Neural Architecture for Fast and Reliable Coagulation Assessment in Clinical Settings: Leveraging Thromboelastography

Yulu Wang[1], **Ziqian Zeng**[2], **Jianjun Wu**[2], **Zhifeng Tang**[†1]
[1]Zhejiang University
[2]Huzhou Institute of Zhejiang University
{yulu.wang, zeng155, tangzhifeng}@zju.edu.cn, wujj@hizju.org

Abstract

In an ideal medical environment, real-time coagulation monitoring can enable early detection and prompt remediation of risks. However, traditional Thromboelastography (TEG), a widely employed diagnostic modality, can only provide such outputs after nearly 1 hour of measurement. The delay might lead to elevated mortality rates. These issues clearly point out one of the key challenges for medical AI development: Making reasonable predictions based on very small data sets and accounting for variation between different patient populations, a task where conventional deep learning methods typically perform poorly. We present Physiological State Reconstruction (PSR), a new algorithm specifically designed to take advantage of dynamic changes between individuals and to maximize useful information produced by small amounts of clinical data through mapping to reliable predictions and diagnosis. We develop MDFE to facilitate integration of varied temporal signals using multi-domain learning, and jointly learn high-level temporal interactions together with attentions via HLA; furthermore, the parameterized DAM we designed maintains the stability of the computed vital signs. PSR evaluates with 4 TEG-specialized data sets and establishes remarkable performance -- predictions of $R^2 > 0.98$ for coagulation traits and error reduction around half compared to the state-of-the-art methods, and halving the inferencing time too. Drift-aware learning suggests a new future, with potential uses well beyond thrombophilia discovery towards medical AI applications with data scarcity.

Introduction

Extracting meaningful information from limited and diverse datasets is a common predicament in training prediction models, even more so for those applications where failure directly affects human lives (Suresh et al., 2018). This compounds the issue as real-time updating is required, and the differences between instances can be so significant that the failings of standard deep learning arise. Literature identified two main strategies: few-shot learning and domain adaptation. Few-shot learning (An et al., 2025; Huang et al., 2024b), provided the ability to generalize from limited samples but is difficult due to the high variance in sample quality and distribution; whereas domain adaptation (Wang et al., 2024a; Feng et al., 2023) is able to transfer knowledge from one dataset to another provided that stable distributions are used during training, however, stable distributions are rare in clinical scenarios with dynamic and patient-specific instances. Thus, limited data with real-time drift adaptation has not been addressed properly. Examples of physiological monitoring challenges are evident in thromboelastography (TEG), used for evaluating coagulation. As a method to diagnose coagulation status by observing the viscoelasticity of blood, this method is helpful for gaining a real-time view of clot formation events, thus aiding greatly in treatment decisions during surgical situations. Conventional TEG testing requires time to complete, with results that are too late to inform accurate critical care decisions. A typical conventional TEG test would take nearly an hour, and an additional 15-minute delay increases the risk of trauma patients dying by approximately 10% (Gayet-Ageron et al., 2018). Predictive models for TEG need to be constructed. In fact, it reflects a broader problem: most current AI methods extract much less useful knowledge given very few observations; they also struggle with shifting distributions caused by the need to adjust for different populations. All of these contribute to poor model performance when there are changes in individual cases (Liu and Hauskrecht, 2017).

In recent years, the new neural network architectures provide advanced mechanisms to deal with varied kinds of data. For example, the Transformer variant (Liu et al., 2024a; Nie et al., 2023) can capture long-range dependency, while Kolmogorov-Arnold Networks (KANs) (Huang et al., 2024a; Genet and Inzirillo, 2024; Liu et al., 2024b). exhibits excellent capability of approximating complex functions, however, these models assume a static underlying distribution and sufficient training data, which is unrealistic in dynamic

---

[†] Corresponding author

real-world scenarios with limited training data. The theoretical gap lies in the insurmountable learn-theoretic requirements for accurate predictions in restricted environments. Traditional PAC-learning bounds are loose under severe sample complexity restrictions (Cohen-Addad et al., 2025), and drift detection methods require sufficient data for stable baselines, which can harm the performance of adaptive learning systems.

To surmount the challenges, we propose an innovative drift-aware learning paradigm—Physiological State Reconstruction (PSR). PSR presents a novel mathematical scheme that simultaneously achieves feature extraction, prediction, and adaptation in part-time series data, ensuring reliable inference in limited data scenarios compared to existing methods. PSR includes three key components:

- **Multi-Domain Feature Extraction (MDFE)**: A method to achieve the time-frequency domain decomposition while preserving the signal's original features.
- **Hierarchical Learning Architecture (HLA)**: An integration of KANs and attention mechanisms that enhances approximation for complex physiological functions while maintaining interpretability.
- **Dynamic Adaptation Module (DAM)**: A real-time adaptation mechanism that facilitates incremental learning from minimal observation.

PSR introduces drift-aware learning to address AI deployment challenges in data-scarce, safety-critical conditions, enabling real-time adaptation and contributing to: (1) A new paradigm for safety-critical AI that improves responsiveness to data changes; (2) The theoretical basis of making reliable forecasts to achieve efficient information mining; (3) Utilize few case records to deliver clinical-grade care.

Related Works

Multi-domain Integration in Time Series. Multi-domain feature extraction and fusion are essential for improving time series forecasting. While single-domain methods like ARIMA identify stable trends, they miss nonlinear dynamics in complex datasets (Li et al., 2023). Researchers are now integrating time and frequency domains. For instance, CTFNet combines convolutional mapping with time-frequency decomposition, reducing forecasting error by 64.7% (Zhang et al., 2024). TFMSNet uses multi-scale processing for effective feature fusion across 70 datasets (Song et al., 2025). These techniques are crucial for capturing intricate physiological data features that PSR aims to utilize.

Prior Work on TEG Modeling. Fast coagulation assessment and effective TEG modeling must be real-time. The Biological Mechanism-Driven Model (BPTM), using blood protein concentration, predicts TEG output to understand plasma coagulation in emergencies, effectively describing the relationships between blood proteins and patient coagulation (Ghetmiri et al., 2024). Besides, the knowledge that not all models can retain all the biological processes was also gained via KAN-based models—TimeKAN (Huang et al., 2024a) and TKAN (Genet and Inzirillo, 2024), Multilayer Perceptrons-based models—TimeMixer (Wang et al., 2024b), DLinear (Zeng et al., 2023), and FreTS (Yi et al., 2023), Transformer-based models—iTransformer (Liu et al., 2024a) and PatchTST (Nie et al., 2023), they generally omit most biological conditions and only integrate the related time series characteristics.

