

Graph-enabled Reinforcement Learning for Time Series Forecasting with Adaptive Intelligence

Thanveer Shaik, Xiaohui Tao, Haoran Xie, Lin Li, Jianming Yong, and Yuefeng Li

Abstract—Reinforcement learning is well known for its ability to model sequential tasks and learn latent data patterns adaptively. Deep learning models have been widely explored and adopted in regression and classification tasks. However, deep learning has its limitations such as the assumption of equally spaced and ordered data, and the lack of ability to incorporate graph structure in terms of time-series prediction. Graphical neural network (GNN) has the ability to overcome these challenges and capture the temporal dependencies in time-series data. In this study, we propose a novel approach for predicting time-series data using GNN and monitoring with Reinforcement Learning (RL). GNNs are able to explicitly incorporate the graph structure of the data into the model, allowing them to capture temporal dependencies in a more natural way. This approach allows for more accurate predictions in complex temporal structures, such as those found in healthcare, traffic and weather forecasting. We also fine-tune our GraphRL model using a Bayesian optimisation technique to further improve performance. The proposed framework outperforms the baseline models in time-series forecasting and monitoring. The contributions of this study include the introduction of a novel GraphRL framework for time-series prediction and the demonstration of the effectiveness of GNNs in comparison to traditional deep learning models such as RNNs and LSTMs. Overall, this study demonstrates the potential of GraphRL in providing accurate and efficient predictions in dynamic RL environments.

Index Terms—Graph Neural Networks, Reinforcement Learning, Intelligent Monitoring, Bayesian Optimization

I. INTRODUCTION

The emergence of Machine Learning (ML) in healthcare signifies a paradigm shift towards automating clinician tasks and augmenting patient care capabilities [1]. Amidst the evolving ML landscape, Federated Learning has gained traction for preserving data privacy while constructing sophisticated server models [2]. Reinforcement Learning (RL), another ML strategy, has demonstrated substantial improvements in prediction performance and decision-making tasks [3], [4]. RL’s application is particularly noteworthy in controlling autonomous systems, such as robots and drones, training them to make optimal decisions in real-time based on environmental sensor data.

Thanveer Shaik and Xiaohui Tao are with the School of Mathematics, Physics & Computing, University of Southern Queensland, Toowoomba, Queensland, Australia (e-mail: Thanveer.Shaik@usq.edu.au, Xiaohui.Tao@usq.edu.au).

Haoran Xie is with the Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, Hong Kong (e-mail: hrxie@ln.edu.hk)

Lin Li is with the School of Computer and Artificial Intelligence, Wuhan University of Technology, China (e-mail: cathylin@whut.edu.cn)

Jianming Yong is with the School of Business at the University of Southern Queensland, Queensland, Australia (e-mail: Jianming.Yong@usq.edu.au)

Yuefeng Li is with the School of Computer Science, Queensland University of Technology, Brisbane, Australia (e-mail: y2.li@qut.edu.au).

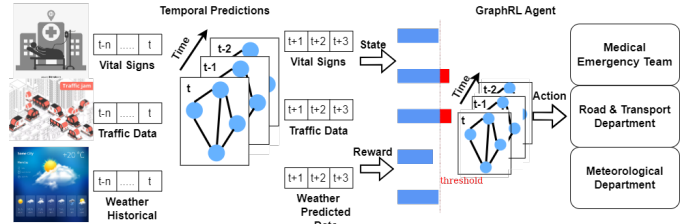


Fig. 1: Graphical Abstract

In various sectors, including healthcare, traffic, and weather forecasting, Early Warning Systems (EWS) play a pivotal role. They analyze real-time monitoring data and issue alerts for potential issues, facilitating proactive responses. RL-based EWS can adapt over time, refining their predictions and supporting clinical decision-making. This adaptability has proven effective in applications like predicting hospital readmissions and sepsis detection.

Time-series data modeling, vital in monitoring and predicting future states, has seen advancements with deep learning models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks [5]. These models are adept at capturing temporal dependencies, yet face limitations in handling irregularly structured data and integrating complex graph structures. This study introduces the GraphRL framework, an innovative amalgamation of RL and Temporal Graphical Convolutional Networks (T-GCN), aiming to surpass the constraints of traditional deep learning models in time-series prediction. GraphRL’s design facilitates handling complex temporal structures and incorporates additional information such as node and edge attributes, as depicted in Fig. 1. The GraphRL framework’s core contributions include:

- A versatile framework capable of providing early warnings and monitoring in complex settings.
- A customizable RL environment designed for effective forecasting in dynamic domains like healthcare and traffic systems.
- A novel approach to virtual monitoring of predicted states in RL, enhancing decision-making and intervention capabilities.

Our comparative analysis with state-of-the-art models across various datasets showcases GraphRL’s superior performance, underscoring its potential as a versatile solution for time-series prediction challenges.

The paper is organized as follows: Section II reviews existing literature on self-learning systems and prediction tasks. Section III outlines the research problem. The proposed

GraphRL framework and its algorithm are detailed in Section IV. Section V describes the datasets and baseline models used for evaluation. The performance of the predictive RL environment and agent is compared with baseline models in Section VI. Section VII discusses the fine-tuning of the framework’s hyperparameters using Bayesian optimization. The paper concludes in Section VIII.

II. RELATED WORKS

A. Self Learning Systems

Self-learning systems, particularly those utilizing Reinforcement Learning (RL), have seen significant advancements in various applications. For instance, Shin et al. [6] introduced a dual-agent framework in mobile health, effectively demonstrating user modeling and behavior intervention strategies. This work underscores the potential of RL in personalizing user experiences, a concept that aligns with our GraphRL framework’s goal of adaptive learning in dynamic environments. Similarly, Taylor et al. [7] applied RL in modeling maladaptive eating behaviors, further showcasing RL’s versatility in behavior prediction and modification. Chen et al. [8] developed the MIRROR framework, emphasizing the rapid learning capabilities of RL in human behavior modeling. These advancements set a precedent for our work in complex sequential decision-making tasks. Zhou et al.’s [9] CalFit app and Li et al.’s [10] method in autonomous driving highlight RL’s efficacy in personalized goal setting and complex urban scenario navigation, respectively, which are foundational to our GraphRL framework’s approach in handling dynamic and intricate patterns in data.

B. Early Detection of Patient Deterioration

In the healthcare domain, early detection of patient deterioration is vital. Traditional vital signs monitoring, as discussed by Asimwe et al. [11] and evaluated by Scully et al. [12], Baig et al. [13], and others, has laid the groundwork for our study. These works highlight the importance of continuous monitoring and early warning systems (EWS), which are integral to GraphRL’s objective. The limitations in existing methods, such as the need for manual calculations and the inability to handle large, unstructured data effectively, are addressed in our framework through the integration of GNNs, which can process complex temporal data more efficiently.

C. Vital Signs Prediction

Vital signs prediction has been explored through various machine learning models. Alghatani et al. [14] and Youssef et al. [15] demonstrated the use of traditional machine learning in mortality prediction and vital signs forecasting, respectively. Harerimana et al.’s [16] work with multi-head transformers and Xie et al.’s [17] DeepVS model highlight the potential of deep learning in this domain. However, these methods often assume equally spaced and ordered data and lack the ability to incorporate complex graph structures, limitations our GraphRL framework aims to overcome.

