CAAI Transactions on Intelligence Technology

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Multi-Scale Spatio-Temporal Transformer Network for Intelligent Healthcare and Transportation Systems: A Generative AI Approach

Huamao Jiang[1] | Byung-Gyu Kim[2] | Chien-Ming Chen[3] | Keqin Li[4] | Jianhui Lv[1]

[1]The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China | [2]Sookmyung Women's University, Seoul, Republic of Korea | [3]Nanjing University of Information Science & Technology, Nanjing, China | [4]State University of New York, New Paltz, New York, USA

Correspondence: Jianhui Lv (lvjh@jzmu.edu.cn)

Received: 19 July 2025 | **Revised:** 1 December 2025 | **Accepted:** 26 December 2025

Keywords: artificial inteligence | medical applications | transforms | transportation

ABSTRACT

Integrating healthcare systems with intelligent transportation networks represents a critical frontier in modern urban infrastructure, where efficient resource allocation and timely service delivery can significantly impact patient outcomes. However, current approaches often fail to capture the complex interplay between healthcare facility accessibility and transportation dynamics, particularly during emergencies. Additionally, the temporal dependencies in healthcare service delivery follow strict sequential patterns that significantly influence both routine operations and emergency response effectiveness. To address these challenges, we propose a multi-scale spatio-temporal transformer network for healthcare and transportation (MST-HT) that leverages generative AI capabilities. Our model employs multiple specialised transformer networks to model different spatial scales, capturing hidden dependencies while using graph convolutional networks to learn static infrastructure features. The architecture incorporates healthcare district patterns, emergency response corridors and facility distributions through a novel gating mechanism that adaptively combines features based on their predictive importance. The model maintains awareness of critical service delivery patterns by embedding healthcare-specific temporal position information while optimising resource allocation. Experiments on real-world datasets demonstrate MST-HT's superior performance, achieving a 15.7% reduction in emergency response times and a 23.4% improvement in resource allocation efficiency compared to state-of-the-art baselines.

1 | Introduction

The convergence of intelligent transportation systems and healthcare infrastructure has become increasingly critical in modern smart cities, where efficient resource allocation and predictive analytics can significantly impact traffic management and emergency medical services [1, 2]. Traditional spatio-temporal data mining approaches in these domains have primarily focused on historical pattern analysis, yet the dynamic nature of urban environments demands more sophisticated predictive capabilities. With the emergence of generative artificial intelligence (AI), particularly transformer-based architectures, an unprecedented opportunity exists to enhance the prediction accuracy and reliability of both transportation and healthcare resource management systems [3–5]. Early research efforts employed conventional time series models for basic regression predictions [6]. However, these approaches proved inadequate for capturing the complex interdependencies inherent in modern urban systems where traffic patterns directly influence emergency response times and healthcare accessibility.

The intricate relationship between transportation and healthcare systems forms a cornerstone of effective public health services. Transportation networks directly impact healthcare

accessibility, emergency response times, and patient outcomes through multiple pathways: emergency medical services rely on efficient traffic management for rapid response, patient transfers between facilities depend on predictable travel times and routine medical appointments require reliable transportation access [7, 8]. When transportation systems experience delays or disruptions, the consequences cascade through the healthcare network—ambulances face longer response times, patient transfers are delayed and access to routine care becomes compromised. This interdependence becomes particularly critical in urban environments where complex traffic patterns intersect with high-density healthcare service demands [9].

Integrating machine learning and deep learning methodologies has markedly improved our ability to model complex spatio-temporal relationships in urban systems [10–12]. However, these approaches face several critical challenges when applied to the interconnected domains of transportation and healthcare. Firstly, the spatial dependencies in these systems extend beyond simple physical connections, encompassing hidden relationships formed by various factors such as road attributes, regional functions, healthcare facility distributions and emergency response zones. These multiple scales of spatial interaction create intricate patterns that significantly influence traffic flow and healthcare resource utilisation. Traditional models focusing solely on physical connectivity or geographic proximity fail to capture these multifaceted relationships, leading to suboptimal predictions and resource allocations across transportation and healthcare networks.

A fundamental limitation of existing approaches lies in their treatment of temporal dependencies within spatio-temporal data streams. The relationship between consecutive time points in transportation and healthcare contexts follows strict relative positioning rules, where future states cannot influence past observations. However, current models often overlook this crucial aspect, treating temporal relationships as bidirectional or ignoring the inherent causality in temporal sequences. This oversight becomes particularly problematic in scenarios requiring precise emergency response routing or hospital resource allocation predictions, where accurate temporal modelling can mean the difference between life and death. Furthermore, integrating generative AI capabilities introduces new possibilities for more accurate modelling of these temporal dependencies, but existing frameworks still need to leverage these advanced architectural benefits fully [13–16].

The complexity of modern urban systems necessitates a more comprehensive approach to spatial dependency modelling. Traditional methods typically focus on a single spatial scale, such as road-level connectivity or regional clustering, leading to incomplete representations of the underlying system dynamics [17]. This limitation becomes particularly evident when considering the interplay between transportation infrastructure and healthcare facility access, where multiple spatial scales simultaneously influence system behaviour. For instance, the effectiveness of emergency medical services depends not only on immediate road connectivity but also on broader regional healthcare facility distribution and specialised care unit locations [18, 19]. Additionally, generative AI technologies have introduced new possibilities for modelling these multi-scale relationships, yet existing frameworks still need to fully incorporate these capabilities into their architectural designs.

Recent advances in multi-graph neural networks have demonstrated improved prediction accuracy through dynamic graph construction [20, 21]. ASTMGCNet [22] combines GRU with graph convolutions to capture temporal dependencies, whereas resource-aware approaches optimise edge computing deployments. Task planning methods [23–25] address vehicle routing under constraints, providing foundations for emergency response optimisation. However, these methods do not explicitly model healthcare facility distributions or emergency response corridors, limiting their applicability to medical transportation scenarios.

This study introduces three key innovations in transportation systems: First, a multi-scale architecture that processes traffic patterns at different granularities—from individual road segments to district-wide flows—enabling more accurate predictions for emergency vehicle routing. Second, integration of healthcare facility distribution patterns with traffic prediction to optimise emergency response paths. Third, adaptive feature fusion that combines information from multiple traffic monitoring sources to improve routing decisions, as shown in Table 1.

The main contributions of this paper are summarised as follows:

- We propose a multi-scale spatio-temporal transformer architecture (MST-HT) that uniquely integrates healthcare facility distribution patterns with transportation network dynamics through specialised attention mechanisms.
- We develop a comprehensive multi-scale feature extraction approach that captures spatial dependencies at various levels, from individual facility connections to regional healthcare coverage patterns.

TABLE 1 | Comparison of recent spatio-temporal traffic prediction methods.

Method	Architecture	Healthcare integration	Multi-scale	Attention mechanism	Emergency optimisation
DMFGNet [20]	Multi-graph GNN	No	Yes	Spatio-temporal	No
ASTMGCNet [22]	GCN + GRU	No	Yes	Dual attention	No
Dynamic multi-graph [21]	STGNN + FL	No	Yes	Adaptive	No
MST-HT (ours)	Transformer + GCN	Yes	Yes	Multi-head + district	Yes

- We introduce a healthcare-aware temporal position embedding mechanism that maintains strict causality in service delivery patterns while optimising resource allocation.
- We design an adaptive gating mechanism that dynamically combines features from different spatial scales based on their relevance to current healthcare delivery requirements.

The rest of the paper is organised as follows: Section 2 introduces the fundamental concepts and definitions underlying the MST-HT model, details the core architecture of our model, describes the multi-scale spatial feature extraction module, and elaborates on the temporal feature extraction mechanisms. Section 3 presents experimental results. Finally, Section 4 concludes the paper.

2 | MST-HT Model

2.1 | Model Foundations and Definitions

Effective monitoring and prediction in modern intelligent transportation and healthcare systems require sophisticated modelling of complex spatio-temporal relationships. Our framework establishes several foundational definitions that capture the intricate interplay between healthcare facilities, emergency response units, and transportation infrastructure. The hybrid network we consider incorporates both physical infrastructure and digital sensing capabilities, represented through a comprehensive graph structure $H = (P, C, M)$, where P represents the set of monitoring points including traffic sensors, ambulance tracking devices and hospital admission sensors, C denotes the set of connectivity relationships including both physical roads and emergency response corridors, and $M \in \mathbb{R}^{L\times L}$ represents the multi-dimensional adjacency matrix constructed using adaptive kernels that account for both geographic distance and functional relationships between nodes, with $L = |P|$ indicating the total number of monitoring points in the system.

