

Application of Aibased Traffic Prediction Model For Urban Congestion Management

Abhishek

Pursuing M.tech from UIET, MDU Rohtak

ABSTRACT

Urban traffic congestion remains one of the most persistent and economically devastating challenges confronting modern metropolitan areas worldwide. With rapid urbanisation, the proliferation of private vehicles, and the inadequacy of existing road infrastructure, cities across India and the developing world experience unprecedented levels of traffic delay, fuel wastage, and environmental degradation. Traditional traffic management approaches, which rely primarily on fixed signal timings and reactive response strategies, are fundamentally inadequate for the dynamic and nonlinear nature of contemporary urban traffic flows.

This research paper investigates the development and application of Artificial Intelligence (AI)based traffic prediction models for proactive urban congestion management. Specifically, the study proposes a hybrid deep learning framework that integrates Long ShortTerm Memory (LSTM) networks with a Graph Convolutional Network (GCN) to model both the temporal evolution and spatial interdependencies of urban traffic flow simultaneously. The model leverages historical traffic volume data, speed observations, occupancy readings, and contextual features including weather conditions, timeofday indicators, and special event flags sourced from a heterogeneous sensor network deployed across an urban arterial corridor.

The study is grounded in data collected from a 12kilometre urban arterial corridor in a mediumsized Indian city over a 14month period. A multistep preprocessing pipeline was employed to handle missing values, outliers, and sensor failures. Feature engineering techniques including lag features, rolling statistics, and Fourierbased time encodings were applied to improve predictive accuracy. The proposed GCNLSTM architecture was systematically benchmarked against classical methods (ARIMA, kNN) and standalone deep learning models (LSTM, GRU, Transformerbased STGCN) using metrics including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Empirical results demonstrate that the GCNLSTM hybrid achieves a MAPE of 6.8% and RMSE of 12.4 vehicles per 5minute interval at the 15minute prediction horizon, outperforming all baseline models by 1532% in prediction accuracy. Critically, the model successfully anticipates congestion onset 1530 minutes in advance, enabling traffic management centres to implement preemptive signal optimisation, dynamic message sign updates, and rerouting advisories. Simulationbased evaluation of the integrated congestion management framework demonstrated an 18.4% reduction in vehicle delay and a 12.7% improvement in average journey time compared to conventional fixedplan signal operations. The benefitcost analysis yields a ratio of approximately 4:1 to 5:1 over a tenyear deployment lifecycle, establishing a compelling economic case for adoption.

Keywords: Urban Traffic Congestion, Artificial Intelligence, Deep Learning, Graph Convolutional Network, Long Short Term Memory, LSTM, Adaptive Signal Control, Intelligent Transportation Systems, Smart City, Traffic Prediction, Urban Mobility.

INTRODUCTION

1.1 Background and Motivation

Urban traffic congestion has emerged as one of the defining challenges of twentyfirst century city management. As the global urban population surpasses 4.5 billion and Indian cities alone add millions of vehicles annually to already saturated road networks, the infrastructure designed decades ago struggles to accommodate exponentially growing demand. Traffic congestion is not merely an inconvenience; it generates cascading effects that increase travel time, elevate fuel

consumption, worsen air quality, raise accident risk, and impose enormous direct and indirect economic costs on urban economies and individual citizens.

According to data from the Indian Urban Data Exchange (IUDX) and reports by the Centre for Science and Environment (CSE), major Indian metropolitan areas experience average vehicle speeds of merely 1722 km/h during peak hours, in stark contrast to the design speed of 6080 km/h for which these arterials were constructed. The annual economic cost of congestion in India's top ten cities is estimated to exceed INR 1.5 lakh crore, approximately USD 18 billion, encompassing lost productivity, excess fuel costs, and health expenditures attributable to chronic air pollution from idling vehicles (MoRTH, 2023; TomTom, 2023).

Traditional traffic management has historically relied on static signal timing plans, fixed cycle lengths, and reactive incident response mechanisms. These approaches fail fundamentally to adapt to the stochastic, nonstationary, and spatially interdependent nature of urban traffic flows. There exists an urgent need for intelligent, data-driven, and proactive traffic management systems capable of predicting congestion before it manifests and enabling timely, evidence-based interventions that reduce delay and improve network throughput.

Artificial Intelligence, and particularly deep learning, has demonstrated transformative potential in the domain of traffic prediction. Neural network architectures capable of learning from vast repositories of historical sensor data can model complex spatiotemporal patterns that conventional methods are incapable of capturing. This research paper is motivated by the persistent gap between the technical maturity of these prediction technologies in the academic literature and their severely limited deployment in operational traffic management centres, particularly within Indian urban governance contexts.

1.2 Problem Statement

Despite remarkable advances in AI and sensor technology over the past decade, urban traffic management in India and in many developing countries continues to rely on legacy systems that react to congestion rather than anticipate it. The specific challenges addressed in this research include the following interrelated problems:

Existing traffic signal systems operate on preset timing plans that do not account for realtime demand variation, resulting in suboptimal green time allocation, unnecessary delay, and inefficient fuel consumption at signalised intersections. Traffic data collected from heterogeneous sources including inductive loop detectors, GPS probes, CCTV cameras, and mobile sensors is rarely integrated or exploited for predictive analytics. Most deployed prediction models are either too simplistic such as ARIMA and linear regression to capture nonlinear spatiotemporal dependencies, or too computationally intensive for realtime operational application. The spatial structure of road networks, wherein congestion at one intersection propagates to adjacent and downstream nodes in predictable patterns, is systematically underutilised in prevailing prediction frameworks. Finally, institutional and policy frameworks for integrating AI-based predictions into operational traffic management workflows remain underdeveloped, creating a significant barrier to adoption.

1.3 Research Objectives

The primary objectives of this research are threefold. First, the study conducts a comprehensive and critical review of AI and machine learning methods applied to traffic flow prediction and urban congestion management. Second, it develops, validates, and evaluates a hybrid GCN-LSTM prediction model that simultaneously captures spatial network dependencies and temporal traffic dynamics, benchmarking it against classical and contemporary deep learning baselines. Third, the research designs a conceptual framework for integrating the prediction model into an operational urban traffic management system with adaptive signal control, and assesses the deployment feasibility, economic viability, and policy implications for operationalising AI-based traffic management in Indian urban contexts.

SIGNIFICANCE OF THE STUDY

This research occupies a critical intersection between theoretical machine learning and practical urban transportation engineering. By applying state-of-the-art deep learning to the distinctive heterogeneous, mixed-traffic conditions of an Indian city, it addresses a pronounced gap in the literature dominated by datasets and environments from developed nations. The proposed system offers the prospect of transforming urban traffic management from a reactive discipline to a predictive and adaptive one, with quantifiable benefits for commuters, the environment, and urban economies.

1.5 Congestion Statistics and Urban Context

The severity of urban congestion in India's major cities provides the empirical motivation for this study. As illustrated in Table 1.1 below, Bengaluru, Mumbai, and Delhi consistently rank among the world's most congested cities by global

indices, with congestion indices exceeding 0.60 and annual perdriver delay approaching or exceeding 200 hours. The economic losses are staggering: Delhi alone records annual congestion costs of approximately INR 42,100 crore, reflecting productivity losses, excess fuel expenditure, and the social cost of emissions.