D. Temporal Graphical Convolutional Networks (T-GCN)

The integration of T-GCN within our GraphRL framework is pivotal. T-GCNs, known for their ability to capture temporal dependencies and complex relationships in graph-structured data, offer significant enhancements in processing time-series data [18]. This technology addresses limitations in traditional deep learning models by effectively managing irregular time intervals and integrating additional contextual information (such as node and edge attributes) for richer data representation [19]. The inclusion of T-GCN in GraphRL allows for a more nuanced understanding and prediction of dynamic systems [20], making it highly suitable for applications in healthcare monitoring, traffic forecasting, and weather prediction. The capability of T-GCNs to handle non-linear and complex temporal patterns [21] aligns with the core objectives of GraphRL, pushing the boundaries of current self-learning systems in real-world scenarios.

In summary, while existing works in self-learning systems, patient deterioration detection, and vital signs prediction have laid a strong foundation, our GraphRL framework aims to address their limitations by introducing a novel approach that combines the strengths of GNNs and RL. This approach allows for a more sophisticated handling of temporal dependencies and real-time monitoring, which is crucial in dynamic environments such as healthcare, traffic management, and weather forecasting.

The motivation behind the use of RL in our framework primarily arises from the need to tackle the challenges of multi-step time series prediction, where traditional supervised learning approaches may encounter limitations. Although supervised learning methods like GNN+Bert and GNN+TCN are indeed common and effective for time series forecasting, RL offers a unique advantage in dealing with situations where errors can accumulate over time, especially in dynamic environments. RL enables our predictive GraphRL Environment not only to forecast future states but also to actively influence decision-making, a capability particularly valuable in applications such as healthcare monitoring and the gaming industry.

III. RESEARCH PROBLEM

The research problem addresses deep learning challenges in predicting future states of a complex and dynamic Reinforcement Learning (RL) environment and adaptively learning latent behavior patterns of data.

Definition 1 (Vital Parameters and Time-Series Forecasting): In the context of our framework, we consider a set V of n vital parameters, denoted as $V_t = \{v^1, v^2, \dots, v^n\}$, which represent continuous time-series data reflecting the health status of a subject S . These vital parameters are dynamic and change over time, providing valuable insights into the subject’s well-being. To facilitate time-series forecasting, we segment these continuous vital parameters into time windows, denoted as T , which encompasses data points from the past (V_{t-2}, V_{t-1}, V_t) and extends into the future ($V_{t+1}, V_{t+2}, \dots, V_{t+n}$). Non-linear models are trained on historical data within the time windows $\{V_{t-2}, V_{t-1}, V_t\}$ to predict future values $\{V_{t+1}, V_{t+2}, \dots, V_{t+n}\}$.

Definition 2 (Learning Agents and Markov Decision Process): Following the training phase, subject S is associated with a group of learning agents that operate based on the principles of the Markov Decision Process (MDP). This MDP is a 5-tuple denoted as $M = (S, A, P, R, \gamma)$, and it forms the foundation for continuous monitoring and pattern learning of the vital parameters $V_t = \{v^1, v^2, \dots, v^n\}$. Here's a breakdown of each component:

- S represents a finite state space, where $s_t \in S$ signifies the state of an agent at a specific time t ,
- A is the set of actions available to each agent, and $a_t \in A$ represents the action taken by the agent at time t ,
- P is a Markovian transition function $P(s, a, s')$ that quantifies the probability of the agent transitioning from state s to state s' while executing action a ,
- R is a reward function $R: S \times A \rightarrow \mathbb{R}$ that provides an immediate reward $R(s, a)$ for the action a performed in state s ,
- γ is a discount factor, ranging between 0 and 1, which emphasizes immediate rewards over future rewards.

$$R(s_t, a_t) = \sum_{t=0}^{\infty} \gamma^t r_t, \quad (1)$$

This equation returns the immediate reward $R(s, a)$ for the action taken in state s , as defined in Eq. 1.

IV. GRAPHRL FRAMEWORK

To address the research problem, a novel graphical neural network (GNN) enabled reinforcement learning (RL) framework is proposed. In the graph-enabled RL framework (GraphRL), two GNNs are deployed: one for forecasting time-series data and another for Q-function approximation as shown in Fig. 2. The proposed framework is demonstrated in Fig. 2 in which the interaction between an environment and an AI agent is illustrated. As discussed in the research problem, finite MDP is adopted to formulate the process of modelling current and past states.

A. Predictive GraphRL Environment

The primary objective of the proposed study is to learn from the past and current states of a dynamic environment and predict the future states of the complex environment. To achieve this objective, we propose a predictive monitoring environment which is responsible for defining the observation space with state $s_t^i \in S$ where $i = 0, 1, 2, \dots, n$, action space with actions $a_t^j \in A$ where $j = 1, 2, 3, \dots, m$, and rewards R for each action taken by the agent as it transitions from a state s_t to s_{t+1} in a real-world scenario. For example, consider a subject in a dynamic environment whose current state is denoted as $V_t = v^1, v^2, \dots, v^n$ at time t . Similarly, the subject holds historical data of their state at times $t-1, t-2, t-3, \dots, t-n$. In traditional reinforcement learning formulations, the monitoring environment is a static entity that cannot forecast future states, which might affect the subjects in the environment.

1) *T-GCN Forecast:* Forecasting the future states of a subject before a few time steps can revolutionise the most dynamic industries such as gaming, healthcare, and so on by identifying the deteriorating state of the subject in the environment. To predict the future state in a reinforcement learning environment, a temporal graph convolutional network (T-GCN) is adopted. The graphical network is trained with past and current states at their timestamps in a supervised approach as shown in Eq. 3. The training process also includes the features leading to those states.

$$y = f\left(b + \sum_{i=1}^n v_i \cdot w_i\right) \quad (2)$$

$$y(v) = \sum_{i=1}^n \text{Activation1}(b + w_i v_i) \quad (3)$$

$$y(v) = \text{Activation1}\left(\frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}}\right)$$

Eq. 2 describes a basic neural network neuron, with y as the output, f as the activation function, b as the bias, v_i as the input features, and w_i as the weights; while Eq. 3 involves an 'Activation1' function, computing a weighted sum of inputs and normalizing these outputs, possibly into a probability distribution, akin to a softmax function.

2) *Static Spatial-temporal modelling:* A two-layered graphical network is adopted for Spatial-temporal modelling, a spatial modelling layer is based on a graphical convolutional network (GCN), and a temporal layer based on recurrent neural networks (RNN) is configured. The spatial layer is responsible to capture spatial features among nodes which are input states $s_t^i \in S$. This can be achieved by constructing Fourier transform filter and it acts on the graph nodes and its first-order neighbourhood. In this study, a static graph temporal signal is adopted in which the node positions in the graph remain the same and the label information is dynamic. The spatial layer is to set the static graph with nodes as input states $s_t^i \in S$. The two-layered GCN model is defined in Eq. 4.

$$f(X, A) = \sigma(\hat{A} \text{Relu}(\tilde{A} X W_0) W_1) \quad (4)$$

Where X is the input matrix, A represents the graph adjacency matrix, \hat{A} and \tilde{A} represent the preprocessing step and self-connection structure respectively. W_0, W_1 represents weights of the first and second layers of ST-GCN, and $\sigma(\cdot)$. $\text{Relu}()$ is an activation function.

