Traffic accident prediction via three-dimensional convolution autoencoder and victim-party demographic data

Walden Ip[a], Grigorev Artur[a], Mihaita Adriana-Simona[a]

[a] University of Technology Sydney, Australia>
waldenyukyin.ip@uts.edu.au, waldenyukyin.ip@uts.edu.au, adriana-simona.mihaita@uts.edu.au,

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

April 14, 2024

Keywords: accident prediction, spatial-temporal modelling, artificial intelligence, deep learning

1 INTRODUCTION

Background and challenges: Predicting how many accidents will occur in the future is of critical importance for traffic management centres for planning and timed interventions to save human lives. Several attempts have been made to predict the likelihood of crashes (Xu *et al.*, 2013), the severity of the incident Mihaita *et al.* (2019) or its duration Grigorev *et al.* (2024). Many of past studies have utilised restricted data sets (mainly traffic logs) due to difficulty in accessing multiple complementary data sources. Often, various features to be used for traffic accident prediction are of spatial versus temporal nature and difficult to model together. Some studies have used deep learning approaches to address the complex nature of traffic accident causation, including GSNets by Wang *et al.* (2021) and HeteroConvLSTM by Yuan *et al.* (2018) which have used a combination of road network information, environmental characteristics, historical collision records, and traffic states. However, as the spatio-temporal resolution of the data sets increases, zero-inflated issues can occur; often the human component (such as victim data) is not considered either, which makes the problem of traffic accident prediction very challenging.

Objective: This paper proposes a new spatial-temporal framework to predict the number of accidents in an urban area, by integrating multiple heterogeneous data sets such as road structure, intersection density, weather, historical collisions, traffic information and victim and party demographics data. **Firstly** we collect all data sets for the area of Los Angeles which is our case study. **Secondly**, we propose a deep learning framework and efficient pipeline based on a three-dimensional convolutional autoencoder (3DCAE) that determines the optimal size and sequential combinations of input for predicting the risk level of potential accidents in a subarea at a given point in time. The framework is further improved based on victim and demographic party data, which indicates that the age of 23 is common among victims (mostly younf between 18-35 years old). **Thirdly**, we show that our proposed architecture outperforms other baseline models such as SDAE (stacked denoising autoencoder), MLP (Multilayer Perceptron) and ConvLSTM (convolutional long short-term memory network) in terms of root mean square error and symmetrical mean absolute percentage error.

2 CASE STUDY

The main area selected for this study is Los Angeles (district 7 of California), as shown in Fig. 1a) for which the following data sets have been collected:

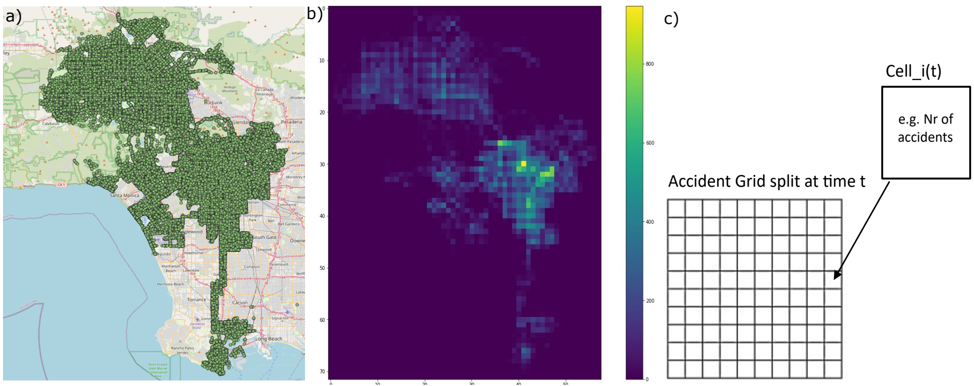

Figure 1 – *Incident records across Los Angeles Map, b) Grid Heatmap based on number of accidents per cell, c) area grid split into cells at time i.*

1. Traffic Accident Data [temporal data; 93 features]: was collected from the State-wide Integrated Traffic Records System (SWITRS) under the California Highway Patrol division; it ranges from 2014 to 2019 and contains several features such as time, location, collision type, severity, reason for collision and the road condition.

2. Weather Data-temporal data [5 features] was collected from the National Weather Service through the Synoptic Data web service on a daily basis which contain features of precipitation, wind speed, wind direction, air temperature and visibility. Four key weather stations have been selected at the following locations; Van Nuys Airport, Santa Monica, Pasadena and Long Beach.

3. Traffic Speed and Flow Data [temporal data, 5 features] have been collected from the Caltrans Performance Measurement System (PeMS) Data Clearinghouse under the California Department of Transportation. It also contains metadata on the vehicle detector station locations with each detector being placed across the freeway network of the state of California.