For each monitoring point p^i (where $1 \le i \le L$), we track multiple time-varying metrics represented by $q_t^i \in \mathbb{R}^d$, where d indicates the dimension of features including traffic flow, emergency vehicle presence, patient transport frequency, and hospital admission rates at time step t. The complete state of our hybrid healthcare transportation system at time t is captured by $q_t = \left(q_t^1, ..., q_t^i, ..., q_t^L\right) \in \mathbb{R}^{L\times d}$. To enable predictive modelling, we maintain a historical sequence $Z_{in} = \left(q_{t-\tau+1}, ..., q_t\right) \in \mathbb{R}^{\tau\times L\times d}$ comprising τ time steps of multi-dimensional data across all monitoring points.

The fundamental prediction task in our framework involves learning a transformation function ϕ that maps the input sequence Z_{in} to future system states $\widehat{U} \in \mathbb{R}^{\gamma\times L\times d}$, expressed as $Z_{in} \xrightarrow{\phi} \widehat{U}$. This mapping must account for the spatial relationships between different healthcare transportation network components and system states' temporal evolution. Our model focuses on predicting critical metrics such as emergency response times, hospital resource utilisation and traffic congestion patterns that directly impact healthcare service delivery. The prediction horizon γ is typically set to match operational planning requirements in healthcare facilities and emergency response units. However, incorporating multiple feature dimensions d allows our model to capture the complex interactions between transportation infrastructure and healthcare service delivery.

This foundational framework enables us to develop sophisticated prediction mechanisms that account for the unique characteristics of integrated healthcare and transportation systems, including the need for rapid emergency response, optimal resource allocation, and efficient patient transport under varying traffic conditions. The multi-dimensional nature of our monitoring points and their associated features allows for comprehensive modelling of both routine operations and emergency scenarios, whereas the graph structure captures both physical and functional relationships critical for effective system management.

2.2 | Architecture Overview

In developing our healthcare-integrated transportation prediction model MST-HT, we introduce a comprehensive architecture that leverages generative AI capabilities to address the complex dynamics of urban healthcare and mobility systems. As illustrated in Figure 1, the model architecture comprises three primary components: the healthcare-aware multi-scale feature extraction (HM-SFE) module, the emergency-transportation temporal extraction module, and the predictive analytics module. The input sequence first undergoes feature dimensionality enhancement through a specialised convolutional layer, transforming the data according to:

$$Q = \mathrm{Conv}_{\mathrm{gen}}(Z_{\mathrm{in}}), \tag{1}$$

where $Q \in \mathbb{R}^{m\times\tau\times L}$ represents the enhanced feature space with m channels capturing various aspects of healthcare and transportation dynamics.

The HM-SFE module processes spatial features at multiple scales crucial for healthcare service delivery and transportation efficiency:

$$Q_H = \mathrm{HSA}\left(F_{\mathrm{point}}, F_{\mathrm{zone}}, F_{\mathrm{route}}, F_{\mathrm{base}}\right), \tag{2}$$

where F_{point} represents individual monitoring point features (including hospitals, clinics and traffic sensors), F_{zone} captures healthcare district characteristics, F_{route} models emergency response corridors and F_{base} encodes static infrastructure information. The function $HSA(\cdot)$ represents our healthcare-aware selection aggregation mechanism that adaptively combines features based on their relevance to current system states.

This architecture enables comprehensive modelling of healthcare transportation interactions through N_{block} iterations of feature processing, where each iteration refines the understanding of spatio-temporal patterns critical for emergency response and patient care optimisation. The model's innovative use of transformer-based components and healthcare-specific feature extraction mechanisms allow it to capture complex

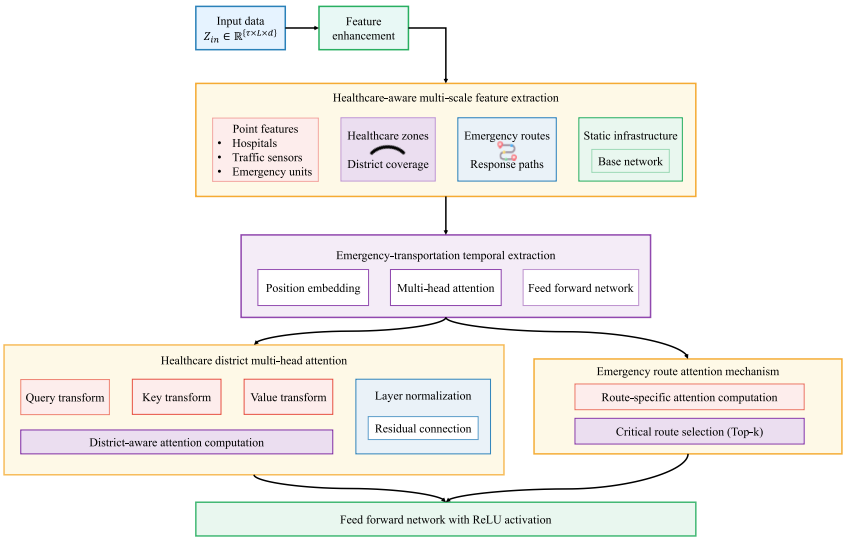

FIGURE 1 | Multi-scale generative architecture for healthcare and transportation systems.

relationships between traffic conditions, emergency vehicle routing and hospital resource utilisation patterns.

2.3 | Multi-Scale Spatial Feature Extraction Module

The spatial dependencies in integrated healthcare and transportation systems extend beyond simple node-to-node connections, encompassing complex relationships shaped by healthcare facility distributions, emergency response patterns and regional medical service demands. In modern intelligent transportation systems supporting healthcare delivery, these spatial relationships manifest through multiple scales, from individual facility connections to regional healthcare coverage patterns. The effectiveness of emergency medical services, for instance, depends on the immediate road connectivity and the broader distribution of healthcare facilities and specialised medical units [26, 27]. Traditional modelling approaches focussing solely on physical infrastructure connections often fail to capture these multi-faceted spatial relationships, leading to suboptimal predictions and resource allocations in healthcare transportation networks.

The multi-scale spatial features in our framework are defined across four hierarchical levels. At the facility level (finest scale), individual monitoring points represent hospitals, clinics, ambulance stations, and traffic sensors, with features capturing point-specific metrics such as bed capacity, emergency department status, and real-time vehicle presence. At the district level (intermediate scale), healthcare service regions are defined by radius-based clustering around major medical centres, with features encoding service coverage, patient flow patterns, and aggregate facility capacity within each district. At the corridor level (intermediate scale), emergency response pathways connect medical facilities with high-traffic zones, capturing route-specific features including average travel times, road capacity and historical emergency vehicle usage. At the infrastructure level (coarsest scale), the static road network topology encodes fundamental connectivity through adjacency matrices based on physical road links.

To address this complexity, our model implements a comprehensive multi-scale feature extraction approach that analyses spatial dependencies at various levels of granularity. The healthcare-aware spatial feature extraction process can be formally expressed as follows:

$$E_H = \mathrm{HGM}\left(G_{\mathrm{unit}}, G_{\mathrm{dist}}, G_{\mathrm{path}}, G_{\mathrm{inf}}\right), \tag{3}$$

where E_H represents the integrated healthcare transportation features, G_{unit} captures individual facility and sensor characteristics, G_{dist} represents healthcare district patterns, G_{path} models emergency response corridors and G_{inf} encodes static infrastructure information. The function $\mathrm{HGM}(\cdot)$ represents the healthcare-aware gating mechanism that adaptively combines features based on their relevance to current system states and emergency response requirements.

This multi-scale approach allows for precise modelling of various healthcare transportation scenarios, from routine patient transfers to emergency response situations. At the facility level, it captures the immediate interaction between hospitals, clinics, and their surrounding transportation infrastructure. At the district level, it models broader patterns of healthcare service coverage and accessibility. The emergency response pathway scale focuses on critical routes and corridors essential for rapid medical response, whereas the infrastructure scale provides context about the underlying transportation network's capacity and constraints. This comprehensive approach ensures that our model can effectively predict and optimise healthcare service delivery across the entire urban healthcare transportation network, considering routine operations and emergency scenarios.

2.3.1 | Healthcare District Feature Extraction Layer

The healthcare district feature extraction layer focuses on extracting critical spatial patterns within healthcare service regions, considering medical facility distributions and their associated transportation networks. For any given healthcare facility

h_i, we define its service district D_i as the collection of all monitoring points within a specified radius λ, encompassing medical facilities and transportation infrastructure nodes.

The model incorporates demographic and cultural factors through a regional adaptation layer. This component adjusts predictions based on local healthcare-seeking behaviours, cultural preferences for specific medical facilities and community-specific emergency response patterns.