Temporal modelling is based on the RNN variant gated recurrent unit (GRU) [22] which has a simple structure and faster training ability. In the GRU model, an update gate z_t controls the degree of information retrieved from the previous state and a reset gate r_t controls the degree of ignoring the status information at the previous moment are configured as shown in Eq. 5.

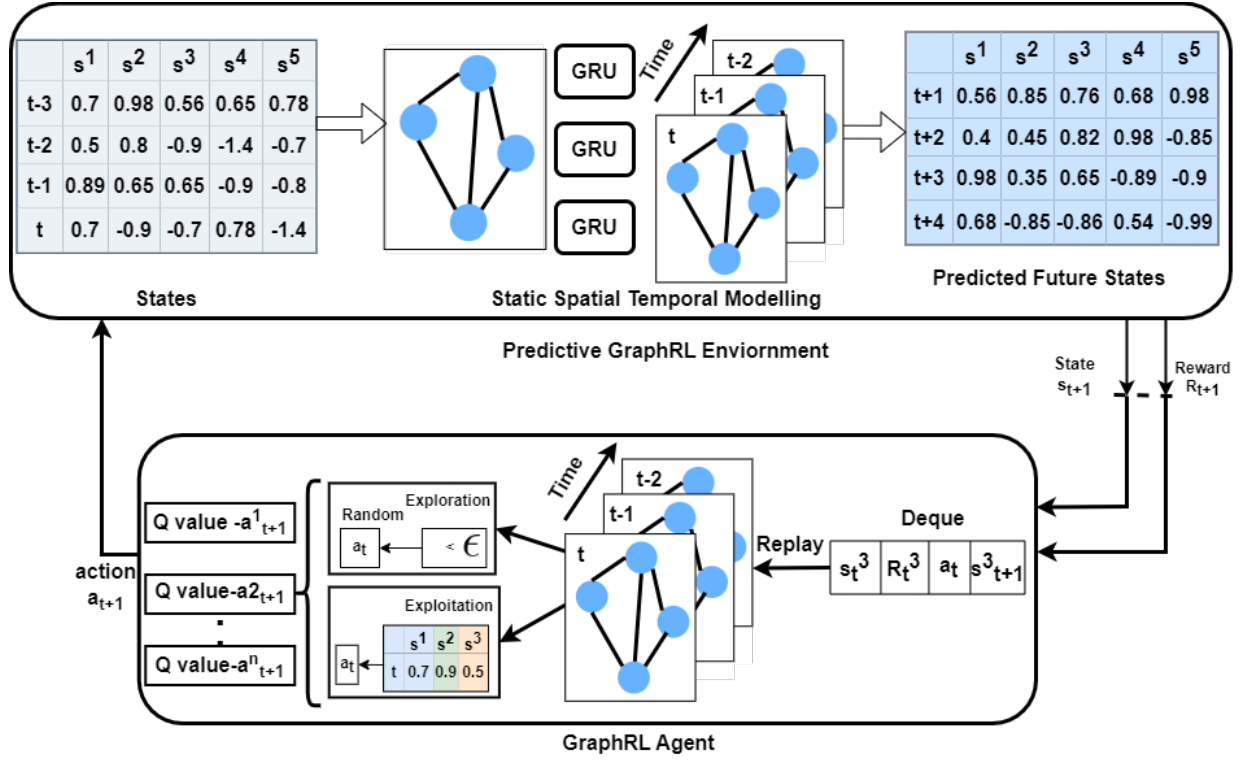


Fig. 2: GraphRL Framework

$$\begin{aligned}
 z_t &= \sigma_g(W_z[f(X_t, A), h_{t-1}] + b_z) \\
 r_t &= \sigma_g(W_r[f(X_t, A), h_{t-1}] + b_r) \\
 \hat{h}_t &= \phi_h(W_h[f(X_t, A), h_{t-1}](r_t \odot h_{t-1}) + b_h) \\
 h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h}_t
 \end{aligned} \quad (5)$$

x_t : input vector, h_t : output vector, \hat{h}_t : candidate activation vector, z_t : update gate vector, r_t : reset gate vector, W, U, b : parameter matrices and vector, σ_g : The original is a sigmoid function, ϕ_h The original is a hyperbolic tangent.

In the training process, the T-GCN model predicts the future states at time $t + 1, t + 2, t + 3, \dots, t + n$ as \hat{Y}_t and compares with the real-data Y_t . This determines the loss function of the graph network [23] as shown in Eq. 6. To avoid over-fitting problems in the training process, the loss function is optimised with L2 regularisation L_{reg} and a hyperparameter λ .

$$loss = \|\hat{Y}_t - Y_t\| + \lambda L_{reg} \quad (6)$$

B. Predictive GraphRL Environment Algorithm

Algorithm 1 outlines the creation of the Predictive GraphRL Environment, a crucial component of the proposed GraphRL framework. This environment leverages the T-GCN, chosen for its effectiveness in capturing spatial-temporal dynamics essential for complex systems. The T-GCN's ability to forecast future states, in addition to analyzing current and historical data, makes it invaluable for critical applications like health monitoring and traffic management, where early detection and timely response are key. A significant feature of this

Algorithm 1 Predictive GraphRL Environment

Ensure: **Input:** time series data $\mathcal{D} = \{s_{t-n}, \dots, s_{t-2}, s_{t-1}, s_t\}$; a set of labels $\mathcal{K} = \{1, 2, \dots, K\}$

Ensure: **Output:** Predicted time series data of \mathcal{K} , a set of labels, in the form of states $\{s_{t+1}, s_{t+2}, s_{t+3}, s_{t+4}\}$

- 1: Define $forecast_model \leftarrow T - GCNModel$
- 2: $Train(forecast_model) \leftarrow forecast_model(\mathcal{D})$
- 3: $\{s_{t+1}, s_{t+2}, s_{t+3}, s_{t+4}\} \leftarrow forecast_model(predict)$
- 4: **Initialization** : $observation_space = \{s_t^i \in S\}, action_space = \{a_t \in A\}, rewardR$
- 5: **Set** $monitor_length = N$
- 6: **if** $action$ is appropriate **then**
- 7: $R \leftarrow +reward$
- 8: **else**
- 9: $R \leftarrow -reward$
- 10: **end if**
- 11: $monitor_length \leftarrow N - 1$
- 12: $s_{t+1} \leftarrow s_t(monitor_length)$
- 13: **if** $N = 0$ **then**
- 14: $done \leftarrow True$
- 15: **else**
- 16: $done \leftarrow False$
- 17: **end if**
- 18: $visualize(a_t, R, vital\ signs)$
- 19: $initial_state \leftarrow s_t[0]$ ▷ reset environment

algorithm is its reward mechanism, which is instrumental in guiding the learning process. By awarding rewards based on the suitability of the agent's actions, the environment ensures that the agent's policy is aligned with the primary objectives of accurate forecasting and effective intervention. This design was motivated by the need for a proactive system, capable of not only forecasting but also informing real-time decision-making. The algorithm is meticulously structured to set up the observation space, action space, and reward policy based on predicted states. The initial lines (1-3) justify the use of the T-GCN model, especially for its applicability in

dynamic and nonlinear data contexts like vital sign monitoring. The subsequent lines (4-5) are dedicated to initializing the environment, forming the basis for RL-driven decision-making. The reward policy, detailed in lines 6-10, aligns with standard RL practices, promoting actions that yield beneficial outcomes. Finally, lines 11-19 focus on continuous monitoring and adaptation, a critical aspect for applications that demand real-time responsiveness, such as in healthcare scenarios. This algorithm represents a significant step in advancing the field of RL, moving from passive observation to an active role in shaping decisions.