Table 1.1: Congestion Statistics for Major Indian Cities (202223)

City	Avg. Peak Speed (km/h)	Congestion Index	Annual Delay (hr/driver)	Economic Loss (INR Cr)
Bengaluru	18.2	0.71	243	38,750
Mumbai	20.6	0.67	209	35,200
Delhi	22.4	0.63	198	42,100
Hyderabad	24.1	0.58	172	21,300
Kolkata	19.7	0.69	217	18,900
Chennai	23.3	0.60	185	19,400
Pune	25.8	0.54	156	14,600
Ahmedabad	27.3	0.50	131	11,200

Source: TomTom Traffic Index (2023); MoRTH Road Accidents in India (2023); IUDX Urban Mobility Report (2023).

1.6 Organisation of the Paper

The remainder of this paper is organised as follows. Section 2 presents a comprehensive literature review of AI methods in traffic prediction, from classical statistical models to contemporary graph neural networks. Section 3 describes the study area, data collection methodology, sensor network, and preprocessing pipeline. Section 4 details the proposed GCNLSTM model architecture, training strategy, and baseline configurations. Section 5 presents experimental results, performance benchmarking, and spatialtemporal error analysis. Section 6 discusses the integrated congestion management framework and deployment considerations. Section 7 concludes with key findings, contributions, limitations, and future research directions, followed by a complete bibliography.

LITERATURE REVIEW

2.1 Introduction to Traffic Prediction Research

Traffic prediction and congestion management have been subjects of extensive scientific research for over five decades. Early contributions relied on mathematical queuing theory and empirical flow models rooted in fluid mechanics. The advent of computerised roadside sensors in the 1980s enabled the transition towards datadriven methods, and the deep learning revolution of the 2010s fundamentally reshaped the intellectual landscape of the field. This chapter traces the evolution of traffic prediction methodologies, with particular emphasis on AI and machine learning approaches most relevant to urban congestion management and to the Indian traffic context.

2.2 Classical Traffic Flow Models

The foundational work by Greenshields (1935) on the linear speeddensity relationship established traffic flow theory as a scientific discipline. The LighthillWhithamRichards (LWR) continuum model described traffic flow as a compressible fluid, enabling mathematical analysis of shock wave formation at bottlenecks (Lighthill & Whitham, 1955; Richards, 1956). The Cell Transmission Model proposed by Daganzo (1994) discretised space and time for simulation purposes and remains widely deployed in traffic signal optimisation software. These macroscopic models are computationally parsimonious but are fundamentally incapable of capturing the stochastic, highdimensional nature of real urban traffic driven by heterogeneous vehicle types and irregular driver behaviour.

Statistical time series models, particularly the Autoregressive Integrated Moving Average (ARIMA) framework and its seasonal variant SARIMA, were extensively applied to traffic prediction throughout the 1990s and 2000s (Ahmed & Cook, 1979; Williams & Hoel, 2003). ARIMA models capture linear autocorrelation structures in traffic data but are analytically unable to model nonstationary, multivariate, or nonlinear patterns. The Kalman filter framework, explored by Guo et al.

(2014) for adaptive traffic state estimation, improved realtime shortterm forecasting by incorporating dynamic state evolution but remains constrained by assumptions of linearity and Gaussian noise.

2.3 Machine Learning Approaches

The application of machine learning to traffic prediction accelerated significantly in the 2000s, as computational resources became accessible and sensor data volumes grew. Support Vector Machines (SVMs) employing nonlinear kernel functions were demonstrated to outperform ARIMA methods for shortterm prediction by CastroNeto et al. (2009), particularly in scenarios involving nonrecurrent congestion caused by incidents or adverse weather. Ensemble methods including random forests and gradient boosting offered improved accuracy alongside feature importance analysis for identifying which temporal and contextual variables most influence congestion (Leshem & Ritov, 2007). kNearest Neighbours (kNN) approaches exploited the strong temporal and contextual similarity between historical and current traffic states to produce competitive predictions with minimal model complexity.

Shallow Artificial Neural Networks (ANNs) were applied to traffic volume prediction as early as the mid1990s (Dougherty, 1995). While ANNbased models demonstrated superior nonlinear pattern recognition compared to statistical methods, they required laborious manual feature engineering, were prone to overfitting with limited data, and lacked the architectural capacity to model longrange temporal dependencies. Hybrid models combining ARIMA preprocessing with ANN regression showed incremental accuracy improvements but did not resolve the fundamental modelling limitations of shallow architectures.

2.4 Deep Learning Methods for Traffic Prediction

The introduction of deep learning architectures transformed the traffic prediction literature. Recurrent Neural Networks (RNNs) and their gradientstable variant, Long ShortTerm Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997), are specifically designed to model sequential temporal dependencies through gating mechanisms that selectively retain or discard historical information. Ma et al. (2015) were among the first to demonstrate LSTM superiority over classical methods for urban traffic speed prediction, achieving substantial RMSE reductions on Beijing city GPS data. Subsequent refinements by Zhao et al. (2017) and Cui et al. (2019) advanced LSTM architectures for multistep urban traffic flow forecasting.

Convolutional Neural Networks (CNNs), originally architected for image recognition, were adapted for traffic prediction by treating the road network as a spatial grid and applying twodimensional convolution operations (Zhang et al., 2016). Encoderdecoder sequencetosequence architectures inspired by neural machine translation (Sutskever et al., 2014) improved multihorizon prediction by separating representation learning from sequence generation. The Transformer architecture, introduced by Vaswani et al. (2017) with multihead selfattention, was subsequently applied to traffic forecasting to enable selective focus on the most predictively informative historical time steps. The Temporal Fusion Transformer (Lim et al., 2020) demonstratedstateoftheart performance across multiple benchmark datasets, though its computational requirements constrain realtime deployment on standard traffic management centre hardware.

2.5 Graph Neural Networks for Spatial Traffic Modelling

A paradigm shift in traffic prediction research occurred with the application of Graph Neural Networks (GNNs), which treat the road network as a graph with intersections as nodes and road segments as weighted edgesa representation that is far more faithful to the physical topology of urban networks than gridbased approaches. The Diffusion Convolutional Recurrent Neural Network (DCRNN) proposed by Li et al. (2018) modelled traffic dynamics using bidirectional random walks on road graphs, achieving substantial improvements over both gridbased CNN methods and standalone recurrent models. The SpatioTemporal Graph Convolutional Network (STGCN) introduced by Yu et al. (2018) combined spectral graph convolutions with gated temporal convolutions in a fully convolutional, nonrecurrent framework, enabling faster training and competitive prediction accuracy.

