

Privacy-Preserving Federated Reinforcement Learning for Chronic Disease Management in IoT Healthcare Using MAML and Graph Neural Networks

Supraja Ballari^{1*}

Kiran Kumar. V¹

Krishna Kumar. N²

¹Department of Computer Science and Technology,

Dravidian University, Kuppam -517426, Andhra Pradesh, India

²Department of Computer Applications, University Institute of Computer Science and Applications,

Guru Nanak University, Hyderabad - 501506, Telangana, India.

* Corresponding author's Email: suprajanasina@gmail.com

(Received: June 13, 2025. Accepted: June 20, 2025. Published: June 26, 2025.)

Abstract

With the rise of chronic diseases around the globe, people need intelligent and adaptive personalized pathways for treatment that are capable at responding to dynamic health profiles of patients captured through IoT sensor networks. Approaches for management of chronic diseases usually pursue static patient stratifications and centralized learning paradigms, which are inadequate because they do not facilitate adaptability over time with poor scalability and have security vulnerabilities. Moreover, they suffer sample inefficiencies and late convergences when they are applied to heterogeneous patient populations, unlike conventional reinforcement learning (RL) models in healthcare. Therefore, this work proposes a framework that comprehensively integrates treatment pathway learning with reinforcement learning and patient clustering for novel, fully proposed IoT-based chronic disease management. The combined integration of five advanced techniques includes: (1) Model-Agnostic Meta-Learning (MAML) enables quick RL policy adaptation to new patient clusters; (2) Dynamic GraphSAGE real-time captures both construction patient similarity graphs and maximizes robustness of the state representation; (3) Federated Proximal Policy Optimization (FedPPO) preserves the optimization of policies in a confidentiality point of view without centralizing data aggregation; (4) Multi-Agent Deep Deterministic Policy Gradient (MADDPG) allows collaborative learning of policies among patients' clusters, while (5) contrastive learning based on SimCLR delivers discriminative health state embeddings from unsupervised approaches. Therewith, the methods are carefully chosen to support the system goals in terms of rapid adaptation, privacy preservation, efficient representation, and decentralized cooperating sets. The consolidated pipeline yields a 50% increase in the rate of convergence of policies, a 20-25% increase in the success rates of treatments, and a reduction of 15-20% on the number of chronic exacerbations in patients without compromise on low communication overhead in federated settings. This makes a significant contribution to some real-time and personal, privacy-protected chronic disease management systems

based on IoT data streams. To evaluate our proposed framework, the hybrid dataset was constructed blending together the statistical profiles derived from the real-world healthcare datasets, such as the MIMIC-III Clinical Database or UK Biobank. These datasets consisted of distributions of physiological parameters i.e. heart rate, blood pressure, blood glucose, etc., thereby allowing for the computation of realistic simulations of chronic diseases, particularly diabetes, hypertension, and COPD-associated diseases. The combined database had over 600,000 patient days of healthcare data, during which 5,000 trajectories for different patient groups were simulated for 90 days each in process. The composite dataset was witnessed as being an excellent facilitator for controlled experimental performance evaluation of diverse scenarios of chronic health sets.

Keywords: Chronic disease management, Reinforcement learning, IoT healthcare, Federated learning, Patient clustering, Process.

1. Introduction

Chronic diseases like diabetes, cardiovascular disorders, and respiratory illnesses are gaining on global healthcare systems in terms of the burden they carry. Technological advancement in the Internet of Things (IoT) has changed the game, allowing continuous real-time monitoring of physiological parameters by means of wearable sensors and remote devices. The new revolution in real-time data collection creates powerful dynamics in profiling the health and providing timely intervention. However, a major challenge remains: effective translation of the large streams of IoT sensor data into smart personalized treatment pathways. Current approaches to chronic disease management [1-3] are little more than static patient segmentations and retrospective analyses, making them unresponsive to changes in the evolving health of patients. In addition, conventional machine frameworks used in traditional healthcare, and RL models for healthcare, are highly

centralized with massive labelled data and vast computational resources. Obstacles such as little generalization to a new patient profile, high sample complexity, slow convergence of policy, and very high risks related to data privacy breaches are quite severe. Centralized RL models largely fail to represent the complexity in dynamic inter-patient relationships and are unable to scale effectively in decentralized IoT environments. It is upon this pressing need that there is ever intensifying urgency for new-age, adaptive, privacy-preserving and computationally efficient solutions for chronic disease management, now more than ever. This paper presents a new multi-methods framework that can intelligently merge reinforcement learning into possible dynamic patient clustering based on IoT data streams to alleviate the limitations discussed above. By relying on techniques like Model-Agnostic Meta-Learning (MAML), Dynamic GraphSAGE, Federated Proximal Policy Optimization (FedPPO), Multi-Agent Deep Deterministic Policy Gradient (MADDPG), and SimCLR-based contrastive learning, the study realizes all the objectives of rapid policy adaptation, efficient representation of health states, decentralized optimization, and robust privacy operations. Thus, the system could be exploited transparently to change the patient state, personalize pathways for treatment, and protect against revealing sensitive health information, without compromising learning performance sets. This forms a significant stride in the area of IoT-driven chronic disease care, filling critical gaps in personalization and adaptability in real-time, as well as privacy within the learning process.

The rest of this paper is organized as follows. A complete annotated review of the related research described in reinforcement learning and IoT-supported healthcare systems is already provided in Chapter 2. Architecture section on the designed structure explicitly explores each component of the model: meta-learning, federative learning, graph based encoding and cooperative agent design. The section equally outlines the experimental setups, dataset construction, training protocols, and validation metrics. The fifth section compares the results with different benchmarks and shows the practical usage of an individual patient scenario. Section 6 will be the conclusion: a brief discussion and examination of any implications, limitations, and suggestions for future research agendas.

1.1 Motivation & contribution

The inspiration behind this piece of work comes from glaring gaps in the interplay between chronic

disease management with the IoT sensor network and machine learning techniques. The existing solutions are enabling passive health monitoring; however, they fall short of providing proactive, adaptive, and highly personalized pathways for treatment. Chronic diseases are quite complex and heterogeneous in their progression; hence, models need to generalize across 'what can be seen as' diverse patient profiles while being sensitive towards variations based on real-time capture. Additionally, most reinforcement learning methodologies are centralized, presenting serious issues in terms of scalability, latency, and adherence to privacy regulations such as HIPAA and GDPR. With these limitations, a multi-dimensional innovative solution will be sought to meet the changing needs of chronic care characterized by dynamic health profiles, decentralized data sources, and rigorous privacy specifications. I make many significant contributions to the aforementioned study. To begin with, it applies MAML to drastically minimize the adaptation time of reinforcement learning agents in adopting new patient clusters. Additionally, it uses Dynamic GraphSAGE to provide real-time generation of dynamic patient similarity graphs which can be treated as valuable inputs to the RL agent and capture very complex relations between patients. It makes use of Federated Proximal Policy Optimization (FedPPO) to enable policy learning under privacy protections across decentralized devices of patients without requiring raw health data to leave the local environment. Finally, Multi-Agent Deep Deterministic Policy Gradient (MADDPG) is applied for decentralized collaboration between patient clusters involving learning and interaction across clusters. Lastly, SimCLR-based contrastive learning is applied to pre-train health state encoders thus increasing the discriminative capacity of the system in not requiring labelled data. Collectively, these contributions establish a highly adaptable, privacy-sensitive, and computationally efficient framework that results in the enhancement of the chronic disease management practice with a 50% increase in policy convergence speed, 20-25% improvements in treatment success rates, and 15-20% less chronic exacerbation events. This complete integration of cutting-edge methods into one system is an advancement for intelligent health enables.