C. GraphRL Agent

In this study, the Deep Q-Networks (DQN) algorithm is used. The Deep Q-Networks (DQN) algorithm, developed by Google’s DeepMind, was initially designed for playing Atari games. This algorithm enabled the AI to learn game strategies directly from visual input, without requiring pre-programmed rules or prior game-specific training. In this algorithm, the Q-Learning functions are approximated using the proposed T-GCN model, and the learning agent is rewarded based on the graph network prediction of the right action for the current state.

1) *Q-Function Approximation*: T-GCN model used in this study to approximate the Q-Function for each action in the action space as shown in Fig. 2. The model is configured with parameters such as the relu activation function, mean square error as loss function, and Adam optimiser. The model gets trained with the state and its corresponding action. The learning agent performs an action $a_t \in A$ for a transition from state s_t to s'_t and achieves a reward R for the action. In this transition process, the maximum of the Q-function in Eq. 7 is calculated, and the discount of the calculated value uses a discount factor γ to suppress future rewards and focus on immediate rewards. The discounted future reward is added to the current reward to get the target value. The difference between the current prediction from the neural networks and the calculated target value provides a loss function. The loss function is a deviation of the predicted value from the target value and it can be estimated from Eq. 8. The square of the loss function allows for the punishment of the agent for a large loss value.

$$Q^\pi(s, a) = E_\pi \left\{ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, \pi(s_t)) \mid s_0 = s, a_0 = a \right\} \quad (7)$$

$$loss = \left(\underbrace{R + \gamma \cdot \max(Q^{\pi^*}(s, a))}_{target_value} - \underbrace{Q^\pi(s, a)}_{predicted_value} \right)^2 \quad (8)$$

2) *Exploration and Exploitation*: The concepts of exploration and exploitation are at odds with each other. Exploration involves randomly selecting actions that have not been performed before to uncover more possibilities and enhance the agent’s understanding. Exploitation, on the other hand, entails selecting actions based on past experiences and knowledge to

maximize rewards. To balance the trade-off between exploration and exploitation, different strategies such as the greedy algorithm, epsilon-greedy algorithm, optimistic initialization, and decaying epsilon-greedy algorithm are employed. This study proposes controlling the exploration rate by multiplying the decay by the exploration rate. This approach reduces the number of explorations during execution as the agent learns patterns and maximizes rewards to achieve high scores. As the T-GCN model is retrained with previous experiences in the replay, the decay is multiplied by the exploration rate depending on the agent’s ability to predict the right actions. All these parameters are defined as hyper-parameters for DQN learning agents.

Algorithm 2 Learning Agent

```

1: Initialize  $\gamma, \epsilon, \epsilon_{decay}, \epsilon_{min}, memory = \emptyset, batch\_size$ 
2: Define  $model \leftarrow T - GCN\_model$ 
3:  $memory \leftarrow append(s_t, a_t, R, s_{t+1})$ 
4: if  $np.random.rand() < \epsilon$  then ▷ Exploration
5:    $action\_value \leftarrow random(a_t)$ 
6: else ▷ Exploitation
7:    $action\_value \leftarrow model.predict(s_t)$ 
8: end if
9:  $minibatch \leftarrow random(memory, batch\_size)$ 
10: for  $s_t, a_t, R, s_{t+1}, done$  in  $minibatch$  do
11:    $target \leftarrow R$ 
12:   if not done then
13:      $target \leftarrow R + \gamma \cdot \max(model.predict(s_{t+1}))$ 
14:   end if
15:    $target\_f \leftarrow model.predict(s_t)$ 
16:    $target\_f[a_t] \leftarrow target$ 
17:    $model.fit(s_t, target\_f)$ 
18: end for
19: if  $\epsilon \geq \epsilon_{min}$  then
20:    $\epsilon \leftarrow \epsilon \cdot \epsilon_{decay}$ 
21: end if

```

3) *GraphRL Agent Algorithm*: Algorithm 2 introduces the GraphRL Agent, presents the functionality of the GraphRL Agent within a complex action-state environment, utilizing T-GCN for Q-function approximation. This integration enables effective handling of spatial-temporal data complexities, enhancing decision-making. The algorithm’s design is founded on a strategic balance between exploration and exploitation, achieved via an epsilon-greedy strategy, crucial for adaptive learning and continual improvement in decision-making. It starts with initializing key parameters (Line 1), defining a T-GCN model for handling complex data structures (Line 2), and storing memories for experience replay (Line 3). The agent’s learning process involves iterative learning from minibatches of experiences, computing target Q-values, and adjusting its policy (Lines 8-21), with a dynamic adjustment of the exploration rate (Lines 16-18). The predict() function, pivotal in the exploitation phase, utilizes the T-GCN model’s predictions to guide actions, showcasing the algorithm’s advanced approach in navigating dynamic environments through a blend of exploration and strategic exploitation.

4) *Implementation Algorithm*: Algorithm 3 serves as the comprehensive implementation of the GraphRL framework, intricately combining the Predictive GraphRL Environment (Algorithm 1) with the GraphRL Agent (Algorithm 2). This pivotal algorithm orchestrates the real-time interactions between the agent and the environment, thus forming the operational core of the framework. It outlines the simulation scope, input parameters (subjects \mathcal{C} , vital signs \mathcal{V} , and episodes

Algorithm 3 Proposed GraphRL Framework Implementation

Require: Input:
 $\mathcal{C} = \{1, 2, \dots, C\}$: set of subjects
 $\mathcal{V} = \{1, 2, \dots, V\}$: set of vital signs
 $\mathcal{M} = \{1, 2, \dots, M\}$: number of episodes

Ensure: Output: Rewards achieved by Agents in each episode.

```

1:  $env \leftarrow ForecastingEnvironment()$  ▷ Algorithm 1
2:  $agent \leftarrow LearningAgent()$  ▷ Algorithm 2
3: for episode  $m \in \mathcal{M}$  do
4:    $state \leftarrow env.reset()$ 
5:    $score \leftarrow 0$ 
6:   for time in range(timesteps) do
7:      $a_t \leftarrow agent.action(s_t)$ 
8:      $s_{t+1}, R, done \leftarrow env.step(a_t)$ 
9:      $agent.memorize(s_t, a_t, R, s_{t+1})$ 
10:     $s_t \leftarrow s_{t+1}$ 
11:    if done then
12:       $print(m, score)$ 
13:       $break$ ;
14:    end if
15:  end for
16:   $agent.replay(batch\_size)$ 
17: end for

```

\mathcal{M}), and the output in the form of cumulative rewards. The initialization phase prepares the environment env and the agent for interaction. The episodic loop, encompassing the agent’s action-response cycle, is vital for continuous learning and adaptation. Crucial to this process is the $memorize()$ function, which stores experiences for later recall during experience replay, allowing the agent to learn from past actions and refine its decision-making strategy. This algorithm thus encapsulates the dynamic and iterative nature of the GraphRL framework, highlighting the importance of memory and experience in the realm of advanced reinforcement learning, and showcasing its functionality in complex, evolving environments.

