M2NN: Multi-task Multi-view Neural Network Using Congestion Heatmap Imagery for Incident Prediction on Road Segments

Thanh Tran[a], Dan He[a], Jiwon Kim[a,*] and Mark Hickman[a]

[a] School of Civil Engineering, The University of Queensland, St. Lucia, QLD 4072, Australia
thanh.tran@uq.edu.au
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 28, 2024

Keywords: Traffic incident prediction, Congestion heatmap imagery, Hybrid deep learning

1 INTRODUCTION

The advancement in traffic data collection technologies, including loop detectors and GPS devices, has significantly transformed traffic management systems. These technologies enable the generation of vast data volumes, catalyzing the development of sophisticated machine learning (ML) models tailored for real-time Traffic Incident Prediction (TIP). Traditionally, these models have focused on 'link-level' incident prediction, assessing the risk on specific road segments by analyzing data from upstream and downstream links. Although valuable, such models often overlook broader traffic dynamics that influence incident probabilities on a larger scale. In contrast, 'network-wide' incident prediction models provide a holistic view of risks across a city's traffic network, supporting effective incident management and resource allocation for sub-areas (Tran *et al.*, 2023). However, these models face challenges in handling large data volumes and computational demands, which may limit their predictive accuracy and detail in managing incidents on specific road segments or smaller regions. Recent research has increasingly leveraged ML, especially deep learning and Graph Neural Networks (GNNs), to handle the complex patterns in traffic data effectively. GNNs align with the graph structures of road networks, as demonstrated by Deep Spatio-Temporal Graph Convolutional Network for Traffic Accident Prediction (DST-GCN) for links (Yu *et al.*, 2021). Further, ML researchers (Wang *et al.*, 2021a,b) have employed a grid-based approach to segment traffic networks into 'cells', each representing a specific area, attempting to predict regional risk for these cells. However, these models adopt a grid or image representation of the city network, which does not accurately reflect the real-world interconnected road structures, thereby limiting their practical applicability. Moreover, while the integration of imagery and numeric data has been demonstrated in traffic demand prediction (Wang *et al.*, 2024), it remains largely unexplored for incident prediction. This integration could leverage diverse data sources, such as HERE or Google Congestion Maps, providing broader views on traffic network structures and conditions. Therefore, there is a critical need for models that not only utilize the complementary aspects of imagery (e.g., link congestion heatmap images) and numeric data (e.g., loop detector data) but also predict at a more granular level, such as individual links or sub-areas, rather than abstract grid cells. Such models would provide clearer, more precise predictions for traffic operators, closely aligning with the actual configuration of transportation networks. In response to existing limitations, our research introduces Multi-task Multi-view Neural Networks (M2NN), a novel approach that diverges from grid-based predictions and traditional models reliant solely on numeric data to enhance model applicability across real-world networks. M2NN achieves fine-grained predictions for links and allows for flexible application across diverse sub-areas in the large network. We innovatively integrate sub-area incident prediction as a sub-task within our multi-task learning framework. This design enables M2NN to simultaneously learn and predict incident risks at both the link and sub-area levels, effectively capturing localized patterns and providing a comprehensive risk overview by leveraging complementary link-level and sub-area-level data. Also, M2NN incorporates congestion heatmap imagery of links alongside numeric data, such as loop detector data and information vectors—a synergy not extensively explored in current research. This multi-view approach enriches the encoding process, enhancing our model's predictive performance. Preliminary results with diverse real-world data sets demonstrate M2NN's enhanced fine-grained predictive capability, aligning the complementary use of imagery and numeric data in improving traffic demand analysis (Wang *et al.*, 2024) and expanding the use of map images (e.g., satellite images) for TIP models.