4. Road Structure Data-spatial data has been collected from LA GeoHub and this data set represents the spatial component of our modelling, including features such as: street centreline (25 features), intersections (12 features), and freeway exit locations (31 features).

5. Victim and Party Data [temporal data, 13 features] is a rare to find data set which was also collected from SWITRS and complements the traffic accident data by common case ID. It contains information on individual victims of the collision cases such as age, role and degree of injury, as well as the total number of parties in a collision case, number of injured, fatality rate and movement preceding a collision.

Data Preparation: The data sets have underwent an extensive process of cleaning and formatting including anomaly detection, outlier and duplicate removal, etc. In order to aggregate features into given subsections of the Los Angeles study area, it was necessary to first represent the study area as a grid of i rows x j columns of cells which were bounded by latitude and longitude (see Fig. 1c), with each cell being set to a 1km x 1km size. The 1kmx1km size was selected after we conducted a sensitivity analysis and comparing sparsity levels on different cell sizes ranging from 200mx200m up to 2kmx2km(to be included in full paper upon acceptance); this resulted in the area being fragmented in i=72 rows and j=58 columns. Due to the nature of traffic accident risk prediction being unbalanced and creating the zero-inflated issue, a dynamic masking of only non-zero accident risk cells for both predicted and ground truth were used to alleviate the artificial reduction of calculated error. We provide a heatmap example in Fig. 1b) when using the 1kmx1km grid split and colouring by the number of accidents per grid at time t.

3 METHODOLOGY

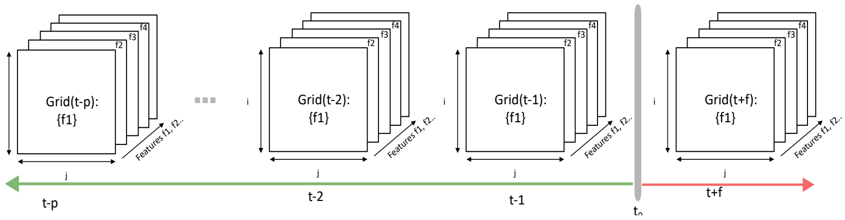

Figure 2 – *4D tensor representation of model input and output across time*

Feature generation: Given the data set preparation and the grid representation, the problem definition that we want to solve is to be able to predict the total number of accidents that will occur at a specific location in the city, after a number of days into the future, by leveraging all historical information in space and time; this is modelled as an image regression task, with a first step being the feature preparation. A grid (split into 1kmx1km cells) is then characterised by a set of features f_k, $k \in \{1...N\}$ at a specific moment in time $t \in \{1...2191\}$ (as we have 2191 days in our data set). Fig. 2 shows an example of a stacked features per grid which we call a 3D tensor representation of the total number of incidents per grid at a time t_0 $(Grid(t_0)\{f_1, f_2, ...f_N\})$. The time component is split from the current time t_0 in a past horizon $(t-1, t-2..., t-p)$ and a future horizon $(t+1, t+2, ...t+f)$, where p is the size of the training horizon that will be fed into the model architecture (how many days from the past) and f the future horizon at which we are interested to predict (how many days into the future). The entire feature data set is then split in a $60\% - 20\% - 20\%$ for training, validation and testing.