D. Bayesian Optimisation

Bayesian optimisation is a global optimisation method that uses a probabilistic model to guide the search for the optimal solution. The model is updated as new data points are sampled and evaluated, allowing the algorithm to improve its predictions over time by fine-tuning the L_{reg} , λ and minimising the loss function defined in Eq. 6. The basic idea behind Bayesian optimisation is to model the objective function, $f(x)$, as a Gaussian process (GP). The GP model is used to predict the objective function value at any point x , given the observations of the function at other points. The prediction is given by the posterior distribution of the GP, which is a Gaussian distribution with mean and variance given by Eq. 9.

$$\begin{aligned} \mu(x) &= k(x, X)^T (K + \sigma^2 I)^{-1} y \\ \sigma^2(x) &= k(x, x) - k(x, X)^T (K + \sigma^2 I)^{-1} k(X, x) \end{aligned} \quad (9)$$

Where X is the matrix of previously sampled points, y is the vector of corresponding function values, K is the Gram matrix of the covariance function evaluated at X , and σ^2 is the noise level in the function evaluations. The next point to sample is chosen based on an acquisition function, which balances the trade-off between exploration and exploitation. Common acquisition functions include the probability of improvement and the expected improvement. Given a set of observed points (X, y) and a Gaussian process prior, Bayesian optimisation seeks the point x^* that minimises the loss function value, given

by Eq. 10. The optimisation process continues iteratively, sampling new points and updating the GP model until a stopping criterion is met.

$$EI(x) = \mathbf{E}[max(0, f(x) - f(x^*))] \quad (10)$$

Where x^* is the current best point.

V. EXPERIMENT

The primary objective of this study is to overcome deep learning challenges such as the assumption of equally spaced and ordered data [24] and the lack of ability to incorporate graph structure where the data has a complex temporal structure [25]. These challenges are particularly relevant in domains such as health, weather, and traffic where it is important to analyze temporal patterns and make accurate forecasts for early warning systems. However, traditional deep learning models often fail to capture these complex patterns, limiting their effectiveness in these critical domains.

The proposed GraphRL framework is evaluated on three different forecasting applications: heart rate prediction, traffic forecast, and weather forecast. The framework predicts future events in the form of states and optimizes actions based on those predictions. The observation space is customized for each application and actions for the agent are pre-defined. The agent receives a reward for correctly predicting a state and communicating with the relevant team. The proposed approach is a generic framework that can be applied to monitor and predict time-series data and train an RL agent to learn the latent patterns of the monitoring process. The baseline models for comparison include traditional deep learning models such as GRU, LSTM, and RNNs.

A. Datasets

The GraphRL framework’s testing with datasets from healthcare, traffic, and weather domains was a deliberate strategy to assess its versatility and robustness in handling diverse time-series data. The choice of these varied domains was intended to demonstrate the framework’s adaptability and efficacy in different contexts. Each domain poses unique challenges: healthcare data’s complexity and sensitivity, traffic data’s dynamic patterns requiring real-time analysis, and weather data’s intricate interplay of environmental factors. Successfully navigating these distinct datasets underscores the framework’s capability for widespread real-world application. Additionally, using datasets from different fields facilitates a thorough evaluation of the framework, ensuring its versatility and effectiveness across various problem types and data structures. This comprehensive approach is vital for a tool designed for extensive applications in data analysis and prediction. Three datasets utilized for evaluating the GraphRL framework as shown in Fig. I, each from a different domain: healthcare, traffic, and weather. In healthcare, the WESAD dataset, containing electrocardiogram (ECG) and photoplethysmogram (PPG) data from 17 participants, offers a rich source of biometric time-series data for pattern recognition analysis. The Los Angeles (LA) Traffic dataset, sourced from the Los

Angeles Department of Transportation (LADOT), provides real-time urban traffic data like traffic counts and speeds, while the Large-Scale Traffic and Weather Events (LSTW) dataset, with data across the United States, uniquely combines traffic conditions and weather events, posing a multifaceted challenge for the framework.

TABLE I: Datasets

| Dataset | Domain | Key Features | Statistics | Suitability |
|-----------------|------------|---|-------------------------------|--|
| WESAD [26] | Healthcare | Physiological data (ECG, PPG), motion data (accelerometers) | 17 participants | Rich biometric time-series data, ideal for testing pattern recognition in health-related data. |
| LA Traffic [27] | Traffic | Traffic counts, speeds, travel times | Data from LADOT | Real-time urban traffic data, useful for analyzing and predicting dynamic traffic flows. Combines traffic and weather data, challenging the framework to handle complex, multifactorial scenarios. |
| LSTW [28] | Weather | Traffic conditions, weather events | Data across the United States | |

B. Baseline Models

In our study, we selected three baseline models for comparison, each epitomizing state-of-the-art approaches in multi-agent forecasting, graph neural networks, and traffic prediction. These models were chosen based on their innovative methodologies and proven effectiveness in areas closely aligned with our research objectives.

- **ELMA Method** [29]: Developed by Li et al. [29], the ELMA method utilizes graph neural networks for forecasting multi-agent activities, particularly adept at handling spatiotemporal data. Its novelty lies in the use of energy-based learning, making it an excellent benchmark against our framework, which similarly leverages graph-based techniques in complex environments.
- **Self-Supervised Technique** [30]: This technique is pioneering in self-supervised learning for predicting multi-agent driving behavior. Its relevance to our study comes from its focus on behavior prediction in diverse scenarios, using self-supervised domain knowledge—an advanced trend in multi-agent learning.
- **Internet Traffic Prediction Study** [31]: It involves internet traffic prediction using distributed multi-agent learning, employing LSTM and GRU models. GRU’s superior performance in their study provides a valuable point of comparison for our research, which focuses on sophisticated learning techniques in traffic prediction.

Each of these models represents a significant stride in their respective fields. Their selection for comparison in our study is justified by their alignment with our research goals and their benchmark status in handling complex, dynamic datasets. By comparing our GraphRL framework against these models, we aim to demonstrate our approach’s novelty and effectiveness in diverse real-world applications.