Concept Drift Handling Techniques. Concept drift handling techniques highly challenge time series analysis in the context of health care, and most recent approaches focus on actively detecting drift and passively adapting models via shifting models according to changes in data (Liu et al., 2023), some active methods fuse manifold projection and statistical process control to obtain better results (Wang et al., 2023). A hybrid feature extraction algorithm can detect drift occurring in a stream dataset more quickly because of incremental learning (Yu et al., 2022). Incremental learning also can cause the models to fit different sets of data according to data variations, for in-stance, StreamWNN increases accuracy as data are gradually added (Melgar-Garcia et al., 2023), and OneNet decreases errors by more than 50% (Wen et al., 2023). The importance of adapting to drift and keeping the model's capability has come up again. From PSR's point of view, it advocates to increase the DAM.

Preliminaries

This section presents the Physiological State Reconstruction (PSR) framework, composed of two stages—the offline pretraining stage and the online prediction stage—which is provided with three main modules to conduct estimation.

Offline Pretraining & Online Prediction. Let L_{pred} represent the number of historical TEG data points used in the online stage for accurate output estimations. During offline, pretraining the model scans the historical TEG series $\mathrm{X}(t_1),\ldots,\mathrm{X}(t_{L_{\text{pre}}})$ with an N-point sliding window $\mathrm{X}_N(i)=[\mathrm{X}(t_i),\ldots,\mathrm{X}(t_{i+N-1})]$ and its timestamps $T_N(i)=[t_i,\ldots,t_{i+N-1}]$. The feature extractor converts each window into a matrix $\mathrm{X}_{\text{multi}}(i)=\mathrm{MDFE}(\mathrm{X}_N(i),\mathrm{T}_N(i))\in\mathbb{R}^{N\times F_0}$, capturing salient time and frequency domain cues, with F_0 representing the total features derived from TEG data. The Hierarchical Learning Architecture (HLA) predictor f_θ^{pre}, combined with the Dynamic Adaptation Module (DAM), is trained on these features, and its worst observed error

$$\delta_0=\sup_{i\le L_{\text{pre}}}\left|f_\theta^{\text{pre}}(\mathrm{X}_{\text{multi}}(i))-y(i)\right| \;\geq\; 0 \tag{1}$$

serves as the baseline for later online updates. After being pretrained, the model moves on to the online prediction process. A TEG curve $\hat{\mathrm{X}}\in\mathbb{R}^M$ needs to be reconstructed given a partial sequence of recorded TEG traces $\mathrm{X}=[\mathrm{X}(t_1),\ldots,\mathrm{X}(t_m)]\in\mathbb{R}^m$, where $M>m$.

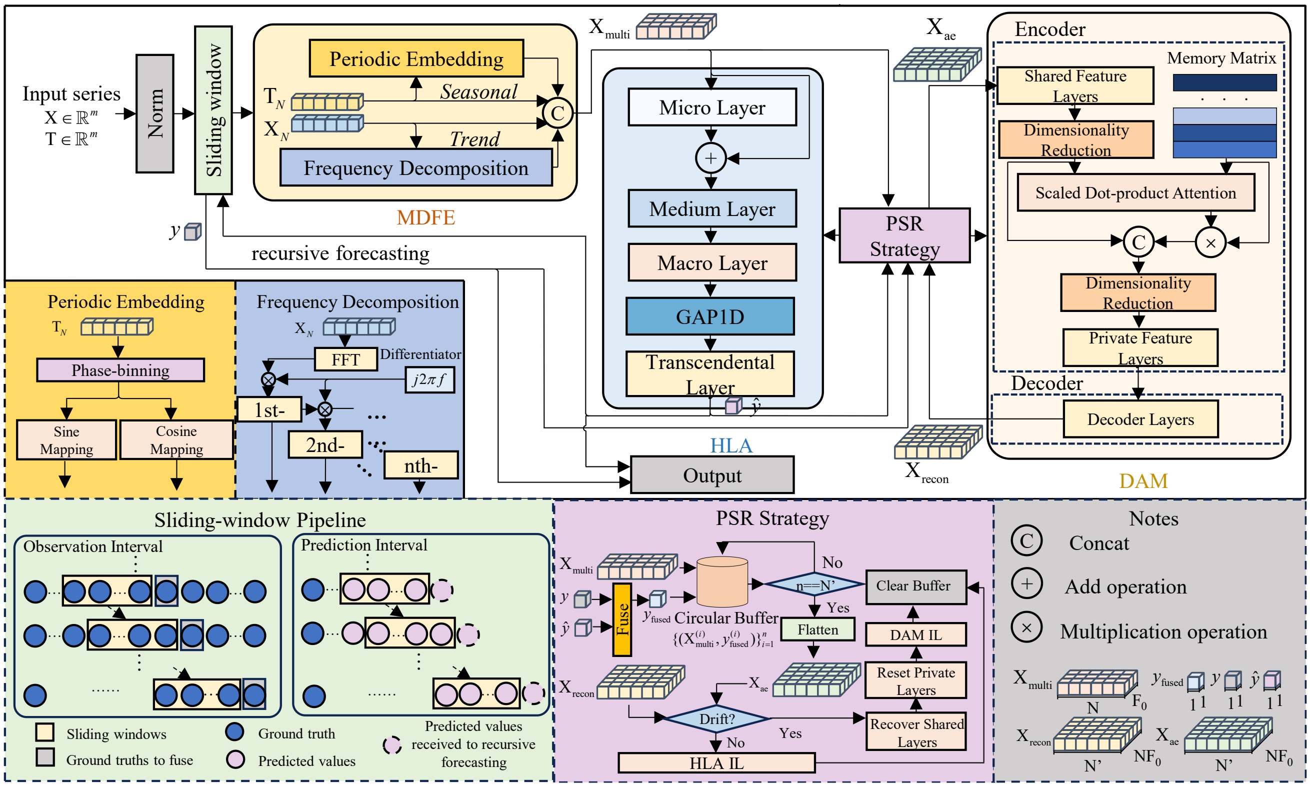


Figure 1: Overall structure of PSR. This figure illustrates the PSR framework. MDFE extracts key features across time and frequency domains. HLA combines Kolmogorov-Arnold networks with attention mechanisms for improved accuracy. DAM enables real-time learning from minimal data while preventing catastrophic forgetting.

