Application of MAML Meta-Learning in Rapid Adaptation and Effectiveness Evaluation of Personalized Nursing Plans

Ruirui Wang, Kui Chen
Department of Nursing, Luohe Medical College, Luohe 462002, Henan, China
Corresponding author: Ruirui Wang (e-mail: 18639515917@163.com)

ABSTRACT Personalized nursing plans are important for improving patient outcomes, satisfaction, and efficient use of medical resources, yet significant individual differences make rapid plan adaptation and early effectiveness evaluation difficult. Existing methods usually require long-term follow-up or large numbers of labeled samples, limiting practical deployment. To address these issues, this paper adopts a framework based on Model-Agnostic Meta-Learning to model personalized nursing adjustment as a few-shot learning task. A multi-nursing dataset is constructed, a shared network generation scheme is designed, and cross-task features are extracted through MAML. A dynamic weight module is introduced to fuse short-term and long-term data for real-time prediction. Simulation experiments show that MAML achieves a median improvement rate of 75.0% in key physiological indicators, which is 12.7% higher than CNN, and more than 60% of patients obtain scores above 8 points. The prediction accuracy based on 24-hour early data reaches 89.7%, and that based on 48-hour data reaches 92.8%, outperforming comparative methods. These results demonstrate that MAML enables rapid adaptation and efficient evaluation of personalized nursing plans under limited data conditions.

INDEX TERMS model-agnostic meta-learning, personalized nursing plan, few-shot learning, effectiveness evaluation, adaptive prediction

I. INTRODUCTION

PERSONALIZED care plans are key to improving patient outcomes, satisfaction, and efficient utilization of medical resources, which is especially critical for industry workers who face unique occupational health challenges, such as musculoskeletal strain and dust exposure. In the context of precision medicine, addressing patients' diverse physiological, psychological, and social needs is critical. Research has shown that patientspecific care plans can significantly improve chronic disease management outcomes and patient experience. However, current personalized care practices face two key bottlenecks. First, there is the vast diversity among patients (involving heterogeneity across multiple dimensions, such as genetics, medical history, lifestyle, and comorbidities); second, there is a critical need for timely decision-making. These two factors pose significant challenges to the speed of adaptation and effectiveness of treatment plan generation systems.

Existing technologies struggle to effectively address these challenges. Traditional machine learning methods (such as ResNet-based time series models or matrix factorization recommendation algorithms) typically require retraining or fine-tuning on a large number of labeled samples (often hundreds) when adapting to new patients. This results in lengthy adaptation cycles (often taking several days) and makes it difficult to meet the demands of real-time decision-making. While deep transfer learning (such as fine-tuning CNNs) has improved data reuse, it still suffers from slow convergence (often requiring dozens of iterations), weak generalization, and a tendency to overfit on small samples when using very small samples (only 5-10 data points during initial hospitalization). More importantly, existing methods generally lack effective early-stage efficacy prediction mechanisms. Effectiveness assessment relies heavily on long-term follow-up, lasting weeks or even months. This makes it difficult to promptly identify ineffective or high-risk regimens and dynamically optimize them, potentially leading to delayed intervention and wasted resources.

To address these limitations in adaptation and efficacy evaluation, this paper employs a solution based on the MAML framework. Through meta-training, optimal model initialization parameters are learned from a large number of historical nursing tasks. This initialization enables the model to quickly adapt to personalized regimens for new

patient tasks, requiring only minimal patient data and a few gradient updates. Compared to traditional fine-tuning methods, MAML's advantage lies in its ability to rapidly generalize models from a small number of samples using explicit optimization.

The MAML framework in this paper utilizes three key technical innovations: 1) It transforms patient differences into diverse task distributions, designing a dual-branch solution generator (spatiotemporal feature extraction + multi-objective decision layer) to uniformly process heterogeneous medical data. 2) It introduces a dynamic weight evaluation module to integrate short-term metrics with long-term predictions. 3) A constrained gradient update algorithm can be developed to ensure that the adaptation process meets medical safety constraints. Constrained projection is used to ensure that the parameters of the generated solution meet safety requirements, and a dual threshold mechanism (loss change rate and indicator fluctuation) is combined in the convergence judgment to improve efficiency and safety.

II. RELATED WORKS

Personalized care is gaining increasing attention in the medical field, especially in the management of health problems [1]. A large number of studies have confirmed the effectiveness of personalized solutions. Mikkola I [2,3]'s research showed that there is a positive correlation between personalized care plans and clinical outcomes in patients with type 2 diabetes. Studies have found that implementing personalized care plans can significantly improve patients' health. This result emphasizes the importance of developing medical plans based on the patient's specific needs and circumstances. Pillon N J [4] pointed out that providing patients with tailored care plans can more effectively address metabolic problems related to obesity. This personalized perspective not only focuses on the patient's physiological characteristics, but also takes into account the living environment and habits, thereby formulating a more comprehensive treatment strategy. Krishna K B [5]'s research emphasized the importance of individual differences and specific needs of patients in making medical decisions, indicating that personalized care is not only applicable to common diseases, but also to special medical fields. Pathak K [6] discussed the application of 3D printing technology in biomedicine and pointed out the potential of this technology in promoting personalized care. Through additive manufacturing, medical staff can provide patients with customized medical devices and implants. This method not only improves the accuracy of treatment, but also enhances the patient's treatment experience. In summary, personalized care plans are of great value in improving treatment outcomes and increasing patient satisfaction. At the same time, high-quality personalized care requires a deep understanding of patient heterogeneity and the rapid generation of customized solutions. However, traditional approaches (such as rule-based expert systems or static recommendation algorithms) face bottlenecks such as low solution generation efficiency and difficulty adapting to new patients in real time when faced with highly dynamic patient states and multidimensional characteristics. Meta-learning, particularly the MAML framework, offers new solutions for addressing these challenges of rapid adaptation and knowledge transfer with limited samples [7,8]. Shao Y [9] proposed a two-layer constraint method to improve the generalization ability of MAML in few-shot classification. It showed that the use of two-layer constraints can effectively improve the performance of the model on new tasks, especially in the case of data scarcity, and provide a more robust solution for few-shot classification. Tian Y [10] used a model combining CNN and MAML to predict circRNAdisease associations. This demonstrates the application potential of MAML in the biomedical field and reflects the advantages of meta-learning in the analysis of complex biological data. Thayammal M C [11] showed that by utilizing MAML, the LSTM (Long Short-Term Memory) model can adapt to new environmental conditions more quickly when processing time series data, thereby improving the accuracy of predictions. Sun J [12] proposed a network security situation prediction method based on MAML and BiGRU (Bi-directional Gated Recurrent Unit), which highlights the importance of MAML's rapid adaptation in dynamic and complex security environments. Bala P [13] proposed a MAML-based framework that focuses on few-shot sentiment adaptation in low-resource NLP (Nature Language Processing). It shows that MAML can effectively deal with data scarcity and improve the adaptability of sentiment analysis models in new fields. Yi G [14]'s research focused on multi-omics disease classification based on MAML, demonstrating a new method for applying meta-learning to biomedical data analysis and promoting the understanding of complex disease mechanisms. The above literature shows that MAML has demonstrated strong adaptability and effectiveness in multiple fields. This work aims to fill this gap and establish a meta-learning nursing program that supports real-time optimization and effect tracking.

