimates. As indicated in Equation (3), these multimodal components are concatenated into a single patient-state vector which is given to the temporal prediction model.

This involves predicting future events and synchronizing the digital twin.

The two-layer gated recurrent unit (GRU) network with an attention mechanism is used as the AI prediction engine to analyse the temporal variation of the patient-state sequence. Hidden size of the first GRU is 128, and hidden size of the second GRU is 64. The dropout rate is applied between layers which is set at 0.30 to prevent overfitting. Patient-specific feature-fusion brings together the static clinical information with the learned physiological, behavioural and contextual information. The attention mechanism puts more weight on observations that greatly affect health deterioration. The attention-weighted temporal representation is computed by

$$\boldsymbol{z_{i,t}} = \sum_{\tau=t-w+1}^{t} \boldsymbol{\alpha_{i,\tau}} \boldsymbol{h_{i,\tau}}, \quad \boldsymbol{\alpha_{i,\tau}} = \frac{exp(e_{i,\tau})}{\sum_{j=t-w+1}^{t} exp(e_{i,j})}, \quad \textbf{(4)}$$

where $\boldsymbol{h_{i,\tau}}$ is the GRU hidden state of patient $\boldsymbol{i}$ at time $\boldsymbol{\tau}$. The coefficient $\boldsymbol{\alpha_{i,\tau}}$ represents the normalized attention weight, while $\boldsymbol{e_{i,\tau}}$ is the relevance score assigned to the corresponding observation. As shown in Equation (4), the representation $\boldsymbol{z_{i,t}}$ is obtained by combining the hidden states according to their clinical relevance. A multi task output layer is used to compute the probabilities of metabolic deterioration, asthma exacerbation, cardiovascular abnormality, and physiological abnormality. Forecasts are made for 6 hours, 12 hours and 24 hours. The prediction function is given by

$$\hat{Y}_{i,t+h} = f_{\theta}(S_{i,t-w+1:t}), \quad h \in \{6,12,24\} hours, \quad (5)$$

where $\hat{\boldsymbol{Y}}_{\boldsymbol{i,t+h}}$ denotes the predicted risk vector for patient $\boldsymbol{i}$ at horizon $\boldsymbol{h}$. The term $\boldsymbol{w}$ represents the observation-window length, while $\boldsymbol{f_{\theta}}$ denotes the GRU-attention model with trainable parameters $\boldsymbol{\theta}$. As seen in Equation (5), the future risk estimate is created from the sequence of personalized patient states in the recent past. As the new measurements become available, a digital−twin synchronization module will compare the predicted physiological state with the observed state. The synchronization error is determined by the formula

$$E_{i,t} = \frac{1}{M} \sum_{m=1}^{M} \left| x_{i,t}^{m} - \hat{x}_{i,t}^{m} \right|, \quad (6)$$

where M is the number of monitored variables. The terms $x_{i,t}^{m}$ and $\hat{x}_{i,t}^{m}$ represent the observed and predicted values of variable mmm. The mean absolute difference (MAD) between the newly created synthetic observations and the digital-twin estimates is calculated using equation (6). The parameters of the patient-specific digital-twin are then adjusted based on

$$\phi_{i,t+1} = \phi_{i,t} + \eta(x_{i,t} - \hat{x}_{i,t}), \quad (7)$$

where $\boldsymbol{\phi_{i,t}}$ represents the patient-specific parameters at time t, and $\boldsymbol{\eta}$ is the synchronization rate. The difference between $\boldsymbol{x_{i,t}}$and $\hat{\boldsymbol{x}}_{\boldsymbol{i,t}}$ is the state-correction term. The digital twin is then updated in Equation (7) to ensure that its model stays up to date with the latest observed patient condition. The deterioration probability forecast is then translated into a risk level through the use of

$$R_{i,t} = \begin{Bmatrix} Low, & 0 \le p_{i,t} < 0.30 \\ Moderate, & 0.30 \le p_{i,t} < 0.70 \\ High, & 0.70 \le p_{i,t} \le 1. \end{Bmatrix}, \quad (8)$$

The probabilities less than 0.30 are considered as low risk, the probabilities between 0.30 and below 0.70 are considered as moderate risk and the probabilities greater than or equal to 0.70 are considered as high risk according to Equation (8). These thresholds are tested for their suitability during the simulation by sensitivity analysis.

