Multi-Architecture Temporal Models for Early Symptom-Based Disease Prediction

Timely diagnosis remains a challenge when relying on patient-reported symptoms. The sequence and evolution of these symptoms hold crucial information that, if modeled correctly, can transform early detection. Most existing classifiers treat patient symptoms as static features, ignoring how they evolve over time. To address this, we evaluate four temporal modeling approaches on a clinical dataset of 246,000 records spanning over 300 symptoms and 700 diseases. First, incremental logistic regression (LR) processes cumulative symptom counts across seven timesteps, achieving 81% accuracy (weighted F1 = 0.83), with performance improving steadily from 4.1% at $t = 1$. Second, a Long Short-Term Memory (LSTM) network modeling full symptom trajectories reaches 83% accuracy, progressing from 6.9% to 83% across timesteps.

Building on these temporal baselines, we introduce two complementary architectures that balance interpretability with temporal depth. The Hybrid Residual Logistic Sequence (HRLS) starts with LR as a coefficient-based, interpretable baseline and learns temporal corrections via an LSTM. Specifically, HRLS predictions are computed as:

$$\hat{y} = \arg\max_{c} \ \text{softmax}\left(\sum_{t=1}^{T} L_t + \Delta \right)_c, \tag{1}$$

where L_t are the LR logits at timestep t and Δ is the LSTM-predicted residual capturing temporal patterns missed by the linear model. By directing the LSTM to model only what the linear baseline misses, HRLS preserves interpretability while capturing complex temporal symptom dynamics. While full evaluation is still in progress, initial results indicate that HRLS provides a practical compromise as a bridge between classical and deep learning methods.

Finally, the Temporal Graph Neural Network (TemporalGNN) models each timestep as a bipartite graph linking active symptom nodes to disease nodes. Per-graph features are extracted using GraphSAGE convolutions and aggregated temporally via an LSTM, capturing the evolution of symptom–disease relationships. In a preliminary study on a balanced subset, class-weighted TemporalGNN training boosted accuracy from 39% to 91% within eight epochs, highlighting its potential to learn complex relational patterns even under computational limitations.

To evaluate interpretability, we leverage SHAP values and propose Temporal Explanation Robustness (TER), defined as the Spearman correlation between feature importance and accuracy drops under perturbations, ensuring explanations remain reliable under real-world shifts. Future directions include full-scale HRLS evaluation, extended TemporalGNN training, and integration of multimodal clinical data to facilitate practical clinical decision-making.