III. DESIGN OF A PERSONALIZED NURSING PROGRAM FRAMEWORK BASED ON MAML

A. DATA CONSTRUCTION AND UNIFIED MODELING

Data collection strictly adheres to a structured protocol. Physiological indicators are collected in real time using standard medical equipment. Medical histories are extracted from EHRs (electronic health records), and rehabilitation behaviors are based on patient compliance data and subjective feedback. This study was approved by the Ethics Committee of Luohe Medical College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. In this work, a meta-task is defined as the challenge of generating and evaluating a personalized nursing plan for a single, individual patient. This formulation treats each new patient as a new task, which aligns with the reality of personalization. To structure this problem for meta-learning,

this paper categorize these patient-specific tasks based on a combination of three axes: disease type, population demographics, and primary nursing goal. Each meta-task is encapsulated as an independent few-shot learning problem, thus translating individual differences between patients into diverse task distributions. In this framework, for each patient task, a small amount of data collected during the initial admission phase is defined as the support set Si, which is used to fine-tune model parameters to quickly adapt to the patient's characteristic distribution. The complete record of the nursing plan's effects after implementation constitutes the query set Qi, which is used for gradient calculation in the outer loop optimization phase to guide the update of global initialization parameters. The data features of Si and Qi are multimodally encoded, using a temporal convolutional network (TCN) combined with a multi-head self-attention mechanism [15,16] to capture the temporal dependencies and cross-variable interactions of patient physiological parameters. For unstructured text data, the pre-trained Med-BERT (Bidirectional Encoder Representations from Transformers) [17] is used for embedding representation to preserve semantic details. After encoding, the features of all modalities are concatenated and normalized through a unified feature fusion layer to form the input representation vector for the task. To ensure the stability and generalization of Si gradient updates, the inner loop uses first-order approximation MAML optimization. Within the inner loop, parameters are updated based on a task-specific loss function (weighted multi-objective form). The outer loop aggregates the gradient directions of multiple patient tasks on the Qi and updates the global initialization parameters, ensuring that the updated parameters achieve optimal or suboptimal performance within a small number of gradient steps when facing new patients. To prevent overfitting to specific population characteristics during meta-training, the task sampling strategy introduces stratified random sampling [18,19] to ensure that each meta-iteration includes patient tasks from different disease categories, age groups, and underlying health conditions, thereby improving the transferability of global parameters.

The final task formalization process can be abstracted as follows: given a set of tasks, where each task T_i corresponds to a patient, the objective optimization in the meta-training phase is as follows:

$$\min_{\theta} \sum_{T_i \sim p(\tau)} L_{Q_i}(U_\theta^k(S_i)) \tag{1}$$

θ represents the global initialization parameter, U_θ^k represents the parameter state after performing k sub-gradient updates on S_i, and L_{Q_i} represents the multi-objective loss calculated on Q_i. The key elements of the personalized nursing meta-task are mapped in Table 1.

TABLE 1. Key Elements of the Personalized Nursing Meta-Task

Number	Element	Data source	Role in task
1	Patient characteristics (Input)	Vital signs monitoring, laboratory tests, medical history, nursing assessment forms	Support input features for rapid model parameter adaptation
2	Intervention parameters (Input)	Nursing intervention records	Combined with patient characteristics to generate personalized plans
3	Outcome indicators (Output)	Physiological monitoring systems, patient feedback, complication records	Target variables for query set, used for outer loop optimization and evaluation

As shown in Table 1, the patient characteristics used as input, Si comprehensively reflects the patient's health status and individual differences. By efficiently encoding and integrating this preliminary data, the model can quickly capture the distribution of patient characteristics with a small sample size and fine-tune personalized parameters. Plan parameters are composed of nursing intervention records. This information is closely integrated with patient characteristics to form the foundation for generating care plans tailored to individual needs. The effect index, as the output of Qi , provides multi-dimensional data on the implementation effect of the solution, which is used to guide the global parameter optimization and solution quality evaluation of the model in the outer loop phase.

B. EXAMPLE META-TASK

Input: Patient characteristics (e.g., age=65, diabetes history, HbA1c=8.5)

Output: Nursing plan vector [medication_dosage=0.6, exercise_intensity=0.8, monitoring_frequency=0.9]

Actionable Plan: Administer metformin 500mg twice daily, recommend brisk walking 30min/day, and monitor blood glucose every 6 hours.

C. META-TRAINING AND CROSS-TASK FEATURE EXTRACTION

In the data preprocessing phase, a standardization process is implemented, continuous variables are normalized by Z-score [20], time series data are processed by cubic spline interpolation to handle missing values [21], and categorical variables are implemented using one-hot encoding [22].

The formula for fusing heterogeneous features using the feature alignment module based on the attention mechanism is as follows [23]:

$$h_f = \sum_{i=1}^{n} \alpha_i v_i, \quad \alpha_i = softmax\left(\frac{q k_i^T}{\sqrt{d}}\right) \tag{2}$$

q comes from the patient ID (Identity document) embedding layer, and the key-value pairs (ki,vi) correspond to

metadata, sensor feature vectors, and subjective evaluation vectors, respectively.

The solution generator design adopts a dual-branch architecture (Table 2 shows hyperparameter configuration). Architecture 1 uses a spatiotemporal feature extraction network, and a bidirectional LSTM layer is used in the time dimension to process physiological time series data [24]. The spatial dimension uses 1D-CNN to extract local feature patterns [25]. The spatiotemporal features are adaptively weighted through a gated fusion module, as shown in the following formula:

$$g = \sigma(W_g[h_{lstm}; h_{cnn}] + b_g) \tag{3}$$

Wg refers to the weight matrix, and bg refers to the bias term.

Architecture 2 is the multi-objective decision layer, which receives fused features and outputs three prediction heads in parallel (discrete operation classifier, continuous parameter regressor, and effect predictor). The loss function is designed as a weighted combination:

$$L = \lambda_1 L_{cls} + \lambda_2 L_{reg} + \lambda_3 L_{eff} \tag{4}$$

$\lambda 1$, $\lambda 2$, $\lambda 3$ are assigned values of 0.4, 0.3, and 0.3, respectively.

Note*: ReLU (Rectified Linear Unit).

In Table 2, for spatiotemporal feature extraction, the bidirectional LSTM layer has 256 hidden units per layer, effectively capturing the contextual information of time series data. The CNN layer uses two convolutional layers with 64 and 128 kernels, respectively, and a kernel size of 3×1, to extract features in the temporal dimension. After entering the multi-objective decision layer, the dimension of the discrete operation head is 128, mapped to the number of operation categories. A fully connected layer combined with a softmax layer is used for multi-class classification. The continuous parameter head also has a dimension of 128, which maps to the number of parameters. A fully connected layer and ReLU activation function are used to enhance nonlinear features. The effect prediction head has a dimension of 128, which maps to the number of effect indicators. A fully connected layer and sigmoid activation function are used for binary classification prediction. Regarding optimization parameters, the inner loop learning rate α is fixed at 0.01, and the outer loop learning rate β is set to 0.001, using the Adam optimizer to ensure stable and efficient model training. Furthermore, the batch size is 32, meaning that 32 tasks can be sampled in each outer loop to improve training efficiency and model generalization.

The parameter update mechanism during the meta-training and meta-adaptation phases is shown in Figure 1.

In Figure 1, the parameter separation strategy is implemented via a structured gradient mask. During backpropagation, a parameter constraint matrix is applied to the inner loop computation. Gradients are prohibited from flowing at locations corresponding to shared feature extraction layers, while gradient updates are permitted at locations corresponding to task-specific layers.