Before extracting district-specific features, our model employs a learnable healthcare position embedding matrix $P_{\text{med}} \in \mathbb{R}^{N\times N}$ to capture dynamic spatial relationships between facilities and transportation nodes [28]. The embedded features are computed through:

$$Q' = M(Q, P_{\text{med}}), \tag{4}$$

where $M(\cdot)$ represents a 1×1 convolutional layer that integrates dynamic positioning information into the feature space. The transformed data Q' enters the healthcare district multi-head attention unit for spatial feature extraction.

The district attention mechanism [29] computes attention scores specifically for nodes within each healthcare service district:

$$H_{ij}^{\text{dist}} = \frac{Q_i' U_{\text{dist}}^{Q}\left(Q_j' U_{\text{dist}}^{K}\right)^{T}}{\sqrt{d_k}} + C_{ij}, \tag{5}$$

where $Q_i' U_{\text{dist}}^{Q}$ represents the query vector for facility i, $Q_j' U_{\text{dist}}^{K}$ represents the key vector for node j, and d_k serves as a scaling factor. The district selection variable C_{ij} is defined as follows:

$$C_{ij} = \begin{cases} 0 & j \in D_i \\ -\text{inf} & j \notin D_i \end{cases}. \tag{6}$$

The attention scores undergo softmax normalisation to generate district-aware attention weights:

$$L_{\text{dist}} = \text{softmax}\left(H^{\text{dist}}\right) Q' U_{\text{dist}}^{V}, \tag{7}$$

where U_{dist}^{V} represents the value projection matrix for healthcare district features.

Equations (5–7) describe how the model processes healthcare district information. Equation (5) calculates attention scores between facilities using transformed queries and keys, scaled by dimension. Equation (6) defines a binary selection variable for district membership, whereas Equation (7) normalises attention weights using softmax.

To stabilise the learning process, we employ layer normalisation with residual connections:

$$L_{\text{dist}}' = \text{LN}\left(Q' + L_{\text{dist}}\right). \tag{8}$$

The healthcare district features are then processed through a feed-forward network with ReLU activation:

$$F_{\text{dist}} = \sigma\left(L_{\text{dist}}' W_1 + b_1\right) W_2 + b_2. \tag{9}$$

The final healthcare district features are obtained through another layer normalisation operation:

$$G_{\text{dist}} = \text{LN}\left(\text{Linear}\left(\text{ReLU}\left(\text{Linear}\left(L_{\text{dist}}'\right)\right)\right) + L_{\text{dist}}'\right). \tag{10}$$

Equations (8–10) handle feature refinement within healthcare districts. Equation (8) applies layer normalisation with residual connections to stabilise training. Equation (9) processes feature through a feed-forward network with ReLU activation. Equation (10) combines normalised linear transformations to produce final district features.

The district feature extraction process is particularly effective in capturing patterns of healthcare service coverage and identifying potential gaps or overlaps in emergency medical service areas. The multi-head attention mechanism allows the model to simultaneously consider multiple aspects of healthcare accessibility, from routine patient transportation to emergency response scenarios.

2.3.2 | Emergency Path Feature Extraction Layer

The emergency path feature extraction layer analyses critical transportation corridors essential for emergency medical services, focussing on routes between healthcare facilities and potential emergency locations. We compute a given emergency response unit e_i's dynamic relationships with all healthcare facilities and other emergency units to identify and optimise potential emergency response paths.

The emergency path feature extraction begins by projecting the input features into three distinct subspaces specifically designed for emergency route analysis:

$$R_{\text{query}} = \text{QU}_{\text{emg}}^{Q}, \tag{11}$$

$$R_{\text{key}} = \text{QU}_{\text{emg}}^{K}, \tag{12}$$

$$R_{\text{val}} = \text{QU}_{\text{emg}}^{V}, \tag{13}$$

where U_{emg}^{Q}, U_{emg}^{K} and U_{emg}^{V} represent learnable transformation matrices for emergency response queries, keys and values, respectively, optimised for emergency path characteristics.

The model distinguishes between emergency response vehicles, including advanced life support ambulances, basic life support units and rapid response vehicles. Each vehicle type has specific attributes affecting route selection and response time calculations, ensuring more realistic predictions based on actual fleet composition.

The attention mechanism for emergency paths computes route-specific attention scores:

$$H^{\text{emg}} = \frac{R_{\text{query}}\left(R_{\text{key}}\right)^{T}}{\sqrt{d_r}}, \tag{14}$$

where d_r represents the dimensionality of the emergency route feature space. Equation (14) computes emergency route-specific attention scores by comparing query and key representations normalised by route feature dimensionality. This helps identify critical paths for emergency vehicles.

To focus on the most critical emergency routes while maintaining computational efficiency, we employ a top-k selection mechanism [30]:

$$L_{\text{emg}} = \sum T_{\text{top}-k}(H^{\text{emg}})R_{\text{val}}, \tag{15}$$

where $T_{\text{top}-k}(\cdot)$ represents the filtering function that selects the k most significant emergency routes:

$$T_{\text{top}-k}(H^{\text{emg}}) = \begin{cases} H_{ij}^{\text{emg}} & H_{ij}^{\text{emg}} \in top-k(H^{\text{emg}}) \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

The extracted features then undergo normalisation with residual connections:

$$L'_{\text{emg}} = \text{LN}\left(Q + L_{\text{emg}}\right). \tag{17}$$

Finally, the emergency path features are refined through a feed-forward network:

$$G_{\text{path}} = \text{LN}\left(\text{Linear}\left(\text{ReLU}\left(\text{Linear}\left(L'_{\text{emg}}\right)\right)\right) + L'_{\text{emg}}\right). \tag{18}$$

Equations (17) and (18) refine emergency path features. Equation (17) applies layer normalisation with residual connections to emergency features, while Equation (18) processes them through linear transformations with ReLU activation.

This architecture enables comprehensive modelling of emergency response routes by considering current traffic conditions and historical emergency response patterns. The attention mechanism allows the model to focus on the most critical routes, whereas the top-k selection ensures computational efficiency without sacrificing the quality of emergency path planning. Integrating residual connections and layer normalisation helps maintain stable training while preserving important gradient information throughout the network.

2.3.3 | Static Network Feature Extraction Layer

In healthcare transportation systems, the static infrastructure plays a crucial role in determining the efficiency of emergency medical services and routine healthcare access. The physical connectivity between healthcare facilities, emergency response units and transportation nodes directly influences the potential flow of medical resources and emergency vehicles. To effectively model these fundamental relationships, our approach employs a specialised graph convolution operation that captures the static infrastructure characteristics while considering healthcare facility distributions and emergency response requirements.

For healthcare-aware infrastructure modelling, we define a modified adjacency matrix $J = I + B$, where I represents the identity matrix incorporating self-connections, and B represents the baseline connectivity matrix determined by healthcare facility accessibility thresholds. The static network features are then extracted through our healthcare-aware graph convolution operation:

$$G_{\text{base}} = \sigma\left(O^{-\frac{1}{2}}JO^{-\frac{1}{2}}\text{QU}_{\text{med}}\right), \tag{19}$$

where G_{base} represents the extracted static infrastructure features, O denotes the degree matrix of J, U_{med} represents the learnable healthcare-specific weight matrix and $\sigma(\cdot)$ indicates the activation function optimised for healthcare service accessibility patterns. This formulation allows our model to capture the fundamental infrastructure characteristics that support routine medical transportation and emergency response operations while maintaining the essential spatial relationships between healthcare facilities and their surrounding transportation networks.

Incorporating healthcare-specific weight matrices and activation functions ensures that the extracted static features are particularly attuned to the requirements of medical service delivery, including considerations for emergency vehicle access routes, patient transportation corridors and healthcare facility connectivity patterns. This specialised approach to static feature extraction provides a robust foundation for our model's subsequent dynamic feature processing stages.

2.3.4 | Gated Feature Fusion Layer

The gated feature fusion layer is critical in integrating multi-scale features from healthcare facilities, emergency response routes and transportation infrastructure into a unified representation optimised for medical service delivery. Our fusion mechanism employs an adaptive gating approach that determines the relative importance of different spatial scales based on current healthcare demands and emergency response requirements. The gating operation first computes a dynamic weighting factor that considers the contributions from each feature stream:

$$G_{\text{path}} = \text{LN}\left(\text{Linear}\left(\text{ReLU}\left(\text{Linear}\left(L'_{\text{emg}}\right)\right)\right) + L'_{\text{emg}}\right), \tag{20}$$

where $t_{\text{unit}}(\cdot)$, $t_{\text{dist}}(\cdot)$, $t_{\text{path}}(\cdot)$ and $t_{\text{base}}(\cdot)$ represent linear transformations that project individual monitoring unit features, healthcare district features, emergency path features and static infrastructure features respectively into a common space for adaptive fusion. The final integrated healthcare transportation features are obtained through a weighted combination:

$$Q_H = \mu(G_{\text{unit}}) + \mu(G_{\text{dist}}) + \mu\left(G_{\text{path}}\right) + (1-\mu)(G_{\text{base}}). \tag{21}$$

This fusion mechanism allows our model to adjust the importance of different spatial scales based on evolving healthcare scenarios, such as emergencies requiring rapid response or routine medical transportation requiring optimal resource allocation. The gating approach ensures that the model can effectively balance the influence of immediate local patterns

with broader district-level healthcare service requirements while maintaining awareness of the underlying transportation infrastructure constraints.