2 METHODOLOGY

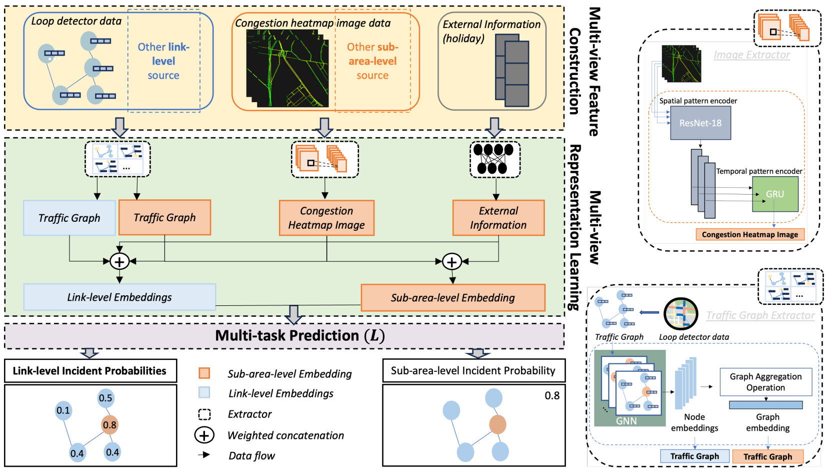
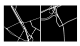
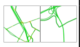


Figure 1 – *Multi-task Multi-view Neural Networks*

Link-level Incident Prediction Definition: Consider a set S^c encompassing all sub-areas within a radius c of a given study network. We have m distinct link-level data sources, each representing unique subgraphs for every sub-area. Our multi-view multi-task traffic incident prediction model uses a set of subgraphs $\mathcal{G}_t^{s_j} = G_1^{s_j}, G_2^{s_j}, \ldots, G_{m\,t}^{s_j}$ ($s_j \in S^c$) associated with the sub-area s_j at the current timestamp t, sub-area-level data source, which is a series of map images (such as congestion heatmap or satellite images) $\mathcal{M}_t^{s_j} = M_t^{s_j}, M_{t-1}^{s_j}, \ldots, M_{t-n_t}^{s_j}$, and an external information vector X_t that encodes information like calendar data corresponding to the same timestamp. Our model generates two levels of outputs: 1) binary indicator $y_t^{s_j}$ that is 1 if at least one traffic incident occurs within s_j during time period $[t + T_p, t + T_p + T_d]$ and 0 otherwise; 2) multi-class classifier $p_{t,1}^{s_j}, p_{t,2}^{s_j}, \ldots, p_{t,l}^{s_j}$, each predicting the probability of incidents on the l links within the sub-area s_j, with $p_l^{s_j} = 1$ indicating incident occurs on link j, and 0 otherwise. For the scenario $y_t^{s_j} = 0$, we have $p_{t,1}^{s_j}, p_{t,2}^{s_j}, \ldots, p_{t,l}^{s_j} = \{0, 0, ..., 0\}$. Here, T_p is the prediction horizon that marks the start of the prediction window, and T_d is the window's duration. The input subgraph set $\mathcal{G}_t^{s_j}$, created using time-varying traffic features from the interval $[t - T_b, t]$, and the sequence of congestion heatmap images of links $\mathcal{M}_t^{s_j}$, also derived within the same interval, provide spatial and temporal data. This data is enriched by contextual features like calendar data (e.g., day of the week, public holiday) included in the feature vector $X_t^{s_j}$. Figure 1 illustrates our M2NN model, which comprises three modules: *Multi-view Feature Construction* (**MFC**), *Multi-view Representation Learning*(**MRL**), and *Multi-task Prediction* (**MP**), respectively. In our experiments, **MFC** module assembles cross-level input data from any given sub-area s into structured and unstructured representations: traffic graphs G for links' loop detector data, congestion heatmap images $\mathcal{M}_t^s$ from the HERE platform, each image M_t^s reflecting typical traffic conditions at time t, augmented with vectors for temporal context X_t^s. These multi-aspect representations are then processed by **MRL** module using specialized extractors: *Image Extractor* employs a pre-trained Convolutional Neural Network, i.e., ResNet-18, for capturing spatial embeddings of the sequence of input images M_t^s, and Gated Recurrent Units (GRUs) for sequential encoding of these spatial embeddings; *Traffic Graph Extractor* obtains graph and node ('sub-area-wide' and link) embeddings, via Graph Neural Networks; and *External Information Extractor* encodes X_t^s into embeddings using a Multi-layer Perceptron. Next, **MP** module leverages these multi-aspect embeddings to predict incident risks (probabilities) at both link and sub-area levels, optimizing prediction accuracy via a multi-task strategy that minimizes a joint or overall objective function L. L incorporates the consistency loss, $L_{\text{consistency}}$, crucial in ensuring alignment between link-level and sub-area-level incident predictions. $L_{\text{consistency}}$ is formulated as the mean squared error between the aggregated probabilities of link-level incidents within a specific sub-area and the corresponding sub-area-level incident probability: $L_{consistency} = \left\| \sum_{i \in s} \hat{y}_{n+1,link}^{t_i} - \hat{y}_{n+1,area}^{t_i} \right\|^2$. The overall objective is defined as $L = \frac{1}{B} \sum_{b=1}^{B} (\theta_1 L_{link}^b + \theta_2 L_{area}^b + \theta_3 L_{consistency}^b)$. Here, θ_1, θ_2, and θ_3 are hyperparameters that balance the contributions of link-level, area-level, and consistency losses within the overall objective function, denoted by L. B indicates the total number of training batches. This L within **MP** module allows M2NN to train end-to-end, optimizing the accuracy of link-level predictions. It achieves this by effectively leveraging the complementarity between macroscopic (sub-area-wide) and microscopic (link-specific) perspectives, which enhances incident prediction across a large network.