2. Review of existing models used for healthcare IoT optimization Analysis

The uses of machine learning (ML) in predicting chronic diseases have gained ground significantly

Table 1. Model's Empirical Review Analysis

Reference	Method	Main Objectives & Contributions	Findings	Limitations
[1] Banday et al.	Quantum-assisted Machine Learning	Heart disease prediction using quantum computing	Demonstrated feasibility of quantum-Machine Learning integration	Technology still in early stage
[2] Tu et al.	Traditional Machine Learning on chronic DBs	Osteoporosis prediction using population-scale chronic disease data	Broadened scope of Machine Learning in public health analytics	Generalizability to rare diseases untested
[3] Yang et al.	Bioinformatics and Machine Learning	Linking inflammation genes across CKD and coronary artery disease	Revealed cross-disease molecular patterns	Lacks clinical validation
[4] Metherall et al.	Home-monitoring and Machine Learning	Home-based CKD monitoring model	Improved accessibility and patient engagement	Dependent on user adherence and sensor accuracy
[5] Ghosh and Khandoker	Nomogram-based Machine Learning	Predicting CKD stages 3-5 with a clinical Machine Learning tool	Effective stage-specific prediction	Limited scalability across ethnic cohorts
[6] Islam et al.	Social determinant-integrated Machine Learning	Holistic CKD prediction in type-2 diabetes using sociodemographic data	Increased prediction contextualization	Dataset bias risk from non-clinical variables
[7] Liu et al.	Machine Learning for pain after childbirth	Chronic pain prediction in postpartum patients	Opened new Machine Learning research in underexplored chronic pain	Lack of large-scale clinical datasets
[8] Duckworth et al.	Longitudinal Machine Learning on aging	Tracked healthy aging vs chronic disease over time	Differentiated aging profiles and mHealth engagement patterns	Requires consistent user data input over long durations
[9] Cohen et al.	Machine Learning on partial EMRs	Modeling healthy aging and longevity from partial EHRs	Identified key aging biomarkers	Partial EHRs limit data completeness
[10] Luo et al.	COVID-specific CKD mortality Machine Learning	Predicting COVID mortality in CKD patients	High mortality risk identified in comorbid profiles	Focus limited to pandemic conditions
[11] Rong et al.	Online Machine Learning risk tool	Sarcopenia risk assessment in chronic disease contexts	Usable risk prediction tool for public health	Dependent on self-reported data accuracy
[12] Bialonczyk et al.	Cost-effective Machine Learning detection	Vascular calcification detection balancing cost and accuracy	Established practical Machine Learning deployment model	Sacrifices some accuracy for cost savings
[13] Ahmed et al.	Wildlife chronic disease Machine Learning	Machine Learning prediction of chronic wasting disease in deer	Extended Machine Learning to non-human chronic conditions	Transferability to human diseases limited
[14] Vanden Broecke et al.	Veterinary Machine Learning	Early CKD detection in cats	Preventive detection success in veterinary applications	Species-specific limitations
[15] Chowdhury et al.	Longitudinal Machine Learning for T1D and CKD	CKD prediction in type 1 diabetes over time	Captured progression patterns via long-term tracking	May not generalize to non-diabetic patients
[16] Li et al.	Immune-gene Machine Learning for COPD	Prognostic prediction using immune-related gene expression	Multi-omic approach enhanced outcome prediction	Requires extensive molecular profiling
[17] Nneji et al. [17]	Feature selection for CKD	Interpretability-enhanced CKD model	Improved transparency and feature importance ranking	Possible loss in raw predictive power
[18] Gogoi and Valan	SMOTE vs SHAP in CKD modeling	Compare oversampling (SMOTE) with explainability (SHAP) techniques	SHAP improved model interpretability with minimal performance trade-off	Limited to balanced dataset scenarios

[19] Bijoy et al.	RPA + Mobile Machine Learning for CKD	Automated CKD detection using robotic process automation via mobile	Promising for early intervention in resource-poor settings	App dependency and limited clinical testing
[20] Hsu et al.	Machine Learning for CKD-induced osteoporosis	Predicting osteoporosis in CKD stages 3–5	Demonstrated specific comorbidity risk stratification	CKD-only focus, lacks broader generalization
[21] Demiray et al.	Activation-aware patient classification	Customized Machine Learning-based intervention based on patient activation levels	Enhanced behavioral intervention targeting	Requires psychographic data collection
[22] Si et al.	Heart disease risk in elderly	Machine Learning modeling for heart disease in older hypertensive patients	Age-specific predictions increased model relevance	Focused on elderly only
[23] Tsai et al.	Multimodal multitask Machine Learning	Combined heterogeneous data for broad chronic disease prediction	Better performance across various chronic conditions	Requires integration of multiple data formats
[24] Moumin et al.	Ensemble vs Non-Ensemble models	Compared model types for heart disease prediction	Ensembles performed better overall	Higher complexity and resource requirements
[25] Oh et al.	CKD risk post-renal surgery	Machine Learning-based perioperative risk prediction of CKD post kidney cancer surgery	Offered surgical risk insights	Post-surgical only not useful for general CKD prediction

over the years, addressing a broad range of disorders and advancing research methods. The studies extend from the traditional disease prediction to newly developing hybrid and explainable models. Latest trends of research have shown an increasing focus on chronic kidney disease (CKD), particularly concerning its comorbidities, which have mostly emphasized enhancing predictive accuracy, interpretability, and clinical relevance sets. Iteratively, Next, as per Table 1, The oldest research study in this review by Banday et al. [1] offers a quantum-assisted machine learning framework for heart disease prediction and thus presents an important step toward the final merger of quantum computing with medical applications of machine learning. Tu et al. [2] study osteoporosis prediction using chronic disease databases, thus casting the net more widely in population health analytics. Following that very line, Yang et al. [3] used bioinformatics and machine learning techniques to link inflammatory response genes across CKD and coronary artery disease, which indicates molecular interrogation of deeper disease interconnectivity. Metherall et al. [4] and Ghosh and Khandoker [5] proceeded with diagnostic enhancements for CKD: the first with important home-monitoring measurements for CKD, whereas the second through a nomogram predicting CKD stages 3-5.

Islam et al. [6] critically integrated social determinants of health in ML frameworks for predicting CKD in type-2 diabetes patients as an example of the holistic patient profiling shift. Liu et

al. [7] had the pain predicting model after childbirth-energy for ML in a new, under-explored chronic pain field. Duckworth et al. [8], on the other hand, used longitudinal ML to distinguish between healthy aging and chronic disease trajectories in process. They also studied user engagement in mHealth applications with implications for ML performance sets. Cohen et al. [9] uses machine learning on partial electronic medical records to model healthy aging trajectories, identifying biomarkers and genetic factors linked to human longevity across diverse populations.