The Index sets are: $I_{\text{obs}} = \{1,\ldots,m\}$, $I_{\text{pred}} = \{m+1,\ldots,M\}$, where $\beta(i) \in (0,1)$ smoothly increases across I_{obs}. The predictor is initialized with $f_\theta^{\text{online}} \leftarrow f_\theta^{\text{pre}}$. For each $i = 1,\ldots,M$:

Sliding-Window & Feature Extraction. Over the data, create sliding windows of N $\mathrm{T}_N(i) = [t_i,\ldots,t_{i+N-1}] \in \mathbb{R}^N$, $\mathrm{X}_N(i) = [\mathrm{X}(t_i),\ldots,\mathrm{X}(t_{i+N-1})] \in \mathbb{R}^N$. Multi-channel features are obtained: $\mathrm{X}_{\text{multi}}(i) = \mathrm{MDFE}(\mathrm{X}_N(i), \mathrm{T}_N(i)) \in \mathbb{R}^{N\times F_0}$, $\hat{y}(i) = f_\theta^{\text{online}}(\mathrm{X}_{\text{multi}}(i)) \in \mathbb{R}$.

Fusion Target. The fusion target should be defined:

$$y_{\text{fused}}(i) = \begin{cases} \beta(i)\cdot\hat{y}(i) + (1-\beta(i))\cdot y(i), & \text{if } i+N \in I_{\text{obs}}, \\ \hat{y}(i), & \text{otherwise.} \end{cases} \quad (2)$$

The fusion stage aligns predictions with observed values, so that estimates better represent reality.

Circular Buffer & Recursion. Store $(\mathrm{X}_{\text{multi}}(i), y_{\text{fused}}(i))$ in a circular buffer of size. Only use $y_{\text{fused}}(i)$ for recursive forecasting if there are not enough entries less than *N'*.

Concept Drift Detection. Upon the buffer being filled up, flatten the features: $\mathrm{X}_{\text{adapt}} \in \mathbb{R}^{N'\times(N\cdot F_0)}$. DAM reconstruction yields $\mathrm{X}_{\text{recon}}$. The reconstruction loss is: $\mathcal{L}_{\text{rec}} = \|\mathrm{X}_{\text{adapt}} - \mathrm{X}_{\text{recon}}\|_2^2 / N'NF_0$. If the loss exceeds the threshold ε, that implies concept drift has occurred in the model. We need to reset DAM's shared layers and initialize its private layers currently; otherwise, only updating HLA. Thresholds are obtained analytically.

Incremental Learning Gains & Error Sequence. The incremental learning capabilities are evaluated through:

$$\begin{aligned} & g_\theta(X) = f_\theta(\mathrm{MDFE}(X,T)), \\ & \gamma_i = \sup_X |g_{\theta_{i-1}}(X) - y_{\text{fused}}(i)| - \sup_X |g_{\theta_i}(X) - y_{\text{fused}}(i)| \ge 0, \\ & \lambda_i = \sup_X |g_{\theta_i}(X) - y_{\text{fused}}(i)| - \sup_X |g_{\theta_i^+}(X) - y_{\text{fused}}(i)| \ge 0, \\ & \delta_i = \delta_{i-1} - (\gamma_i + \lambda_i) \ge 0. \end{aligned} \quad (3)$$

With $\gamma_i, \lambda_i \ge 0$, the sequence $\{\delta_i\}$ is nonincreasing, ensuring $\delta_i \ge 0$. Assuming $g_\theta(X)$ is *L*-Lipschitz in *X*.

Theorem 1. *(Adaptive Learning Convergence)*
Under established assumptions, for any $1 \le n \le M$:

$$\sum_{i=1}^{n} |\hat{y}(i) - y(i)| \le \sum_{i=1}^{n} [1+\beta(i)] L [\delta_0 - \sum_{j=1}^{i-1} (\gamma_j + \lambda_j)]. \quad (4)$$

This theorem establishes error bounds for our framework, ensuring manageable prediction errors as data grows. Nonnegative terms indicate model improvement with each update. The proof of Equation (4) is in Appendix A. A tighter $\mathcal{O}(\sqrt{T})$ *regret* result (proved in Appendix C.3) shows that the cumulative absolute error satisfies

$$\sum_{t=1}^{T} |\hat{y}_t - y_t| \le (\delta_0 + G)\sqrt{T}. \quad (5)$$

Using adaptive steps and memory terms from DAM, our bound tightens Equation 4's $\mathcal{O}(T)$ to $\mathcal{O}(\sqrt{T})$, explaining why PSR oscillates back to steady state quickly even during drift. The TEG curve was reconstructed $\hat{\mathrm{X}} = [[\mathrm{X}(t_i)]_{i\in I_{\text{obs}}}, [\hat{y}(i)]_{i\in I_{\text{pred}}}] \in \mathbb{R}^M$.

Methodology

Multi-Domain Feature Extraction

The Multi-Domain Feature Extraction (MDFE) module is vital to the Physiological State Reconstruction (PSR) framework for processing TEG data using two operators for feature extraction.

Periodic Embedding Operator (*P*): This operator is applied to capture the biological rhythm (seasonal features) with phase-binning and cycle length C_j and period constants P_j for $j = 1,\ldots,l$: $V_j(i) = \lfloor \mathrm{T}_N(i) / P_j \rfloor \bmod C_j \in \mathbb{Z}^N$. The corresponding periodic matrices are constructed as: $A_j(i) = [\sin(2\pi V_j(i) / C_j), \cos(2\pi V_j(i) / C_j)] \in \mathbb{R}^{N\times 2}$. Then we use the constructed periodic matrix to concatenate them and build the periodic feature matrix: $A(i) = [A_1(i), \ldots, A_l(i)] \in \mathbb{R}^{N\times 2l}$, which can refine the evaluation result on rhythms of physiological change to represent the coagulation status.

Frequency Decomposition Operator (*F*): The operator decomposes a signal into the frequency domain using Fast Fourier Transform: $F_N(i) = \mathrm{FFT}(\mathrm{X}_N(i)) \in \mathbb{C}^N$. For each frequency bin $f_n = n / N$ and derivative order $k = 0,\ldots K-1$, features are calculated: $b_{k,n}(i) = (j2\pi f_n)^k F_N(i)[n]$. These coefficients reflect trend features, the real parts of them are preserved: $B(i) \in \mathbb{R}^{N\times K}$, $B(i)_{n,k} = \Re[b_{k,n}(i)]$.

The outputs from each operator create the feature set: $\mathrm{X}_{\mathrm{multi}}(i) = [\mathrm{T}_N(i), \mathrm{X}_N(i), A(i), B(i)] \in \mathbb{R}^{N\times F_0}$, where $F_0 = 2 + 2l + K$. This process captures both seasonal and trend patterns.