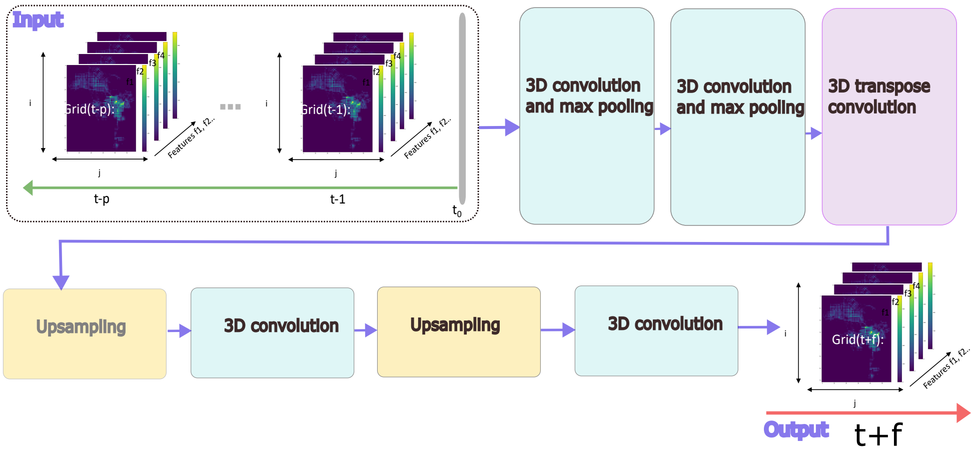

Figure 3 – *High-level overview of the 3DimCAE architecture.*

Model architecture and implementation: Using the above set-up we propose a new framework using a three dimensional convolutional autoencoder (3DimCAE) which was trained using all spatial-temporal features. Fig. 3 represents the model architecture consisting of a series of three dimensional convolution and max pooling layers, a three dimensional transpose convolution, and a series of upsampling and three dimensional convolution layers to restore to the required output size. With regard to model training, all models used a batch size of 16 and the Adam optimiser was used with a learning rate of 0.001 over 15 epochs. We further performed hyper-parameter tuning, optimisation and regularisation during model training to improve prediction performance. The performance has been evaluated by using the root mean squared error (RMSE) and symmetric mean absolute percentage error (sMAPE) against three other baseline competitive models such as: a) a stacked denoising autoencoder (SDAE) as per (Chen *et al.*, 2016), b) a Multilayer Perceptron with RelU activation function (MLP) as per Popescu *et al.* (2009) and c) a convolutional long short-term memory network, as per Shi *et al.* (2015).

4 RESULTS

Table 1 – *Accident risk prediction performance comparison.*

Victim/Party Demographic Data	Yes		No	
Model	RMSE	MAPE	RMSE	MAPE
MLP	0.04953	12.48%	0.05090	12.62%
SDAE	0.03128	12.46%	0.03149	12.61%
ConvLSTM	0.00136	12.49%	0.00137	12.64%
3DimCAE	0.00018	12.46%	0.00028	12.62%

Victim and part demographics data indicated that from all victims 61.4% were pedestrians; majority of them were aged between 18-35 with an average age of 23 at the time of the accident. Such a data set represents a strong statement in how regular road users are extremely vulnerable to accidents, therefore our experimental set-up included tests with and without this data set to outlien its importance.

Table 1 presents the prediction results of the proposed framework using 3DimCAE agains all baseline models, evaluated against RMSE and MAPE on a setup of p=14 days in the past and f=3 days into the future (the best scenario set-up in terms of performance). Our results are two fold: a) we show that the 3DImCAE framework outperforms the other models in terms of both RMSE (0.00018 against 0.04953 for MLP) and SMAPE (12.46% against 12.49% for ConvLSTM) and b) using the victim data improve the performance of all models. **Observation:** Although the dynamic masking we have previously deployed in data preparation mitigates the impact of the zero-inflated issue, resulting in artificially lowered error rates, the disparities in RMSE and sMAPE scores between models with and without demographic data are minimal. This phenomenon is primarily attributable to the sparse occurrence of daily traffic accidents, which leads to small prediction numbers and contributes to the zero-inflated issue to a lesser extent.

Future directions: Leveraging demographic information aids in identifying high-risk areas and provides additional context for risk prediction, extending current methodologies to address the irregular and sparse nature of traffic accidents. The methodology demonstrated herein, tailored to the Los Angeles study area, offers potential applicability to larger geographic regions, augmenting traditional spatial-temporal and historical accident data for more generalized risk prediction beyond specific locales.

References

Chen, Quanjun, Song, Xuan, Yamada, Harutoshi, & Shibasaki, Ryosuke. 2016. Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**(2), 338–344.

Grigorev, Artur, Mihăiţă, Adriana-Simona, Saleh, Khaled, & Chen, Fang. 2024. Automatic Accident Detection, Segmentation and Duration Prediction Using Machine Learning. *IEEE Transactions on Intelligent Transportation Systems*, **25**(2), 1547–1568.

Mihaita, A.S., Liu, Z., Cai, C., & Rizoiu, M.A. 2019. Arterial incident duration prediction using a bi-level framework of extreme gradient-tree boosting. *Proc. of ITS World Congress (ITSWC 2019), Singapore*, Oct.

Popescu, Marius-Constantin, Balas, Valentina, Perescu-Popescu, Liliana, & Mastorakis, Nikos. 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, **8**(07).

Shi, Xingjian, Chen, Zhourong, Wang, Hao, Yeung, Dit-Yan, Wong, Wai-Kin, Woo, Wang-Chun, & Observatory, Hong Kong. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in Neural Information Processing Systems*, **28**.

Wang, Beibei, Lin, Youfang, Guo, Shengnan, & Wan, Huaiyu. 2021. GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**(5), 4402–4409.

Xu, Chengcheng, Wang, Wei, & Liu, Pan. 2013. Identifying crash-prone traffic conditions under different weather on freeways. *Journal of Safety Research*, **46**(9), 135–144.

Yuan, Zhuoning, Zhou, Xun, & Yang, Tianbao. 2018. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data. *Knowledge Discovery and Data Mining*, **18**(7), 984–992.