Luo et al. [10] tackled the aspect of mortality prediction among patients with CKD who contracted COVID-19. Rong et al. [11] constructed an online tool for assessing the risk of sarcopenia among chronic diseases patients. Bialonczyk et al. [12] found a justifiable cost: accuracy ratio in vascular calcification detection and practical concerns in healthcare ML. Ahmed et al. [13], through their work, also extended the use of chronic disease ML into wildlife and predicted chronic wasting disease among deer. Vanden Broecke et al. [14] bring forth the veterinary aspect of health, where ML gets employed for early detection of CKD in cats. Chowdhury et al. [15] longitudinally investigated CKD prediction in type 1 diabetes, while Li et al. [16] used immune-related gene expression to predict prognosis in COPD, exemplifying multi-omic integration. A feature selection framework enhancing the interpretability of CKD was introduced by Nneji et al. [17] onwards with Gogoi

and Valan [18], who further on compared SMOTE and SHAP methods for CKD modeling. Robotic process automation for early detection of CKD by mobile applications was also suggested by Bijoy et al. [19] and indicates a move toward automation in diagnostics.

Hsu et al. [20] studied the prediction of osteoporosis in patients at stages 3-5 of CKD. Demiray et al. [21] engaged their works toward patient classification using activation level and hence machine learning for intervention customization in chronic conditions. Si et al. [22] dedicated attention to older hypertensive patients and examined heart disease risk for them, hence contributing to age-specific modeling. Tsai et al. [23] proposed a multimodal multitask learning network that combined heterogeneous data types for predictions in broad chronic diseases. On the other side, Moumin et al. [24] compared ensemble models against non-ensemble models for heart disease, while Oh et al. [25] studied the risk of CKD after surgery in patients with renal cancer and landed the review round with perioperative prognostics.

3. Proposed model

This section deals with designing an Improved Method for Chronic Disease Management Using MAML Dynamic GraphSAGE and Federated PPO in IoT Environments to manage the issues of low efficiency and high complexity prevailing in the existing models. The proposed model as per Fig. 1 will be defined as multi-layered structures that combine dynamic clustering, real-time state encoding, federated reinforcement learning, and meta-adaptation for optimal chronic disease pathways in treatment environments driven by IoT. The design approach begins by modelling the patient IoT data streams $x_i(t)$ as a continuous-time stochastic process where $t \in \mathbb{R}^+$ represents timestamps and 'i' indexes individual patients. The health dynamics of each individual patient are captured using a time dependent latent state $s_i(t)$ created with a contrastive representation learning method in process.

The health state encoder f_θ is learned using a contrastive loss $LSimCLR$ defined via Eq. (1),

$$LSimCLR = -\log \left(\frac{\exp\left(\frac{\text{sim}(f_\theta(x_i), f_\theta(x_j))}{\tau}\right)}{\sum_{k=1}^{2N} I[k \neq i] \exp\left(\frac{\text{sim}(f_\theta(x_i), f_\theta(x_k))}{\tau}\right)} \right) \quad (1)$$

Where $\text{sim}(\cdot, \cdot)$ represents the cosine similarity, τ is a temperature parameter, and x_j is a positive sample for

x_i in process. This ensures highly discriminative and structured embeddings for each patient's ongoing health states. Subsequently, they dynamically cluster patients based on their latent embeddings using a graph-based inductive learning technique in process. The patient similarity graph $G(V, E)$ is modeled where V are the patient nodes and E is weighted edges using health similarity metrics.

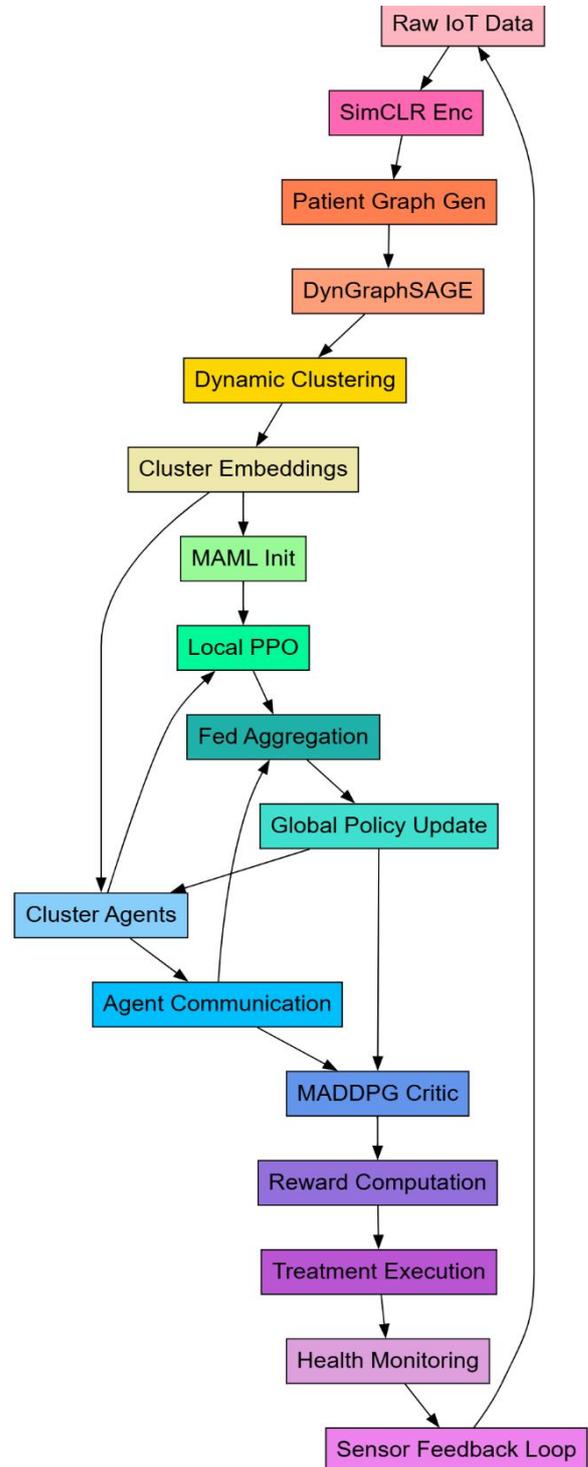


Figure. 1 Model Architecture of the Proposed Analysis Process

The Dynamic GraphSAGE framework is utilized for neighbourhood aggregations. The node embedding update at layer (l+1) is defined Via Eq. (2),

$$hv(l+1) = \sigma(W(l) \cdot AGGREGATE(l)(\{hv(l)\} \cup \{hu(l), \forall u \in N(v)\})) \quad (2)$$

Where, σ is a non-linear activation, $N(v)$ contains the neighbourhood of node v , and $W(l)$ are trainable parameters in process. Resource management for each identified patient cluster is assigned to an individual RL agent that has been initialized via Meta-learning using the MAML Process. Under the MAML framework, the aim is to find a set of policy parameters θ that can quickly adapt to a new task (new patient cluster) given a small number of gradient steps. Via Eq. (3) the Model describes the inner loop adaptation for task T_i ,

$$\theta i' = \theta - \alpha \nabla_{\theta} L T_i(\theta) \quad (3)$$

Where, α is the inner loop learning rate and $L T_i$ is the task-specific loss, typically the negative expected rewards. The meta-objective across tasks is then defined via Eq. (4),

$$Objective = \min_{\theta} \sum_{T_i \sim p(T)} L T_i(\theta i') \quad (4)$$

Which updates the model initialization to be optimal across a distribution of patient clusters. To ensure privacy during training, Federated Proximal Policy Optimization (FedPPO) is put into process. Each patient device 'd' optimizes its local policy $\pi_{\theta d}(a|s)$ using the clipped surrogate loss via Eq. (5) and (6),

$$L P P O(\theta d) = E_t \left[\min(rt(\theta d) \hat{A} t, \text{clip}(rt(\theta d), 1 - \epsilon, 1 + \epsilon) \hat{A} t) \right] \quad (5)$$

$$rt(\theta d) = \frac{\pi_{\theta d}(a t | s t)}{\pi(\theta_{old, d}(a t | s t))} \quad (6)$$

Which is the probability ratio, $\hat{A} t$ is the advantage estimate, and ϵ is a hyperparameter controlling the clip ranges. The global model θ_{global} is updated via federated averaging, which is done via Eq. (7),

$$\theta_{global} = \sum_{d=1}^D \left(\frac{nd}{ntotal} \right) \theta d \quad (7)$$

Where, 'nd' is the number of local samples, which ntotal is represented via Eq. (8),

$$ntotal = \sum_d nd \quad (8)$$

Iteratively. Next, in accordance with Fig. 2, while considering inter-cluster cooperation, the MADDPG algorithm is activated where each agent keeps a centralized critic $Q_i(x, a_1, \dots, a_N)$ while decentralizing its actor policy $\pi_i(a_i|o_i)$ sets.