3 RESULTS

Table 1 – *Performance showing the impact of congestion heatmap imagery & multi-task learning.*

Model	Acc@5%	Acc@10%	MRR	NDCG@10%
M2NN w/o image data	0.4583	0.6406	0.4466	0.5416
M2NN w/ road map image (B/W) 	0.4648	0.6476	0.4377	0.5576
DSTGCN	0.4723	0.6433	0.4443	0.5538
M2NN w/ congestion heatmap images 	0.4948	0.6880	**0.4776**	0.5776
M2NN	**0.5313**	**0.7031**	0.4687	**0.5888**

Effectiveness of M2NN architecture: We employed the MSGNN incident predictor on traffic data from loop detectors in Brisbane and Gold Coast, Australia, supplemented by Traffic Pattern Map images (representing the typical congestion heatmap given a specific time of day and day of week) from the HERE platform. Following the undersampling guidelines from Yu *et al.* (2021), we constructed 1600 cases with half incident and half non-incident cases and a 15-minute prediction horizon (T_p), 5-minute prediction window (T_d) and $T_b = 30$ minutes, distributed in a 7:1:2 train-validation-test split. To assess the model's precision in identifying specific incident links from multiple possibilities in any given sub-areas, we adopted evaluation metrics from Hong *et al.* (2023), each scaled between 0 (poorest performance) and 1 (optimal performance): Top-k Accuracy (Acc@K%) evaluates whether true incidents rank within the top k-percentile of predicted risks, accommodating varying link counts across sub-areas. Mean Reciprocal Rank (MRR) measures the average inverse rank at which actual incidents are predicted, emphasizing the model's ability to prioritize incident links effectively. Normalized Discounted Cumulative Gain (NDCG@K%) quantifies the model's ranking quality, comparing the predictions to an ideal ranking, with an emphasis on the most relevant K-percentile of links, providing an intricate assessment of prediction accuracy in diverse road networks. An ablation study, summarized in Table 1, highlights the M2NN model's dependency on visual data for improved performance. Models without imagery underperformed, validating the hypothesis that visual context significantly enhances prediction accuracy. Incremental improvements were observed with black and white imagery, but the integration of congestion heatmap imagery showed a substantial increase in performance metrics since it provides network-wide traffic conditions along with network structures, which are further complemented by numeric data such as loop detector data. Full M2NN configuration, which integrates congestion heatmap images with multi-task learning, outperformed all while DSTGCN, although moderately better than the M2NN without image data, still fell short, indicating its limited ability to utilize complex contextual data effectively. **How M2NN compares with existing model?** We applied M2NN and DSTGCN to different sub-areas to obtain risk predictions at the time of two actual incidents. In Figure 2, the first scenario (left) demonstrates the M2NN model's nuanced spatial accuracy by aligning its risk prediction with the actual incident location at a complex intersection. In contrast, the DSTGCN model's prediction deviates slightly, highlighting the enhanced spatial discernment of M2NN. The right scenario shows M2NN accurately identifying high risk on a link that is directly connected to the actual incident location, while DSTGCN's prediction is displaced further from the location where the road splits. This suggests a potential underestimation of environmental and contextual risk factors, such as road structures, by DSTGCN, which M2NN effectively considers via its multi-view and multi-task learning modules. These scenarios demonstrate M2NN's superior spatial accuracy, attributable to its integrated multi-view and multi-task learning modules, which effectively synthesize link-specific and broader area-wide data, significantly improving upon DSTGCN's approach that lacks such integration.