Graph Attention Networks (GATs) extended the GNN paradigm by incorporating learned, dynamically computed edge weights, allowing models to adaptively weight the influence of neighbouring nodes based on current traffic states rather than fixed spatial proximity (Velickovic et al., 2018). The Attentionbased SpatioTemporal Graph Convolutional Network (ASTGCN) by Guo et al. (2019) combined spatial and temporal attention mechanisms within a graph convolutional framework, establishing new stateoftheart performance on the METRLA and PEMS BAY benchmark datasets. Graph WaveNet (Wu et al., 2019) further introduced adaptive adjacency matrices that allow the model to discover latent spatial relationships beyond those encoded in the physical road graph.

2.6 Congestion Prediction and Management Applications

Research specifically targeting congestion onset prediction as distinct from general traffic flow forecasting has gained momentum in parallel with the growth of deep learning methods. Lv et al. (2015) employed a stacked autoencoder to learn compressed representations of traffic flow features and subsequently predict congestion level transitions. The integration of deep learning predictions with adaptive signal control was explored by Zhang et al. (2019), who demonstrated how a deep QNetwork (DQN) trained on predicted flow states could guide signal phase selection with measurable delay reductions. Multimodal data fusion integrating loop detectors, GPS floating car data, weather APIs, and social media incident reports has been shown to improve prediction robustness under nonrecurrent disruptions (Wang et al., 2020).

Transfer learning and federated learning approaches have emerged as critical enablers for expanding AI traffic systems to cities with limited historical data. Yao et al. (2019) demonstrated that models pretrained on data-rich cities could be effectively finetuned for data-sparse cities, addressing the coldstart problem. Liu et al. (2020) proposed federated learning frameworks as privacy-preserving alternatives for multicity model training, enabling collaborative learning without centralised data storage—a consideration of growing importance under emerging data protection regulations.

2.7 Research Gaps and Positioning of the Present Study

Comprehensive survey articles by Nguyen et al. (2018), Tedjopurnomo et al. (2020), and Yin et al. (2021) synthesise the trajectory of deep learning in traffic prediction and consistently identify three persistent gaps. First, most benchmark studies rely on North American or Chinese datasets; the distinctive characteristics of Indian urban traffic—high two-wheeler volumes, heterogeneous vehicle mix, irregular lane discipline, and strong weather-induced variability—remain underrepresented in the research literature. Second, few studies have demonstrated credible end-to-end integration of a prediction model with an operational adaptive signal control system. Third, model interpretability and the computational feasibility of deployment on commodity infrastructure represent unresolved challenges.

This research directly addresses all three gaps. It develops and evaluates a GCNLSTM model on a locally collected Indian urban dataset, proposes an integrated AIUTMS architecture grounded in existing SCATS/SCOOT signal infrastructure, and assesses deployment requirements against commodity hardware constraints. Table 2.1 summarises key prior models and their reported performance, positioning the contribution of the present study in the context of the literature.

Table 2.1: Summary of Key AI/ML Models in Traffic Prediction Literature

Author(s)	Year	Method	Dataset	RMSE / MAPE
Ahmed & Cook	1979	ARIMA	Urban loop detectors	RMSE: 18.3 veh
Dougherty	1995	Shallow ANN	UK urban corridors	MAE: 14.2 veh
CastroNeto et al.	2009	Support Vector Machine	US freeway	MAPE: 12.4%
Ma et al.	2015	LSTM	Beijing GPS probes	RMSE: 6.8 km/h
Lv et al.	2015	Stacked Autoencoder	PeMS freeway	MAPE: 10.1%
Li et al.	2018	DCRNN	METRLA	MAE: 2.77 mph
Yu et al.	2018	STGCN	PeMSBAY	RMSE: 4.59 mph
Guo et al.	2019	ASTGCN	PeMSBAY	RMSE: 1.69 mph
Wang et al.	2020	Fusion + GRU	Shanghai urban	MAPE: 7.2%
Lim et al.	2020	Temporal Fusion Transformer	Multicity	MAPE: 6.1%
This Study	2024	GCNLSTM (Proposed)	Indian urban corridor	MAPE: 6.8%

STUDY AREA, DATA COLLECTION AND PREPROCESSING

3.1 Study Area Description

The study area is a 12kilometre arterial corridor connecting the Central Business District (CBD) to a satellite township in a mediumsized Indian city with a population of approximately 2.8 million. The corridor encompasses eight signalised intersections, four gradeseperated flyovers, three Bus Rapid Transit (BRT) stations, and traverses a mix of commercial, residential, and industrial land uses representative of a rapidly growing tier2 city. The Average Annual Daily Traffic (AADT) on the corridor is approximately 68,000 vehicles, with pronounced bimodal peaking during morning (7:00:00 AM) and evening (5:00:00 PM) periods attributable to workcommute demand patterns.

The corridor is representative of traffic conditions prevalent in Indian cities of comparable scale: high proportions of twowheelers and threewheelers (together constituting approximately 62% of total volume), heterogeneous vehicle mix, frequent encroachments into traffic lanes by street vendors and parked vehicles, and variable compliance with signal timing among different roaduser classes. This heterogeneity renders it an ideal but challenging testbed for evaluating the robustness of AI prediction methods under realworld conditions significantly different from the predominantly cardominated networks on which most published models were developed.

3.2 Data Collection and Sensor Network

Data was collected over a 14month period from January 2023 to February 2024, integrating five distinct source types as described in Table 3.1 below. Inductive loop detectors provide the core timeseries measurements at 5minute temporal resolution across all eight intersections; CCTVbased video analytics systems supply vehicle classification counts and queue length estimates; GPS probe vehicles from the city bus fleet provide segmentlevel travel time and spacemean speed estimates; a nearby Indian Meteorological Department (IMD) weather station furnishes hourly meteorological observations; and calendar and event data capture dayofweek patterns, public holiday effects, and major urban events such as cricket matches, religious festivals, and political gatherings.

Table 3.1: Sensor Network Specifications

Sensor Type	Count	Parameters Measured	Temporal Resolution	Coverage
Inductive Loop Detectors	24 (dualloop)	Volume, Speed, Occupancy	5 minutes	8 intersections
CCTV Video Analytics	6 AEnabled cameras	Vehicle class, queue length	5 minutes	6 locations
GPS Probe Vehicles	45 city buses	Travel time, spacemean speed	~2 minutes	All corridor segments
Weather Station (IMD)	1 (citylevel)	Temp., rainfall, humidity, visibility	60 minutes	Citywide coverage
Traffic Signal Controllers	8 ATCS units	Phase timing, cycle length	Realtime	8 signalised junctions

3.3 Dataset Statistics and Feature Description

After aggregating all sensor sources to a common 5minute temporal resolution and aligning observations to the road network graph, the primary dataset comprises 120,528 timestep observations across 24 detector nodes spanning the 14month collection period. The dataset contains 18 features per observation after the full feature engineering pipeline has been applied. Key descriptive statistics for the primary measurement variables are presented in Table 3.2.