This mechanism forces the model to encode common features across tasks into shared residual blocks during meta-training, while task-specific knowledge is retained in the output layer. The MAML adaptation process strictly follows the second-order optimization paradigm. The inner loop samples batches of tasks, calculates the S_i loss, and performs gradient updates:

$$\theta_i' = -\alpha \nabla_\theta L_{\tau_i}(f_\theta, D_i^{support}) \tag{5}$$

$L_{\tau_i}(f_\theta, D_i^{support})$ represents the loss S_i, and the gradient clipping threshold is 10.0.

The outer loop calculates the Q_i loss $L_{\tau_i}(f_{\theta_i'}, D_i^{query})$ based on the updated parameters and optimizes the initial parameters using the second-order derivative:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\tau_i} L_{\tau_i}(f_{\theta_i'}, D_i^{query}) \tag{6}$$

Cross-task generalization enhancement is achieved through gradient alignment regularization [26], which introduces the task gradient direction similarity constraint when calculating the meta-gradient $\nabla_\theta L_{meta}$:

$$R = \lambda \sum_{i \neq j} \left(1 - \frac{\langle \nabla_\theta L_{\tau_i}, \nabla_\theta L_{\tau_j} \rangle}{\|\nabla_\theta L_{\tau_i}\| \|\nabla_\theta L_{\tau_j}\|}\right) \tag{7}$$

D. MODULE AND NURSING PROGRAM PERSONALIZATION ADAPTATION STRATEGY

The dynamic weight effect evaluation module adopts an adaptive fusion mechanism [27] to achieve real-time quantification of the nursing program effect. Its core is to dynamically adjust the weight contribution of short-term rehabilitation indicators and long-term follow-up results according to the intervention stage. Short-term rehabilitation indicators are defined as objective physiological parameters and subjective feedback (24-hour satisfaction questionnaire scores) collected at high frequency within 0-72 hours after intervention. Long-term follow-up results include key endpoint events and stable period assessments 7-30 days after intervention. The weight distribution function is designed to be a time-dependent exponential decay form(In order to prioritize short-term indicators and gradually shift towards long-term results):

$$\begin{cases} \omega_{short}(t) = \omega_0 \cdot e^{-\gamma t} \\ \omega_{long}(t) = 1 - \omega_{short}(t) \end{cases} \tag{8}$$

The initial weight ω_0 is 0.8, the decay coefficient $\gamma = 0.05$, and the duration t is calculated in hours from the start of the intervention. This function ensures that the short-term weight is dominant ($\omega_{short} \approx 0.8$) in the initial intervention period (<24 hours), the weight is linearly shifted (ω_{short} is reduced to 0.4) during the transition period (24 hours <t< 72 hours), and the long-term weight is dominant (t>0.7) during

TABLE 2. Hyperparameter Configuration

Component	Parameter name	Value/Type	Description
Spatiotemporal feature extraction	LSTM layers	2	Bidirectional structure
Spatiotemporal feature extraction	LSTM hidden units	256	Dimension for each layer
Spatiotemporal feature extraction	CNN kernel quantity	[64, 128]	Two layers of convolution
Spatiotemporal feature extraction	CNN kernel size	3×1	Convolution in the time dimension
Multi-Objective decision layer	Discrete action head dimension	$128 \rightarrow$ Number of action categories	Fully connected + softmax
Multi-Objective decision layer	Continuous parameter head dimension	$128 \rightarrow$ Number of parameters	Fully connected + ReLU
Multi-Objective decision layer	Effect prediction head dimension	$128 \rightarrow$ Number of effect indicators	Fully connected + sigmoid
Optimization parameters	α	0.01	Fixed value
Optimization parameters	β	0.001	Adam optimizer
Optimization parameters	Batch task size	32	Number of tasks sampled per outer loop step

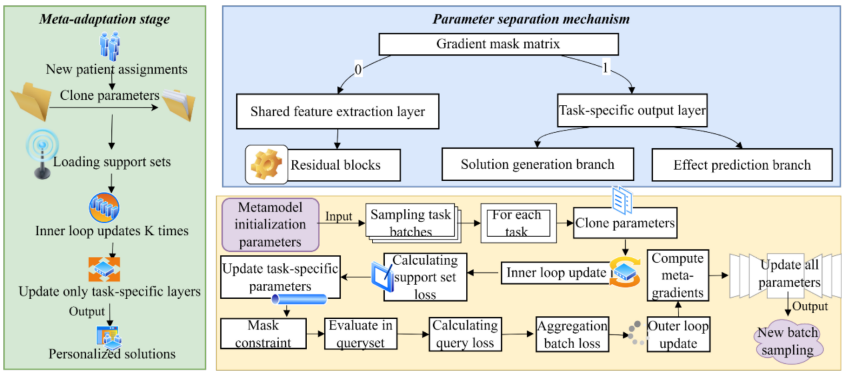

FIGURE 1. Parameter update mechanism during meta-training and meta-adaptation

the stable period ($\omega_{long} > 72$ hours). These parameters were empirically optimized on the validation set to balance early sensitivity and long-term stability. Unlike RNNs, which require sequential processing and may introduce latency, the decay function provides a closed-form solution for weight allocation, critical for rapid decision-making.

The early effect evaluation module predicts long-term effects based on the latent space feature map of the fine-tuned model. When a new patient completes the personalized care plan fine-tuning, the activation values of the intermediate layers of the plan generator are extracted as latent space representations:

$$H_s \in R^{d \times t} \tag{9}$$

d represents the feature dimension. These representations are concatenated with the patient baseline features to form a joint input vector:

$$Z = concat(H_s, X_b) \in R^{(d \times t)+k} \tag{10}$$

This is then fed into a pre-trained effectiveness evaluation subnetwork. This subnetwork employs a threestage architecture. First, a 1D-CNN (kernel size = 3, number of channels = 64) extracts temporal local dependency features. Second, a BiGRU layer (hidden units = 128) captures long-term contextual associations [28,29]. Finally, a self-attention mechanism calculates the feature weight distribution. The output layer generates predictions via two parallel branches: the main branch uses a Sigmoid activation function to output the final effectiveness probability of the solution [30,31], while the risk branch uses Softmax to output the four-category risk probability distribution [32].

To quantify the uncertainty of the prediction, Monte Carlo was introduced to perform 50 forward propagations and calculate the prediction variance [33,34]. When the variance is greater than 0.25, an early warning mechanism is triggered, requiring additional monitoring data after 72 hours to re-evaluate the patient. The evaluation module training uses a multi-objective loss function:

$$L = \lambda_1 \cdot BCE(P_e, Y_e) + \lambda_2 \cdot CE(P_r, Y_r) + \lambda_3 \cdot \|\Theta\|_2^2 \tag{11}$$

BCE represents binary cross entropy, and CE represents multi-class cross entropy. Ye and Yr represent the true labels for the final effect and risk, respectively. Training data was constructed using historical patient data from the first seven days of intervention as input, and the outcome at day 90 of intervention as a supervisory signal. 120,000 training samples were generated using sliding window sampling.

Feature interpretability was achieved using an ensemble gradient method, calculating the contribution of each input feature to the prediction result:

$$\emptyset_i(Z) = (Z_i - Z_i') \times \int_{\alpha=0}^{1} \frac{\partial \epsilon_\theta(Z' + \alpha(Z - Z'))}{\partial Z_i} d\alpha \tag{12}$$

Z' is the baseline input (taken as the mean of the training set). When a high-risk prediction occurs, the system automatically generates a report of the top five key decision factors. This module is optimized simultaneously with meta-training during the training phase, using a segregated validation set (15% of the sample) for early stopping monitoring. During deployment, the weight of early data is dynamically adjusted using a time decay factor, ensuring that the prediction's dependence on early data gradually decreases as the intervention period increases.