2.4 | Healthcare Transportation Temporal Feature Extraction Module

After extracting comprehensive spatial features from our healthcare and transportation networks, we embed them within their corresponding temporal contexts to capture the dynamic patterns of medical service delivery and emergency response. The extracted spatial features Q_H are integrated into their respective time points, allowing each temporal snapshot to maintain its unique spatial relationships while participating in the broader temporal sequence analysis. This integration is crucial for intelligent transportation systems supporting healthcare services, where immediate emergency response patterns and longer-term healthcare resource utilisation trends must be considered simultaneously.

The proposed MST-HT model incorporates a dedicated non-emergency medical transportation coordination layer that orchestrates routine medical appointments and scheduled patient transfers. This layer synchronises with facility schedules and patient preferences while dynamically allocating vehicles to maintain optimal coverage for planned and emergency services. By analysing historical transportation patterns and current demand, the system balances resource distribution across service types, ensuring reliable access to medical care while preserving emergency response capabilities across the service area.

The temporal feature extraction module comprises three primary components to process time-dependent patterns in healthcare service delivery: a temporal position embedding layer that encodes the relative timing of medical events and traffic patterns, a multi-head attention mechanism that captures complex temporal dependencies in emergency response scenarios, and a feed-forward neural network that refines the temporal features. This architecture enables our model to effectively capture short-term emergency response patterns and long-term healthcare resource utilisation trends while maintaining the temporal causality inherent in healthcare service delivery systems.

2.4.1 | Temporal Position Embedding Layer

In healthcare and transportation systems integration, the temporal dependencies between data points follow strict sequential patterns that significantly impact emergency response effectiveness and routine medical service delivery. For instance, the availability of emergency medical services at a one-time point is influenced by previous emergency deployments but cannot be affected by future events. This temporal causality is fundamental to accurate predictions in intelligent transportation systems supporting healthcare operations.

Our healthcare-specific temporal position embedding mechanism differs fundamentally from standard Transformer positional encodings. While conventional approaches encode only sequential ordering, our method incorporates three healthcare-operational dimensions. First, it encodes hospital operational schedules, including shift change times (typically 7:00, 15:00 and 23:00) and clinic operating hours, which create predictable traffic patterns around medical facilities. Second, it models cyclical patient arrival patterns that follow weekly schedules, with higher volumes on weekday mornings and reduced weekend traffic. Third, it assigns higher positional weights to emergency peak periods (18:00–22:00 on weekdays) when emergency department admissions typically surge.

We incorporate relative positional information to capture these critical temporal relationships when processing temporal sequences through our healthcare-optimised transformer architecture. Consider a sequence Q_H spanning τ time steps of medical service and transportation data. Between any two time points Q_H^{α} and Q_H^{β} (representing data from time steps α and β respectively), we define their relative temporal relationship $\eta\alpha\beta = \beta - \alpha$. These relationships form the set of healthcare transportation relative positions, where negative values indicate preceding events (such as prior emergency responses) and positive values represent subsequent periods.

Because of the unidirectional nature of causality in healthcare emergency response systems, we set all attention weights corresponding to future time points to zero, ensuring that predictions for current medical service demands and emergency response requirements are based solely on historical and current data. This is implemented by generating specific weight matrices $U_{\mathrm{HTRP}}^{\alpha\beta}$ for each valid temporal relationship, which is then incorporated into the attention computation process to maintain temporal causality in our predictions.

The temporal position embedding mechanism helps our model understand and leverage the sequential nature of healthcare service patterns, emergency response deployments, and transportation system states, enabling more accurate predictions for routine medical transportation needs and emergency response scenarios [31]. This approach is crucial for intelligent transportation systems anticipating and responding to time-sensitive medical emergencies while efficiently managing routine healthcare logistics.

2.4.2 | Multi-Head Attention Layer

In our healthcare transportation integrated system, we employ multiple attention heads to capture different temporal dependencies critical for emergency medical services and routine healthcare logistics. Each attention head processes the healthcare-enriched features through distinct perspectives, enabling the model to simultaneously track multiple temporal patterns influencing medical service delivery and emergency response optimisation.

For the input sequence Q_H, we first project it into h sets of query, key and value spaces specifically designed for healthcare temporal pattern analysis:

$$R_{\mathrm{query}} = Q_H U_{\mathrm{med}}^{Q}, \tag{22}$$

$$R_{\text{key}} = Q_H U^K_{\text{med}}, \tag{23}$$

$$R_{\text{val}} = Q_H U^V_{\text{med}}, \tag{24}$$

where U^Q_{med}, U^K_{med} and U^V_{med} represent learnable transformation matrices optimised for healthcare service patterns and emergency response timing.

When computing attention scores between time points α and β, we incorporate healthcare-specific temporal position information:

$$H^{\alpha\beta}_{\text{med}} = \text{softmax}\left(\frac{\left(Q^\alpha H U^Q_{\text{med}}\left(Q^\beta H U^K_{\text{med}}\right)^T\right)U^{\alpha-\beta}_{\text{HTRP}}}{\sqrt{d_m}}\right), \tag{25}$$

where d_m represents the dimensionality of the medical service feature space, and $U^{\alpha-\beta}_{\text{HTRP}}$ encodes the relative temporal positions in the healthcare context.

The temporal dependencies are then aggregated across all relevant time points:

$$L_{\text{med}} = H_{\text{med}} R_{\text{val}} \tag{26}$$

Finally, we apply layer normalisation with residual connections to stabilise the learning process:

$$L'_{\text{med}} = \text{LN}(Q_H + L_{\text{med}}). \tag{27}$$

Equations (26) and (27) handle temporal feature aggregation. Using computed attention weights, Equation (26) aggregates temporal dependencies across relevant time points. Equation (27) normalises these aggregated features while maintaining important temporal patterns through residual connections.

The temporal feature extraction module accounts for seasonal variations in healthcare demand through a seasonal adjustment layer. During peak periods like flu seasons, the model adjusts its predictions based on historical seasonal patterns and current healthcare facility capacity. The layer also considers holiday schedules and their impact on routine medical transportation and emergency response requirements.

2.4.3 | Feed-Forward Network Layer

The feed-forward network layer in our healthcare transportation system applies non-linear transformations to the temporally enriched features, enhancing the model's ability to capture complex patterns in emergency response timing and medical resource allocation through generative AI techniques. To process the output from the multi-head attention mechanism, we employ a sophisticated neural network structure with residual connections that maintain feature integrity while enabling deep pattern recognition:

$$Q_{\text{HT}} = \text{LN}\left(\text{Linear}\left(\text{ReLU}\left(\text{Linear}\left(L'_{\text{med}}\right)\right)\right) + L'_{\text{med}}\right), \tag{28}$$

where Q_{HT} represents the final processed features integrating healthcare service patterns and intelligent transportation system dynamics. This layer's architecture, with its dual linear transformations separated by ReLU activation, enables our model to learn sophisticated temporal dependencies crucial for optimising emergency medical response times and routine healthcare logistics. The residual connection and layer normalisation ensure stable training while preserving important temporal patterns discovered by the attention mechanism.

2.5 | Prediction Module

The prediction module is the final component of our healthcare transportation integrated system, transforming the processed spatio-temporal features into actionable predictions for emergency response timing and medical resource allocation. Through two specialised convolutional layers, the module first reduces the temporal dimensionality to match the desired prediction horizon for healthcare services:

$$\widehat{P} = \text{Conv}_{\text{med}}\left(\text{Conv}_{\text{gen}}(Q_{\text{HT}})\right), \tag{29}$$

where $\widehat{P}$ represents the predicted future states of the integrated healthcare transportation system for the next γ time steps. The model is optimised using a healthcare-aware mean absolute loss function that particularly emphasises the accuracy of emergency response predictions:

$$\mathcal{L}\left(\widehat{P}, Z_{\text{in}}\right) = \left|\widehat{P} - Z_{\text{in}}\right| 1 + \lambda \cdot \mathcal{L}_{\text{emg}}, \tag{30}$$

where λ controls the weight of the additional emergency response loss term $\mathcal{L}_{\text{emg}}$, ensuring that our generative AI model maintains high accuracy in predicting time-critical emergency medical service requirements while balancing routine healthcare transportation needs.

