Work Zone Crash Occurrence Prediction based on Planning Stage Work Zone Configurations Using an Artificial Neural Network

TRBAM-22-03731 *Yang Cheng, Keshu Wu, Hanchu Li, Steven T. Parker, Bin Ran, and David A. Noyce*

College of Engineering
UNIVERSITY OF WISCONSIN–MADISON

Motivation

This study proposes an artificial neural network (ANN) based approach to predict the crash occurrence in work zones only using work zone configurations and operational parameters. The goal is to explore whether using simple work zone configuration features available at the planning stage as the input can achieve satisfying work zone crash prediction.

Key Findings

- The proposed approach is able to predict the crash occurrence in work zones based on work zone configurations and operational parameter.
- The proposed approach can provide designers and decision-makers with quick work zone safety evaluation for all feasible work zone configuration and scheduling alternatives and suggest whether extra resources and attention are needed to reduce potential work zone crashes.
- Lane configuration selection plays an important role in work zone safety.

Data Preparation

The data sets used in this study include the statewide detailed work zone records and the crash records in Wisconsin from 2009 to 2020:

- **Wisconsin Lane Closure System (WisLCS):** a comprehensive management and reporting system for lane closures and restrictions on highways.
- **Wisconsin Crash Database:** information on all police-reported crashes in Wisconsin. The Wisconsin DT4000 police report also has a "construction zone flag" to indicate whether a crash occurred in a work zone.

Table 1 Prediction Model Inputs

Feature	Details
Closure Type	Construction, Maintenance, Permit, Special Event, Emergency
Schedule Type	Long Term, Continuous, Weekly, Daily/Nightly
Facility Type	Bridge, Mainline, Ramp, System Interchange
Lane Closure Details	Full Closure, 1/2/3 Left Lanes Closed, 1/2/3 Right Lanes Closed, Flagging Operation, Lane Restriction, Left/Right Shoulder Closed, Median Turn Lane Closed, Off Roadway Left, Off Roadway Right, Passing Lane Closed, Various Lanes Closed
Interval	The time interval of the work zone in days
Length	The length of the work zone in miles

Network Architecture

An ANN-based model is developed to predict whether a crash may happen at a given work zone with specific work zone settings. The input of the ANN model is

$$X = [x_1, x_2, x_3, x_4, x_5, x_6]$$

where

$$x_1: CLOSURE\_TYPE,\ x_2: F\_DURATION,$$
$$x_3: F\_FAC\_TYPE,\ x_4: F\_TRAFFIC\_IMPACT,$$
$$x_5: INTERVAL,\ x_6: LENGTH$$

The output of the model is

$$Y = [y]$$

where

$$y: HAS\_ACC$$

Denoting the model as h, the indicator of the crash can be predicted using the model $\hat{Y} = h(X)$.

The neural network has one input layer containing 6 neurons, two hidden layers containing 200 neurons, and one output layer containing 1 neuron. The results of the two output neurons represent the probabilities of the two outputs 0 and 1. The one with a higher probability will be considered as the final output of the network.