3.5 Personalized Risk Management and Proposed Algorithm

The personalized risk-management module transforms the probabilities of diseases predicted from the model and the global deterioration probability into simulated recommendations for monitoring. Current classification of the risk, expected disease type, prediction horizon and historical classification of the risk trajectory are used to generate the recommendation. Low-risk patients will continue to be monitored on a daily basis. Patients who fall into the moderate-risk category are treated with greater care and have reminders to take their medication and recommendations for their lifestyle regarding physical activity, sleep and stress management. A high-risk classification will alert to a high-risk classification and advise urgent clinical review.

If the high-risk threshold is not met, but the moderate-risk threshold is met for two or more observation windows in a row, then this can lead to increased monitoring. In the same way, if the risk probability increases very quickly, an early-warning notification can be generated. These outputs are for simulation based clinical decision support purposes only. They are not independent diagnosis, prescription and treatment of medicines.

The entire digital-twin updating and risk-management process is shown in Algorithm 1. To start the algorithm, the patient-specific digital twin is first initialized and then multimodal health information is continuously collected. Data collected are pre-processed and presented in a synchronized observation window. The current patient state is then created and utilized to forecast risks of diseases at 6 hours, 12 hours and 24 hours. With further physiology observations, the synchronization error is computed and updated patient-specific digital-twin parameters. Then, Algorithm 1 categorises the risk and creates the associated simulated monitoring recommendation. This is done on a continuous basis when new patients details are received.

Algorithm 1. AI-Assisted Digital-Twin Updating and Predictive Disease-Risk Management

Input: Multimodal patient data $X_{i,t}$, observation-window length w, prediction horizons $h \in \{6,12,24\}$hours and risk thresholds $\tau L = 0.30$ and $\tau L = 0.70$

Output: Updated digital-twin state $S_{i,t}$, disease-risk vector$\hat{Y}_{i,t+h}$, risk category $R_{i,t}$and simulated monitoring recommendation $A_{i,t}$

1: Initialize the patient-specific digital twin

2: Acquire physiological, clinical, behavioural and contextual data

3: Validate, impute, filter and normalize the collected data

4: Temporally align the data and generate the observation window

5: Construct the current patient-state vector S(i,t)

6: Predict disease risks for 6-hour, 12-hour and 24-hour horizons

7: Receive the subsequent observed patient state

8: Calculate the synchronization error E(i,t)

9: Update the patient-specific digital-twin parameters

10: Assign Low, Moderate or High risk using the defined thresholds

11: Generate the corresponding simulated monitoring recommendation

12: Store the updated state and repeat when new data are available

The proposed framework works as a continuous feedback mechanism for the classification model as shown in Algorithm 1 and is not a one-time classification model. Low risk patients are followed routinely. High-risk patients will create an urgent alert and a clinical review recommendation, and patients who are moderate risk will receive enhanced monitoring and reminder recommendations. If a synthetic patient-profile state and associated risk estimates are repeated to adapt to new simulated observations as they are collected.

4. SIMULATION AND EXPERIMENTAL SETUP

Fifty thousand synthetic adolescent patient profiles have been followed for 30 days to assess the architecture. The simulation was written in the Python 3.11 language, with PyTorch for the GRU-attention model, and Scikit-learn for baseline implementation and evaluation. Running on the computing environment were a 12 core Intel Core i7 processor, 32 GB of RAM and a 12 GB NVIDIA GPU with Ubuntu 22.04 installed. Patient generation, data partitioning and model initialization were done using an initial random seed of 42. All results were borne from 10 independent runs where the seed for each run is different.