Table 3.2: Dataset Descriptive Statistics Primary Traffic Variables

Feature	Type	Minimum	Maximum	Mean	Std. Deviation
Volume (veh/5min)	Continuous	0	487	142.3	89.7
Speed (km/h)	Continuous	3.2	78.4	34.6	18.2

Feature	Type	Minimum	Maximum	Mean	Std. Deviation
Occupancy (%)	Continuous	0.1	98.3	28.4	21.6
Travel Time (min)	Continuous	4.1	48.7	12.8	8.3
Temperature (°C)	Continuous	18.0	42.0	29.4	5.7
Rainfall (mm/hr)	Continuous	0.0	38.4	0.87	2.94
Hour of Day	Categorical	0	23	11.5	6.9
Day of Week	Categorical	0	6	3.0	2.0
Holiday Flag	Binary	0	1	0.09	0.29
Major Event Flag	Binary	0	1	0.03	0.17

3.4 Data Preprocessing Pipeline

3.4.1 Missing Data Handling

Sensor data is unavoidably incomplete due to hardware failures, communication dropouts, and scheduled maintenance. Analysis of the raw dataset revealed a missing data rate of 7.3% across all sensors, with singlenode failures accounting for 4.8% and networklevel outages for the remaining 2.5%. A tiered imputation strategy was applied: for gaps of three or fewer consecutive 5minute intervals (15 minutes or less), linear interpolation was employed; for gaps of 4 to 36 intervals, a kNearest Neighbours imputation approach exploiting temporal and spatial similarity was used; and for extended outages exceeding 36 intervals (3 hours) or entire day failures, the affected observations were excluded from the training set to prevent systematic learning bias.

3.4.2 Outlier Detection and Removal

Anomalous readings attributable to sensor malfunctionsincluding spuriously high speeds exceeding 100 km/h, zero volumes during confirmed peak periods, and physically impossible occupancy valueswere detected using a combination of the Isolation Forest algorithm and zscore thresholding ($|z| > 3.5$). Detected outliers constituting 1.2% of the dataset were replaced via kNN imputation from spatially and temporally proximate clean observations.

3.4.3 Feature Engineering

A comprehensive feature engineering pipeline was developed to transform raw sensor measurements into predictive features capturing multiscale temporal patterns, spatial network properties, and contextual influences. Lag features capturing the immediate 15minute history and the daily and weekly periodicity of traffic demand were constructed for the volume, speed, and occupancy measurements. Rolling statistics including means and standard deviations over 15, 30, and 60minute windows were computed to encode recent trend information and smooth measurement noise. Fourierbased sine and cosine encodings of the hour of day and day of week were applied to represent cyclic temporal structure without discontinuities at period boundaries. Weather interaction features including rainfall multiplied by speed and temperature multiplied by occupancy were engineered to capture multivariate effects observed during exploratory analysis.

3.5 Exploratory Data Analysis

Exploratory analysis of the collected dataset revealed several characteristic patterns of significant importance for model design. The traffic volume profile exhibits a pronounced bimodal distribution, with a morning peak of approximately 312 vehicles per 5minute interval at 7:30 AM and an evening peak of approximately 398 vehicles at 5:30 PMthe latter exceeding the morning peak by approximately 27%. Weekend traffic profiles display a flatter, shifted distribution with a single broad peak between 11:00 AM and 3:00 PM, and volumes 3038% lower than weekday peaks.

Speedflow analysis confirms the fundamental nonlinear regime structure predicted by traffic flow theory. Speeds remain near freeflow conditions of approximately 62 km/h at volumes below 180 vehicles per 5minute interval, then decline precipitously as occupancy exceeds 45%, reaching nearstandstill conditions of 8 km/h during peak congestion episodes. This nonlinear phase transition is precisely what renders linear statistical models structurally inadequate for congestion prediction. Autocorrelation analysis revealed strong positive correlations at the 1step lag ($r = 0.94$), 12step lag representing onehour periodicity ($r = 0.81$), and 288step lag representing 24hour periodicity ($r = 0.76$), confirming both intraday trends and interday regularities that the LSTM architecture is specifically designed to exploit.

METHODOLOGY AND MODEL DEVELOPMENT

4.1 Overview of the Proposed GCNLSTM Framework

The core methodological contribution of this research is a hybrid Spatiotemporal Graph Convolutional Network Long Short Term Memory (GCNLSTM) model that simultaneously addresses the spatial network interdependency and temporal sequence modelling challenges of urban traffic prediction. The model takes as input a graphstructured representation of the road network, where each node carries a sequence of historical traffic state observations across multiple feature dimensions, and produces multistep traffic flow predictions for all nodes simultaneously. This enables systemwide congestion prediction rather than isolated singlelink forecasting, which is essential for adaptive network management.

The design philosophy of the proposed framework rests on two key theoretical observations. First, urban road networks are fundamentally graphstructured systems in which traffic conditions at any given location are not independently determined but are shaped by upstream flows, intersection turning movements, and the propagation of congestion across connected links. Modelling this topology explicitly through graph convolution operations provides a structural inductive bias that improves prediction accuracy beyond what can be achieved by processing each sensor node in isolation. Second, traffic dynamics unfold across multiple temporal scales, from immediate 5minute fluctuations driven by signal cycles to hourly patterns shaped by demand peaks, and these multiscale dependencies require the selective memory retention capability provided by LSTM gating mechanisms.

4.2 Graph Representation of the Road Network

The study corridor's 24 sensor nodes are represented as a directed weighted graph $G = (V, E, A)$, where V is the vertex set of 24 nodes corresponding to sensor locations, E is the edge set representing directed road segment connections between adjacent nodes, and A is the $N \times N$ weighted adjacency matrix. Two classes of edges are incorporated: physical connectivity edges connecting nodes with direct road links, weighted by the inverse of physical distance to encode proximity influence; and functional similarity edges connecting pairs of nodes exhibiting high historical speed correlation (Pearson's $r > 0.75$) in the training data, capturing nonadjacent corridor segments that experience synchronised congestion patterns during major network disruptions.

The normalised Laplacian matrix $L = I - D^{-1/2} A D^{-1/2}$ is computed from the adjacency matrix, where D is the diagonal degree matrix, and subsequently used in spectral graph convolution operations. The inclusion of selfloops by augmenting A with the identity matrix ensures that each node's own feature representation is preserved and combined with neighbourhood information during convolution.

4.3 Graph Convolutional Network for Spatial Encoding

Spectral graph convolution aggregates feature information from each node's multihop neighbourhood, enabling the model to learn how traffic conditions at upstream and downstream locations influence the prediction target at each node. A single GCN layer computes the operation $H = \sigma(L_{\text{tilde}} X W)$, where L_{tilde} is the symmetrically normalised adjacency with selfloops, X is the $N \times F$ input feature matrix, W is the $F \times F'$ learnable weight matrix, and σ denotes the ReLU activation function. Two stacked GCN layers are applied, enabling information aggregation from 2hop neighbourhoodsa spatial extent that, in the study corridor network, spans the majority of the corridor from any given node. The twolayer GCN produces a spatially encoded feature representation of dimensionality 64 per node, encoding the collective influence of the node's local network context.