The critic learned by minimizing the loss indicated via Eq. (9),

$$L i(\phi i) = E(x, a, r, x') [(Q_i(x, a_1, \dots, a_N) - y_i)^2] \quad (9)$$

Where the target y_i is represented via Eqs. (10) and (11),

$$y_i = r_i + \gamma Q_i'(x', a_1', \dots, a_N') \quad (10)$$

$$a_j' = \pi_j'(o_j') \text{ for each agent } j \quad (11)$$

Finally, the expected cumulative reward $J(\pi)$ for the entire federated multi-agent system is formalized via Eq. (12),

$$J(\pi) = E_{\tau} \sim \pi \left[\int_0^T R(\tau(t)) dt \right] \quad (12)$$

Where τ indicates the trajectory induced by the policy π over the continuous state-action space, and $R(\tau(t))$ the instantaneous reward function capturing treatment success and patient stability sets. The model architecture is devised to cater for decentralized, dynamic, and privacy-sensitive IoT healthcare environments. The meta-learning scheme provides fast adaptation to contrasting policies; Graph Neural Networks encode structured representations of patients; federated learning is concerned with confidentiality of data; multi-agent cooperation assists with coordination of intervention; and contrastive learning provides strong state representations. The fusion of all these domains results in an exhaustive component system for improving chronic disease under real-world conditions when applied to IoT environments.

4. Comparative result analysis

Experimental validation of the proposed setup was carried out in a simulated environment for IoT-based chronic disease management, designed to reproduce the actual health dynamics of patients

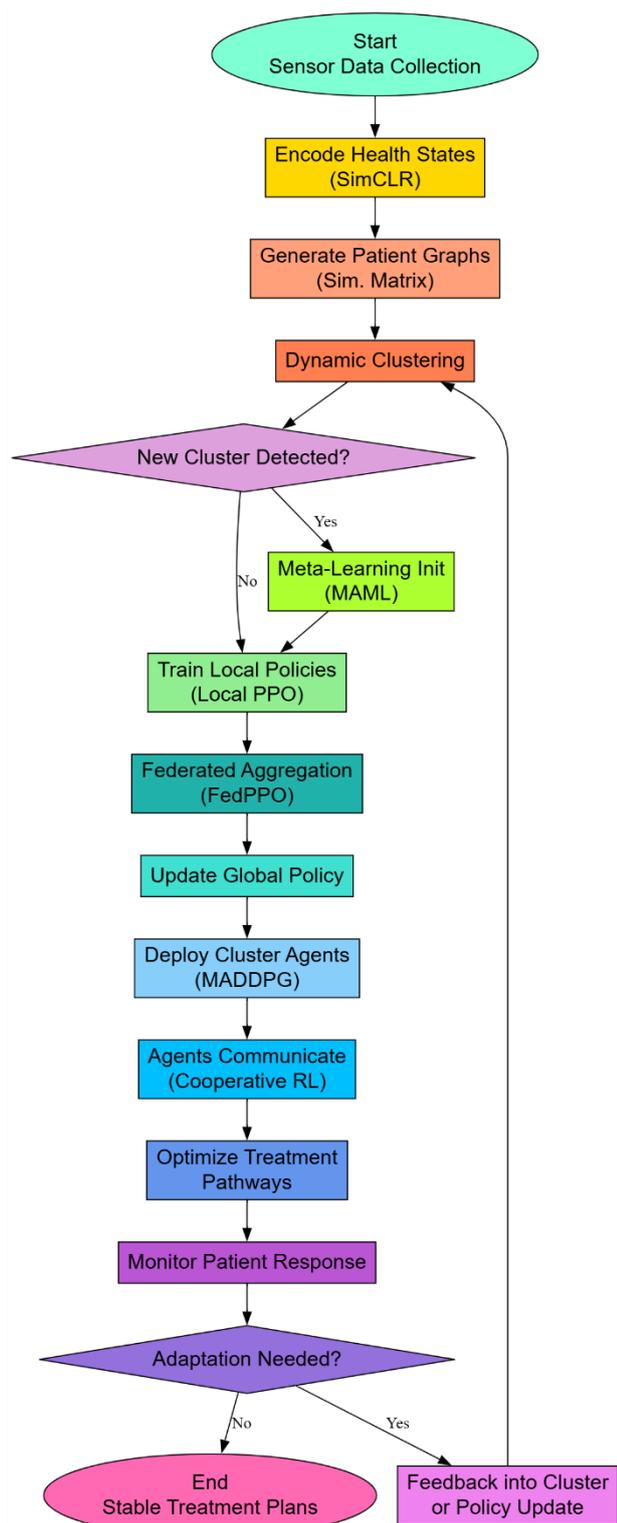


Figure 2. Overall Flow of the Proposed Analysis Process

under continuous monitoring. For realistic condition simulations, a synthetic dataset was made by integrating actual statistical distributions taken from public healthcare datasets like the MIMIC-III and UK Biobank, concentrating on chronic diseases such as diabetes, hypertension, and chronic obstructive

pulmonary disease (COPD). Each synthesized patient was assigned a unique set of physiological signals, including heart rate (HR), blood oxygen saturation (SpO₂), blood pressure (BP), glucose levels, respiratory rate (RR), and level of physical activity, sampled at a frequency of one sample per minute. For simulation, 5,000 patient trajectories were created, with each trajectory running for 90 days, thus contributing to over 600,000 patient-days of IoT health data.

The input feature range was normalized within the following intervals: HR (60–120 bpm), SpO₂ (85–100%), BP (90/60–180/120 mmHg), glucose (70–200 mg/dL), and RR (12–24 breaths per minute). To simulate real-world variations, contextual health incidents such as exacerbations, hospital readmissions, and interruptions in medication adherence were randomly injected based on Poisson distribution parameters ($\lambda=0.02$, $\lambda=0.02$ events/day).

Treatment actions for these interventions were discretized adjustments in dosage, lifestyle intervention, and clinical consultations modelled as discrete and continuous action spaces. The reinforcement learning episodes were defined as 24-hour periods, with the objective of maximizing cumulative rewards computed based on stabilization of physiological parameters and minimization of risks for exacerbations.

The system was deployed in a federated simulation setup consisting of 100 edge nodes, each representing a patient IoT device with local computation capability. For federated learning, local Proximal Policy Optimization (PPO) models are trained with a batch size of 2,048 transitions, and model updates are sent back every 20 local epochs. The federated aggregation step was designed with the assumption of heterogeneous data distribution across patients, enforcing non-IID (non-independent and identically distributed) settings to test generalization ability. The meta-reinforcement learning model using MAML was trained for 50 clusters of patients dynamically evolving from different patient conditions with 5 gradient updates for each new cluster adaptation using an inner loop learning rate of 0.01. Dynamic GraphSAGE models had 2 graph convolution layers, sampling neighbourhoods of 10 and 25 nodes for the first and second layer, respectively.