The decision support module dynamically optimizes care based on early assessment results. When the risk probability exceeds 0.35, a high-risk alert process is triggered. Gradient-weighted class activation mapping is used to identify the dimensions in the latent space that contribute most to risk. Feature inversion is used to calculate the corresponding indicators and generate interpretable alert reports. If the expected improvement rate of key indicators in the long-term outcome prediction is less than 10%, the counterfactual optimization engine is activated to solve within the scenario parameter space. For other scenarios, the original scenario is maintained and the regular monitoring cycle is initiated, forming a closed-loop decision chain of "prediction-warning-optimization."

During the personalized adaptation phase (key parameters and threshold settings are shown in Table 3), a constrained gradient update algorithm is used to achieve fast convergence with a small number of samples. Convergence is determined using a dual-condition joint trigger mechanism, terminating the update when both conditions are met or the maximum number of iterations is reached. This mechanism ensures adaptation efficiency while avoiding resource waste caused by ineffective iterations.

TABLE 3. Key Parameters and Threshold Settings for the Personalized Adaptation Phase

Parameter category	Value	Mechanism	Constraint conditions
Gradient control	0.05	Initial learning rate	$> 10^{-4}$
Gradient control	0.1	Constraint projection strength	[0.05, 0.2]
Gradient control	1.0	Maximum gradient norm	≤ 1.0
Convergence criterion	0.005	Loss change threshold	< 0.005
Convergence criterion	0.02	Metric fluctuation threshold	< 0.02
Convergence criterion	15	Maximum number of iterations	–
Regularization	0.01	Feature smoothing coefficient	–

Table 3 shows the gradient control parameters in the personalized care plan adaptation strategy to ensure the refinement of the later updates, while setting the maximum gradient norm to avoid overfitting. The constraint projection strength allows dynamic adjustment between [0.05, 0.2], and the mandatory parameters meet the medical safety boundary. Convergence judgment uses a dual threshold joint trigger, and the maximum number of iterations is used as a mandatory termination condition. The feature smoothing regularization term suppresses the drift of the shared layer parameters [35].

IV. EXPERIMENTAL VERIFICATION AND ANALYSIS

A. EXPERIMENTAL DATA AND SETTINGS

The simulated dataset generation method strictly adhered to the data collection specifications defined in section Data Construction and Unified Modeling (based on data requirements in real-world nursing scenarios). The simulated dataset contains 15 meta-tasks, each representing a different patient population. Each task includes 5-10 support samples (few-shot design) and 20-30 query samples. The feature vectors have a dimension of 128 and include physiological time series, categorical variables, and subjective feedback. Data synthesis employs Conditional Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), combined with medical expert rules to ensure physiological plausibility (such as the logical correlation between drug dosage and symptom improvement).

While the simulated dataset provides a controlled and medically realistic environment for initial validation, its inherent limitations and the gap between simulation and reality must be recognized. Real-world data presents challenges that are difficult to fully capture in simulation; therefore, the superior performance reported in this study must be interpreted as a proof-of-concept under ideal conditions. The primary objective here is to determine the potential and methodological effectiveness of the MAML framework in this application. The results demonstrate that the model can effectively learn from the task distribution and rapidly adapt to new instances within the same controlled distribution. As stated in the conclusion section, bridging the gap with real-world clinical practice is a crucial next step.

1) Baseline Model Selection

(1) ResNet time series model: The input is the patient's historical 72-hour physiological indicator time series data (sampling frequency 1 hour), and the terminal fully connected layer outputs the nursing plan parameter vector. End-to-end training is performed using mean square error loss;

(2) Matrix decomposition recommendation algorithm: The patient characteristics and nursing plan items are constructed as an interaction matrix. The latent factor vectors of patients and plans are learned through weighted regularized matrix decomposition. The prediction stage generates the top-N recommended plans based on the nearest neighbor.

(3) Fine-tuning CNN: A CNN isomorphic to the MAML solution generator is used as the base model. It is first pre-trained on the meta-training set. When encountering new patients, a small amount of their data (5-10 samples) is used for supervised fine-tuning. The learning rate is set to 0.001,

and the upper limit of fine-tuning rounds is 50. All baseline models are evaluated using the same meta-test task as the MAML framework, and the input data format and feature encoding method are consistent.

2) Verification of the Speed of New Patient Adaptation

The speed of new patient adaptation was verified on a strictly isolated meta-test set. Five unseen simulated patients were randomly selected as the support set for each test task, and the number of gradient updates required for different methods to converge was recorded. Convergence was determined when the crossentropy loss of the solution generator changed by less than 0.1% over three consecutive iterations or when the maximum number of iterations (50) was reached. During the adaptation process, the MAML framework (denoted by F1) employed a second-order optimization strategy, the fine-tuning CNN (denoted by F2) used the Adam optimizer, and the ResNet temporal model (denoted by F3) and matrix factorization algorithm (denoted by F4) were not iteratively adapted, with pre-training results directly output as a reference. After each gradient update, the average loss value for all test tasks was recorded, ultimately generating a convergence curve comparison as shown in Figure 2. The quantitative metrics shown in Table 4 were also compiled.

TABLE 4. Comparison of adaptation speed and convergence rounds.

Method	Average convergence	Standard deviation	Minimum Round	Maximum round	Speed Increase
	cycles (units: round)	(Unit: round)	(Unit: round)	(Unit: round)	Multiple
F1	8.2	1.5	5	12	4.8×
F2	39.6	6.3	28	50	-
F3	-	-	-	-	-
F4	-	-	-	-	-

Note: Speed improvements are calculated using the fine-tuning CNN as a benchmark; "-" indicates iterative adaptation is not applied.

In Figure 2, the convergence curve analysis shows that MAML's loss decreases faster than the fine-tuning CNN in the first five update rounds, verifying that the cross-task prior knowledge extracted during the meta-training phase significantly accelerates the few-shot adaptation process.

The experimental results in Table 4 show that the MAML framework exhibits remarkably fast convergence on the new patient task, requiring an average of only 8.2 gradient updates to reach the convergence threshold (standard deviation ± 1.5 rounds). In comparison, the fine-tuning CNN method requires an average of 39.6 updates (standard deviation ± 6.3 rounds). Due to the lack of an iterative optimization mechanism, the ResNet temporal model and matrix factorization algorithm maintain a constant loss throughout the evaluation cycle, making it impossible to optimize the personalized solution.

3) Improved Effectiveness of the Personalized Solution

The effectiveness of personalized care plans was validated on 500 unseen patient samples from the meta-test set, covering three major disease categories: cardiovascular disease, diabetes, and postoperative rehabilitation. For each patient, personalized care plans were generated using F1-F4, implemented in a unified simulation environment, and the results were recorded. Physiological improvement rate statistics focused on key disease indicators: systolic blood pressure reduction and heart rate stability index were used for cardiovascular disease; fasting blood glucose reduction rate and expected HbA1c (Glycated Hemoglobin A1c) improvement were used for diabetes; and pain index reduction rate and improvement in mobility were used for postoperative rehabilitation.

Outliers were handled using Tukey's rule, resulting in the box plot (Figure 3) showing the distribution of improvement rates for the four methods.