Hierarchical Learning Architecture

The HLA enhances physiological signal prediction using a structured layered approach with input $\mathrm{X}_{\mathrm{multi}}(i) \in \mathbb{R}^{N\times F_0}$.

Micro Layer (Mi-L): Use a Residual Convolutional Neural Network (ResCNN)to extract key local features for coagulation detection through convolutional patterns.

Medium Layer (Me-L): Implement an LSTM network that captures temporal dependence. Thus, the model understands the past and can detect trends.

Macro Layer (Ma-L): Employ a multi-head self-attention mechanism to identify global relationships among features, dynamically weighing their significance.

Transcendental Layer (TL): Extract knowledge from previous layers using KANs to generate precise single-step prediction $\hat{Y}(i) \in \mathbb{R}$, integrating local and global context. The functional model:

$$\hat{Y}(i) = (\mathrm{TL} \circ \text{Ma-L} \circ \text{Me-L} \circ \text{Mi-L})\mathrm{X}_{\mathrm{multi}}(i) \tag{6}$$

Theorem 2 *Let* $D \subset \mathbb{R}^{N\times F_0}$ *be a compact set and* $f : D \to \mathbb{R}$ *a continuous function. Denote by* $\| \cdot \|$ *the Euclidean norm on the ambient vector space. Then for every* $\varepsilon > 0$ *there exist integers* F, d, h, u, K *and a choice of all trainable parameters θ in the HLA such that, for the model mapping* $\hat{Y}(\mathrm{X}_{\mathrm{multi}}, \theta) : D \to \mathbb{R}$, *we have:*

$$\sup_{\mathrm{X}_{\mathrm{multi}} \in D} \left| \hat{Y}(\mathrm{X}_{\mathrm{multi}}, \theta) - f(\mathrm{X}_{\mathrm{multi}}) \right| < \varepsilon \tag{7}$$

Theorem 2 ensures that with the right parameters, the model can approximate any continuous function over D with bounded error and obtain the desired accuracy ε which enables HLA to grasp complex information relationships for predicting results with high accuracy (Refer to Appendix A).

Dynamic Adaptation Module

The Dynamic Adaptation Module (DAM) is a key component of the PSR framework, adapting to changes in physiological signals through three core insights introduced:

1. Physiological Equilibrium Hypothesis: Healthy individuals maintain physiological balance:

$$S = \{s \mid \|\Phi(s) - \mu\| < \varepsilon\}, \tag{8}$$

where $\Phi(s)$ is the physiological state and μ is the equilibrium reference.

2. Pathological Drift Theorem: Both pathological conditions and normal variations among individuals lead to non-linear shifts in physiological parameters:

$$D(t) = KL(P(t) \| P_{baseline}) > \text{threshold} \tag{9}$$

where $P(t)$ is the current parameter distribution and P_{baseline} is the normative profile.

3. Adaptive Convergence Principle: The DAM features a memory-enhanced learning mechanism that improves adaptability to physiological signal changes by retaining relevant historical information for flexible adaptation to new data, used for re-balancing in the DAM:

$$\theta(t+1) = \theta(t) + \eta\nabla L + \beta\nabla M \tag{10}$$

Here, η is the learning rate, ∇L is the loss gradient, and ∇M accounts for information from earlier times. More details and proof of these ideas can be found in Appendix C. Once a buffer is full, multiple channels of input get transformed to $\mathrm{X}_{\mathrm{adapt}} = \mathrm{reshape}(\mathrm{X}_{\mathrm{multi}}, [N', N \cdot F_0]) \in \mathbb{R}^{N' \times (N \cdot F_0)}$. The DAM uses 3 shared KAN layers to extract features: $\mathrm{X}_{\mathrm{shared}} = (\mathrm{KAN}^{(3)}_{\mathrm{shared}} \circ \mathrm{KAN}^{(2)}_{\mathrm{shared}} \circ \mathrm{KAN}^{(1)}_{\mathrm{shared}})\mathrm{X}_{\mathrm{adapt}} \in \mathbb{R}^{N' \times F_{\mathrm{shared}}}$, where F_{shared} specifies the feature dimension that matches basic physiological patterns.

These features are projected to a lower-dimensional space: $Q = X_{\mathrm{shared}} \cdot W_{\mathrm{proj}} + b_{\mathrm{proj}} \in \mathbb{R}^{N' \times E}$, where W_{proj} and b_{proj} are reduction parameters. A learnable Memory Matrix (MM) captures historical data: $M_{\mathrm{read}} = \mathrm{softmax}(QM^{\mathrm{T}})M \in \mathbb{R}^{N' \times E}$. KAN maps inputs to 1-D features as keys, while memory M holds recent target. Current and historical data are integrated as: $\mathrm{X}_{\mathrm{mem}} = [Q, M_{\mathrm{read}}]W_{\mathrm{mem}} + b_{\mathrm{mem}} \in \mathbb{R}^{N' \times F_{\mathrm{mem}}}$.

Then we use 3 private KAN layers to get $\mathrm{X}_{\mathrm{private}} \in \mathbb{R}^{N' \times F_{\mathrm{private}}}$, where F_{private} denotes the dimensionality of individual-specific features. The final output is generated through 3 decoder layers as $\mathrm{X}_{\mathrm{recon}} \in \mathbb{R}^{N' \times (N \cdot F_0)}$. The setup provides increased flexibility while keeping the model robust to the fluctuations in the individual's feature vectors.

Experiments

Simulation Study

This study is about simulating prediction models based on concept drift with the usage of synthetic data; it is shown that the proposed model can capture the variation from changes in the physiological signals.

We generated a synthetic dataset with predefined time series characteristics and introduced concept drift to mimic sudden transitions in human physiological states. As shown in Figure 2, the green line represents ground truth values, while the red dotted line denotes model predictions considering drift detection mechanisms.

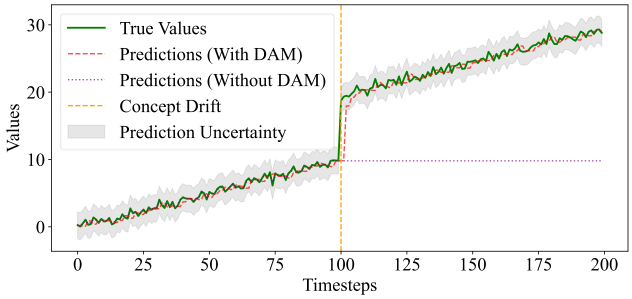


Figure 2: Simulation of Concept Drift Detection.