C. Evaluation Metrics

Mean Absolute Error (MAE) is a widely-utilized regression metric that gauges the average magnitude of errors between predicted and actual values in a dataset. It is calculated by averaging the absolute differences between these values, yielding a singular metric. Root Mean Squared Error (RMSE)

TABLE II: Performance of the proposed framework in health forecasting

| | | 15 Min | 30 Min | 45 Min | 60 Min |
|-------------------------|------|-------------|-------------|-------------|-------------|
| ELMA [29] | MAE | 6.2 | 6.2 | 6.2 | 6.13 |
| | MAPE | 13.91 | 13.91 | 13.91 | 13.91 |
| | RMSE | 8.75 | 8.75 | 8.75 | 8.67 |
| GRU [30] | MAE | 0.95 | 0.95 | 0.97 | 0.98 |
| | MAPE | 5.47 | 5.48 | 5.51 | 5.5 |
| | RMSE | 1.25 | 1.25 | 1.27 | 1.28 |
| GRU-Based | MAE | 1.02 | 1.02 | 1.25 | 1.65 |
| | MAPE | 8 | 3.47 | 4.53 | 5.27 |
| Multi-Agent [31] | RMSE | 2.46 | 2.58 | 2.69 | 3.09 |
| | MAE | 0.56 | 0.87 | 0.68 | 0.7 |
| GraphRL (Ours) | MAPE | 2.8 | 2.9 | 2.65 | 3.98 |
| | RMSE | 1.18 | 1.47 | 1.3 | 1.32 |

is another prominent regression metric, assessing the average magnitude of differences between predicted and actual values. RMSE is computed as the square root of the mean of these squared differences. Mean Absolute Percentage Error (MAPE) represents yet another regression metric, quantifying the average absolute percentage error between predicted and actual values. It is derived by averaging the absolute differences between these values, expressed as a percentage of the actual values. Conversely, Cumulative Rewards is a performance metric specific to reinforcement learning. It measures the total rewards an agent accumulates over a specified timeframe or across a set number of actions, calculated by summing all rewards received during this period.

In the context of the experiments conducted for this study, Python version 3.7.6 served as the programming environment, with the deployment of several packages including TensorFlow, Keras, OpenAI Gym, and stable_baselines3.

VI. RESULTS AND ANALYSIS

In this section, the proposed framework performance in terms of time series forecasting and RL monitoring is compared to the baseline models in each application.

A. Predictive GraphRL Performance

Healthcare Forecasting: The proposed framework is evaluated to monitor health status by predicting future vital signs such as heart rate. Based on the sensor data and other clinical parameters such as ECG, Respiration, the time series prediction of the heart rate is conducted. The predicted values of heart for the next one hour are break-down into different time intervals (15 minutes, 30 minutes, 45 minutes, 60 minutes). Each of these time interval values acts as an observation for the GraphRL agent to monitor and communicate with the appropriate emergency team. The observation space of the vital sign, action space of different emergency teams and rewards for the agent actions in the predictive GraphRL environment are defined based on the modified early warning scores (MEWS) [32]. For the evaluation process, the WESAD dataset

TABLE III: Performance of the proposed framework in traffic forecasting

| | | 15 Min | 30 Min | 45 Min | 60 Min |
|-------------------------|------|-------------|-------------|-------------|-------------|
| ELMA [29] | MAE | 6.73 | 6.73 | 6.73 | 6.72 |
| | MAPE | 6.73 | 15.14 | 15.14 | 15.07 |
| | RMSE | 6.72 | 9.4 | 9.4 | 9.39 |
| GRU [30] | MAE | 1.04 | 1.04 | 1.04 | 1.04 |
| | MAPE | 6.04 | 6.01 | 5.96 | 6.1 |
| | RMSE | 1.36 | 1.36 | 1.36 | 1.36 |
| GRU-Based | MAE | 1.85 | 1.85 | 1.96 | 1.82 |
| Multi-Agent [31] | MAPE | 6.07 | 5.7 | 4.93 | 6.07 |
| | RMSE | 2.88 | 3.21 | 3.43 | 3.54 |
| GraphRL (Ours) | MAE | 0.65 | 0.78 | 0.64 | 0.8 |
| | MAPE | 4.1 | 7.85 | 5.65 | 7.99 |
| | RMSE | 1.27 | 1.22 | 1.26 | 1.41 |

is adopted to conduct time series forecasting of heart rate. The proposed T-GCN in the predictive GraphRL environment performs better than the other baseline frameworks ELMA, GRU, and GRU-Based Multi-Agent as shown in Tab. II. It achieves the lowest MAE, MAPE and RMSE values in all the time intervals.

Traffic Forecasting: The goal of the proposed framework is to predict traffic using the predictive GraphRL environment. The system takes in data with the following features: EventId, Type, Severity, TMC, Description, StartTime, EndTime, TimeZone, LocationLat, LocationLng, Distance, AirportCode, Number, Street, Side, City, County, State, and ZipCode. The observation space includes the current traffic state, which is represented by the traffic events and their severity in a particular region. The actions referred to possible traffic management strategies, such as altering traffic light timings or changing the speed limit. The rewards are defined based on the efficiency of the chosen strategy, such as reduced travel time or decreased congestion. For all the baseline models and the proposed framework, the MAE, MAPE, and RMSE values are reported for forecasting at 15, 30, 45, and 60-minute intervals. As shown in Tab. III, T-GCN outperforms the other models for all the forecasting intervals with the lowest MAE, MAPE, and RMSE values. The second-best performer is the GRU-Based Multi-Agent model, followed by GRU and ELMA.

Weather Forecasting: In weather forecasting, the goal of the proposed framework is to use past weather data to predict future weather events and to optimise actions based on those predictions. In the predictive environment, the observation space is configured based on both the traffic and weather events datasets, including the event type, severity, start time, end time, location (latitude and longitude), and timezone. The actions represent the decisions the RL agent can take based on the observation space. For example, the agent could decide to issue a warning or alert for severe weather, adjust traffic signals or road signs, or change the speed limit on certain roads. The agent could receive a reward for correctly predicting severe weather and issuing a timely warning. Using

TABLE IV: Performance of the proposed framework in weather forecasting

| | | 15 Min | 30 Min | 45 Min | 60 Min |
|-------------------------|------|-------------|-------------|-------------|-------------|
| ELMA [29] | MAE | 6.69 | 6.69 | 6.69 | 6.65 |
| | MAPE | 6.69 | 15.02 | 15.02 | 14.99 |
| | RMSE | 6.69 | 9.39 | 9.39 | 9.34 |
| GRU [30] | MAE | 1.03 | 1.03 | 1.04 | 1.04 |
| | MAPE | 5.96 | 5.94 | 5.93 | 6.02 |
| | RMSE | 1.36 | 1.35 | 1.36 | 1.36 |
| GRU-Based | MAE | 1.65 | 1.65 | 1.85 | 2.02 |
| Multi-Agent [31] | MAPE | 7.32 | 4.71 | 4.89 | 5.86 |
| | RMSE | 2.76 | 2.99 | 3.16 | 3.43 |
| GraphRL (Ours) | MAE | 0.61 | 0.83 | 0.66 | 0.75 |
| | MAPE | 3.95 | 5.88 | 5.15 | 7.99 |
| | RMSE | 1.23 | 1.12 | 1.28 | 1.26 |

TABLE V: Proposed GraphRL Performance

| AI Agents | WESAD | LAM Traffic Forecasting | US Weather Forecasting |
|----------------|--------------|-------------------------|------------------------|
| Q Learning | 43130 | 28840 | 39480 |
| PPO | 39480 | 33945 | 29480 |
| A2C | 41195 | 22845 | 40615 |
| Double DQN | 42615 | 25600 | 33945 |
| DDPG | 44600 | 34590 | 39945 |
| DQN | 41986 | 35219 | 40985 |
| GraphRL | 48790 | 36195 | 53145 |

the proposed GraphRL framework allows modelling the relationships between different weather events and their impact on traffic in a more efficient way than traditional machine learning methods. The GraphRL agent learns from these relationships to make better decisions and improve its predictions over time. Comparing the different models, T-GCN had the best performance across all metrics and different time intervals: 15, 30, 45, and 60 minutes, followed by GRU-Based Multi-Agent, GRU, and ELMA. The results show that the T-GCN model had the lowest MAE, MAPE, and RMSE values for all forecasting intervals, indicating its superior forecasting performance compared to the other models as shown in Tab. IV.