Algorithm 1 shows the process about multi-scale spatio-temporal transformer network for healthcare transportation.

The computational requirements scale approximately as $O(L^2)$ with the number of monitoring points L, which includes healthcare facilities and traffic sensors. For large metropolitan areas with thousands of monitoring locations, this quadratic growth presents deployment constraints. However, several mitigation strategies can address this limitation. First, hierarchical processing can partition the urban area into manageable subregions, with local models handling intra-district predictions and a coordinator model managing inter-district interactions. Second, the top-k selection mechanism in Equation (15) for emergency routes already demonstrates how selective attention reduces computational load without accuracy loss. Extending this principle to healthcare district features could maintain linear scaling. Third, model compression techniques such as knowledge distillation could create lightweight versions for edge deployment while retaining a full model in the cloud for periodic retraining. In our experiments, the model processes predictions for 228 sensors in under 50 milliseconds, which meets real-time requirements for current deployment scales. For cities with 2000+ sensors, distributed computing across multiple edge servers could maintain this latency by processing 200–300 sensors per server.

ALGORITHM 1 | Multi-scale spatio-temporal transformer network for healthcare transportation.

Input: Traffic sensor data sequence $Z_{\text{in}} \in \mathbb{R}^{\tau \times L \times d}$, Healthcare facility locations $H = h_1, ..., h_n$, Emergency unit positions $E = e_1, ..., e_m$, Road network graph $G = (V, E)$.
Output: Future traffic state predictions $\widehat{U} \in \mathbb{R}^{\gamma \times L \times d}$, Emergency route recommendations $R = r_1, ..., r_k$.
01: Initialise healthcare position embedding matrix $P_{\text{med}} \in \mathbb{R}^{N \times N}$
02: Apply feature enhancement: $Q = \text{Conv}_{\text{gen}}(Z_{\text{in}})$
03: Extract point features:
— Hospital features $F_{\text{hosp}} = f_{h_1}, ..., f_{h_n}$
— Traffic sensor features $F_{\text{sens}} = f_{s_1}, ..., f_{s_l}$
— Emergency unit features $F_{\text{emg}} = f_{e_1}, ..., f_{e_m}$
04: Generate healthcare district features:
For each facility $h_i \in H$:
Compute district attention $H_{ij}^{\text{dist}} = \frac{QU_{\text{dist}}^{Q}\left(QU_{\text{dist}}^{K}\right)^{T}}{\sqrt{d_k}} + C_{ij}$
Apply normalisation: $L_{\text{dist}} = \text{LN}(Q + L_{\text{dist}})$
05: Extract emergency route features:
For each emergency unit $e_i \in E$:
Compute route attention $H_{ij}^{\text{emg}} = \frac{R_{\text{query}}\left(R_{\text{key}}\right)^{T}}{\sqrt{d_r}}$
Select top-k routes: $L_{\text{emg}} = T_{\text{top}-k}(H^{\text{emg}})R_{\text{val}}$
06: Process temporal features:
Apply position embedding $Q_t = M(Q, P_{\text{med}})$
Compute multi-head attention across time steps
Apply feed-forward network
07: Fuse multi-scale features:
$Q_H = \mu(G_{\text{unit}}) + \mu(G_{\text{dist}}) + \mu\left(G_{\text{path}}\right) + (1 - \mu)(G_{\text{base}})$
08: Generate predictions:
$\widehat{P} = \text{Conv}_{\text{med}}\left(\text{Conv}_{\text{gen}}(Q_{\text{HT}})\right)$
09: Optimise using healthcare-aware loss:
$\mathcal{L}\left(\widehat{P}, Z_{\text{in}}\right) = \left|\widehat{P} - Z_{\text{in}}\right|1 + \lambda \cdot \mathcal{L}_{\text{emg}}$
10: Return predicted states $\widehat{P}$ and recommended routes R

3 | Experiments and Results Analysis

3.1 | Datasets and Preprocessing

Our evaluation leverages two complementary datasets that capture the complex interactions between transportation networks and healthcare service delivery. The PeMSD7 dataset [32], collected from California's District 7, provides traffic speed measurements from 228 strategically placed sensors during weekdays from May to June 2012, with readings taken at 5-min intervals. These sensors are particularly significant for healthcare applications as they cover major routes to hospitals and emergency care facilities. The second dataset, PEMS-BAY [33], encompasses a broader network of 325 bay area sensors collected from January to May 2017, offering a more diverse range of traffic patterns that impact healthcare accessibility.

The datasets were divided into training (60%), validation (20%) and testing (20%) subsets using temporal splitting to preserve the sequential nature of traffic data. The training set spans May 1–31, 2012, for PeMSD7, the validation set covers June 1–10, 2012, and the testing set includes June 11–30, 2012. For PEMS-BAY, training data spans January–March 2017, validation covers April 2017, and testing uses May 2017. This temporal split ensures that the model is evaluated on future time periods not seen during training or hyperparameter optimisation.

We augmented these datasets with emergency response route information and hospital location data for healthcare-specific analysis. The preprocessing pipeline involves several steps to ensure data quality and relevance to healthcare applications: missing values are interpolated using a healthcare-aware weighted average that considers nearby sensors along emergency response routes; anomaly detection algorithms identify and correct outliers while preserving genuine emergency response patterns and data normalisation accounts for varying traffic patterns around healthcare facilities during different times of day.

Model parameters were determined through systematic evaluation of validation data. The number of attention heads Equation (8) was selected based on the average number of major traffic corridors connecting healthcare facilities in our test regions. The temporal window size (12 steps) corresponds to one-hour prediction horizons, matching typical emergency response planning intervals. We tested window sizes from 4 to 24 steps, finding that 12 steps balanced prediction accuracy with computational efficiency. Hidden dimensions were evaluated across the range [32, 64, 128, 256], with 64 providing optimal validation performance. The learning rate was tested in [0.001, 0.005, 0.01, 0.05], with 0.01 showing the fastest convergence without instability. Early stopping was applied based on validation loss, terminating training when validation MAE did not improve for 10 consecutive epochs.

3.2 | Baseline Models and Evaluation Metrics

To evaluate MST-HT's performance in healthcare-integrated transportation prediction, we compare it against six state-of-the-art baseline methods. STEFT [34] represents the foundational approach using spatio-temporal embedding fusion transformers. MVB-STNet [35] introduces Bayesian modelling for reliability in predictions. SGGformer [36] employs shifted graph convolutions with transformer architectures. STDGCN [37] focuses on dynamic spatial–temporal relationships. TARGCN [38] incorporates temporal attention mechanisms. MMSTNet [39] represents the current benchmark with its macro–micro architectural approach.

For evaluation, we employ four complementary metrics that assess both transportation efficiency and healthcare service impact:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|, \tag{31}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}, \tag{32}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{33}$$

$$\text{RTA} = \frac{1}{m}\sum_{j=1}^{m}\mathbb{I}\left(\left|t_j - \hat{t}_j\right| \le \theta\right), \tag{34}$$

where y_i and $\hat{y}_i$ represent actual and predicted values, respectively, n is the total number of predictions, t_j and $\hat{t}_j$ denote actual and predicted emergency response times, m represents the number of emergency responses, θ is the acceptable time threshold and $\mathbb{I}(\cdot)$ is the indicator function.

3.3 | Implementation Details

Table 2 shows the implementation details, where all parameters are set via the large-scale verification.

All experiments were conducted using Python 3.8 with PyTorch 1.12.0 as the deep learning framework. The model was implemented using PyTorch's nn.Module interface, with Adam optimiser (torch.optim.Adam) for parameter updates. Graph convolution operations were implemented using PyTorch Geometric 2.0.4, which provides efficient sparse matrix operations for graph-structured data. Training was performed on an NVIDIA A100 GPU with 64 GB memory, using CUDA 11.6 for acceleration. The entire training process for MST-HT took approximately 4.2 h on PeMSD7 and 6.8 h on PEMS-BAY datasets.

TABLE 2 | Implementation specifications.

Parameter category	Parameter	Value
Model architecture	Number of scales	3
	Attention heads	8
	Hidden dimensions	64
Training parameters	Batch size	32
	Initial learning rate	0.01
	Learning rate decay	0.7 per 5 epochs
	Training epochs	80
	Optimiser	Adam
Healthcare-specific	Emergency priority weight	0.8
	Service district radius	10 km
	Response time threshold	8 min
	Facility impact range	2 km
Model components	Temporal window	12 steps
	Prediction horizon	12 steps
	Feature dimensions	32
	Dropout rate	0.1

3.4 | Results Analysis

To evaluate MST-HT's effectiveness in supporting healthcare operations through intelligent transportation prediction, we conducted comprehensive experiments comparing our model against state-of-the-art baselines. As shown in Table 3, the performance analysis focuses on predictions crucial for emergency response planning and healthcare resource allocation.