Main Results

Datasets. We evaluated four TEG datasets for real-time coagulation curve reconstruction: HPP and HPC (healthy platelet-poor plasma), HWC (healthy whole blood), and TWA (trauma patient whole blood). These datasets encompass standard physiological conditions and the higher heterogeneity of trauma patients, ensuring robust model training and real-world applicability. Details are in Appendix B.

Baselines. We select 9 representative models to serve as baselines, including (1) Biological Mechanism-Driven model: BPTM (Blood Protein-Based TEG Model) (Ghetmiri et al., 2024); (2) KAN-based models: TimeKAN (Huang et al., 2024a), TKAN (Genet and Inzirillo, 2024), KAN (Liu et al., 2024b); (3) MLP-based models: TimeMixer (Wang et al., 2024b), DLinear (Zeng et al., 2023); (4) Transformer-based models: iTransformer (Liu et al., 2024a), PatchTST (Nie et al., 2023); (5) Frequency-based model: FreTS (Yi et al., 2023).

Experimental Settings. Five-fold cross-validation was adopted: 80% of the data were training and validation sets, and 20% was used as the test set. In terms of training, the 13th gen Intel® Core™ i9-13900HX CPU was utilized for GPU-free clinical real-time use. Every model applied Adam with an L2 loss function. More details are in Appendix D.

Results. Table 1 shows Mean Absolute Errors (MAE), Mean Squared Errors (MSE), and Coefficients of Determination (R^2) for each model on the HPP, HPC, HWC, and TWA, with the first place in bold font, the second place in underlined text, and the third place in dashed text. In all but one instance, PSR outperformed all the baselines by large margins, reducing MAE by 50–80%, and MSE by 45–95% relative to the best baseline (BPTM) while achieving R^2 scores >0.98. In TWA, BPTM only marginally outperforms PSR; however, both methods greatly exceed all others. Overall, PSR achieves an MAE of 0.056 and an MSE of 0.017, about 50% and 33% of BPTM's performances, showcasing exceptional real-time coagulation curve forecasting accuracy despite low sample sizes and non-stationary data.

Ablation Study

To verify the effectiveness of each component of PSR, we provide a detailed ablation study on every possible design in MDFE, HLA, DAM, and PSR Strategy (Table 2).

Study on MDFE. In ablation② (only the periodic embedding operator), removing the MDF component causes the MAE to increase from 0.056 to 0.383, and the R2 to decrease from 0.989 to 0.695. In ablation③ (only frequency decomposition operator), removing TP embedding leads to an MAE of 0.482 and an R2 of 0.382. This shows that TP and MFD both play important roles in the coagulation model.

Study on HLA. Ablation ④–⑦ analyzed the contribution of every single layer:

- Ablation ④: Mi-L + TL (KAN).
- Ablation ⑤: Mi-L + Ma-L + TL (KAN).
- Ablation ⑥: Mi-Layer + Me-L + TL (KAN).
- Ablation ⑦: Mi-L + Me-L + Ma-L + TL (MLP).

This discovery underlines that almost every part is important for precise long-distance predictions, especially Me-L. Only the complete HLA achieved high performance, underscoring the need for integrating Mi-L, Me-L, Ma-L, and TL (KAN) for accurate coagulation reconstruction.

Model	**PSR (Ours)**			BPTM (2024)			TimeKAN (2025)			TKAN (2024)			KAN (2024)			TimeMixer (2024)			DLinear (2023)			iTransformer (2024)			PatchTST (2023)			FreTS (2024)		
Metric	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2
HPP	**0.050**	**0.010**	**0.990**	0.092	0.018	0.982	0.190	0.168	0.839	0.189	0.117	0.883	0.224	0.171	0.832	0.139	0.110	0.896	0.115	0.063	0.937	0.098	0.042	0.960	0.273	0.269	0.743	0.139	0.089	0.916
HPC	**0.034**	**0.004**	**0.996**	0.132	0.050	0.809	0.271	0.344	0.669	0.409	0.305	0.690	0.649	0.733	0.248	0.359	0.517	0.509	0.284	0.248	0.759	0.296	0.311	0.711	0.241	0.131	0.870	0.270	0.272	0.740
HWC	**0.060**	**0.011**	**0.989**	0.342	0.224	0.783	0.146	0.075	0.929	0.239	0.143	0.863	0.225	0.090	0.909	0.187	0.129	0.878	0.270	0.211	0.798	0.147	0.103	0.903	0.391	0.344	0.660	0.188	0.128	0.878
TWA	0.062	0.023	0.979	**0.047**	**0.005**	**0.980**	0.144	0.135	0.872	0.213	0.223	0.789	0.195	0.192	0.818	0.136	0.129	0.877	0.153	0.142	0.865	0.158	0.151	0.856	0.171	0.156	0.852	0.162	0.142	0.864
Overall	**0.056**	**0.017**	**0.989**	0.109	0.048	0.922	0.168	0.161	0.912	0.246	0.215	0.870	0.275	0.261	0.860	0.180	0.190	0.899	0.190	0.164	0.900	0.174	0.160	0.907	0.226	0.191	0.879	0.181	0.157	0.908

Table 1: Main results. The look-back window was set as 55% of the sequence length for HPP, but it was reduced to 45% in the case of longer datasets to minimize redundancy. Across the five folds, we report MAE, MSE and R^2 values.