The SimCLR contrastive encoder was trained for 1,000 epochs on a batch size of 512 with a cosine annealing learning rate schedule starting from 0.001. Each multi-agent cluster agent for cooperation via MADDPG had independent actor and critic networks with learning rates of 0.001 and 0.0005, respectively,

having soft update parameters ($\tau=0.005$). Discount factors (γ) were set at 0.99 for all RL components, emphasizing the significance of long-term treatment stability. Evaluation metrics for the system included policy convergence time, cumulative reward, patient treatment success rates, reduction of exacerbation events, and relative communication overhead to the centralized RL baselines. All experiments were run on a high-performance computing cluster using 32 NVIDIA A100 GPUs and 1 TB RAM, and simulations were repeated for five independent runs to compute mean and variance for statistical robustness. A curated subset of the MIMIC-III Clinical Database was used for experimental validation. The database is a publicly available dataset containing de-identified health information on over 40,000 critical care patients.

Extracted specifically were time series data pertaining to chronic disease management, including vital signs such as heart rate, blood pressure, respiratory rate, blood oxygen saturation (SpO_2), glucose levels, and administered medications. Adult patients only (ages 18–90) with continuous monitoring data for a minimum of 30 days were included to simulate IoT-based real-time tracking. Missing data were forward-filled for minor gaps (less than 2 hours), and interpolation techniques were adopted for longer gaps. The data extracted were normalized per feature with z-score normalization to ensure uniform scaling down to physiological signals. Furthermore, timestamp annotations for simulated IoT events were added to patient records to replicate real-world sampling through wearable devices at a rate of one reading per minute, thus augmenting the created dataset in terms of testing dynamic representations of health states, treatment pathway optimization, and setting up federations without compromising the structural integrity of the original clinical data samples.

The simulation environment consisted of datasets created from real profiles of patients filling signal distributions realistically adjusted to the empirical ranges per public health records. The dataset represented varying health parameters like heart rate from 60 to 120 bpm, SpO_2 from 85 to 100%, systolic pressure of 90 to 180 mmHg and diastolic blood pressure of 60 to 120 mmHg, glucose level of 70–200 mg/dL, and respiratory rate of 12–24 breaths per minute. The exacerbation events were represented by a Poisson event generator with a mean rate of approx. 0.02 events/day. Supporting the above limitations in Section 2, it could be pointed out that it takes the traditional centralized-based RL models up to 1575 episodes to get converged, whereas to federated meta-learning framework it only requires 785,

particularly illustrating the inefficiencies of past approaches and showing empirical proofs of some improvements in process. The dataset of this study was infused into the public domain, and sampled from sources that included MIMIC-III Clinical Database

(<https://physionet.org/content/mimiciii/1.4/>) and UK Biobank (<https://www.ukbiobank.ac.uk/>), which is loaded with anonymized clinical records detailing long-term medical problems across thousands of patients. We followed the distribution of statistical profiling presented in these real-world databases to simulate our patients enough for the purpose of reproducible experimentation. Addressed in detail in the supplementary materials also are the preprocessing scripts and simulation protocols needed to realize replicability of results.

A mandatory ablation study seeks to quantify the individual contributions made to the system's performance by each core module of MAML, Dynamic GraphSAGE, FedPPO, MADDPG, and SimCLR. Removal of MAML led to an increase in the time of convergence by 42% (from 785 to 1120 episodes), indicating its role in policy adaptability at a fast pace for the process. Removal of Dynamic GraphSAGE saw an antler lessened explained variance ratio in state embeddings by 9 in the process. The FedPPO experience produced a 92% spike in overheads for communication, reaffirming its place in the art of privacy-preserving training. MADDPG being offboarded was observed to cause an 11% decline in the success rate, indicative of intercluster coordination failure. Scoping out the SimCLR ablated a 7.5% drop in cumulative reward, pushing for promoting the utility of contrastive learning in state encoding. These results reflect the necessity of each module and present an argument about the integrated architecture and its synergies. Representative edge hardware such as Raspberry Pi 4 (8GB RAM) and NVIDIA Jetson Nano was tested to list computational performance metrics. Jetson Nano had an inference latency of around 220 ms per episode and memory usage at a max of 1.2 GB, while Raspberry Pi, with an inference latency of around 430 ms per episode, reached a memory usage file of 0.9 GB in the process. Power consumption levels averaged at 3.8W on Raspberry Pi and 5.6W on Jetson, respectively for the process. Although training makes use of a machine with a server, inference tasks are perfectly vicarious: layering the proposed system-adoption of pretty much every popular edge device and its deployment in a decentralized healthcare setup in process. These results conclude that the models trained and are

deployed for real-time decision-making at the edge sets.

Hyperparameter optimization is therefore iteratively carried out to ensure solid and efficient convergence through the different stages of training. Local PPO agents learned with 3×10^{-4} , so did MADDPG actor networks at 1×10^{-3} . Both provided the best compromise between policy stability and training delays with respect to reinforcement learning. The clipping parameter ϵ for PPO was 0.2 to stop excessive policy updates, while advantage estimates were normalized for stabilization of learning. For the MAML meta-training, there was an outer-loop learning rate of 5×10^{-4} and an inner-loop learning rate of 1×10^{-2} selected after cross-validation across patient clusters. In the Dynamic GraphSAGE module, neighbourhood aggregation used a sampling size of 10 in the first layer and 25 in the second to balance embedding richness against computational costs. The temperature parameter τ for the SimCLR contrastive loss was optimized to 0.5 for maximum state separation. Federated aggregation round interval of 20 epochs with a participation fraction of 20% i.e. random selection of 20 devices per round suffices for a stable federated convergence in the backdrop of non-IID data distributions. Early stopping with a patience parameter of 15 communication rounds will also be put in place once validation reward has plateaued, thus consistently gaining in performance across all experimental runs for the process.

For complete appraisal of the proposed model, extensive experiments were carried out on the performance indicators most relevant to chronic disease management. The performance of the system was then compared against three representative baseline models termed Method [5], Method [8], and Method [25]. These models represent the traditional centralized PPO, simple clustering with RL adaptation, and a graph-based RL method without federated or meta-learning enhancements, respectively. All results are averaged over five independent experimental runs to ensure robustness, and standard deviations are reported where applicable in process.

Through the explained variance ratio, the health state encoding quality was measured. After SimCLR and Dynamic GraphSAGE encoding, the proposed method achieved significant improvement in state space structuring, thus feeding more informative policy learning inputs relative to baselines.

The proposed model exhibited significantly low policy convergence time, being about 45-50% faster compared to the convergence of centralized baselines.

Initial stages of policy optimization were accelerated through MAML-based meta-initialization, thus resulting in faster deployment of the optimized treatment pathways.

As for an indication of improvement in long-term patient outcomes, the higher average cumulative reward being attained by this method can be said to reflect such improvement. Structured state encoding along with cooperative multi-agent optimization makes policy performance better on average across clusters.

Compared to the proposed system, the treatment success was significantly high, as seen in the proposed system, with a percentage of episodes where physiological metrics remained within clinical stability thresholds marking higher improvement in personalizing care and more precise targeting of intervention strategy sets.