The patient population was split at the individual patient level ensuring 3500 patients in the training set, 750 in the validation set and 750 in the testing set. All measurements were taken over 24 hours in total with 288 measurements taken every 5 minutes. Forecasts were made for 6 hour, 12 hour and 24 hour timescales. Patient-level separation is used to ensure that patient records are not to occur in more than one partition.

The model consisted of two layers of GRU with 128 and 64 hidden units, respectively, followed by a 64-dimensional attention layer, and a multi-task output layer. The training was done with Adam, learning rate of 0.001, and dropout of 0.30. Only 100 epochs were used for training and early stopping was used after 10 epochs without validation loss. Table 2 summarizes the complete configuration.

Table 2. Simulation and Model Parameter Settings

Category	Parameter	Setting
Software	Programming language	Python 3.11
Software	ML framework	PyTorch
Software	Baseline implementation	Scikit-learn
Hardware	Processor	12-core Intel Core i7
Hardware	System/GPU memory	32 GB/12 GB
System	Operating system	Ubuntu 22.04
Reproducibility	Initial seed/runs	42/10
Dataset	Synthetic patient profiles	5,000
Dataset	Training/validation/testing	3,500/750/750
Monitoring	Period/sampling interval	30 days/5 minutes
Monitoring	Window/measurements	24 hours/288
Prediction	Prediction horizons	6, 12 and 24 hours
Model	GRU layers/hidden units	2/128 and 64
Model	Attention dimension	64
Model	Dropout	0.30
Training	Batch size	64
Training	Learning rate	0.001
Training	Maximum epochs/patience	100/10
Training	Optimizer	Adam

The proposed model was benchmarked with logistic regression, random forest, XGBoost, long short-term memory (LSTM) and GRU (gated recurrent unit) without the updating of the digital-twin model. These baselines were linear baseline, ensemble and temporal approaches, and the GRU baseline was designed to isolate the effect of the continuous twin synchronization. The same patient partitions and prediction horizon were used for all models.

Predictive evaluation included accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, precision–recall AUC, false-alert rate and Brier score. Sensitivity, F1-score and precision – recall AUC are highlighted due to the possible misinterpretation of accuracy in case of class imbalance issue. The fidelity of digital-twin was determined through the mean absolute error, root mean square error and synchronization error. Prediction latency, synchronization latency, throughput, processor utilization and memory usage were all evaluated through the computational methods.

Healthcare outcomes simulated were events of deterioration detected, missed, early-warning time, false-alert, emergency escalation and 30-day readmission. These results are comparable behaviour in the synthetic environment, rather than clinically proven.

Ten runs were performed with 95% confidence intervals reported as mean ± standard deviation. A $p<0.05$ was used to perform run-level paired comparisons between the proposed model and each of the baselines. It was reported that Cohen's (d) was used to measure effect size.

5. RESULTS AND DISCUSSION

5.1 Predictive Performance and Digital-Twin Fidelity

The AI-driven digital-twin architecture performed best in the predictive performance test among all the models tested. As summarized in Table 3, it obtained an accuracy of 94.3 ± 0.5%, sensitivity of 93.1 ± 0.6%, specificity of 94.9 ± 0.5%, F1-score of 92.8 ± 0.6% and ROC-AUC of 0.967 ± 0.004. A 95% Confidence interval of accuracy was 93.94 to 94.66 % GRU without updating with digital twin showed the greatest baseline accuracy of 92.1 ± 0.6% and F1 score of 90.2 ± 0.7%. Thus, the continuous twin updating did improve the accuracy by 2.2 % points and the F1-score by 2.6 % points, respectively.

The proposed architecture outperformed all the baseline models with an p-value < 0.001 after paired comparisons were performed over 10 simulation runs. The Cohen's d for the accuracy improvement over GRU with no twin update was 3.99, which would represent a large difference in the simulated environment. More significant differences were noted with XGBoost, random forest and logistic regression. These improvements are attributed to patient-state modelling personalized to the patient, attention-based model temporal learning and continuous synchronization of the observed and predicted patient states in time.