B. GraphRL Agent Performance

The proposed RL agent was enabled with T-GCN and its performance is compared with other traditional RL agents as shown in Tab. V. The table provides a comparison of different AI agents and their performance on three different datasets: WESAD, LAM Traffic Forecasting, and US Weather Forecasting. The performance of each agent is measured by a score, which is the total score achieved by the agent on the task over ten episodes. From the table, it can be seen

that the proposed GraphRL agent is the most efficient agent on the WESAD dataset, as it scored the highest score. The DDPG and Q-Learning agents have the second-highest score on the WESAD dataset. On the LAM Traffic Forecasting dataset, the Q-Learning agent scored the lowest, and the proposed GraphRL agent scored the highest. On the US Weather Forecasting dataset, the A2C agent scored the lowest, while the GraphRL agent scored the highest. The GraphRL agent has outperformed other RL agents in all three predictive and monitoring applications.

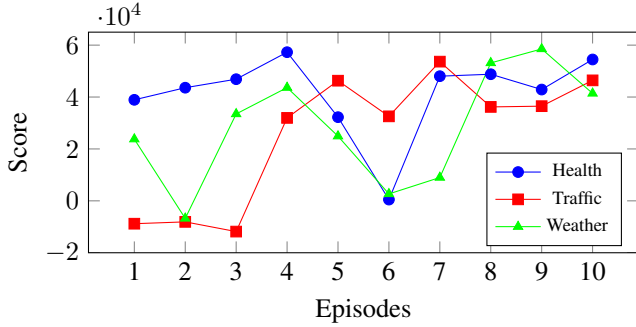


Fig. 3: GraphRL Agent Rewards Distribution

The performance of the GraphRL agent is measured by the episode score, which appears to be the total score achieved by the agent after a certain number of episodes. The breakdown of the proposed agent’s score in each episode of the three applications is presented and compared in Fig. 3. The agent’s performance on the WESAD dataset is relatively consistent, with the scores fluctuating between 32245 and 57280. On the LAM Traffic Forecasting dataset, the agent’s performance is relatively inconsistent, with the scores fluctuating between -11845 and 46295. On the US Weather Forecasting dataset, the agent’s performance is also relatively inconsistent, with the scores fluctuating between -6765 and 58530. This inconsistency of the scores is due to the exploration rate where the algorithm tries exploring all the actions randomly instead of using T-GCN model predictions.

VII. BAYESIAN OPTIMISATION RESULTS

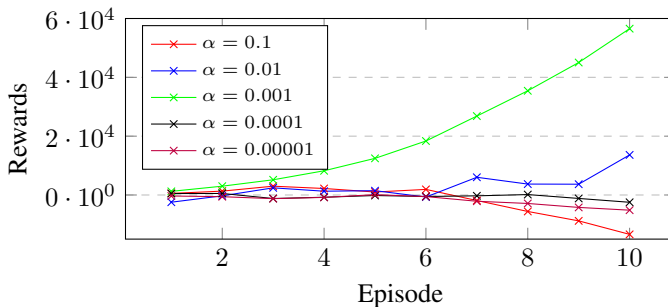


Fig. 4: Bayesian optimisation of α for GraphRL Agent

The results of Bayesian optimisation for the proposed GraphRL agent using different values of the learning rate parameter, α , during different episodes are shown in Fig. 4.

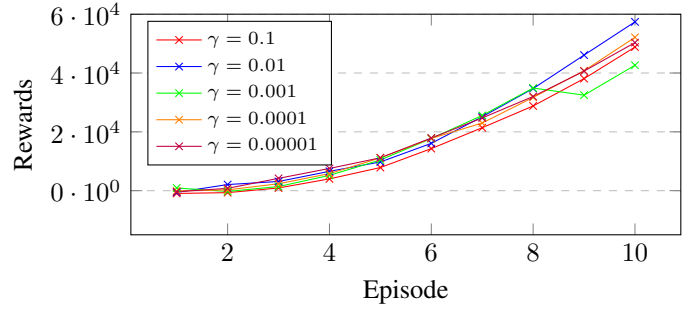


Fig. 5: Bayesian optimisation of γ for GraphRL Agent

The values in the y-axis of the line chart represent the scores or rewards obtained by the agent during each episode. It can be observed that the performance of the agent varies for different values of alpha. For example, in episode 1, the agent performs better with alpha = 0.001 (20630) compared to the other values. Similarly, in episode 10, the agent performs better with alpha = 0.001 (97380) compared to the other values. These results suggest that the optimal value of alpha for the agent is $\alpha = 0.001$, and Bayesian optimisation can be used to find the best value of alpha for a given task.

These results in Fig. 5 show the performance of an RL agent using temporal GCN for Q function approximation, using different values of the discount factor gamma. As we can see, the performance varies greatly depending on the value of the gamma chosen. A high value of gamma (0.95) results in poor performance, while lower values (0.75) result in better performance. This suggests that a lower discount factor is more appropriate, as it gives more weight to immediate rewards and less to future rewards. It also suggests that there is an optimal value of gamma, which would need to be further explored through more extensive experimentation.

VIII. CONCLUSION

The GraphRL framework, introduced in this study, embodies an innovative amalgamation of T-GCN and RL. It is specifically engineered to augment the prediction of future states in dynamic environments. Rigorous evaluations, utilizing an array of datasets such as WESAD, LA Traffic Forecasting, and US Weather Forecasting, have substantiated the framework’s enhanced performance compared to conventional RL agents. Nonetheless, it is imperative to acknowledge that the efficacy of GraphRL is significantly contingent upon the caliber of the input data and necessitates substantial computational resources. The framework’s reliance on data of high quality and structure constitutes a considerable limitation, with its accuracy and effectiveness being closely bound to the data’s integrity. Additionally, the computational requisites, predominantly due to the T-GCN model integration, present challenges in scalability and broader applicability.

Future enhancements of the GraphRL framework will be directed towards surmounting these constraints and broadening its functional scope. Prospective developments entail the incorporation of spatial data processing, aimed at bolstering the framework’s analytical prowess, particularly in processing

data with spatial or geographical dimensions. Investigating a spectrum of graph-based models could yield insights for enhancing both the efficiency and efficacy of the framework. Furthermore, the exploration of real-time adaptive learning strategies presents a promising avenue for subsequent research. Such advancements are anticipated to enable the framework to dynamically adapt to evolving data patterns and environmental shifts. In summation, the GraphRL framework signifies a substantial advancement in the domain of time-series prediction and monitoring. Its adeptness in managing complex temporal data surpasses traditional RL methodologies, heralding innovative applications in sectors such as healthcare, traffic management, and environmental forecasting. As the framework undergoes continued refinement and evolution, it is positioned to emerge as an instrumental component in the progression of predictive analytics and intelligent monitoring systems, with extensive applicability across diverse fields.