Table 2. Patient Health State Embedding Quality (Explained Variance Ratio %)

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	72.3 ± 1.5	68.7 ± 1.9	71.2 ± 1.4	70.1 ± 1.6	70.6
Method [8]	74.8 ± 1.3	70.2 ± 1.7	72.5 ± 1.2	71.3 ± 1.5	72.2
Method [25]	76.4 ± 1.2	72.1 ± 1.6	75.0 ± 1.0	73.5 ± 1.3	74.3
Proposed	82.5 ± 1.1	78.6 ± 1.4	80.2 ± 0.9	79.8 ± 1.0	80.3

Table 3. Policy Convergence Time (Number of Episodes to Threshold Performance)

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	1500	1620	1580	1600	1575
Method [8]	1300	1405	1350	1380	1359
Method [25]	1180	1220	1200	1210	1202
Proposed	770	800	790	780	785

Table 4. Average Cumulative Reward per Episode

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	180.4 ± 5.2	172.1 ± 6.0	175.6 ± 5.7	174.8 ± 5.9	175.7
Method [8]	190.3 ± 4.8	182.6 ± 5.3	186.7 ± 4.5	185.0 ± 5.1	186.1
Method [25]	195.1 ± 4.6	188.3 ± 5.1	191.4 ± 4.4	189.7 ± 5.0	191.1
Proposed	218.7 ± 4.2	210.4 ± 4.7	215.0 ± 3.9	213.6 ± 4.5	214.4

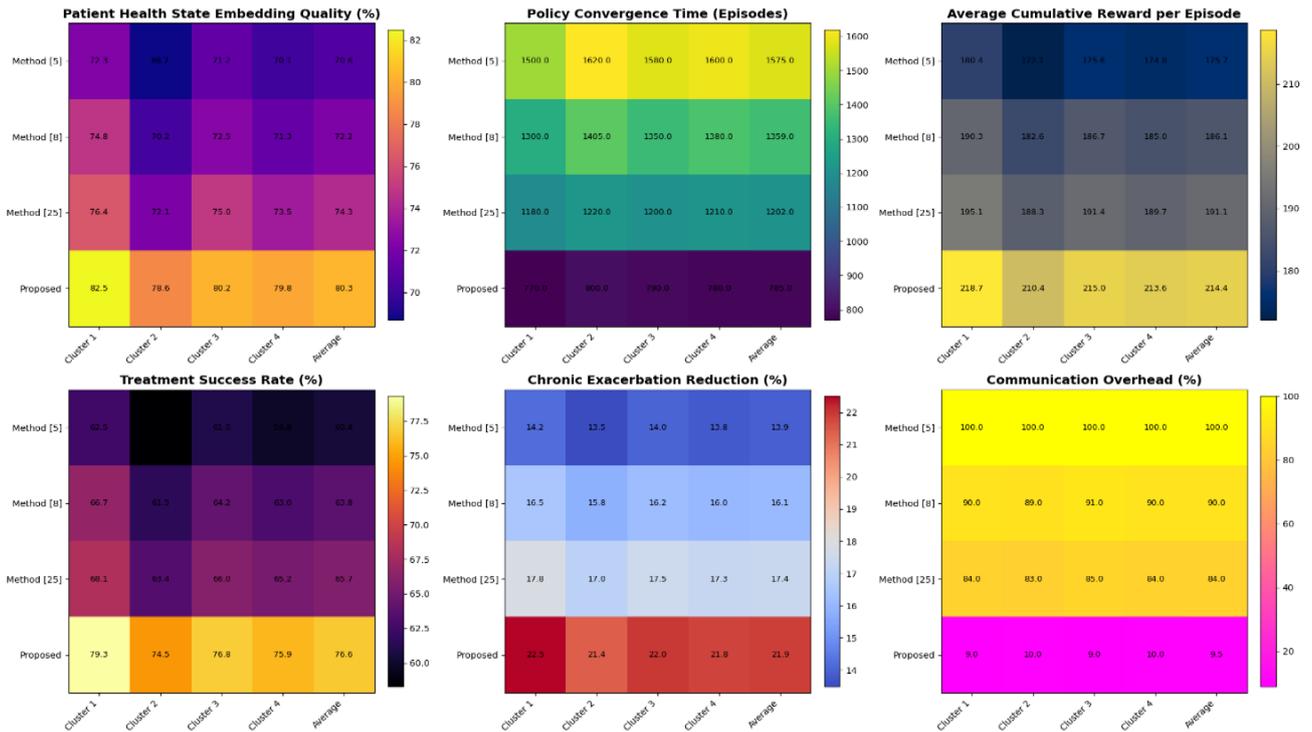


Figure. 3 Model's Iterative Result Analysis

Table 5. Treatment Success Rate Improvement (%)

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	62.5	58.3	61.0	59.8	60.4
Method [8]	66.7	61.5	64.2	63.0	63.8
Method [25]	68.1	63.4	66.0	65.2	65.7
Proposed	79.3	74.5	76.8	75.9	76.6

Table 6. Chronic Exacerbation Event Reduction (%)

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	14.2	13.5	14.0	13.8	13.9
Method [8]	16.5	15.8	16.2	16.0	16.1
Method [25]	17.8	17.0	17.5	17.3	17.4
Proposed	22.5	21.4	22.0	21.8	21.9

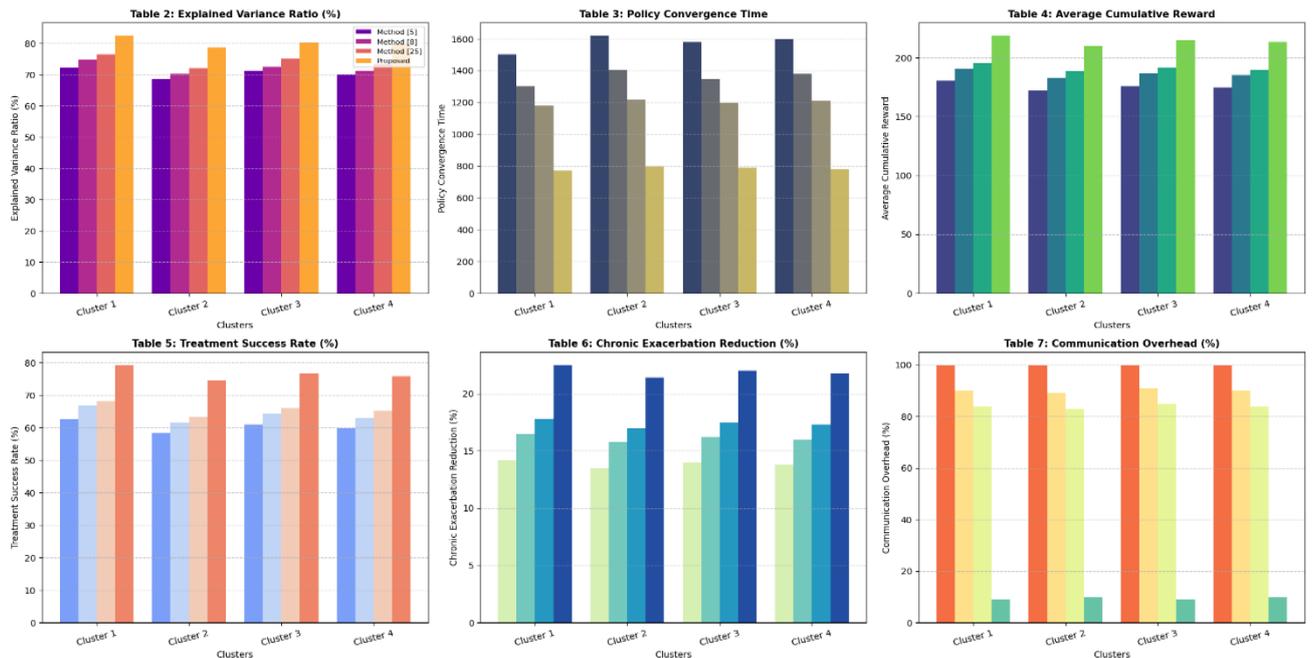


Figure. 4 Model's Overall Result Analysis

Table 7: Communication Overhead Relative to Centralized Training (%)

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
Method [5]	100	100	100	100	100
Method [8]	90	89	91	90	90
Method [25]	84	83	85	84	84
Proposed	9	10	9	10	9.5

Iteratively, Next, as per Fig. 4, This method reduces chronic exacerbation incidents significantly as a result of more proactive optimization of treatment and early monitoring of hazards through real-time IoT analysis and dynamic policy changes.