	MDFE		HLA				DAM		PSR Strategy		HPP			HPC			HWC			TWA			Overall		
	P	***F***	**Mi-L**	**Me-L**	**Ma-L**	**TL**	**MLP/KAN**	**MM**	**DAM**	**fusion**	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE	R^2
①	✓	✓	✓	✓	✓	**KAN**	**KAN**	✓	✓	✓	**0.050**	**0.010**	**0.990**	**0.034**	**0.004**	**0.996**	**0.060**	**0.011**	**0.989**	**0.062**	**0.023**	**0.979**	**0.056**	**0.017**	**0.989**
②	✓	✗	✓	✓	✓	**KAN**	**KAN**	✓	✓	✓	0.261	0.289	0.707	0.453	0.664	0.319	0.285	0.281	0.716	0.407	0.531	0.490	0.383	0.493	0.695
③	✗	✓	✓	✓	✓	**KAN**	**KAN**	✓	✓	✓	0.339	0.408	0.597	0.887	2.161	-1.081	0.613	1.502	-0.468	0.358	0.714	0.310	0.482	1.048	0.382
④	✓	✓	✓	✗	✗	**KAN**	**KAN**	✓	✓	✓	0.154	0.126	0.866	0.550	2.222	-0.951	0.086	0.023	0.977	-	-	-	0.285	0.924	0.604
⑤	✓	✓	✓	✗	✓	**KAN**	**KAN**	✓	✓	✓	0.162	0.121	0.875	0.320	0.328	0.683	0.208	0.323	0.682	-	-	-	0.244	0.285	0.802
⑥	✓	✓	✓	✓	✗	**KAN**	**KAN**	✓	✓	✓	0.299	0.721	0.231	0.516	1.003	0.057	0.154	0.145	0.855	0.208	0.236	0.775	0.256	0.383	0.809
⑦	✓	✓	✓	✓	✓	**MLP**	**KAN**	✓	✓	✓	0.204	0.172	0.818	0.548	0.912	0.084	0.066	0.015	0.985	0.344	0.571	0.452	0.321	0.505	0.701
⑧	✓	✓	✓	✓	✓	**KAN**	**KAN**	✗	✓	✓	0.191	0.190	0.810	0.324	0.613	0.412	0.291	0.698	0.311	0.330	0.653	0.365	0.311	0.611	0.598
⑨	✓	✓	✓	✓	✓	**KAN**	**MLP**	✓	✓	✓	0.125	0.099	0.894	0.110	0.039	0.964	0.114	0.086	0.910	0.140	0.188	0.819	0.130	0.141	0.906
⑩	✓	✓	✓	✓	✓	**KAN**	**MLP**	✗	✓	✓	0.159	0.146	0.851	0.152	0.065	0.938	0.117	0.088	0.909	0.110	0.061	0.942	0.122	0.072	0.956
⑪	✓	✓	✓	✓	✓	**KAN**	✗	✗	✗	✓	0.477	0.911	0.081	0.428	0.526	0.458	0.518	0.720	0.286	0.929	1.977	-0.906	0.747	1.459	0.030
⑫	✓	✓	✓	✓	✓	**KAN**	**KAN**	✓	✓	✗	0.172	0.222	0.783	0.121	0.049	0.951	0.156	0.187	0.816	0.211	0.379	0.644	0.185	0.283	0.815

Table 2: Ablation study results on each component of PSR. Ablations ④ and ⑤ did very poorly on the TWA dataset, leading to their absence from the table (denoted as"–"). Their scores were based on only three non-TWA datasets.

Points Known Portion		25%	35%	45%	55%
HPP	MAE	0.595	0.512	0.256	**0.050**
	MSE	1.153	0.920	0.249	**0.010**
	R^2	-0.078	0.143	0.767	**0.990**
HPC	MAE	0.414	0.234	0.034	**0.034**
	MSE	0.686	0.325	0.004	**0.003**
	R^2	0.296	0.686	0.996	**0.997**
HWC	MAE	0.515	0.293	0.060	**0.058**
	MSE	0.786	0.343	0.011	**0.010**
	R^2	0.216	0.650	0.989	**0.990**
TWA	MAE	0.127	0.063	0.062	**0.037**
	MSE	0.080	0.024	0.023	**0.006**
	R^2	0.925	0.978	0.979	**0.994**
Overall	MAE	0.273	0.163	0.072	**0.041**
	MSE	0.375	0.194	0.036	**0.007**
	R^2	0.800	0.907	0.984	**0.996**

Table 3: A comprehensive performance comparison by varying look-back window.

Study on DAM. Ablations confirm the KAN–memory synergy. Removing the memory matrix (MM) from KAN (⑧) wipes out its recall and collapses performance. Swapping KAN for an MLP (⑨, ⑩) shows that MLPs lack the read/write bias: with MM, they mostly fetch noise; without MM, they train stably but gain only minor generalisation. Only the full KAN + MM setup (①) excels—the Kolmogorov–Arnold keys let MM retrieve the right rows. Thus, as predicted in Appendix C.3, MM helps only with KAN, and neither component shines alone.

Study on PSR Strategy. The PSR study in ablation experiments. As seen in ablation ⑪, removal of the DAM leads to poor long-horizon performance; whereas, ablation ⑫, without the fusion mechanism but with DAM incorporated, only demonstrates a slight decrease in performance when compared to the standard model, implying that incremental update of DAMs and adaptively weight drift-aware fusion is required for robust and precise prediction.

Model Analysis

Varying Look-back Window. Varying the size of the window from PSR (25%-55%). The corresponding results can be found in Table 3 for different datasets, including HPP, HPC, HWC, and TWA, under MAE, MSE, and R^2.

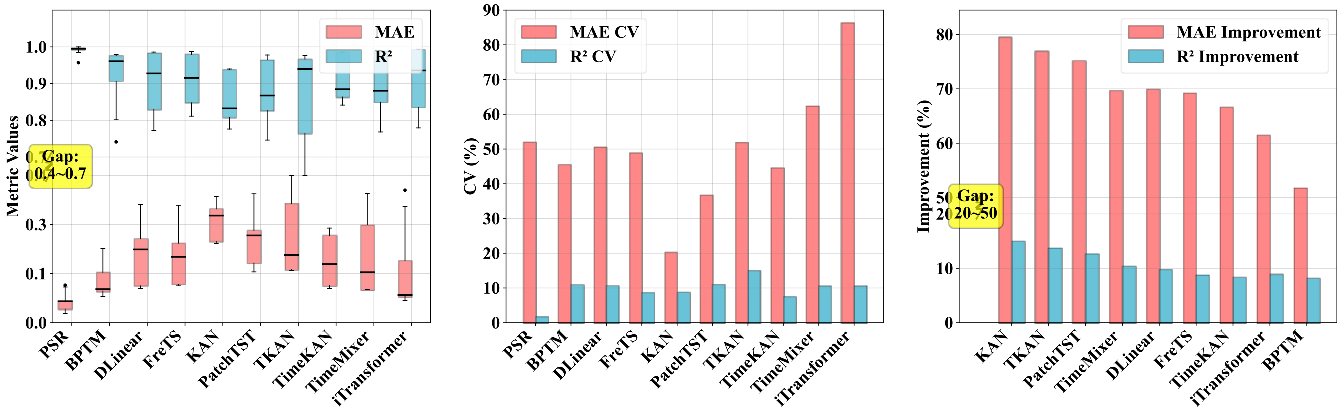
Expanding the look-back window improves accuracy by up to 45%, with diminishing returns beyond that. Increasing from 25% to 35% reduces MAE from 0.273 to 0.163 and raises R^2 from 0.800 to 0.907. A further increase to 45% decreases MAE to 0.072 and elevates R^2 above 0.98. Beyond 45%, gains taper off, indicating this window effectively captures core dynamics for accurate forecasting. The HPP series is poor at 45% ($R^2 = 0.767$) but improves at 55% ($R^2 = 0.990$), so we use 55% for HPP and 45% for HPC, HWC, and TWA. TWA obtained an R^2 of 0.925 and MAE of 0.127 for only 25%, which was due to trauma-induced inflection points, varying clotting response in different persons, and sharp biomarker transitions within the initial phase of data.