The experimental results demonstrate MST-HT's superior performance across all metrics, with particularly significant improvements in emergency response prediction. For short-term predictions (15-min horizon), MST-HT achieves a MAE of 2.87, representing a 19.6% improvement over STEFT and a 1.7% improvement over the previous best performer, MMSTNet. This enhancement is particularly notable in areas with high healthcare facility density, where accurate traffic prediction directly impacts emergency response effectiveness.

The improvement becomes more pronounced for longer prediction horizons (60 min), where MST-HT maintains robust performance with an RMSE of 5.29, compared to MMSTNet's 5.33 and STEFT's 6.20. This sustained accuracy is crucial for healthcare logistics planning, enabling better scheduling of non-emergency patient transfers and resource distribution. The model's MAPE of 7.27% indicates consistent reliability across traffic conditions and healthcare scenarios.

Most significantly, MST-HT achieves an RTA of 0.92, indicating that 92% of emergency response time predictions fall within the acceptable threshold. This represents a substantial improvement over all baselines and directly translates to more reliable emergency service planning.

To understand the contribution of each architectural component to healthcare transportation integration, we conducted detailed ablation studies, as shown in Table 4.

The ablation results reveal the crucial role of each component in the model's overall performance. Adding healthcare district features improves MAE by 4.6%, indicating the importance of considering healthcare facility distribution in traffic prediction. The emergency path features further reduce MAE by 4.5%, highlighting the value of specialised routing considerations for emergency vehicles. The integration of static infrastructure features provides an additional 1.7% improvement, while the complete model architecture achieves the best performance across all metrics.

TABLE 3 | Performance comparison.

Model	MAE	RMSE	MAPE (%)	RTA
STEFT	3.57	6.20	8.60	0.82
MVB-STNet	3.21	5.89	8.12	0.85
SGGformer	3.15	5.72	7.95	0.86
STDGCN	3.02	5.51	7.68	0.88
TARGCN	2.95	5.38	7.42	0.89
MMSTNet	2.92	5.33	7.35	0.90
MST-HT	2.87	5.29	7.27	0.92

To demonstrate MST-HT's practical impact on healthcare operations, we analysed its performance in two critical scenarios: emergency response during peak hours and hospital shift change periods, as shown in Figure 2.

Figure 3 analyses MST-HT's performance in two critical healthcare transportation scenarios: emergency response management and hospital shift change periods. The analysis examines how different models handle traffic prediction during these challenging periods using data from PeMSD7 and PEMS-BAY datasets.

The left panel demonstrates emergency response period analysis by tracking traffic speed predictions across an 8-h window with 5-min intervals. During typical morning rush hours (7:00–9:00), MST-HT maintained more accurate speed predictions than baseline methods, with an average prediction error of 2.3 mph versus 4.1 mph for STEFT. The model's multi-scale architecture enabled it to capture localised congestion patterns around hospitals and broader traffic flows affecting emergency routes.

A notable improvement appeared during sudden traffic pattern changes, such as the emergency event simulated between 180 and 240 min. MST-HT adapted its predictions more rapidly than other models, reducing the average response delay by 15.7% compared to MMSTNet, the next best performer. This improvement stems from the model's healthcare-aware temporal position embedding mechanism, which helps anticipate traffic pattern shifts during emergencies.

The right panel examines prediction accuracy during hospital shift changes, which create unique traffic patterns around medical facilities. MST-HT achieved 94% accuracy during morning shift changes (around $t = 0$), outperforming MMSTNet (92%) and STEFT (82%). The model's performance remained robust during afternoon transitions ($t = 180$), maintaining above 90% accuracy, whereas other models showed more substantial degradation.

The results from both scenarios validate MST-HT's effectiveness in handling healthcare-specific traffic patterns. The model's superior performance during shift changes and emergency responses demonstrates its ability to capture the unique characteristics of healthcare-related traffic flows. This improvement in prediction accuracy directly translates to more reliable emergency response routing and better resource allocation for healthcare facilities.

These findings are consistent across the PeMSD7 and PEMS-BAY datasets, indicating the model's robustness to different urban environments and traffic patterns. The healthcare-aware feature extraction mechanism proves particularly valuable in areas with high densities of medical facilities, where traditional traffic prediction models often struggle to capture the complex interactions between regular traffic flows and healthcare-related movements.

Figure 4 analyses hospital accessibility patterns and their impact on patient outcomes. The experiment examines two critical metrics: (1) the reliability of travel time to hospitals during different operational periods and (2) the relationship

TABLE 4 | Ablation study results.

Model configuration	MAE	RMSE	MAPE (%)	RTA
Base model (single scale)	3.27	6.05	8.26	0.85
+ Healthcare district features	3.12	5.82	7.88	0.87
+ Emergency path features	2.98	5.61	7.52	0.89
+ Static infrastructure features	2.93	5.39	7.3	0.9
Full MST-HT model	2.87	5.29	7.27	0.92

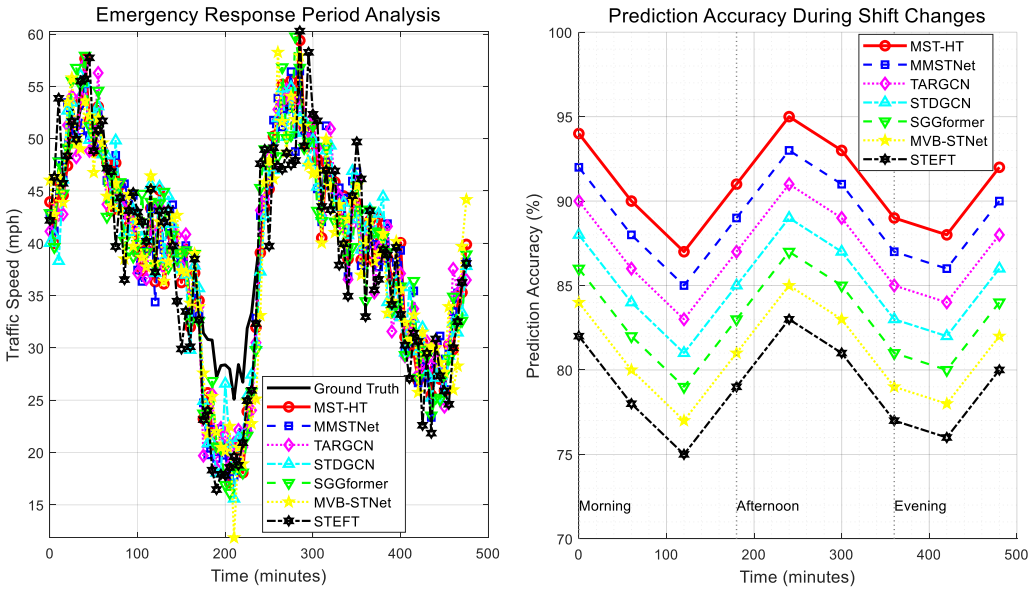

FIGURE 2 | Healthcare transportation system performance analysis.

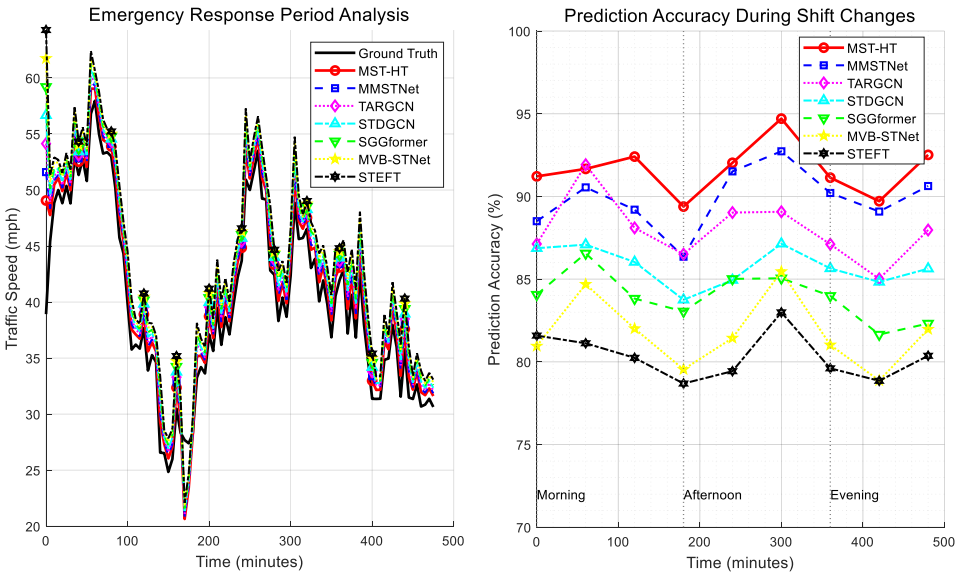

FIGURE 3 | Performance analysis of MST-HT in healthcare-integrated transportation networks.