Table 3. Predictive Performance and Digital-Twin Fidelity of the Proposed Architecture and Baseline Models

Panel A. Predictive Performance

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	ROC-AUC	Accuracy 95% CI	*p*-value	Cohen's *d*
Logistic Regression	81.4 ± 1.1	77.2 ± 1.4	83.5 ± 1.0	76.8 ± 1.3	0.842 ± 0.011	80.61–82.19	<0.001	15.10
Random Forest	86.7 ± 0.9	83.1 ± 1.1	88.4 ± 0.9	82.9 ± 1.0	0.891 ± 0.009	86.06–87.34	<0.001	10.61
XGBoost	89.2 ± 0.8	86.8 ± 0.9	90.3 ± 0.8	86.4 ± 0.9	0.918 ± 0.007	88.63–89.77	<0.001	7.64
LSTM	91.0 ± 0.7	89.1 ± 0.8	91.9 ± 0.7	88.9 ± 0.8	0.937 ± 0.006	90.50–91.50	<0.001	5.43
GRU without Twin Updating	92.1 ± 0.6	90.4 ± 0.7	92.9 ± 0.6	90.2 ± 0.7	0.946 ± 0.005	91.67–92.53	<0.001	3.99
Proposed Digital Twin	**94.3 ± 0.5**	**93.1 ± 0.6**	**94.9 ± 0.5**	**92.8 ± 0.6**	**0.967 ± 0.004**	**93.94–94.66**	Reference	—

Panel B. Digital-twin fidelity

Fidelity Measure	Mean ± SD	95% Confidence Interval
Heart-Rate MAE	3.8 ± 0.3 beats/min	3.59–4.01
Oxygen-Saturation MAE	1.3 ± 0.1 percentage points	1.23–1.37

Blood-Glucose MAE	8.6 ± 0.7 mg/dL	8.10–9.10
Physiological-State RMSE	0.084 ± 0.006	0.080–0.088
Synchronization Error	0.061 ± 0.005	0.057–0.065

The disease-specific discrimination is shown in Figure 2. Metabolic deterioration performed the best with a ROC-AUC of 0.969; cardiovascular abnormality was 0.965 and asthma exacerbation was 0.957. All three curves was largely separated in deterioration and not deterioration states at all the classification thresholds, indicating good separation between classes. The most effective metric for detecting metabolic deterioration was changes in glucose because it yielded relatively distinct temporal patterns of physiological changes associated with glucose and body mass index (BMI) and physical activity. The discrimination for asthma prediction was slightly less as respiratory rate, oxygen saturation, and peak expiratory flow can change during normal activity and respiratory deterioration.

The end result of setting the operating points shown in Figure 2 is the optimal skewing of sensitivity and false-positive rate used when generating alerts. The sensitivities were around 0.88, 0.82 and 0.85 for metabolic deterioration, asthma exacerbation and cardiovascular abnormality at a false-positive rate (FPR) of around 0.05. These points are used to generate early warnings and reduce false warnings.

The results of the digital-twin fidelity evaluation, as shown in Table 3, indicate that the synchronized patient model had a small difference from the patient's subsequent observed synthetic states. The heart-rate MAE was 3.8 beats/min, oxygen-saturation MAE was 1.3 percentage points and blood-glucose MAE was 8.6 mg/dL. The normalized RMSE and synchronization error were 0.084 and 0.061 respectively. The results suggest high degree of agreement within the synthetic setting, but do not confer fidelity for real adolescent patients.