REFERENCES

- [1] A. Zhang, L. Xing, J. Zou, and J. C. Wu, "Shifting machine learning for healthcare from development to deployment and from models to data," *Nature Biomedical Engineering*, July 2022.
- [2] T. Shaik, X. Tao, N. Higgins, R. Gururajan, Y. Li, X. Zhou, and U. R. Acharya, "Fedstack: Personalized activity monitoring using stacked federated learning," *Knowledge-Based Systems*, vol. 257, p. 109929, 2022.
- [3] G. Gao, Q. Gao, X. Yang, M. Pajic, and M. Chi, "A reinforcement learning-informed pattern mining framework for multivariate time series classification," 2022.
- [4] E. M. Forman, S. G. Kerrigan, M. L. Butryn, A. S. Juarascio, S. M. Manasse, S. Ontañón, D. H. Dallal, R. J. Crochiere, and D. Moskow, "Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss?," *Journal of behavioral medicine*, vol. 42, no. 2, pp. 276–290, 2019.
- [5] T. Wang, B. Lu, W. Wang, W. Wei, X. Yuan, and J. Li, "Reinforcement learning-based optimization for mobile edge computing scheduling game," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022.
- [6] E. Shin, S. Swaroop, W. Pan, S. Murphy, and F. Doshi-Velez, "Modeling mobile health users as reinforcement learning agents," *arXiv preprint arXiv:2212.00863*, 2022.
- [7] V. A. Taylor, I. Moseley, S. Sun, R. Smith, A. Roy, V. U. Ludwig, and J. A. Brewer, "Awareness drives changes in reward value which predict eating behavior change: Probing reinforcement learning using experience sampling from mobile mindfulness training for maladaptive eating," *Journal of Behavioral Addictions*, vol. 10, no. 3, pp. 482–497, 2021.
- [8] K. Chen, J. Fong, and H. Soh, "Mirror: Differentiable deep social projection for assistive human-robot communication," *arXiv e-prints*, pp. arXiv–2203, 2022.
- [9] M. Zhou, Y. Mintz, Y. Fukuoka, K. Goldberg, E. Flowers, P. Kaminsky, A. Castillejo, and A. Aswani, "Personalizing mobile fitness apps using reinforcement learning," in *CEUR workshop proceedings*, vol. 2068, NIH Public Access, 2018.
- [10] Z. Li, "A hierarchical autonomous driving framework combining reinforcement learning and imitation learning," in *2021 International Conference on Computer Engineering and Application (ICCEA)*, pp. 395–400, IEEE, 2021.
- [11] S. B. Asimwe, E. Vittinghoff, and M. Whooley, "Vital signs data and probability of hospitalization, transfer to another facility, or emergency department death among adults presenting for medical illnesses to the emergency department at a large urban hospital in the united states," *The Journal of Emergency Medicine*, vol. 58, pp. 570–580, apr 2020.
- [12] C. G. Scully and C. Daluwatte, "Evaluating performance of early warning indices to predict physiological instabilities," *Journal of Biomedical Informatics*, vol. 75, pp. 14–21, nov 2017.
- [13] M. M. Baig, H. G. Hosseini, and M. Linden, "Machine learning-based clinical decision support system for early diagnosis from real-time physiological data," in *2016 IEEE Region 10 Conference (TENCON)*, IEEE, nov 2016.
- [14] K. Alghatani, N. Ammar, A. Rezgui, A. Shaban-Nejad, *et al.*, "Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation," *JMIR medical informatics*, vol. 9, no. 5, p. e21347, 2021.
- [15] A. Youssef Ali Amer, F. Wouters, J. Vranken, P. Dreesen, D. de Kortede Boer, F. van Rosmalen, B. C. van Bussel, V. Smit-Fun, P. Dufloot, J. Guiot, *et al.*, "Vital signs prediction for covid-19 patients in icu," *Sensors*, vol. 21, no. 23, p. 8131, 2021.
- [16] G. Harerimana, J. W. Kim, and B. Jang, "A multi-headed transformer approach for predicting the patient's clinical time-series variables from charted vital signs," *IEEE Access*, vol. 10, pp. 105993–106004, 2022.
- [17] Z. Xie, H. Wang, S. Han, E. Schoenfeld, and F. Ye, "Deepvs: A deep learning approach for rf-based vital signs sensing," in *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 1–5, 2022.
- [18] A. M. T. Elsir, A. Khaled, and Y. Shen, "Hlgt: Hybrid local–global spatio-temporal model for travel time estimation using siamese graph convolutional with triplet networks," *Expert Systems with Applications*, vol. 229, p. 120502, 2023.
- [19] S. Wang, J. Cao, and S. Y. Philip, "Deep learning for spatio-temporal data mining: A survey," *IEEE transactions on knowledge and data engineering*, vol. 34, no. 8, pp. 3681–3700, 2020.
- [20] Z. Chen, Y. Wan, Y. Liu, and A. Valera-Medina, "A knowledge graph-supported information fusion approach for multi-faceted conceptual modelling," *Information Fusion*, vol. 101, p. 101985, 2024.
- [21] F. B. Hüttel, F. Rodrigues, and F. C. Pereira, "Mind the gap: Modelling difference between censored and uncensored electric vehicle charging demand," *Transportation Research Part C: Emerging Technologies*, vol. 153, p. 104189, 2023.
- [22] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.
- [23] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.
- [24] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, and D. Shih, "Machine learning in the search for new fundamental physics," *Nature Reviews Physics*, vol. 4, no. 6, pp. 399–412, 2022.
- [25] P. Almasan, J. Suárez-Varela, K. Rusek, P. Barlet-Ros, and A. Cabellos-Aparicio, "Deep reinforcement learning meets graph neural networks: exploring a routing optimization use case," *Computer Communications*, vol. 196, pp. 184–194, 2022.
- [26] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, pp. 400–408, 2018.
- [27] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.
- [28] S. Moosavi, M. H. Samavatian, A. Nandi, S. Parthasarathy, and R. Ramnath, "Short and long-term pattern discovery over large-scale spatiotemporal data," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2905–2913, 2019.
- [29] Y. Li, P. Wang, L. Chen, Z. Wang, and C.-Y. Chan, "Elma: Energy-based learning for multi-agent activity forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, pp. 1482–1490, 2022.
- [30] H. Ma, Y. Sun, J. Li, and M. Tomizuka, "Multi-agent driving behavior prediction across different scenarios with self-supervised domain knowledge," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3122–3129, IEEE, 2021.
- [31] W. Jiang, M. He, and W. Gu, "Internet traffic prediction with distributed multi-agent learning," *Applied System Innovation*, vol. 5, no. 6, p. 121, 2022.
- [32] V. Signs, "Canberra hospital and health services clinical procedure," 2021.