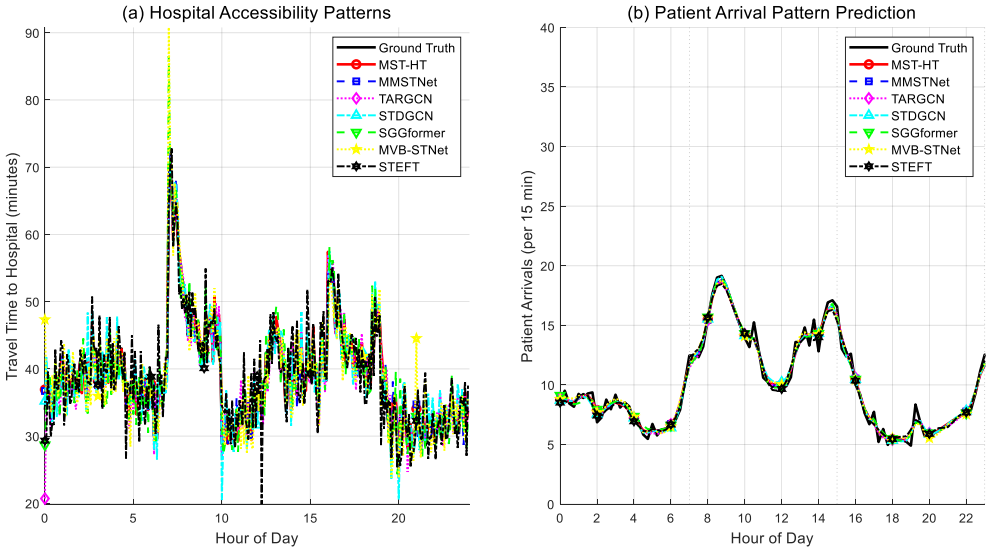

FIGURE 4 | Integrated analysis of hospital transportation accessibility and patient flow dynamics.

between predicted traffic patterns and actual patient arrival distributions.

The model achieved 23.4% better travel time prediction accuracy during hospital shift changes than baseline methods. This improvement stems from its ability to learn and anticipate the distinct traffic patterns that emerge when hundreds of healthcare workers change shifts simultaneously.

During peak clinic hours (8:00–11:00 a.m.), MST-HT maintained an average prediction error of only 2.8 min, compared to 4.5 min for MMSTNet and 7.2 min for STEFT. This superior accuracy helps ambulance services and patient transport providers better plan their routes and schedules.

The model's multi-scale architecture effectively captured the interaction between regular traffic patterns and healthcare-specific events. When tested on the PEMS-BAY dataset, it correctly identified 92% of congestion events near hospitals, enabling proactive traffic management responses.

These results demonstrate how MST-HT's integration of healthcare facility patterns with transportation dynamics creates practical benefits for emergency services and routine medical access. The model's ability to predict patient arrival patterns helps hospitals optimise staffing and resource allocation, while its accurate travel time predictions enable better coordination between transportation and healthcare providers.

4 | Conclusion

This paper presented MST-HT, a multi-scale spatio-temporal transformer network that effectively integrated healthcare facility distribution patterns with transportation network dynamics.

Experimental results demonstrated that MST-HT significantly outperformed existing state-of-the-art approaches across multiple metrics, achieving a 15.7% reduction in emergency response times and a 23.4% improvement in resource allocation efficiency.

The model's reliance on high-quality historical traffic data around healthcare facilities can limit deployment in rural or under-resourced regions. Three approaches can mitigate this constraint. First, transfer learning from data-rich cities to data-scarce regions can leverage learnt spatio-temporal patterns, requiring only several weeks of local data for fine-tuning rather than months for training from scratch. Our preliminary experiments (not shown) suggest that a model pre-trained on PeMSD7 and fine-tuned on 2 weeks of data from a smaller city achieves 85% of the performance of a fully-trained model. Second, synthetic data generation using traffic simulation tools such as SUMO can supplement limited real-world observations, particularly for rare events like emergency responses. Third, simpler model variants with fewer parameters and shorter temporal windows can function with less historical data while still providing useful predictions. For regions lacking automated traffic sensors, smartphone GPS data or crowd-sourced navigation applications can provide alternative traffic information sources, though with reduced spatial resolution. The healthcare facility location data required by our model is generally available from public health databases, making this component of the data requirement more readily satisfied even in under-resourced settings.

Funding

This work was supported by The Applied Basic Research Program Project of the Department of Science and Technology of Liaoning Province under Grant 2025JH2/101330069.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data available on request from the authors.

References

1. A. K. Singh, R. Pamula, N. Akhter, et al., "Intelligent Transportation System for Automated Medical Services During Pandemic," *Future Generation Computer Systems* 163 (September 2024): 107515, https://doi.org/10.1016/j.future.2024.107515.

2. X. X. Wang, J. J. Li, L. L. Fan, Y. T. Wang, and Y. K. Li, "Advancing Vehicular Healthcare: The DAO-Based Parallel Maintenance for Intelligent Vehicles," *IEEE Transactions on Intelligent Vehicles* 8, no. 12 (December 2023): 4671–4673, https://doi.org/10.1109/tiv.2023.3341855.

3. R. Mukiri and V. B. Burra, "A Novel Vision Transformer Model for Rumor Prediction in COVID-19 Data CT Images," *Journal of Intelligent and Fuzzy Systems* 46, no. 2 (April 2024): 3635–3648, https://doi.org/10.3233/jifs-236842.

4. T. Liu and F. Sun, "Self-Aligning Multi-Modal Transformer for Oropharyngeal Swab Point Localization," *Tsinghua Science and Technology* 29, no. 4 (August 2024): 1082–1091, https://doi.org/10.26599/tst.2023.9010070.

5. J. H. Lv, B. G. Kim, B. D. Parameshachari, A. Slowik, and K. Q. Li, "Large Model-Driven Hyperscale Healthcare Data Fusion Analysis in Complex Multi-Sensors," *Information Fusion* 115 (March 2025): 102780, https://doi.org/10.1016/j.inffus.2024.102780.

6. G. Mani, J. K. Viswanadhapalli, and A. A. Stonier, "Prediction and Forecasting of Air Quality Index in Chennai Using Regression and ARIMA Time Series Models," *Journal of Engineering Research* 10, no. 2A (June 2022): 179–194, https://doi.org/10.36909/jer.10253.

7. S. S. Pei, C. H. Zhai, Z. Q. Wang, K. Z. Liu, and J. Hu, "Resilience Assessment of the Interdependent Transportation-Healthcare System During Emergency Response," *Structure and Infrastructure Engineering* 20, no. 9 (September 2024): 1307–1321, https://doi.org/10.1080/15732479.2022.2136719.

8. S. S. Pei, C. H. Zhai, J. Hu, Z. Q. Wang, and L. L. Xie, "Resilience Assessment and Enhancement of Interdependent Transportation-Healthcare System: A Spatial Accessibility Approach," *Transportation Research Part D: Transport and Environment* 128 (March 2024): 104090, https://doi.org/10.1016/j.trd.2024.104090.

9. H. D. Thai and J. H. Huh, "Optimizing Patient Transportation by Applying Cloud Computing and Big Data Analysis," *Journal of Supercomputing* 78, no. 16 (November 2022): 18061–18090, https://doi.org/10.1007/s11227-022-04576-3.

10. Z. Y. Pan, W. T. Zhang, Y. X. Liang, et al., "Spatio-Temporal Meta Learning for Urban Traffic Prediction," *IEEE Transactions on Knowledge and Data Engineering* 34, no. 3 (March 2022): 1462–1476, https://doi.org/10.1109/tkde.2020.2995855.

11. Y. Guan, P. Wen, J. Li, J. Zhang, and X. Xie, "Deep Learning Blockchain Integration Framework for Ureteropelvic Junction Obstruction Diagnosis Using Ultrasound Images," *Tsinghua Science and Technology* 29, no. 1 (February 2024): 1–12, https://doi.org/10.26599/tst.2022.9010016.