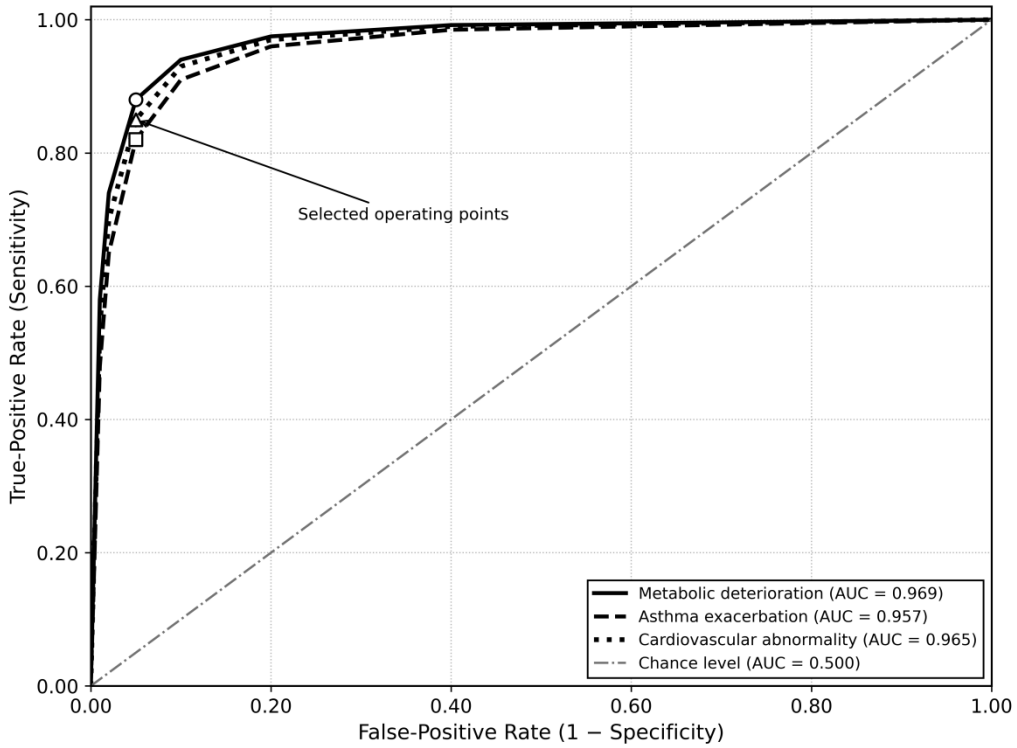

Figure 2. ROC Curves for Metabolic, Asthma and Cardiovascular Risk Prediction

5.2 Early-Warning, Computational and Robustness Outcomes

The proposed architecture enhanced the performance of early-warning over the conventional threshold-based monitoring. Conventional monitoring identified 249 of the 300 simulated deterioration events and failed to identify 51 events, resulting in

an 83.0% detection sensitivity. A proposed architecture was seen to detect 279 events, and to miss 21, for a sensitivity of 93.0%. It thus eliminated 58.8% of the missed deterioration events. The remaining false-alert rate was 7.8/100 patient-days, a decrease from 12.4/100 patient-days, which represents a 37.1% decrease.

The mean EWT went from 2.6 hours with classical monitoring to 7.4 hours with the proposed approach. This gave simulated clinical review an extra 4.8 hrs. Simulated emergency escalations fell by 94 to 71 and simulated 30-day readmissions fell by 68 to 52. These changes are equal to a decrease of 24.5% and 23.5%, respectively. These results are based on a synthesis of patient profiles, and therefore do not represent the actual comparative benefits that can be achieved in clinical settings.

The architecture had an average prediction latency of 42 ms and an average digital twin synchronization latency of 118 ms for monitoring 5,000 synthetic profiles of adolescent patient data. It made about 225 patient predictions per second, it used 46% of the available processor memory and it required 684 MB of memory. Prediction latency rose from 18ms for 500 patients to 24ms for 1,000 patients, 33ms for 2,500 patients up to 42ms for 5,000 patients. The synchronization latency rose to 69, 91 and 118ms with these same sets of patient loads. Memory grew from 312 Mb to 388 Mb, 512 Mb and 684 Mb respectively. The mild growth of these increments is a sign for having a manageable computational performance as the monitored population grew.