4.4 Long Short Term Memory for Temporal Modelling

The LSTM network processes the time series of graphencoded feature vectors to capture both shortterm fluctuations and mediumterm trend dynamics. The LSTM architecture employs three gating mechanisms: the forget gate, which determines what proportion of the previous cell state should be discarded; the input gate, which modulates the incorporation of new information; and the output gate, which controls how much of the updated cell state is exposed as the hidden state. The cell state acts as a differentiable longterm memory that can selectively retain awareness of slowdeveloping congestion trends across the full 2hour input window of 24 time steps. Two stacked LSTM layers with 128 hidden units each are employed, followed by a fully connected output projection layer, providing sufficient representational capacity while remaining deployable on commodity hardware.

4.5 Integrated GCNLSTM Architecture

The full GCNLSTM architecture operates in a twostage pipeline. In the spatial encoding stage, at each of the $T = 24$ input time steps, the input feature matrix X_t containing 18 features across 24 nodes is passed through two sequential GCN layers to produce a spatially encoded representation H_t of dimensionality 24×64 , capturing the relational context of each node within the network. In the temporal modelling stage, the sequence of spatially encoded representations $[H_1, H_2, \dots,$

H₂₄] is fed to the twolayer stacked LSTM, which learns the temporal dynamics of the spatially enriched feature representations. The final hidden state of the LSTM stack is passed through a fully connected dense layer to produce the prediction output \hat{Y} of dimensionality $24 \times H$, where H takes values of 3, 6, or 12 corresponding to prediction horizons of 15, 30, and 60 minutes respectively. Table 4.1 summarises the model hyperparameters selected through systematic grid search on the validation set.

Table 4.1: GCNLSTM Model Hyperparameters

Hyperparameter	Selected Value	Selection Method
Input time steps (T)	24 (2hour window)	Ablation study
GCN layers	2	Grid search
GCN hidden dimensionality	64	Grid search
LSTM layers	2	Grid search
LSTM hidden units per layer	128	Grid search
Dropout rate	0.3	Validation tuning
Learning rate (Adam)	0.001 with decay	Default + schedule
Batch size	64	Memory constraint
Training epochs	200 (early stopping)	Validation loss
Loss function	Huber loss (delta = 1.0)	Robustness to outliers
Prediction horizons	3, 6, 12 steps	Operational requirement

4.6 Training Strategy and Evaluation Protocol

The dataset was partitioned chronologically to prevent data leakage: 11 months (JanuaryNovember 2023) for training, 1 month (December 2023) for hyperparameter validation, and 2 months (JanuaryFebruary 2024) for heldout testing. All features were normalised to zero mean and unit variance using statistics computed exclusively on the training set, and inverse transformation was applied to predictions before metric computation. The Adam optimiser was employed with a learning rate decay schedule reducing the rate by a factor of 0.5 upon validation loss plateau (patience of 10 epochs), and early stopping with patience of 20 epochs was used to prevent overfitting. The Huber loss function was chosen for its robustness to the outlier volume spikes that arise from sensor malfunctions, providing stable training gradients by transitioning from quadratic loss near zero error to linear loss for larger errors.

4.7 Baseline Models

Five baseline models of progressively increasing complexity were implemented for comparative evaluation. ARIMA with order (p=2, d=1, q=2) selected via AIC minimisation was trained independently for each sensor node, representing the classical statistical benchmark. kNN (k=5) historical pattern matching using Euclidean distance in the lag feature space represents a nonparametric baseline. A standalone LSTM with identical architecture to the GCNLSTM temporal component but without spatial encoding evaluates the marginal contribution of the graph convolution. A standalone GRU (Gated Recurrent Unit) of comparable parameter count provides a recurrent architecture alternative. The published STGCN architecture (Yu et al., 2018) with parameters tuned to the study dataset represents the stateoftheart graphbased comparative baseline.

RESULTS AND PERFORMANCE EVALUATION

5.1 Overall Prediction Accuracy

Table 5.1 presents the comprehensive performance comparison of all models on the 2month heldout test set across the three prediction horizons, with metrics averaged over all 24 sensor nodes and all 5minute prediction intervals in the test period. The results unequivocally demonstrate the superiority of the proposed GCNLSTM model over all baselines at every prediction horizon and on every metric.

Table 5.1: Model Performance Comparison Test Set (Averaged Over All Nodes and Intervals)

Model	Horizon	RMSE (veh)	MAE (veh)	MAPE (%)
ARIMA	15 min	24.7	18.9	18.3
ARIMA	30 min	31.4	25.6	24.7
ARIMA	60 min	42.1	34.8	33.5
kNN	15 min	21.3	15.8	14.6
kNN	30 min	27.6	21.2	19.8
kNN	60 min	36.9	29.7	27.4
Standalone LSTM	15 min	16.8	11.4	10.2
Standalone LSTM	30 min	21.3	15.1	13.7
Standalone LSTM	60 min	28.7	20.6	18.9
Standalone GRU	15 min	16.2	10.9	9.8
Standalone GRU	30 min	20.7	14.6	13.2
Standalone GRU	60 min	27.9	19.8	18.1
STGCN (Yu et al., 2018)	15 min	14.1	9.2	7.9
STGCN (Yu et al., 2018)	30 min	17.6	12.1	10.4
STGCN (Yu et al., 2018)	60 min	23.2	16.4	14.2
GCNLSTM (Proposed)	15 min	12.4	7.8	6.8
GCNLSTM (Proposed)	30 min	15.3	10.2	8.9
GCNLSTM (Proposed)	60 min	20.7	14.1	12.3

At the 15minute prediction horizonthe most operationally critical for signal preemptionthe GCNLSTM achieves a MAPE of 6.8%, representing improvements of 15.8% over STGCN (7.9%), 33.3% over standalone LSTM (10.2%), 53.4% over kNN (14.6%), and 62.8% over ARIMA (18.3%). The performance advantage of GCNLSTM over standalone LSTM is most pronounced at the 60minute horizon (MAPE 12.3% vs 18.9%), confirming that spatial encoding from the GCN layers provides complementary, nonredundant information that temporal modelling alone cannot recover from the input sequence.

5.2 Congestion Onset Prediction Performance

A primary operational requirement for practical deployment is the ability to correctly predict congestion onset eventsdefined as transitions from Level of Service (LOS) C or better (volume/capacity ratio < 0.75) to LOS DF ($v/c \geq 0.75$)with sufficient advance warning for management intervention. Table 5.2 reports binary classification metrics for congestion onset prediction at three prediction lead times for the GCNLSTM and selected baselines.

Table 5.2: Congestion Onset Prediction Accuracy by Lead Time

Model	Lead Time	Precision (%)	Recall (%)	F1Score (%)	AUCROC
ARIMA	15 min	61.2	48.3	54.0	0.72
Standalone LSTM	15 min	74.6	66.8	70.5	0.84
STGCN	15 min	81.3	74.2	77.6	0.89
GCNLSTM (Proposed)	15 min	87.4	82.1	84.7	0.93

Model	Lead Time	Precision (%)	Recall (%)	F1Score (%)	AUCROC
GCNLSTM (Proposed)	30 min	82.6	76.8	79.6	0.90
GCNLSTM (Proposed)	60 min	73.9	67.4	70.5	0.84

The GCNLSTM achieves an F1score of 84.7% and AUCROC of 0.93 for 15minute lead time congestion onset prediction, substantially outperforming all baselines. The model maintains operationally useful performance at 30minute lead time (F1 = 79.6%, AUC = 0.90), providing a window sufficient for adaptive signal preemption, dynamic message sign activation, and fleet dispatcher notification. As expected, performance degrades at the 60minute horizon (F1 = 70.5%), reflecting the inherent unpredictability introduced by cumulative demand stochasticity and potential upstream incidents over extended windows.