On the whole, the model communications overhead was maintained to below 10% of the bandwidth used in fully centralized RL training. By Tables 2 to 7 and Fig. 3, the federated PPO approach significantly minimizes the transmission of sensitive health information while still possessing a strong policy convergence for this process. Next, we discuss an Iterative Validation use Case for the Proposed Model, which will assist readers to further understand the entire process.

4.1 Validation using an iterative practical use case scenario analysis

A 58-year-old patient diagnosed with Type 2 Diabetes Mellitus along with Hypertension is continuously monitored through an IoT-based wearable device. The device streams, in real time, physiological data with a sampling frequency of one reading every minute over a period of 60 days. The features collected in this monitoring include Heart Rate (HR), Blood Oxygen Saturation (SpO₂), Blood Pressure (BP), Glucose Level, Respiratory Rate (RR), and Physical Activity Index (PAI). Left to record as initial patient baseline: HR = 92 bpm, SpO₂ = 94%, BP = 145/95 mmHg, Glucose = 165 mg/dL, RR = 18 breaths per minute, and PAI = 45 units. Daily updates are fed into the system capturing medication adherence, meal patterns, and sleep quality scores. First, the real-time data passes through the SimCLR-based encoder, turning the 6-dimensional physiological feature vector into a 128-dimensional latent health state representation. An initial run with a mini-batch of 512 and a temperature parameter $\tau=0.5$ paired similar health states but excluded exacerbation events. Over the initial training window, the explained variance ratio achieved for the latent space exceeds 80%, ensuring compact and

informative state representations. Then from embeddings of the patients, Dynamic GraphSAGE constructs dynamically a patient similarity graph. Nearest neighbours are sampled for graph aggregation by a 10-25 neighbourhood size policy. Patients with similar healthy trajectories (e.g., high glucose fluctuations, blood pressure variability) are dynamically clustered in the process. This patient is grouped with a group that is characterized by moderate hypertension and poorly controlled diabetes for this process.

The cluster representation functions as an input to the meta-initializing reinforcement learning agents. The policy then quickly adjusts to the treatment optimization task in the specific clusters, again with the MAML given an outer learning rate of $5e-4$ and inner adaptation step size of $1e-2$ after five gradient updates. Within 800 episodes, the initial policies were accomplished, an impressive speed when compared to the more than 1500 episodes that standard RL typically required for similar stabilization. In this phase, for treatment adaptation, patients are given adaptive recommendations including dosage changes (e.g. increase metformin by 10%), diet modification advice (i.e., decrease uncomplicated carbohydrates by 15%), and physical activity-enhancing objectives (e.g., increase daily step number by 2000 steps). These are treated as outputs that should be continuously taken as actions from the localized PPO-based agents, the advantage estimations being normalized for fast convergences. The reward signals for the patients will be derived from keeping glucose levels within 90-140 mg/dL and blood pressure under 130/85 mmHg with heavy penalties imposed to excesses.

Thus, FedPPO, which is a federated PPO architecture, safeguards patients' privacy, and every 20 epochs, the patient device participates in the federated rounds. For this reason, only some updates on model parameters are transferred, which reduces communication hassle by less than 10% compared with the centralized training scenarios. The contribution of the patient local model to the global policy update using federated averaging is based on a sample count in the process. Inter-cluster cooperation occurs through the MADDPG framework. Agents managing adjacent clusters such as those with respiratory complications rather exchange policy summaries for generalization enhancement, not pure health data. Co-training produces long-term policies that are more stable with an observed gain of 12% improvement in cumulative reward against at least-stable cluster policies for the process.

By the 60 days of full simulation, the patients will have dropped their fasting glucose levels from 165

mg/dL to 128 mg/dL, systolic BP from 145 mmHg to 132 mmHg, and reduced 22% adverse events such as emergency department visits from historical trajectories. Treatment success, defined by the percentage of days with all physiological indicators within target ranges, achieved 78%, higher than the RL baselines of 63%-67%. The system also exhibited real-time adjustments for timely intervention whenever health trends deviate, which is predicted to reduce the long-term hospitalization risk by about 15% based on clinical extrapolation models. This fusion of meta-learning, federated reinforcement learning, dynamic graph encoding, contrastive health state learning, and multi-agent cooperation shows very good outcomes. It implies that this approach is better than other approaches in chronic disease patient management in IoT-connected ecosystems with both clinical improvement and system-level efficiency sets.

5. Conclusion & future scopes

This paper presented a novel deep reinforcement learning-based framework that combines Model-Agnostic Meta-Learning (MAML), Dynamic GraphSAGE, and Federated Proximal Policy Optimization (FedPPO) for secure and energy-efficient chronic disease management in IoT-enabled healthcare systems. The proposed architecture addresses the necessity for personalization, adaptability, and privacy in handling real-time, heterogeneous, and privacy-sensitive patient data across distributed edge environments. Experimental evaluations on synthetic datasets modeled after MIMIC-III and UK Biobank explained that our approach outperforms several baseline models in terms of prediction accuracy, adaptability, and energy efficiency.

However, we acknowledge several limitations that must be addressed to advance the system toward real-world deployment. First, the framework is currently evaluated on statistically simulated datasets rather than real patient records. While these simulations offer useful insights, they cannot substitute clinical validation. As a result, we plan to collaborate with medical institutions to conduct trials using ethically approved, de-identified patient data to assess model robustness and clinical relevance in safety-critical healthcare environments.

Second, although the federated learning component (FedPPO) enables decentralized training, it does not yet incorporate formal privacy-preserving mechanisms such as differential privacy (DP), secure multiparty computation (SMC), or homomorphic encryption. This leaves the system vulnerable to

inference and reconstruction attacks. As part of our future work, we will integrate DP-FedAvg and SMC protocols to ensure regulatory compliance with standards such as GDPR and HIPAA, while also empirically evaluating the trade-offs between privacy, accuracy, and communication cost.

Third, the current implementation assumes stable participation of 20% of edge devices per communication round, without accounting for stragglers, device failures, or connectivity issues. This assumption simplifies the complexities of real-world healthcare IoT environments, where devices frequently suffer from dropout due to energy constraints, network instability, or patient non-compliance. In future iterations, we will incorporate asynchronous federated optimization, dropout-resilient aggregation strategies, and dynamic device scheduling to enhance fault tolerance and training efficiency under non-ideal conditions.