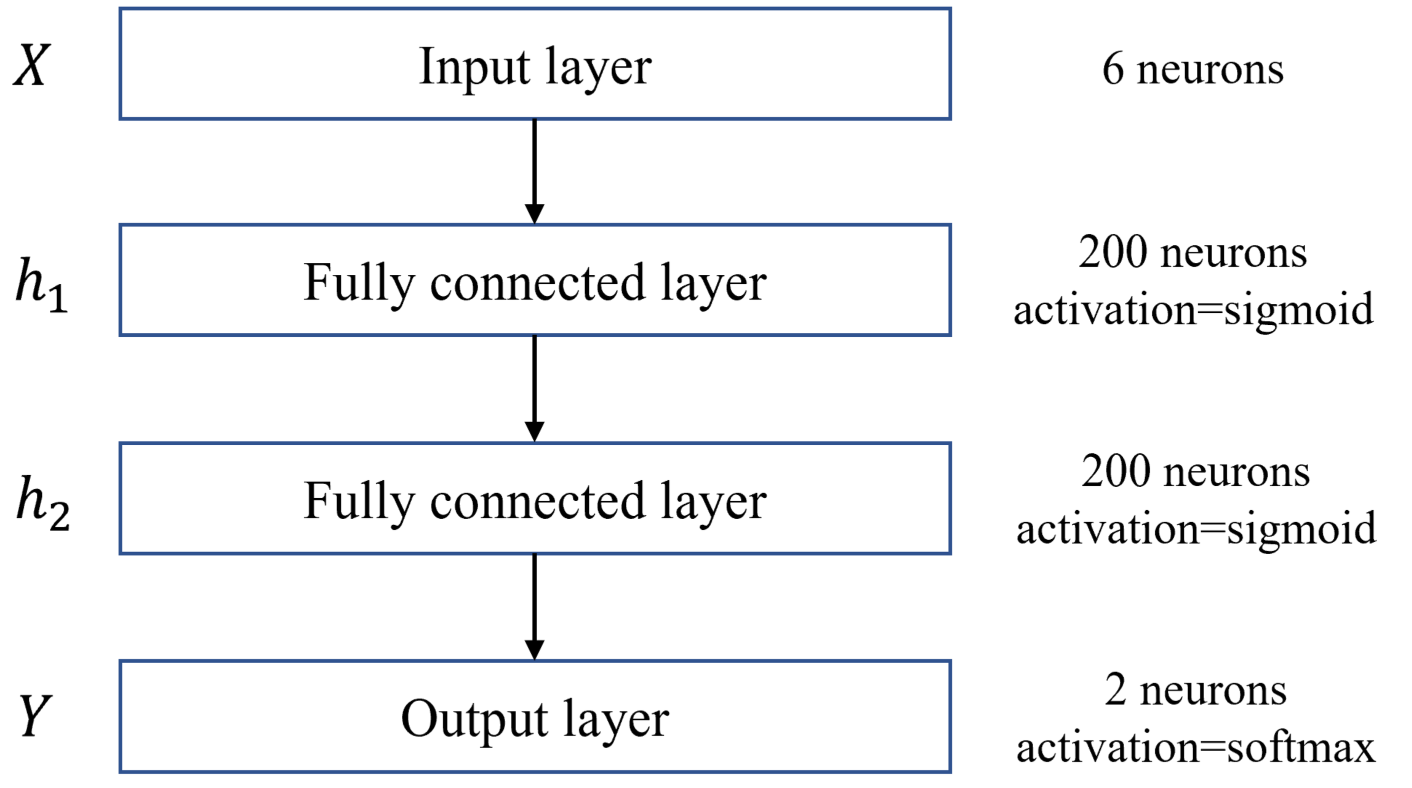

Figure 1 ANN Architecture

Model Training

The overall dataset is randomly split into the training set (70%) and the testing set (30%). 80% of the training set is used for training, and the rest 20% is used as the validation set. The model was trained for 300 epochs with a learning rate of 0.005. Both the accuracy and the loss of the training set and validation set are very close, indicating the model is neither overfitting nor underfitting.