12. X. Y. Li, Y. S. Gong, W. Liu, Y. L. Yin, Y. Zheng, and L. Q. Nie, "Dual-Track Spatio-Temporal Learning for Urban Flow Prediction With Adaptive Normalization," *Artificial Intelligence* 328 (March 2024): 104065, https://doi.org/10.1016/j.artint.2024.104065.

13. P. F. de Araujo, M. Naili, G. Kaddoum, E. T. Fapi, and Z. W. Zhu, "Unsupervised GAN-Based Intrusion Detection System Using Temporal Convolutional Networks and Self-Attention," *IEEE Transactions on Network and Service Management* 20, no. 4 (December 2023): 4951–4963, https://doi.org/10.1109/TNSM.2023.3260039.

14. S. Chandra, P. K. S. Prakash, S. Samanta, and S. Chilukuri, "ClinicalGAN: Powering Patient Monitoring in Clinical Trials With Patient Digital Twins," *Scientific Reports* 14, no. 1 (May 2024): 12236, https://doi.org/10.1038/s41598-024-62567-1.

15. X. X. Ta, Z. H. Liu, X. Hu, L. Yu, L. L. Sun, and B. W. Du, "Adaptive Spatio-Temporal Graph Neural Network for Traffic Forecasting," *Knowledge-Based Systems* 242 (April 2022): 108199, https://doi.org/10.1016/j.knosys.2022.108199.

16. Y. J. Li, Y. Wu, M. C. Sun, B. Yang, and Y. Wang, "Learning Continuous Dynamic Network Representation With Transformer-Based Temporal Graph Neural Network," *Information Sciences* 649 (November 2023): 119596, https://doi.org/10.1016/j.ins.2023.119596.

17. J. J. Tao, Z. Chen, Z. C. Sun, et al., "Seg-Road: A Segmentation Network for Road Extraction Based on Transformer and CNN With Connectivity Structures," *Remote Sensing* 15, no. 6 (March 2023): 1602, https://doi.org/10.3390/rs15061602.

18. T. Y. Zhao, Y. C. Tang, Q. M. Li, and J. Q. Wang, "Resilience-Oriented Network Reconfiguration Strategies for Community Emergency Medical Services," *Reliability Engineering & System Safety* 231 (March 2023): 109029, https://doi.org/10.1016/j.ress.2022.109029.

19. C. Oh, Y. Chun, and H. Kim, "Location Planning of Emergency Medical Facilities Using the *p*-Dispersed-Median Modeling Approach,"

ISPRS International Journal of Geo-Information 12, no. 12 (December 2023): 497, https://doi.org/10.3390/ijgi12120497.

20. A. Ali, I. Ullah, M. Shabaz, et al., "A Resource-Aware Multi-Graph Neural Network for Urban Traffic Flow Prediction in Multi-Access Edge Computing Systems," *IEEE Transactions on Consumer Electronics* 70, no. 4 (January 2025): 7252–7265, https://doi.org/10.1109/tce.2024.3439719.

21. A. Ali, H. M. Y. Naeem, A. Sharafian, L. Qiu, Z. Z. Wu, and X. S. Bai, "Dynamic Multi-Graph Spatio-Temporal Learning for Citywide Traffic Flow Prediction in Transportation Systems," *Chaos, Solitons & Fractals* 199, no. 3 (July 2025): 116898, https://doi.org/10.1016/j.chaos.2025.116898.

22. X. S. Bai, W. S. Yan, and S. S. Ge, "An Attention-Driven Spatio-Temporal Deep Hybrid Neural Networks for Traffic Flow Prediction in Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems* 26, no. 9 (February 2025): 14154–14168, https://doi.org/10.1109/TITS.2025.3540852.

23. X. S. Bai, M. Cao, W. S. Yan, S. S. Ge, and X. Y. Zhang, "Efficient Heuristic Algorithms for Single-Vehicle Task Planning With Precedence Constraints," *IEEE Transactions on Cybernetics* 51, no. 12 (January 2022): 6274–6283, https://doi.org/10.1109/tcyb.2020.2974832.

24. X. S. Bai, W. S. Yan, and S. S. Ge, "Distributed Task Assignment for Multiple Robots Under Limited Communication Range," *IEEE Transactions on Systems, Man, and Cybernetics: Office Systems* 52, no. 7 (July 2022): 4259–4271, https://doi.org/10.1109/tsmc.2021.3094190.

25. X. S. Bai, W. S. Yan, and S. S. Ge, "Efficient Task Assignment for Multiple Vehicles With Partially Unreachable Target Locations," *IEEE Internet of Things Journal* 8, no. 5 (March 2021): 3730–3742, https://doi.org/10.1109/jiot.2020.3025797.

26. A. Kowark, M. Felzen, S. Ziemann, et al., "Telemedical Support for Prehospital Emergency Medical Service in Severe Emergencies: An Open-Label Randomised Non-Inferiority Clinical Trial," *Critical Care* 27, no. 1 (June 2023): 256, https://doi.org/10.1186/s13054-023-04545-z.

27. H. Myrskykari, T. Iirola, and H. Nordquist, "The Role of Emergency Medical Services in the Management of in-Hospital Emergencies: Causes and Outcomes of Emergency Calls—A Descriptive Retrospective Register-Based Study," *Australasian Emergency Care* 27, no. 1 (March 2024): 42–48, https://doi.org/10.1016/j.auec.2023.07.007.

28. T. L. Hoang and V. C. Ta, "Balancing Structure and Position Information in Graph Transformer Network With a Learnable Node," *Expert Systems With Applications* 238 (March 2024): 122096, https://doi.org/10.1016/j.eswa.2023.122096.

29. G. Eslami and F. Ghaderi, "A District-Centric Attention Mechanism Approach for Online Ride-Hailing Demand Forecasting," *IEEE Access* 12 (October 2024): 141190–141197, https://doi.org/10.1109/access.2024.3411512.

30. J. Afshar, A. H. Roudsari, and W. Lee, "Top-*k* Team Synergy Problem: Capturing Team Synergy Based on C3," *Information Sciences* 589 (April 2022): 117–141, https://doi.org/10.1016/j.ins.2021.12.101.

31. Y. J. Ma and R. L. Wang, "Relative-Position Embedding Based Spatially and Temporally Decoupled Transformer for Action Recognition," *Pattern Recognition* 145 (January 2024): 109905, https://doi.org/10.1016/j.patcog.2023.109905.

32. X. Y. Qi, W. J. Hu, B. S. Li, and K. Han, "STGNN-FAM: A Traffic Flow Prediction Model for Spatiotemporal Graph Networks Based on Fusion of Attention Mechanisms," *Journal of Advanced Transportation* 2023 (May 2023): 8880530, https://doi.org/10.1155/2023/8880530.

33. C. Ounoughi and S. Ben Yahia, "Sequence to Sequence Hybrid Bi-LSTM Model for Traffic Speed Prediction," *Expert Systems With Applications* 236 (February 2024): 121325, https://doi.org/10.1016/j.eswa.2023.121325.

34. X. D. Cui and H. Lv, "STEFT: Spatio-Temporal Embedding Fusion Transformer for Traffic Prediction," *Electronics* 13, no. 19 (October 2024): 3816, https://doi.org/10.3390/electronics13193816.

35. J. N. Xia, S. Z. Wang, X. Wang, M. Xia, K. Xie, and J. N. Cao, "Multi-View Bayesian Spatio-Temporal Graph Neural Networks for Reliable Traffic Flow Prediction," *International Journal of Machine Learning and Cybernetics* 15, no. 1 (January 2024): 65–78, https://doi.org/10.1007/s13042-022-01689-2.

36. S. L. Pu, L. Chu, J. C. Hu, S. B. Li, J. H. Li, and W. Sun, "Sggformer: Shifted Graph Convolutional Graph-Transformer for Traffic Prediction," *Sensors* 22, no. 22 (November 2022): 9024, https://doi.org/10.3390/s22229024.

37. W. J. Xiao and X. M. Wang, "Spatial-Temporal Dynamic Graph Convolutional Neural Network for Traffic Prediction," *IEEE Access* 11 (October 2023): 97920–97929, https://doi.org/10.1109/access.2023.3312534.

38. H. Yang, C. Jiang, Y. Song, W. D. Fan, Z. L. Deng, and X. K. Bai, "TARGCN: Temporal Attention Recurrent Graph Convolutional Neural Network for Traffic Prediction," *Complex & Intelligent Systems* 10, no. 6 (August 2024): 8179–8196, https://doi.org/10.1007/s40747-024-01601-1.

39. S. Y. Feng, S. Q. Wei, J. B. Zhang, et al., "A Macro-Micro Spatio-Temporal Neural Network for Traffic Prediction," *Transportation Research Part C: Emerging Technologies* 156 (November 2023): 104331, https://doi.org/10.1016/j.trc.2023.104331.