The results of the robustness test are displayed in Figure 3. The F1-score was found to be 92.0% when 5% of observations were not available, as indicated in Figure 3(a). It declined to 90.9% at 10%, 89.8% at 15% and 88.6% at 20% sensor-data unavailability. The greatest decrease under the worst scenario was 4.2 percentage points. As the percentage of data lost from the sensors increased, synchronization error also rose from 0.061 (when they were all observed) to 0.066, 0.073, 0.081 and finally 0.092.

The effect of measurement noise is shown in figure 3(b). The F1-score decreased from 92.8% without added noise to 92.2% at 2% noise, 91.3% at 5%, 90.1% at 8% and 89.2% at 10%. Synchronization error increased from 0.061 to 0.064, 0.070, 0.078 and 0.086 across these conditions. The performance drop-off was more gradual and did not drop below 89% even at the highest noise level.

In the case of irregular sampling, the proposed architecture has obtained an F1 score of 88.9% and synchronization error of 0.089. The F1-score dropped to 87.8% and the synchronization error rose to 0.101 when the sensor was disconnected for a short period of time. Once sensors were able to collect the performance, the digital twin was resynchronized and the performance recovered. The results given in Figure 3 overall show that the proposed architecture maintained acceptable predictive performance when considering sensor-data unavailability and measurement noise.

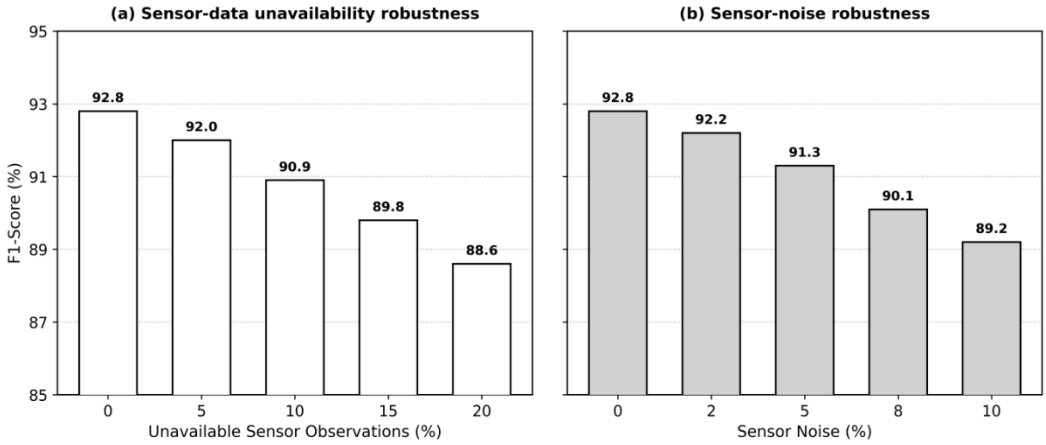

Figure 3. Robustness under Sensor-Data Unavailability and Noise

6. ETHICAL, PRIVACY, SECURITY AND CLINICAL-SAFETY CONSIDERATIONS

There was no recruitment of patients, identification of medical records nor direct clinical intervention for the simulation stage. Ethical approval and consent should be obtained from participants/guardians for further evaluations with information regarding adolescent health. Patient information needs to be protected by encryption when this is stored and transmitted. The

information should be restricted and shared on the basis of job responsibilities, and there should be a record of who accesses the data and what changes are made to the model and what action is taken in response to an alert.

The performance of models should be evaluated by age, sex, socioeconomic and clinical groups for possible bias. Any risk predictions should be left for qualified clinical review. The proposed architecture is designed to be used to support healthcare professionals and it is not meant for independence in diagnosing disease, prescribing drugs and starting treatment.