The experimental results in Figure 3 show that the MAML solution achieved a median improvement rate of 75.0% in physiological indicators, significantly exceeding the 62.3% achieved by the fine-tuning CNN, the 55.6% achieved by the ResNet time series model, and the 49.2% achieved by the matrix factorization algorithm. Patient Engagement and Adherence were assessed using the composite score from the 24-hour engagement questionnaire. This score (on a 0-10 scale) objectively reflects the patient's initial interaction with the care plan, incorporating metrics such as compliance rates and self-reported ease of execution. The distribution of these adherence scores is presented in Figure 4.

In Figure 4, the analysis of adherence scores revealed a significantly higher level of initial engagement with the MAML-generated plans. The mean adherence score for the MAML regimen was 8.2 ± 0.8, with over 60% of simulated patients achieving a high adherence score of 8 or higher. In contrast, less than 60% achieved this level under the baseline approaches. A statistical test (Kruskal-Wallis, $p < 0.001$) confirmed a significant difference between the groups, indicating that the personalized plans from MAML were not only physiologically effective but also more readily adopted and followed by patients in the simulation.

B. PREDICTIVE ASSESSMENT ACCURACY

The validation of predictive assessment accuracy focused on the ability of early-stage (24-48 hours) indicator data to predict the ultimate therapeutic effect. For 500 patients in the meta-test set, key recovery indicators were recorded 24 hours (S1) and 48 hours (S2) after implementation of the nursing plans generated by the four methods. The final efficacy criterion was the overall improvement rate at the end of the plan implementation cycle ($\geq$60% was considered effective). The prediction model used a BiGRU temporal network, which took as input the sequence of indicators for the S1/S2 time periods and output a final effectiveness probability prediction. The threshold was set at a probability $\geq$0.5 for effectiveness. The prediction models were trained independently for each of the four methods (with consistent hyperparameters) and 5-fold crossvalidation was used to

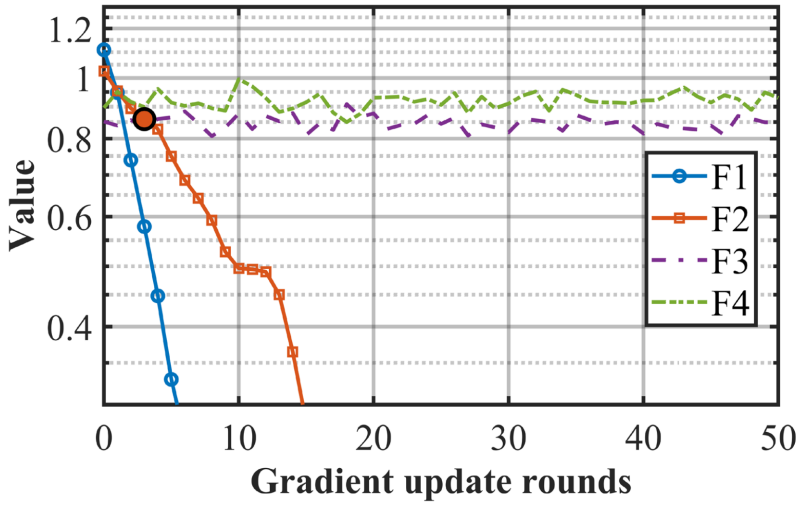

FIGURE 2. Convergence curve comparison of different methods on the novel patient task

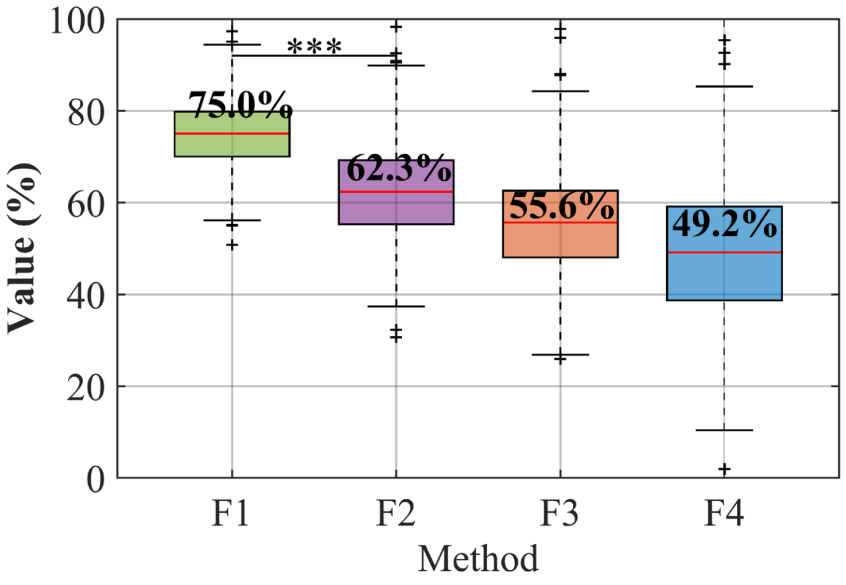

FIGURE 3. Comparison of Improvement Rates for Key Physiological Indicators. *** indicates $p < 0.001$

eliminate randomness, ultimately generating the accuracy bar chart shown in Figure 5.

Figure 5 shows that the early data generated by the MAML scheme demonstrates excellent predictive capabilities. The accuracy of predicting final validity for 24-hour data reached 89.7±2.1%, and for 48-hour data, it increased to 92.8±1.7%. This performance is significantly higher than that achieved by the Fine-tuning

CNN (24-hour: 74.3±3.5%, 48-hour: 81.6±2.9%), the ResNet temporal model (24-hour: 68.2±4.1%, 48-hour:

75.4±3.7%), and the matrix factorization algorithm (24-hour: 71.5±3.8%, 48-hour: 78.3±3.2%). The data reveals three key patterns: All methods achieved significantly higher prediction accuracy at 48 hours than at 24 hours, indicating that 48-hour data is more valuable for prediction. Furthermore, MAML significantly outperformed other methods at both time points and exhibited the smallest standard deviation, demonstrating the most stable prediction results.

1) Verification of the Effectiveness of the Dynamic Weight Module

The effectiveness of the dynamic weight module was verified using a controlled variable method, comparing the full MAML framework (including the dynamic weight module) with a simplified version (MAML-ND) that removed the dynamic weight module on the same meta-test set. The two sets of experiments shared the same protocol generator architecture and meta-training parameters, differing only in whether the designed dynamic weighting mechanism was applied during the efficacy evaluation phase. Prediction