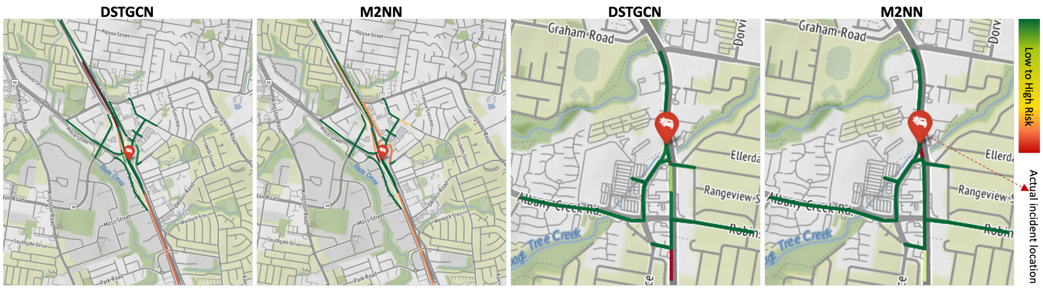


Figure 2 – *Link-level risk predictions comparison regardless of sub-area networks.*

4 DISCUSSION & FUTURE RESEARCH

Our results demonstrate that M2NN surpasses existing approaches in predicting traffic incidents on network links through a multi-task learning approach that integrates multi-level data sources. This method enhances prediction accuracy by enabling coordinated predictions across link and sub-area levels. It opens avenues for further research into synergistic prediction tasks, such as incident duration and severity, highlighting the benefits of combining numeric and imagery data.

References

Hong, Ye, Zhang, Yatao, Schindler, Konrad, & Raubal, Martin. 2023. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C.*

Tran, Thanh, *et al.* 2023. MSGNN: A Multi-structured Graph Neural Network model for real-time incident prediction in large traffic networks. *Transportation Research Part C: Emerging Technologies.*

Wang, Beibei, *et al.* 2021a. GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting. *In: AAAI Conference on Artificial Intelligence.*

Wang, Qingyi, *et al.* 2024. Deep hybrid model with satellite imagery: How to combine demand modeling and computer vision for travel behavior analysis? *Transportation Research Part B: Methodological.*

Wang, Senzhang, *et al.* 2021b. Traffic Accident Risk Prediction via Multi-View Multi-Task Spatio-Temporal Networks. *IEEE Transactions on Knowledge and Data Engineering.*

Yu, Le, Du, Bowen, Hu, Xiao, Sun, Leilei, Han, Liangzhe, & Lv, Weifeng. 2021. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing.*