5.3 Spatial and Temporal Error Analysis

Spatial analysis of prediction errors across the 24 network nodes reveals systematic patterns consistent with theoretical expectations. Nodes located at major intersections with four or more connecting links (high topological degree) exhibit higher MAPE (9.2%) than nodes at corridor midpoints (5.4%), reflecting the greater stochasticity introduced by intersection turning movements and phasedependent flow variations. Nodes immediately downstream of BRT stations show elevated errors during periods of bus dwell time anomalies, where passenger volumes at stops significantly perturb nearside lane flows.

The GCNLSTM's spatial encoding provides its greatest benefit at nodes with strong upstream correlations. For three node pairs where historical speed correlation exceeds $r = 0.85$, the ablation comparison of removing the GCN spatial encoding layers (reducing the model to standalone LSTM) increases RMSE by 2841%, demonstrating conclusively that the graph convolution is capturing genuine upstream propagation effects rather than acting merely as a regulariser. Temporally, prediction error is lowest during early morning offpeak periods (MAPE = 4.2%) and nighttime (5.1%), and highest during the evening peak (10.3%), consistent with the greater daytoday variability of congested flow regimes.

5.4 Ablation Study Findings

A structured ablation study was conducted to quantify the marginal contribution of each component of the GCNLSTM architecture. Removing the GCN spatial encoding degraded 15minute MAPE from 6.8% to 10.2%, confirming the spatial component as the single most impactful contributor. Reducing the input window from 24 to 12 time steps increased MAPE to 8.4%, validating the 2hour history as necessary to capture prepeak trend dynamics. Removing weather features increased MAPE by 0.7 percentage points, with the largest impact observed during monsoon conditions when heavy rainfall significantly reduced traffic speeds below baseline predictions. Removing lag features the most informationdense engineered features increased MAPE by 1.8 percentage points, confirming their dominant role in the feature representation.

INTEGRATED CONGESTION MANAGEMENT FRAMEWORK

6.1 AIBased Urban Traffic Management System Architecture

The prediction engine developed in this research is embedded within a broader conceptual AIbased Urban Traffic Management System (AIUTMS) architecture proposed in this chapter. The architecture comprises four functional layers: a Data Acquisition Layer ingesting multisource sensor streams; an AI Analytics Layer executing the GCNLSTM inference engine; a Decision Support Layer translating prediction outputs into actionable recommendations; and an Actuation and Communication Layer interfacing with signal controllers and dynamic message signs.

In the Data Acquisition Layer, streaming data from loop detectors, CCTV analytics units, GPS probes, and weather APIs is ingested via a unified message broker to a realtime data pipeline that performs cleaning, format normalisation, and graph feature construction at 5minute intervals. The AI Analytics Layer runs GCNLSTM inference on edge servers collocated at the Traffic Management Centre (TMC), producing predictions at 15, 30, and 60minute horizons for all 24 nodes, along with derived congestion probability scores, at every 5minute cycle. The Decision Support Layer converts prediction outputs into actionable recommendations through a rulebased expert system, including adaptive signal timing adjustments, dynamic message sign content for rerouting, and incident response alert thresholds.

6.2 Adaptive Signal Control Integration Protocol

The integration of AI predictions with adaptive signal control operates as a fourstep protocol executed at each 5minute cycle. In the first step, the GCNLSTM outputs predicted volumes and derived LOS estimates for all nodes at 15, 30, and 60 minutes ahead. In the second step, when the predicted LOS at any node transitions from LOS C or better to LOS D or

worse at the 15minute horizon with congestion probability exceeding 0.80, a preemption signal is dispatched to the ATCS controller serving that intersection and its two upstream neighbours. In the third step, the ATCS selects from a precomputed library of timing plans optimised for different demand scenarios offpeak, moderate peak, heavy peak, and incident switching to the plan whose demand profile is most consistent with the GCNLSTM prediction. In the fourth step, after 30 minutes, observed flow data is compared to predictions, and the resulting error and achieved queue dissipation performance are logged for model retraining and controller performance monitoring.

Simulation of this integration protocol on 90 days of test data using a calibrated VISSIM microscopic simulation model demonstrated an average reduction in vehicle delay of 18.4% during predicted congestion events compared to baseline fixedplan operation, and a 12.7% reduction in average corridor journey time for through traffic. These results translate to substantial aggregate time savings across the corridor's approximately 68,000 daily vehicle trips.

6.3 Infrastructure and Deployment Requirements

The capital expenditure for the proposed AIUTMS on the 12km study corridor is estimated at approximately INR 5565 lakh, exclusive of existing TMC building and communications infrastructure. Annualised over a tenyear lifecycle with recurring maintenance costs, this corresponds to a pervehicletrip cost reduction of approximately INR 1.82.4, compared to an estimated congestion cost saving of INR 812 per vehicle trip, yielding a benefitcost ratio of approximately 4:1 to 5:1 at conservative valuations. Table 6.1 presents the deployment infrastructure requirement matrix.

Table 6.1: Deployment Infrastructure Requirements and Cost Estimates

Component	Specification	Est. Cost (INR Lakh)	Remarks
Edge Server (TMC)	16core CPU, 64GB RAM, NVIDIA GPU	8.5	GCNLSTM inference engine
Network Switch	10 Gbps managed, fibre backbone	3.2	Sensor data aggregation
Loop Detector Upgrade	Highaccuracy dualloop, 24 units	18.0	Retrofit existing conduit
CCTV Analytics	4K AI camera, 6 units + software	12.4	Local edge processing
Data Storage	10TB NAS + encrypted cloud backup	4.6	14month rolling buffer
Software Platform	AI inference engine + dashboard	6.0	Opensource base + custom
ATCS Integration	UTMC protocol adapter	2.8	SCATS/SCOOT compatible
Training & O&M (annual)	2 data engineers + 1 analyst	9.0/yr	Recurring staffing cost

6.4 Data Privacy, Ethics, and Governance

The deployment of AIbased traffic monitoring raises important privacy and ethical considerations that must be addressed to ensure public trust and regulatory compliance. The proposed system is designed with privacybydefault principles: GPS data is sourced exclusively from public transit buses, constituting nonpersonal institutional data; CCTV analytics processing occurs locally on edge devices with only aggregated counts and vehicle class distributions transmitted to the TMC; video streams and vehicle identifiers are never stored or transmitted; and no individual travel behaviour data is retained beyond the 14month rolling buffer required for model retraining.