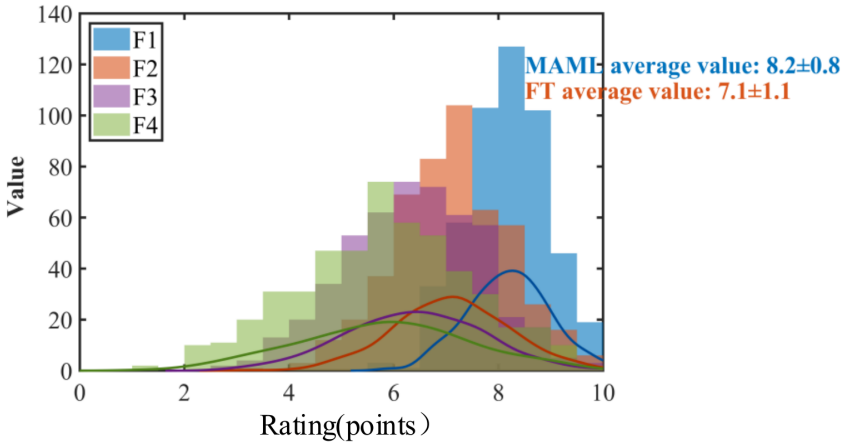

FIGURE 4. Distribution of Subjective Patient Satisfaction Scores

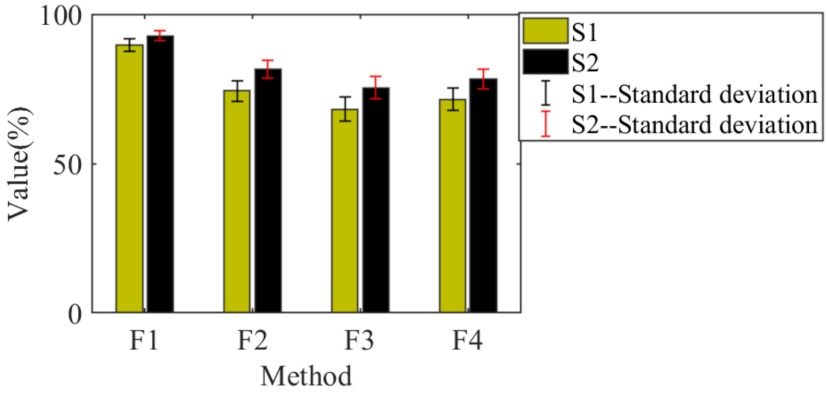

FIGURE 5. Early Prediction Accuracy

accuracy was calculated according to the final efficacy judgment criteria defined in section Predictive Assessment Accuracy, with predictions recorded at five time points: 0 hours (upon initial adaptation), 12 hours, 24 hours, 36 hours, and 48 hours after protocol implementation. The indicator data collected at each time point were processed using the same BiGRU prediction model to output the final efficacy probability. To eliminate the influence of temporal correlation, stratified sampling was used to construct 100 independent test cohorts from the 500 patients, each containing 5 patients with the same disease. The contribution curves shown in Figure 6 were generated by taking the mean ±95% confidence interval.

Figure 6 shows that the dynamic weight assessment module significantly improves early prediction capabilities. At the 24-hour time point, the prediction accuracy of the complete framework reached 89.7±2.1%, a 12.1 percentage point improvement compared to the simplified version (77.6±2.8%) ($p<0.001$). Over time, the difference between the two groups gradually narrowed, remaining at 7.7 percentage points at 48 hours (92.8±1.7% vs 85.1±2.2%). Contribution curve analysis shows that the module's increased contribution rate during the 0-24 hour period is primarily due to its adaptive integration of short-term recovery metrics and the gradually increasing weighting of long-term metrics.

2) Verification of Generalization Ability Across Diseases and Populations

The model's generalization ability was verified using stratified sampling. The meta-test set was strictly divided by disease type (cardiovascular disease, diabetes, and postoperative rehabilitation) and population characteristics (age ≥65 years defined as the elderly group, 40-64 years defined as the middle-aged group, and 18-39 years defined as the young group). Fifty patients were randomly selected from each group, and the performance scores of the personalized care plans generated by the MAML framework on four core indicators were recorded. To eliminate dimensionality differences, Min-Max normalization was used to convert each indicator to the [0, 1] range. The baseline value was taken from the global minimum/maximum values across all subgroups. After normalization, the indicators were displayed using a polar coordinate system, generating the multidimensional performance radar chart shown in Figure 7. Each subplot corresponds to a specific disease type,

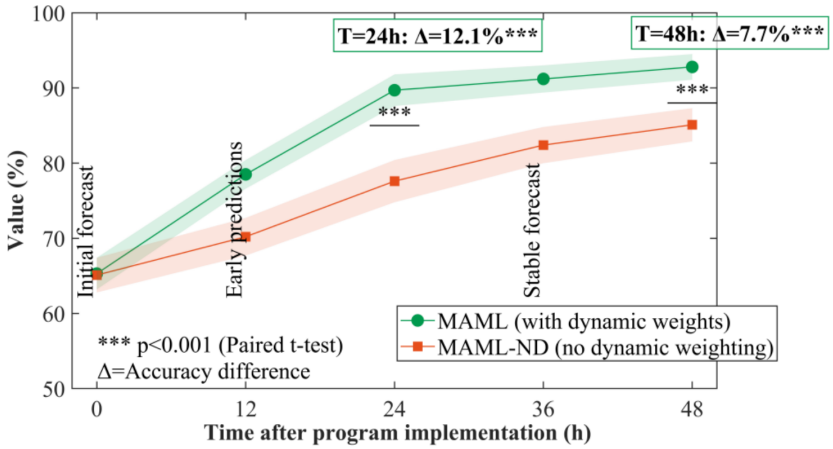

FIGURE 6. Contribution of the Dynamic Weight Assessment Module to Improved Prediction Accuracy

and the three trajectory lines represent the comprehensive performance profiles of different population groups.

Figure 7 (a). Cardiovascular Disease

Figure 7 (b). Diabetes

Figure 7 (c). Postoperative Rehabilitation

In Figure 7, the MAML framework maintains excellent performance across a wide range of disease types and populations. g1-g4 represent the improvement rate of physiological indicators, prediction accuracy, adaptation convergence rounds, and patient satisfaction, respectively. Figure 7(a) shows that the MAML framework performs best overall for middle-aged patients, followed by the elderly. Figure 7(b) shows that the elderly have the highest improvement rate (0.82), but the younger group has a relatively low prediction accuracy (0.75). Figure 7(c) shows balanced performance across all populations, with a satisfaction score of 0.93 for middle-aged patients. The elderly group performs robustly in diabetes, but still lags behind the middle-aged group in satisfaction. The data shows that different populations exhibit differences in various indicators for the same disease. Middle-aged patients generally perform best, while younger patients lag slightly behind older patients. This demonstrates the MAML framework's strong generalization capabilities across different disease types and age groups, providing an intuitive reference for optimizing personalized care plans.

C. COST AND SCALABILITY ANALYSIS

The cost and scalability evaluation was conducted using a unified hardware platform (NVIDIA Tesla V100 GPU, 32GB RAM) and software environment (PyTorch 1.8, CUDA 11.1). Training cost measurement records the wallclock time of a complete meta-training cycle (100 epochs), including data loading, forward propagation, gradient calculation, and parameter update. This is repeated five times and the mean $\pm$ standard deviation is taken. Inference latency testing focuses on the new patient adaptation phase. 100 randomly selected meta-test tasks are used to record the end-to-end processing time of a single gradient update (from data input to parameter update completion), including the execution time of the dynamic weight evaluation module. Scalability metrics are verified using a controlled variable method, with peak memory usage monitored in real time using nvidia-smi. The task scale scalability test gradually increased the number of concurrent tasks (10-100 tasks), recording the system throughput (tasks/second) and response time growth rate. The data dimension scalability test increased the feature dimension from 20 to 100 and recorded the change rate of inference latency. This ultimately generated the training/inference time comparison shown in Figure 8 and the scalability metrics summary in Table 5.

Figure 8 shows that the training time for MAML is 6.8 hours, the fine-tuned CNN is 4.2 hours, the ResNet is 3.5 hours, and the matrix factorization is 1.2 hours. The latency for a single adaptation is 0.12 seconds for MAML, 0.38 seconds for fine-tuned CNN, 0.05 seconds for ResNet, and 0.02 seconds for matrix factorization. The data revealed that MAML took the longest to train (6.8 hours), 5.6 hours longer than matrix factorization, indicating that its meta-learning architecture requires higher initialization costs. During inference, MAML's adaptation latency (0.12 seconds) was significantly lower than that of fine-tuned CNN (0.38 seconds), but higher than that of matrix factorization (0.02 seconds) and ResNet (0.05 seconds). This time-cost distribution suggests that MAML's strategy of combining high training investment with low inference cost is particularly well-suited for practical scenarios requiring frequent, real-time adjustments. While initial training takes longer, it can provide efficient, personalized adaptation for each patient.