Model Efficiency. Table 4 shows that the training process of the PSR takes 0.035s/0.26MB per step—slower than iTransformer (0.024s/0.85MB)—but still much smaller than TKAN (0.539s/25.5MB). For inference, it uses 0.041s/0.051MB, which is slower than ultralight methods like PatchTST (5×10^{-5}s/0.0019MB) and DLinear (2×10^{-5}s), but real time nevertheless. PSR alone features adaptive inference, however, it can take an extra 0.28s/5.05MB when using extra DAM updates during inference. Overall, it supports near real-time with top accuracy despite only moderate computational resources required.

Model Interpretability. SHAP quantified feature contributions for one-step clot strength predictions. For N-step windows, averaged 8 inputs (time, clot strength, sin/cos(time), FFT, and its 3 derivatives) into features. Computed on 100 random samples. Details in Appendix F.

Models	Training Time (s/iter)	Training ΔRSS (MB/iter)	Pure Inference Time(s/step)	Pure Inference ΔRSS (MB/step)	Adaptive Inference Time (s/step)	Adaptive Inference ΔRSS (MB/step)
PSR	0.035	**0.26**	4.10e-2	5.10e-2	**0.28**	**5.05**
TimeKAN	0.043	2.54	8.89e-5	5.74e-3	-	-
TimeMixer	0.039	1.44	6.97e-5	2.23e-3	-	-
TKAN	0.539	25.48	2.87e-3	8.60e-2	-	-
KAN	0.089	5.91	8.51e-4	1.47e-2	-	-
iTransformer	0.024	0.85	3.16e-5	2.65e-3	-	-
FreTS	0.019	7.93	5.70e-5	4.80e-3	-	-
PatchTST	0.010	0.71	5.06e-5	**1.88e-3**	-	-
DLinear	**0.002**	0.37	**2.22e-5**	1.90e-3	-	-

Table 4: Comparative analysis of model performance and inference efficiency. This table compares models based on their training times (seconds per iteration) and resource consumption. ΔRSS (Change in Resident Set Size) indicates the difference in memory usage (MB) during operations, reflecting additional memory utilized during training and inference.



(a) Distribution Analysis (b) Variability Analysis (c) Wilcoxon Signed-Rank Test

Figure 3: Overall statistical evaluation of model performance.

Statistical Robustness Analysis

To ensure reliability, we performed statistical analyses beyond mean comparisons, as shown in Figure 3, which includes distribution characteristics, performance stability, and significance testing using 5-fold cross-validation.

Distribution Analysis (Figure 3a). A Box plot shows different distributions of performance. PSR has a much better central tendency with the mini-mum of outlier appearance than competitors for both MAE and R^2, while the other methods exhibit wide distribution and more outlying values especially for larger errors.

Variability Analysis (Figure 3b). PSR showed impressive stability by calculating the coefficient of variation, $CV=(\sigma/\mu)\times 100\%$, where σ is the standard deviation, μ is mean. For PSR, the CV was 1% for R^2 and 52% for MAE.

Statistical Significance (Figure 3c). Wilcoxon signed-rank tests confirmed significant improvements for PSR across all comparisons, with MAE reductions of 51% to 79% and R^2 improvements of 8% to 15% compared to the best baselines, indicated by p-values<0.05.

Note: Due to the small sample size (n=5 folds), we employed non-parametric Wilcoxon signed-rank tests alongside coefficient of variation analysis to provide robust evidence of statistical and practical significance. All statistical tests were conducted with $\alpha = 0.05$.

Conclusion

This paper introduces Physiological State Reconstruction (PSR), a drift-aware framework of real-time coagulation assessment with minimum data. MDFE, HLA, and DAM come along as the basic parts of it. The combination of them forms a new medical AI algorithm. It has proven good effects with these evaluations: $R^2 > 0.98$, a decrease in MAE over 50%, and the diagnostic time being half the speed of traditional Thromboelastography. Moreover, it delivers up to sub-second inference speeds. Outside of coagulation assessments, the model is a pioneer in drift-aware learning as a mechanism that is formed for highly dynamic, data-scarce medical use cases, filling essential gaps for adaptive deployment of deep learning-based AI.

Ethics Statement

This research adheres to established ethical guidelines for artificial intelligence and machine learning studies, including those outlined by the AAAI Code of Ethics and Professional Conduct. All datasets utilized in this work are publicly available from online sources, ensuring no involvement of human subjects, personal data collection, or experimental interventions that could raise privacy or consent concerns. Detailed descriptions of the datasets, including their sources, licenses, and any preprocessing steps, are provided in the Appendix to promote transparency and reproducibility.

We have assessed potential biases in the datasets. The study does not pose risks of harm to individuals or society, and we encourage responsible use of the proposed methods to avoid unintended application. No conflicts of interest exist, and all work complies with relevant data protection regulations, including open-source licensing terms.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 12474469) and the Huzhou Science and Technology Plan Project (No. 2024GZ15). We would also like to thank our colleagues for their valuable discussions and the reviewers for their constructive feedback.

References

An, H.; Oh, E.; Kim, Y.; Kim, S.; Park, D.; and Oh, C. 2025. ID.EARS: One-Ear EEG Device with Biosignal Noise for Real-Time Gesture Recognition and Various Interactions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 1–18. Yokohama, Japan: Association for Computing Machinery.

Cohen, M. J.; Bir, N.; Rahn, P.; Dotson, R.; Brohi, K.; Chesebro, B. B.; Mackersie, R.; Carles, M.; Wiener-Kronish, J.; and Pittet, J.-F. 2009. Protein C Depletion Early after Trauma Increases the Risk of Ventilator-Associated Pneumonia. *Journal of Trauma-Injury Infection and Critical Care* 67(6): 1176–1181. doi.org/10.1097/TA.0b013e3181c1c1bc.

Cohen-Addad, V.; Lattanzi, S.; and Schwiegelshohn, C. 2025. Almost Optimal PAC Learning for k-Means. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 2019–2030. Prague, Czechia: Association for Computing Machinery.