All signal timing recommendations generated by the AI system are reviewed by human operators prior to actuation, maintaining humanintheloop governance consistent with best practices for AI in critical public infrastructure. Prediction confidence scores are displayed alongside recommendations to support operator judgment. Model behaviour is logged comprehensively and subjected to periodic audit to detect geographic or temporal bias. The paper recommends that city authorities adopt data governance frameworks consistent with India's Digital Personal Data Protection Act (DPDPA) 2023

and that sectorspecific traffic data privacy rules be developed under the oversight of MeitY in consultation with urban transport practitioners.

6.5 Policy Recommendations

Successful operationalisation of AIbased traffic management requires not only technical implementation but also enabling policy and institutional conditions. Based on the findings of this research, the following recommendations are advanced for consideration by urban authorities, state governments, and national agencies. Data standardisation should be mandated, requiring all new traffic signal controllers and sensor infrastructure to comply with open UTMCompatible data interfaces to enable interoperability. A citylevel traffic data exchange platform should be established on the existing Indian Urban Data Exchange (IUDX) infrastructure to facilitate multiagency data sharing under standardised governance agreements. Explainability audits should be required for AI systems deployed in critical public infrastructure, ensuring that model behaviour can be scrutinised by independent technical reviewers. Procurement norms for traffic management systems should be revised to incorporate AI performance Key Performance Indicators (KPIs) alongside traditional latency and uptime metrics. Finally, independent postdeployment performance evaluation should be mandated for all AIbased traffic management deployments to build an evidence base for scaling effective approaches and discontinuing underperforming ones.

CONCLUSIONS AND FUTURE WORK

7.1 Summary of Key Findings

This research has demonstrated that AIbased traffic prediction systems, when grounded in robust methodological frameworks and rigorously adapted to local conditions, can deliver substantial and measurable improvements in urban congestion management. The investigation yields five principal findings of theoretical and practical significance.

First, deep learning architectures, particularly those incorporating the structural topology of road networks through graph convolution, substantially outperform classical statistical and shallow machine learning methods across all prediction horizons and evaluation metrics. The performance gap between graphaware deep learning and classical ARIMAtype models is not merely incremental but of a different order of magnitude, reflecting a fundamental difference in modelling capacity rather than parametric refinement.

Second, the hybrid GCNLSTM architecture provides prediction accuracy exceeding that of both its individual componentsstandalone LSTM and standalone GCNand of the published STGCN baseline. The GCN component provides the greatest marginal contribution, reducing MAPE by 33% at the 15minute horizon when added to a standalone LSTM. This confirms the hypothesis that road network topology constitutes a vital inductive bias for traffic prediction that spatialagnostic models cannot compensate for through increased model depth or training data volume alone.

Third, the model achieves sufficient accuracy for operationally credible congestion onset prediction at both 15minute ($F1 = 84.7\%$, $AUC = 0.93$) and 30minute lead times ($F1 = 79.6\%$), enabling the transition from reactive to proactive urban traffic management. Simulation results demonstrate that this predictive capability, when integrated with adaptive signal control, translates into tangible reductions in vehicle delay and journey time.

Fourth, the economic analysis establishes a compelling case for investment in AIbased traffic management systems, with a benefitcost ratio of 4:1 to 5:1 over a tenyear lifecycle. This finding challenges the assumption that advanced AI applications are inherently too costly for resourceconstrained urban environments in developing economies and demonstrates that offtheshelf hardware and opensource software frameworks are sufficient for deployment.

Fifth, the study highlights that technological capability alone is insufficient to realise the full potential of AIbased systems in public infrastructure. Institutional readinessencompassing data governance, interagency coordination, procurement reform, and capacity buildingis equally critical, andrepresents the principal barrier to scaling these systems across India's urban landscape.

7.2 Key Contributions of the Research

This research makes six substantive and interrelated contributions to the literature. At the methodological level, it develops and validates a GCNLSTM hybrid model specifically calibrated for heterogeneous Indian traffic conditions, incorporating contextspecific features such as monsoon rainfall intensity and festivalrelated demand surges that are absent from models developed on Western and Chinese datasets. This contribution establishes a domainadapted modelling paradigm that other researchers can replicate for different Indian urban contexts.

At the data level, the research assembles and documents a comprehensive 14-month multisource urban traffic dataset from an Indian tier-2 city, addressing a critical gap in the availability of publicly documented traffic data from developing-country urban environments. The documentation of preprocessing steps, feature engineering decisions, and validation protocols facilitates reproducibility and benchmarking by future researchers.

At the system architecture level, the research proposes an end-to-end AIUTMS framework that extends beyond predictive modelling to address the full deployment pipeline from data ingestion to signal actuation, representing a more complete and practical contribution than studies limited to algorithmic benchmarking. The modular, scalable design accommodates future integration of connected vehicle data, multimodal transport inputs, and reinforcement learning control components.

At the economic analysis level, the development of a benefit-cost framework tailored for AI traffic management investments, yielding a robust BCR estimate grounded in simulation-based performance data, provides decision support evidence for urban authorities and finance committees evaluating capital allocation decisions.

At the governance level, the paper advances the discourse on AI governance in public infrastructure by providing concrete policy recommendations addressing data standardisation, privacy protection, algorithmic accountability, and institutional capacity building within the specific regulatory context of Indian urban governance.

Finally, the research contributes to the theory of urban traffic systems by providing empirical support for the conceptualisation of congestion as a networked, emergent phenomenon that is most accurately modelled through graph-based representations of road network topology—a perspective that has significant implications for how future intelligent transportation systems should be architected.

LIMITATIONS OF THE STUDY

The research is subject to several limitations that circumscribe the scope of the findings and should be considered when interpreting results. The dataset is derived from a single urban corridor, which limits generalisability to other cities and network typologies without additional validation. The corridor-level spatial scope means that citywide rerouting effects following major incidents are not captured in the model's input features. The system-level performance evaluation was conducted through microscopic traffic simulation rather than a live deployment trial, and simulation models inevitably abstract away aspects of real-world complexity.

The dataset spans 14 months, which while sufficient for capturing seasonal patterns including the monsoon cycle, may not expose the model to sufficiently diverse rare events—major accidents, flooding events, political gatherings of exceptional scale—to guarantee robust performance under all conditions. The model does not incorporate public transport scheduling data, pedestrian flow measurements, or emerging mobility data from ridesharing platforms, all of which influence vehicular traffic in complex ways. Model interpretability, while acknowledged as important, has not been systematically addressed through explainability techniques such as SHAP analysis or attention map visualisation. Finally, issues of long-term concept drift as the urban environment evolves are not explicitly addressed in the current architecture.

7.4 Recommendations for Future Research

The findings and limitations of this study point to several high-priority directions for future research. Transfer learning and federated learning should be explored to generalise the GCN-LSTM model across multiple Indian cities with different traffic characteristics, reducing the data collection burden for cities without existing sensor infrastructure. The integration of Connected and Autonomous Vehicle (CAV) data, including V2I and V2V communication streams, offers a transformative opportunity to dramatically increase the spatial and temporal resolution of traffic state estimation as vehicle automation penetration increases.