Table 5 shows that MAML requires the highest memory (9.2GB), four times the memory usage of matrix factorization (2.3GB), reflecting its high computational resource consumption. In throughput testing, matrix factorization achieved the highest processing power (42.5 tasks/s), with

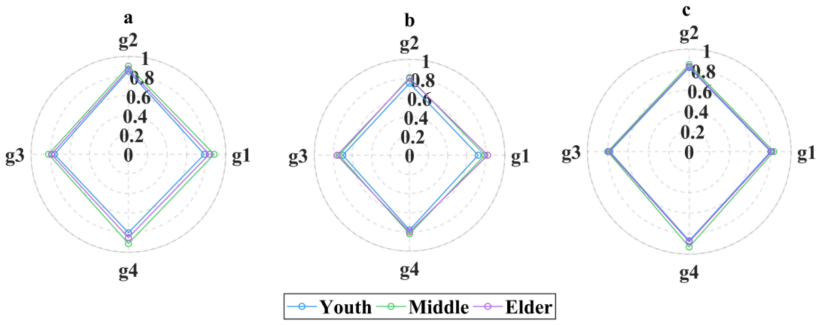

FIGURE 7. Multidimensional Performance of the MAML Framework Across Disease Types and Populations

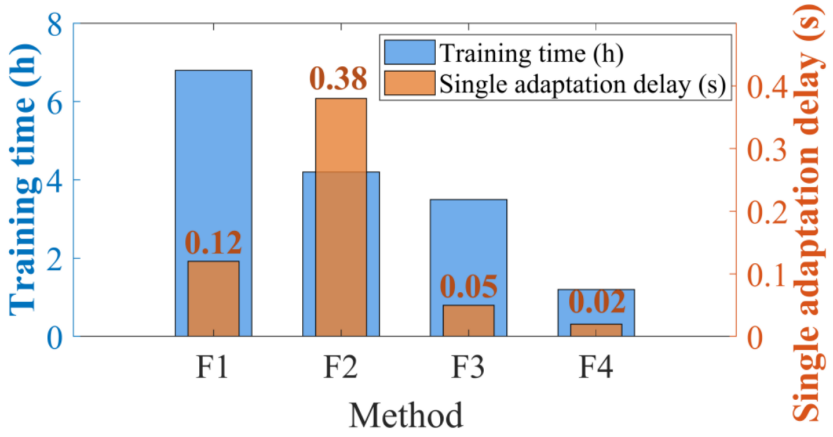

FIGURE 8. Comparison of Model Training and Inference Time

TABLE 5. Comparison of Scalability Metrics of Different Models

Model	Training time (h)	Single adaptation delay (s)	Memory usage (GB)	Throughput (tasks/s, for 50 tasks)	100-dimension feature delay growth rate
F1	6.8 ± 0.5	0.12 ± 0.01	9.2	8.3	+18.7%
F2	4.2 ± 0.3	0.38 ± 0.03	7.1	6.1	+25.4%
F3	3.5 ± 0.2	0.05 ± 0.003	5.8	15.2	+32.9%
F4	1.2 ± 0.1	0.02 ± 0.001	2.3	42.5	+8.3%

MAML (8.3 tasks/s) outperforming fine-tuned CNN (6.1 tasks/s), demonstrating its practicality in multi-tasking. The key latency growth metric reveals that when the feature dimension is expanded to 100, MAML latency increases by only 18.7%, significantly lower than fine-tuned CNN (25.4%) and ResNet (32.9%), demonstrating its architecture's superior stability in high-dimensional feature scenarios.

Although MAML incurs higher training costs, it outperforms simple regression models in scenarios with heterogeneous patient profiles and limited data.

V. CONCLUSION

This article solves the challenges of high patient variability, slow adaptation speed, and long evaluation cycle in personalized care in the industry by using a meta learning solution based on MAML. Experimental validation demonstrates that MAML converges on new patient tasks with an average of only 8.2 gradient updates, a 4.8x improvement over fine-tuning CNNs. The generated solutions achieve a median improvement of 75.0% in key physiological indicators (a 25.8% improvement over matrix factorization), with over 60% of patients reporting high patient satisfaction (a score greater than 8). The final efficacy prediction accuracy based on early (24-hour) data reached 89.7% (increasing to 92.8% after 48 hours), significantly outperforming the baseline method. The dynamic weighting module further improved the accuracy of early assessments by 12.1 percentage points by integrating short-term indicators with long-term predictions. This method demonstrated its efficiency and scalability in a simulation environment.

AUTHOR CONTRIBUTIONS

Conceptualization –RuiRui Wang and Kui Chen; methodology – RuiRui Wang and Kui Chen; investigation – RuiRui Wang and Kui Chen; writing-original draft preparation – RuiRui Wang and Kui Chen. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by: The application research of narrative nursing education in nursing teaching under the background of ideological and political courses.

Project number: 2023-LYZJGYB007.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Luohe Medical College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] K. B. Johnson, W. Q. Wei, D. Weeraratne, M. E. Frisse, K. Misulis, K. Rhee, et al., "Precision medicine, AI, and the future of personalized health care," Clinical and Translational Science, vol. 14, no. 1, pp. 86-93, 2021, doi: 10.1111/cts.12884.

[2] I. Mikkola, M. Hagnäs, J. Hartsenko, M. Kaila, and K. Winell, "A personalized care plan is positively associated with better clinical outcomes in the Care of Patients with Type 2 diabetes: a cross-sectional real-life study," Canadian Journal of Diabetes, vol. 44, no. 2, pp. 133-138, 2020, doi: 10.1016/j.jcjd.2019.05.003.

[3] I. Mikkola, S. Morgan, K. Winell, J. Jokelainen, L. Frittitta, E. Heikkala, et al., "Association of personalised care plans with monitoring and control of clinical outcomes, prescription of medication and utilisation of primary care services in patients with type 2 diabetes: an observational real-world study," Scandinavian Journal of Primary Health Care, vol. 40, no. 1, pp. 39-47, 2022, doi: 10.1080/02813432.2022.2036458.

[4] N. J. Pillon, R. J. F. Loos, S. M. Marshall, and J. R. Zierath, "Metabolic consequences of obesity and type 2 diabetes: Balancing genes and environment for personalized care," Cell, vol. 184, no. 6, pp. 1530-1544, 2021, doi: 10.1016/j.cell.2021.02.012.

[5] K. B. Krishna, B. A. Kogan, MazurT, P. Hoebeke, G. Bogaert, and P. A. Lee, "Individualized care for patients with intersex (differences of sex development): part 4/5," Considering the Ifs, Whens, and Whats regarding sexual-reproductive system surgery. Journal of Pediatric Urology, vol. 17, no. 3, pp. 338-345, 2021, doi: 10.1016/j.jpurol.2021.02.011.

[6] K. Pathak, R. Saikia, A. Das, D. Das, M. A. Islam, P. Pramanik, et al., "3D printing in biomedicine: Advancing personalized care through additive manufacturing," Exploration of Medicine, vol. 4, no. 6, pp. 1135-1167, 2023, doi: 10.37349/emed.2023.00200.