7. LIMITATIONS

The assessment was based on synthetic youth patient profiles and was not based on external clinical information. The progression of the disease considerably differs from what is assumed and may not fully reflect the physiological variability, the presence or absence of comorbidities or unexpected events that may occur in real adolescents. The architecture had also not been used in the context of a working smart-hospital environment. Thus, this integration into clinical workflows, clinician's response to alerts, alert fatigue and the patient's acceptance of these could not be explored. The predictions and health-care outcomes are therefore theoretical rather than from clinical experience. Before using in practice, validation data from actual adolescent records and prospective evaluations in hospital settings are needed.

8. CONCLUSION AND FUTURE WORK

AI-enabled digital-twin system for personalized adolescent healthcare monitoring and prediction in smart hospitals was presented. The architecture integrates physiological, clinical, behavioural and contextual data to create continuously-refreshed patient-specific representations. The GRU-attention model can be trained for metabolic, respiratory and cardiovascular risk prediction over 6 hour, 12 hour and 24 hour time periods. The simulation was quite accurate, easy to compute, and robust to sensor-data unavailability and measurement noise. The results suggest that the architecture provides the capacity to enable ongoing monitoring of adolescents and early recognition of risk.

The architecture needs to be validated in actual clinical data of adolescents from an independent dataset and prospective study conducted in actual hospital settings in future work. Future development should include support for integration with electronic health records (EHR) systems, and for providing explanations of the risks being predicted and how they are being assessed as fair or unfair among demographic and clinical groups. Federated learning can be the key to developing a privacy-preserving model across multiple hospitals. It is also important to evaluate clinical usability, alert safety, compatibility with the workflow processes and regulatory requirements prior to practical use.

REFERENCES

[1] Baker, Stephanie B., Wei Xiang, and Ian Atkinson. "Internet of things for smart healthcare: Technologies, challenges, and opportunities." *Ieee Access* 5 (2017): 26521-26544.

[2] Barricelli, Barbara Rita, Elena Casiraghi, and Daniela Fogli. "A survey on digital twin: Definitions, characteristics, applications, and design implications." IEEE access 7 (2019): 167653-167671.

[3] Björnsson, Bergthor, et al. "Digital twins to personalize medicine." Genome medicine 12.1 (2019): 4.

[4] Bruynseels, Koen, Filippo Santoni de Sio, and Jeroen Van den Hoven. "Digital twins in health care: ethical implications of an emerging engineering paradigm." Frontiers in genetics 9 (2018): 31.

[5] Che, Zhengping, et al. "Recurrent neural networks for multivariate time series with missing values." Scientific reports 8.1 (2018): 6085.

[6] Choi, Edward, et al. "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism." Advances in neural information processing systems 29 (2016).

[7] Corral-Acero, Jorge, et al. "The 'Digital Twin'to enable the vision of precision cardiology." European heart journal 41.48 (2020): 4556-4564.

[8] Islam, SM Riazul, et al. "The internet of things for health care: a comprehensive survey." IEEE access 3 (2015): 678-708.

[9] Liu, Ying, et al. "A novel cloud-based framework for the elderly healthcare services using digital twin." IEEE access 7 (2019): 49088-49101.

[10] Miotto, Riccardo, et al. "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records." Scientific reports 6.1 (2016): 26094.

[11] Rahmani, Amir M., et al. "Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach." *Future Generation Computer Systems* 78 (2018): 641-658.

[12] Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." NPJ digital medicine 1.1 (2018): 18.

[13] Shickel, Benjamin, et al. "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis." IEEE journal of biomedical and health informatics 22.5 (2017): 1589-1604.

[14] Sun, Tianze, Xiwang He, and Zhonghai Li. "Digital twin in healthcare: Recent updates and challenges." Digital health 9 (2023): 20552076221149651.

[15] Venkatesh, Kaushik P., Marium M. Raza, and Joseph C. Kvedar. "Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation." NPJ digital medicine 5.1 (2022): 150.