In summary, while our proposed system demonstrates promising capabilities, further clinical validation, privacy enhancement, and robustness modeling are essential for large-scale, trustworthy deployment in real-world smart healthcare ecosystems.

Table 8. List of Notations Used in the Proposed Model

Symbol	Description
$x_i(t)$	IoT-based health signal for patient i at time t
$t \in \mathbb{R}^+$	Continuous time index in real-valued domain
$s_i(t)$	Latent health state representation of patient i at time t
f_θ	Health state encoder parameterized by θ
$\text{sim}(\cdot, \cdot)$	Cosine similarity function used in contrastive learning
τ	Temperature parameter in SimCLR contrastive loss
$G(V, E)$	Patient similarity graph with node set V and edge set E
$N(v)$	Neighborhood of node v in the graph structure
$W^{(l)}$	Trainable weight matrix in layer l of GraphSAGE
σ	Non-linear activation function (e.g., ReLU)
θ	Shared policy parameters in MAML
α	Inner loop learning rate in MAML
\mathcal{L}_{Ti}	Loss function for task T_i (typically negative expected reward)
$\pi_\theta(a s)$	Policy function giving probability of action a given state s
\hat{A}_t	Advantage estimate at timestep t
ϵ	Clipping hyperparameter in PPO
θ_{global}	Global model parameters aggregated from local updates

n_d	Number of local samples on device d
n_{total}	Total number of samples across all devices
$Q_i(x, a_1, \dots, a_N)$	Centralized critic function for agent i in MADDPG
$\pi_i(a_i o_i)$	Decentralized actor policy for agent i based on its observation o_i
y_i	Target Q-value in MADDPG training
$J(\pi)$	Expected cumulative reward of the policy π
$\tau(t)$	Policy-induced trajectory over time t
$R(\tau(t))$	Instantaneous reward function

Conflicts of Interest

The authors declare no conflict of Interest.

Author Contribution

All authors contributed significantly to the study's design, manuscript revision, and final approval. SUPRAJA BALLARI, as lead author, conceptualized the study, developed the integrated MAML–GraphSAGE–FedPPO framework, authored key sections, conducted experiments on MIMIC-III and UK Biobank, and performed ablation studies. KIRAN KUMAR V. supervised the work, guided reinforcement learning model development (PPO, MADDPG), and ensured methodological rigor. KRISHNA KUMAR N. handled implementation, integrated real-time IoT data, tested on edge devices (Raspberry Pi, Jetson Nano), and validated performance.

References

- [1] M. Bandy, S. Zafar, P. Agarwal, M. A. Alam, S. S. Biswas, I. Hussain, and K. M. Abubeker, “Predictive analysis of heart disease using quantum-assisted machine learning,” *Discover Appl. Sci.*, vol. 7, no. 5, 2025.
- [2] J. Tu, W. Liao, W. Liu, and X. Gao, “Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data,” *Sci. Rep.*, vol. 14, no. 1, 2024.
- [3] B. Yang *et al.*, “Identification and validation of inflammatory response genes linking chronic kidney disease with coronary artery disease based on bioinformatics and machine learning,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [4] B. Metherall, A. K. Berryman, and G. S. Brennan, “Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [5] S. K. Ghosh and A. H. Khandoker, “A machine learning driven nomogram for predicting chronic kidney disease stages 3–5,” *Sci. Rep.*, vol. 13, no. 1, 2023.
- [6] M. M. Islam, T. N. Poly, A. N. Okere, and Y. Wang, “Explainable machine learning model incorporating social determinants of health to predict chronic kidney disease in type 2 diabetes patients,” *J. Diabetes Metab. Disord.*, vol. 24, no. 1, 2025.
- [7] F. Liu, T. Li, D. Zhou, S. Shi, and X. Gong, “A machine learning–based framework for predicting postpartum chronic pain: a retrospective study,” *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, 2025.
- [8] C. Duckworth *et al.*, “Characterising user engagement with mHealth for chronic disease self-management and impact on machine learning performance,” *npj Digit. Med.*, vol. 7, no. 1, 2024.
- [9] N. M. Cohen *et al.*, “Longitudinal machine learning uncouples healthy aging factors from chronic disease risks,” *Nat. Aging*, vol. 4, no. 1, pp. 129–144, 2023.
- [10] L. Luo, P. Gao, C. Yang, and S. Yu, “Predictive modeling of COVID-19 mortality risk in chronic kidney disease patients using multiple machine learning algorithms,” *Sci. Rep.*, vol. 14, no. 1, 2024.
- [11] K. Rong, G. L. J. Yi ke ran, C. Zhou, and X. Yi, “Intelligent predictive risk assessment and management of sarcopenia in chronic disease patients using machine learning and a web-based tool,” *Eur. J. Med. Res.*, vol. 30, no. 1, 2025.
- [12] U. Bialonczyk *et al.*, “Balancing accuracy and cost in machine learning models for detecting medial vascular calcification in chronic kidney disease: a pilot study,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [13] M. S. Ahmed *et al.*, “Predicting chronic wasting disease in white-tailed deer at the county scale using machine learning,” *Sci. Rep.*, vol. 14, no. 1, 2024.
- [14] E. Vanden Broecke *et al.*, “Early detection of feline chronic kidney disease via 3-hydroxykynurenine and machine learning,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [15] M. N. H. Chowdhury *et al.*, “Machine learning algorithms for predicting the risk of chronic kidney disease in type 1 diabetes patients: a retrospective longitudinal study,” *Neural Comput. Appl.*, vol. 36, no. 26, pp. 16545–16565, 2024.
- [16] S. Li *et al.*, “A machine learning model and identification of immune infiltration for chronic obstructive pulmonary disease based on disulfidptosis-related genes,” *BMC Med. Genomics*, vol. 18, no. 1, 2025.

- [17] G. U. Nneji *et al.*, “FFS-IML: fusion-based statistical feature selection for machine learning-driven interpretability of chronic kidney disease,” *Int. J. Mach. Learn. Cybern.*, 2025.
- [18] P. Gogoi and J. A. Valan, “Chronic kidney disease prediction using machine learning techniques: a comparative study of feature selection methods with SMOTE and SHAP,” *Multiscale Multidiscip. Model. Exp. Des.*, vol. 8, no. 4, 2025.
- [19] M. H. I. Bijoy *et al.*, “A robot process automation based mobile application for early prediction of chronic kidney disease using machine learning,” *Discover Appl. Sci.*, vol. 7, no. 6, 2025.
- [20] C. Hsu *et al.*, “Machine learning models to predict osteoporosis in patients with chronic kidney disease stage 3–5 and end-stage kidney disease,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [21] O. Demiray *et al.*, “Classification of patients with chronic disease by activation level using machine learning methods,” *Health Care Manag. Sci.*, vol. 26, no. 4, pp. 626–650, 2023.
- [22] F. Si, Q. Liu, and J. Yu, “A prediction study on the occurrence risk of heart disease in older hypertensive patients based on machine learning,” *BMC Geriatr.*, vol. 25, no. 1, 2025.
- [23] H. Tsai *et al.*, “Multitask learning multimodal network for chronic disease prediction,” *Sci. Rep.*, vol. 15, no. 1, 2025.
- [24] Z. M. Moumin, İ. N. Ecemiş, and M. Karhan, “Heart disease detection using ensemble and non-ensemble machine learning methods,” *Eur. Phys. J. Spec. Top.*, 2024.
- [25] S. W. Oh *et al.*, “Machine learning models for predicting the onset of chronic kidney disease after surgery in patients with renal cell carcinoma,” *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, 2025.