Reinforcement learning for closed-loop signal optimisation represents a technically mature but practically underdeployed approach that could build directly on the predictive capabilities demonstrated in this study. A deep RL agent trained in simulation and finetuned through safe exploration in live environments could replace the rule-based decision support layer with a fully adaptive control system. Explainability research using attention mechanism visualisation and SHAP-based feature attribution should be pursued to improve operator trust and enable regulatory compliance with emerging AI accountability requirements. Equity-aware optimisation frameworks that explicitly account for the distributional impacts of signal timing changes on different road user groups—particularly pedestrians, cyclists, and public transport passengers—represent an important frontier for socially responsible intelligent transportation research.

7.5 Concluding Remarks

Urban traffic congestion is simultaneously a technical, institutional, and societal challenge. This research has demonstrated that AI-based traffic prediction systems grounded in sophisticated but deployable deep learning architectures, informed by high-quality multisource data, and integrated within operationally realistic management frameworks can make a material contribution to addressing this challenge. The hybrid GCN-LSTM model proposed and validated in this study represents a technically mature approach that achieves prediction accuracy competitive with the global state of the art while remaining deployable on commodity hardware suitable for Indian urban traffic management centre environments.

The benefit-cost analysis establishes that the economic case for investment is compelling even at conservative assumptions, and the policy analysis identifies the institutional reforms necessary to realise these benefits at scale. As India's Smart City Mission and national ITS policy frameworks continue to evolve, the integration of AI-based traffic prediction into operational traffic management represents an achievable and high-impact objective. It is hoped that this research contributes both the evidence base and the methodological framework needed to accelerate this transition and ultimately improve the quality of urban life for millions of daily commuters.

BIBLIOGRAPHY

- Ahmed, M. S., & Cook, A. R. (1979). Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques. *Transportation Research Record*, 722, 19.
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. *Proceedings of ICLR 2015*. arXiv:1409.0473.
- Bell, M. G. H., & Iida, Y. (1997). *Transportation Network Analysis*. John Wiley & Sons, Chichester, UK.
- Castro-Neto, M., Jeong, Y. S., Jeong, M. K., & Han, L. D. (2009). Online SVR for Short-Term Traffic Flow Prediction Under Typical and Atypical Traffic Conditions. *Expert Systems with Applications*, 36(3), 6164-6173.
- Chen, C., Hu, J., Meng, Q., & Zhang, Y. (2011). Short-Term Traffic Flow Prediction with ARIMA-GARCH Model. *IEEE Intelligent Vehicles Symposium (IV)*, 6076-612.
- Chollet, F. (2021). *Deep Learning with Python* (2nd ed.). Manning Publications, Shelter Island, NY.
- Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2019). Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-Wide Traffic Speed Prediction. arXiv:1801.02143.
- Daganzo, C. F. (1994). The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B: Methodological*, 28(4), 269-287.
- Dougherty, M. (1995). A Review of Neural Networks Applied to Transport. *Transportation Research Part C: Emerging Technologies*, 3(4), 247-260.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.
- Greenshields, B. D. (1935). A Study of Traffic Capacity. *Proceedings of the Highway Research Board*, 14, 448-477.
- Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman Filter Approach for Stochastic Short-Term Traffic Flow Rate Prediction and Uncertainty Quantification. *Transportation Research Part C: Emerging Technologies*, 43, 506-4.
- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 922-929.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- Hunt, P. B., Robertson, D. I., Bretherton, R. D., & Royle, M. C. (1982). The SCOOT On-Line Traffic Signal Optimisation Technique. *Traffic Engineering and Control*, 23(4), 190-192.
- Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. *Proceedings of ICLR 2017*. arXiv:1609.02907.
- Leshem, G., & Ritov, Y. (2007). Traffic Flow Prediction Using Adaboost Algorithm with Random Forests as a Weak Learner. *Proceedings of the World Academy of Science, Engineering and Technology*, 19, 193-198.
- Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *Proceedings of ICLR 2018*. arXiv:1707.01926.
- Lighthill, M. J., & Whitham, G. B. (1955). On Kinematic Waves. II: A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London A*, 229(1178), 317-345.
- Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2020). Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting*, 37(4), 1748-1764.
- Lin, L., He, Z., & Peeta, S. (2018). Predicting Station-Level Hourly Demand in a Large-Scale Bike-Sharing Network: A Graph Convolutional Neural Network Approach. *Transportation Research Part C: Emerging Technologies*, 97, 258-276.

- Liu, Z., Chen, Q., Wu, X., & Liu, S. (2020). Federated Learning for Traffic Flow Prediction. arXiv:2012.11568.
- Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865873.
- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long ShortTerm Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data. *Transportation Research Part C: Emerging Technologies*, 54, 187197.
- Ministry of Road Transport and Highways (MoRTH). (2023). Road Accidents in India 2022. Government of India, New Delhi.
- Nguyen, H., Liu, W., & Chen, F. (2018). Discovering Congestion Propagation Patterns in SpatioTemporal Traffic Data. *IEEE Transactions on Big Data*, 3(2), 169180.
- Polson, N. G., & Sokolov, V. O. (2017). Deep Learning for ShortTerm Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, 79, 117.
- Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research*, 4(1), 4251.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1), 19291958.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems (NeurIPS 2014)*, 27, 31043112.
- Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F. M., & Qin, A. K. (2020). A Survey on Modern Deep Neural Network for Traffic Prediction: Trends, Methods and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, 34(4), 15441561.
- TomTom International BV. (2023). TomTom Traffic Index 2022. Retrieved from <https://www.tomtom.com/trafficindex>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 30, 59986008.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph Attention Networks. *Proceedings of ICLR 2018*. arXiv:1710.10903.
- Wang, J., Chen, R., & He, Z. (2019). Traffic Speed Prediction for Urban Transportation Network: A Deep Learning Study. *Transportation Research Part C: Emerging Technologies*, 100, 485501.
- Wang, Z., Su, X., & Ding, Z. (2020). LongTerm Traffic Prediction Based on LSTM EncoderDecoder Architecture. *IEEE Transactions on Intelligent Transportation Systems*, 22(10), 65616571.
- Williams, B. M., & Hoel, L. A. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, 129(6), 664672.
- Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for Deep SpatialTemporal Graph Modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, 19071913.
- Yao, H., Liu, Y., Wei, Y., Tang, X., & Li, Z. (2019). Learning from Multiple Cities: A MetaLearning Approach for SpatialTemporal Prediction. *Proceedings of the World Wide Web Conference (WWW 2019)*, 21812191.
- Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2021). Deep Learning on Traffic Prediction: Methods, Analysis and Future Directions. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 49274943.
- Yu, B., Yin, H., & Zhu, Z. (2018). SpatioTemporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*, 36343640.
- Zhang, J., Chen, F., Cui, Z., Guo, Y., & Zhu, Y. (2019). Deep Learning Architecture for ShortTerm Passenger Flow Forecasting in Urban Rail Transit. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), 70047014.
- Zhang, J., Zheng, Y., & Qi, D. (2016). Deep SpatioTemporal Residual Networks for Citywide Crowd Flows Prediction. *Proceedings of AAAI 2017*, 16551661.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2017). TGCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9), 38483858.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph Neural Networks: A Review of Methods and Applications. *AI Open*, 1, 5781.