Feng, Y.; Zhu, H.; Peng, D.; Peng, X.; and Hu, P. 2023. ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7264–7273. Ottawa, ON, Canada: Association for Computing Machinery.

Gayet-Ageron, A.; Prieto-Merino, D.; Ker, K.; Shakur, H.; Ageron, F.-X.; and Roberts, I. 2018. Effect of Treatment Delay on the Effectiveness and Safety of Antifibrinolytics in Acute Severe Haemorrhage: A Meta-Analysis of Individual Patient-Level Data from 40 138 Bleeding Patients. *The Lancet* 391(10116): 125–132. doi.org/10.1016/S0140-6736(17)32455-8.

Genet, R., and Inzirillo, H. 2024. TKAN: Temporal Kolmogorov-Arnold Networks. *arXiv preprint. arXiv:2405.07344*.

Ghetmiri, D. E.; Venturi, A. J.; Cohen, M. J.; and Menezes, A. A. 2024. Quick Model-Based Viscoelastic Clot Strength Predictions from Blood Protein Concentrations for Cyber-Medical Coagulation Control. *Nature Communications* 15(1): 314. doi.org/10.1038/s41467-023-44231-w.

Huang, S.; Zhao, Z.; Li, C.; and Bai, L. 2024a. TimeKAN: KAN-Based Frequency Decomposition Learning Architecture for Long-Term Time Series Forecasting. In *International Conference on Learning Representations*.

Huang, W.; Hu, J.; Bi, X.; and Xiao, B. 2024b. Anatomical Prior Guided Spatial Contrastive Learning for Few-Shot Medical Image Segmentation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 5211–5220. Melbourne, VIC, Australia: Association for Computing Machinery.

Li, Y.; Wu, K.; and Liu, J. 2023. Self-Paced ARIMA for Robust Time Series Prediction. *Knowledge-Based Systems* 269: 110489. doi.org/10.1016/j.knosys.2023.110489.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2024a. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *International Conference on Learning Representations*.

Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024b. KAN: Kolmogorov-Arnold Networks. *arXiv preprint. arXiv:2404.19756*.

Liu, Z.; and Hauskrecht, M. 2017. A Personalized Predictive Framework for Multivariate Clinical Time Series via Adaptive Model Selection. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, 1169–1177. Singapore: Association for Computing Machinery.

Liu, Z.; Godahewa, R.; Bandara, K.; and Bergmeir, C. 2023. Handling Concept Drift in Global Time Series Forecasting. In *Forecasting with Artificial Intelligence*, edited by M. Hamoudia; S. M akridakis; and E. Spiliotis, 163–189. Cham, Switzerland: Springer Nature.

Melgar-Garcia, L.; Gutierrez-Aviles, D.; Rubio-Escudero, C.; and Troncoso, A. 2023. A Novel Distributed Forecasting Method Based on Information Fusion and Incremental Learning for Streaming Time Series. *Information Fusion* 95: 163–173. doi.org/10.1016/j.inffus.2023.02.023.

Moore, H. B.; Moore, E. E.; Chapman, M. P.; McVaney, K.; Bryskiewicz, G.; Blechar, R.; Chin, T.; Burlew, C. C.; Pieracci, F.; West, F. B.; Fleming, C. D.; Ghasabyan, A.; Chandler, J.; Silliman, C. C.; Banerjee, A.; and Sauaia, A. 2018. Plasma-First Resuscitation to Treat Haemorrhagic Shock during Emergency Ground Transportation in an Urban Area: A Randomised Trial. *The Lancet* 392(10144): 283–291. doi.org/10.1016/S0140-6736(18)31553-8.

Nesterov Y. 2003. *Introductory Lectures on Convex Optimization: A Basic Course*. New York, NY, USA: Springer.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. *arXiv preprint. arXiv:2211.14730*.

Song, X.; Zhang, X.; Tian, W.; and Zhu, Q. 2025. TFMSNet: A Time Series Forecasting Framework with Time–Frequency Analysis and Multi-Scale Processing. *Computers and Electrical Engineering* 123: 110260. doi.org/10.1016/j.compeleceng.2025.110260.

Suresh, H.; Gong, J. J.; and Guttag, J. V. 2018. Learning Tasks for Multitask Learning: Heterogeneous Patient Populations in the ICU.

In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 802–810. London, United Kingdom: Association for Computing Machinery.

Wang, N.; Di, Z.; He, H.; Jiang, Q.; and Li, X. 2024a. A Simple and Provable Approach for Learning on Noisy Labeled Medical Images. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 4397–4405. Melbourne, VIC, Australia: Association for Computing Machinery.

Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J. Y.; and Zhou, J. 2024b. TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting. In *International Conference on Learning Representations*.

Wang, S.; Luo, C.; and Shao, R. 2023. Unsupervised Concept Drift Detection for Time Series on Riemannian Manifolds. *Journal of the Franklin Institute* 360(17): 13186–13204. doi.org/10.48550/arXiv.2406.17813.

Wen, Q.; Chen, W.; Sun, L.; Zhang, Z.; Wang, L.; Jin, R.; and Tan, T. 2023. Onenet: Enhancing Time Series Forecasting Models under Concept Drift by Online Ensembling. *Advances in Neural Information Processing Systems*, volume 36, 69949–69980. Curran Associates, Inc.

Wilks S. S. 1938. Shortest average confidence intervals from large samples. *The Annals of Mathematical Statistics* 9(3): 166-175. doi.org/10.1214/aoms/1177732308.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; Lian, D.; An, N.; Cao, L.; and Niu, Z. 2023. Frequency-Domain MLPs Are More Effective Learners in Time Series Forecasting. In Oh, A.; Neumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 76656–76679. Curran Associates, Inc.

Yu, E.; Song, Y.; Zhang, G.; and Lu, J. 2022. Learn-to-Adapt: Concept Drift Adaptation for Hybrid Multiple Streams. *Neurocomputing* 496: 121–130. doi.org/10.1016/j.neucom.2022.05.025.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are Transformers Effective for Time Series Forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 11121–11128. Palo Alto, CA: AAAI Press.

Zhang, Z.; Chen, Y.; Zhang, D.; Qian, Y.; and Wang, H. 2024. CTFNet: Long-Sequence Time-Series Forecasting Based on Convolution and Time-Frequency Analysis. *IEEE Transactions on Neural Networks and Learning Systems* 35(11): 16368–16382. doi.org/10.1109/TNNLS.2023.3294064.