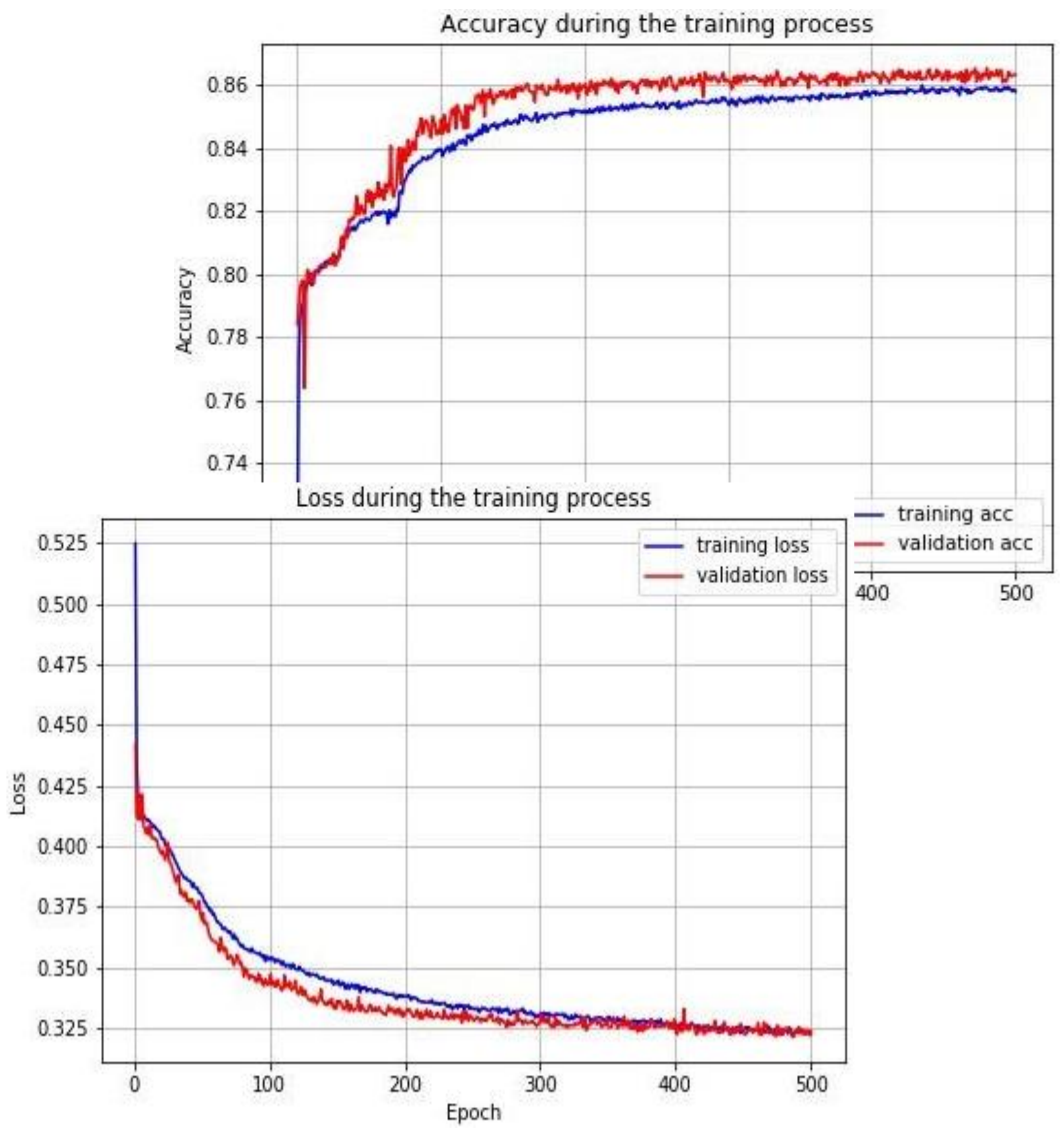

Figure 2 Training and Validation Accuracy and Loss

Prediction Result Analysis

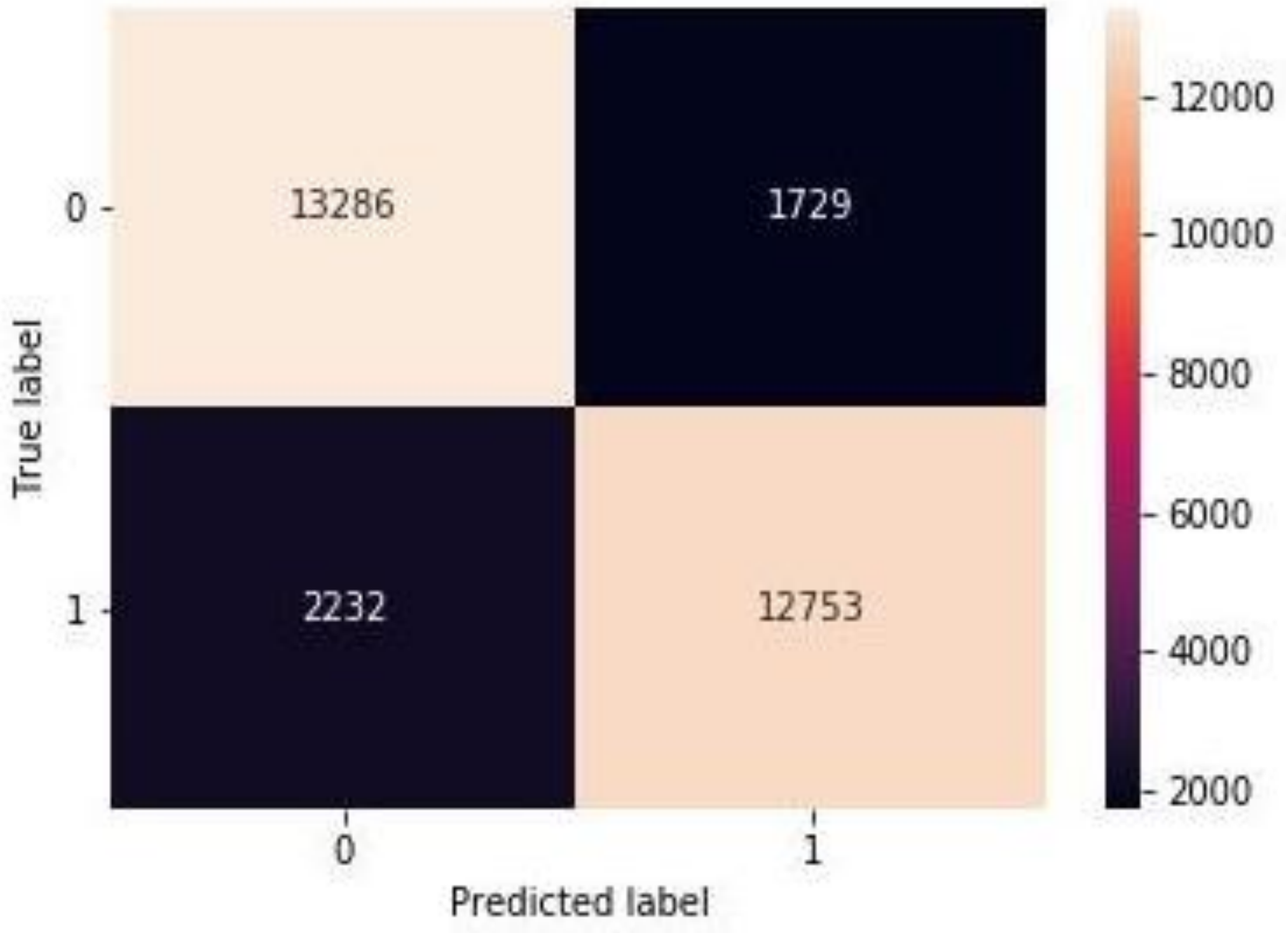

Figure 4 Confusion Matrix

$$Precision = \frac{TP}{TP+FP} = \frac{12753}{12753+1729} = 88.06\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{12753}{12753+2232} = 85.11\%$$

The overall prediction accuracy is 86.79%.

The Support Vector Machines (SVM) and the Decision Trees (D.T.) were used as benchmarks with different sizes of randomly selected data in five independent runs. The proposed approach consistently performed the best.

Table 2 Comparison between different models

Models	Accuracy (%)		
	Data size = 60000	Data size = 100000	Data size = 200000
SVM	79.06	79.00	79.01
Decision Trees	84.51	84.65	84.88
ANN	86.28	86.67	86.83

Discussion and Conclusions

This study shows that work zone crashes are associated with and can be predicted by the work zone configurations, in addition to the work length/period. Work zone engineers would be able to select a safer work zone configuration even at the planning stage by optimizing the two variables. In addition, a number of topics are worthy of mention regarding work zone crash risk prediction in general.

- **Work Zone Crash Identification.** The construction zone indicator from the crash report is generally the only data source to identify work zone crashes. However, it is not accurate enough to be the ground truth. This study uses space-time criteria to identify work zone crashes. The impact, although insignificant, will be quantified in our future work.
- **Work Zone Sample Distribution.** The distribution of work zones across different attributes are not even, which leads to a skewed distribution of work zone crashes. Two derived features, the length and time interval, are able to reduce the impacts.
- **Relationship between Work Zones.** Multiple crashes could occur in one work zone. One work zone without any crashes would only produce one record in the dataset. Those factors lead to the imbalance between "positive" and "negative" samples in the integrated dataset. This study randomly selects the equal number from the two categories to train the model, which turns to be effective in addressing this issue.

Acknowledgement

The Wisconsin Lane Closure System and WisTransPortal Crash Database were developed through sponsorship and collaboration with the Wisconsin Department of Transportation.