Group Number 151 Akshay Mohan 2023aa05315 Monisha G 2023aa05536 Peyala Samarasimha Reddy 2023aa05072 Sreelakshmi Ajith 2023aa05316			
Traffic Flow Prediction			
	PAPER 1	**PAPER 2**	**PAPER 3**
Title of the paper	Dynamic Spatio-temporal Graph-based CNNs for Traffic Flow Prediction	A Hybrid Deep Learning Framework for Long-Term Traffic Flow Prediction	Deep Convolutional Mesh RNN for Urban Traffic Passenger Flows Prediction
Authors	Kenchen, Fei Chen, Baisheng Lai, Zhongming Jin, Yong Liu, Kai Li, Long Wei, Pengfei Wang, Yandong Tang, Jianqiang Huang, Xian-Sheng Hua	Yiqun Li And Guibin Wang, Songjian Chai, Zhengwei Ma	Zhene Zou, Hao Peng, Lin Liu, Guixi Xiong, Bowen Du, Md Zakirul Alam Bhuiyan, Yuntao Long, DaLi
Year of publication	2020	2021	2018
Architecture of Deep Learning (including the number of layers, types of layers, activation functions, and any unique features)	Number of Layers: 3 Types of Layers: 1) STC layers 2) ReLU layer 3) fully connected layers Activation Functions: 1) ReLU (Rectified Linear Unit) Unique Features: 1)In spatial convolution of STC layer, the order K of polynomial approximation is set to be 5	Number of Layers: 1) Wavelet Transform Layer (DWT) 2) Convolutional Layers 3) LSTM Layers Types of Layers: 1) Discrete Wavelet Transform (DWT) 2) Convolutional Layers 3) Activation Layers (ReLU) 4) LSTM Cells (with forget, input, output gates, and cell state) Activation Functions: 1) ReLU (Rectified Linear Unit) 2) Sigmoid (σ) 3) Tanh (hyperbolic tangent) Unique Features: Integration of Discrete Wavelet Transform (DWT) Combination of CNN and LSTM Hierarchical architecture	Layers: 1) Recurrent Neural Networks (RNNs) 2) Convolutional Neural Networks (CNNs) 3) Graph Convolutional Networks (GCNs) Activation Functions: 1) ReLU (Rectified Linear Unit) 2) Sigmoid (σ) 3) Tanh (hyperbolic tangent) Unique Features: Residual connections (e.g., in ST-ResNet), integration of external factors like weather and holidays. Training: Techniques like backpropagation through time (BPTT) for RNNs.
Training procedures (e.g, training strategy, including optimization algorithms, learning rates,	Feature Engineering: 1) STC Layers: i) Spatial Convolution : Traffic data is spatially filtered to identify spatial features such as	Feature Engineering: Wavelet Transform (DWT): Decomposes time-domain data into low-frequency (CAn) and high-frequency (CD1 to	feature engineering and regression are prominently used in the context of traffic passenger flow prediction

	geographical patterns or regional factors. ii) Temporal Convolution: This layer learns temporal dependencies that help us understand how traffic changes with time. 2) Graph Prediction Stream i) Convolutional Layers : They learn the intricate relationships and interactions within the traffic network. It is particularly useful for capturing the influence of one node on its neighbors, for example, road sections or intersections. 3) Auxiliary Information Encoding i) Fully Connected Layers : Additional information about context can be encoded in these layers to improve prediction accuracy by including parameters such as weather, holidays, and other external factors that may influence traffic. Regression i) ReLU Layers: This allows the model to learn more sophisticated mappings from input features to output predictions since it introduces non-linearity in the model. They produce only non-negative outputs which are important since negative flow does not make sense in case of traffic data. Training Schedule: Pre-training: The pre-training of the graph learning sub-network is assigned a	CDn) components. Convolutional Layers: Extract spatial features from the input data using filters. LSTM Network: Handles long-term temporal dependencies in time-series data. Regression: Task: Predicts traffic flow values (continuous numerical values). CNN-LSTM: Combines spatial feature extraction (via CNN) with temporal feature learning (via LSTM) to make accurate traffic flow predictions.	

	dynamic (or learning) e schedule in the entire framework for the first 10 epochs giving the following time interval 4,6,8,10 epochs respectively or by the schedule: 10,000 epochs or 5.2 days (as the last splitter). Joint Training: Then the whole model is trained jointly for 100 epochs. Optimization Algorithm: The model uses Stochastic Gradient Descent (SGD) with momentum. Momentum helps accelerate gradients vectors in the right directions, leading to faster converging. Initial Learning Rate: For the first 50 epochs, the learning rate is set to 10−2 (0.01). Reduced Learning Rate: For the last 50 epochs, the learning rate is reduced to 10-3 (0.001).		
batch sizes, and regularization techniques)	The training process uses a batch size of 64. The weight decay in SGD is 0.0005.	The Batch size of 50 and 60.	The Batch size is 32.
Evaluation / Performance metric used	For evaluation below Metrics are used 1) Root Mean Squared Error (RMSE) 2) Mean Absolute Percentage Error (MAPE) 3) Mean Absolute Error (MAE)	For evaluation below Metrics are used 1) Root Mean Squared Error (RMSE) 2) goodness of t (R-Square) 3) Mean Absolute Error (MAE)	For evaluation below Metrics is used 1) Root Mean Squared Error (RMSE)

Name of Dataset used. If a public dataset, provide the URL.	Two public datasets are used 1) METR-LA[19] 2) TaxiBJ[11]	England traffic flow dataset	Beijing Subway Dataset New York Bikes Dataset:
Conclusion: You must end the comparison with a proper conclusion highlighting your observations.			