[7] A. Vettoruzzo, M. R. Bouguelia, J. Vanschoren, T. Rögnvaldsson, and K. C. Santosh, "Advances and challenges in meta-learning: A technical review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 7, pp. 4763-4779, 2024.

[8] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning approaches for few-shot learning: A survey of recent advances," ACM Computing Surveys, vol. 56, no. 12, pp. 1-41, 2024, doi: 10.1145/3659943.

[9] Y. Shao, W. Wu, X. You, C. Gao, and N. Sang, "Improving the generalization of MAML in few-shot classification via bi-level constraint," IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 7, pp. 3284-3295, 2022, doi: 10.1109/TCSVT.2022.3232717.

[10] Y. Tian, Q. Zou, C. Y. Wang, and C. Jia, "MAMLCDA: a meta-learning model for predicting circRNA-disease association based on MAML combined with CNN," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 7, pp. -4335, 2024, doi: 10.1109/JBHI.2024.3385352.

[11] M. C. Thayammal, "MAML-Enhanced LSTM for Air Quality Time Series Forecasting," Water, Air, & Soil Pollution, vol. 235, no. 12, pp. 1-15, 2024.

[12] J. Sun, C. Li, and Y. Song, "A network security situation prediction approach based on MAML and BiGRU," Journal of Intelligent & Fuzzy Systems, vol. 47, no. 3-4, pp. 307-319, 2024, doi: 10.3233/JIFS-241801.

[13] P. Bala, H. Tyagi, and V. Parashar, "Few-Shot Sentiment Adaptation: A MAML-Based Framework for Low-Resource NLP," International Journal on Computational Modelling Applications, vol. 2, no. 2, pp. 65-75, 2025, doi: 10.63503/j.ijcma.2025.94.

[14] G. Yi, J. Ju, and K. Jo, "Few-Shot Learning for Multi-Omics Disease Classification with a MAML-Based Model," The Transactions of the Korea Information Processing Society, vol. 14, no. 7, pp. 555-561, 2025.

[15] C. Tong, L. Zhang, H. Li, and Y. Ding, "Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting," IET Generation, Transmission & Distribution, vol. 16, no. 8, pp. 1680-1696, 2022, doi: 10.1049/gtd2.12394.

[16] Z. Zhang, W. Li, Y. Wang, Z. Wang, X. Sheng, and T. Zhou, "Multi-Scale Temporal Convolutional Networks and Multi-Head Attention for Robust Log Anomaly Detection," Information Technology and Control, vol. 53, no. 3, pp. 813-832, 2024, doi: 10.5755/j01.itc.53.3.35704.

[17] N. Liu, Q. Hu, H. Xu, X. Xu, and M. Chen, "Med-BERT: A pretraining framework for medical records named entity recognition," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5600-5608, 2021, doi: 10.1109/TII.2021.3131180.

[18] T. D. Nguyen, M. H. Shih, D. Srivastava, S. Tirthapura, and B. Xu, "Stratified random sampling from streaming and stored data," Distributed and Parallel Databases, vol. 39, no. 3, pp. 665-710, 2021, doi: 10.1007/s10619-020-07315-w.

[19] R. Iliyasu and I. Etikan, "Comparison of quota sampling and stratified random sampling," Biom. Biostat. Int. J. Rev, vol. 10, no. 1, pp. 24-27, 2021, doi: 10.15406/bbij.2021.10.00326.

[20] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," International Journal of Informatics and Information Systems, vol. 4, no. 1, pp. 13-20, 2021.

[21] H. Song, R. Yuan, Y. Lv, H. Liu, and Y. Li, "Cubic spline interpolation-based refined composite multiscale dispersion entropy and its application to bearing fault identification," Structural Health Monitoring, vol. 22, no. 4, pp. 2651-2673, 2023, doi: 10.1177/14759217221134050.

[22] B. Ranković, R. R. Griffiths, H. B. Moss, and P. Schwaller, "Bayesian optimisation for additive screening and yield improvements–beyond one-hot encoding," Digital Discovery, vol. 3, no. 4, pp. 654-666, 2024, doi: 10.1039/d3dd00096f.

[23] Z. Huang, Y. Wei, X. Wang, W. Liu, T. S. Huang, and H. Shi, "Alignseg: Feature-aligned segmentation networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 550-557, 2021.

[24] J. Liao, D. Liu, G. Su, and L. Liu, "Recognizing diseases with multivariate physiological signals by a DeepCNN-LSTM network," Applied Intelligence, vol. 51, no. 11, pp. 7933-7945, 2021, doi: 10.1007/s10489-021-02309-2.

[25] T. A. Adesuyi, B. M. Kim, and J. Kim, "Snoring sound classification using 1D-CNN model based on multi-feature extraction," International Journal of Fuzzy Logic and Intelligent Systems, vol. 22, no. 1, pp. 1-10, 2022, doi: 10.5391/IJFIS.2022.22.1.1.

[26] M. Ju, J. Luo, Z. Wang, and H. Luo, "Adaptive feature fusion with attention mechanism for multi-scale target detection," Neural Computing and Applications, vol. 33, no. 7, pp. 2769-2781, 2021, doi: 10.1007/s00521-020-05150-9.

[27] D. G. Wessels and D. F. Malan, "A limit equilibrium model to simulate time-dependent pillar scaling in hard rock bord and pillar mines," Rock Mechanics and Rock Engineering, vol. 56, no. 5, pp. 3773-3786, 2023, doi: 10.1007/s00603-023-03239-w.

[28] J. Bi, K. Xu, H. Yuan, J. Zhang, and M. Zhou, "Network attack prediction with hybrid temporal convolutional network and bidirectional GRU," IEEE Internet of Things Journal, vol. 11, no. 7, pp. 12619-12630, 2023.

[29] M. M. Abdelgwad, T. H. A. Soliman, A. I. Taloba, and M. F. Farghaly, "Arabic aspect based sentiment analysis using bidirectional GRU based models," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 9, pp. 6652-6662, 2022, doi: 10.1016/j.jksuci.2021.08.030.

[30] A. Kumar and S. Singh Sodhi, "Classification of data on stacked autoencoder using modified sigmoid activation function," Journal of Intelligent & Fuzzy Systems, vol. 44, no. 1, pp. 1-18, 2023, doi: 10.3233/JIFS-212873.

[31] B. A. Frasin, S. R. Swamy, and J. Nirmala, "Some special families of holomorphic and Al-Oboudi type bi-univalent functions related to k-Fibonacci numbers involving modified Sigmoid activation function," Afrika Matematika, vol. 32, no. 3, pp. 631-643, 2021, doi: 10.1007/s13370-020-00850-w.

[32] S. H. Oh, "Probabilistic Target Encoding of Neural Networks with Softmax Output Nodes," International Journal OF Contents, vol. 18, no. 4, pp. 102-108, 2022.

[33] G. L. Jones and Q. Qin, "Markov chain Monte Carlo in practice," Annual Review of Statistics and Its Application, vol. 9, no. 1, pp. 557-578, 2022, doi: 10.1146/annurev-statistics-040220-090158.

[34] A. G. Wills and T. B. Schön, "Sequential Monte Carlo: a unified review," Annual Review of Control, Robotics, and Autonomous Systems, vol. 6, no. 1, pp. 159-182, 2023, doi: 10.1146/annurev-control-042920-015119.

[35] L. Ding, T. Hu, J. Jiang, D. Li, W. Wang, and Y. Yao, "Random smoothing regularization in kernel gradient descent learning," Journal of Machine Learning Research, vol. 25, no. 284, pp. 1-88, 2024.