CETRA 2026 27-30 May 2026, Bol, Island Brač, Croatia
9th International Conference on Road and Rail Infrastructure

UNDERSTANDING CONGESTION-PRONE ROAD LINKS THROUGH TIPPING POINT DETECTION AND CLASSIFICATION

Zhiyuan Lin[1], Ronghui Liu[1], Ruifan Tang[1], Jake Clarkson[2], Steve Walker[2], Nurettin Cirakli[2]
[1]University of Leeds, Institute for Transport Studies, Leeds, United Kingdom
[2]National Highways, Birmingham, United Kingdom

Abstract

Traffic congestion and delay impose substantial economic costs and are expected to intensify under many future demand scenarios. A traffic tipping point (TP) marks the threshold at which a road link transitions from free flow to a congested regime. Although such transitions can often be identified visually from speed profiles, the scale of modern network data requires automated detection. Building on automated TP labelling using National Highways 15 minute link level flow and speed data, this paper focuses on a downstream task: classifying links by how frequently they experience TPs, using only non-traffic link attributes. TP frequency labels are generated using a crosschecked detection model combining two complementary methods: (i) maximum curvature of a sigmoid fitted to the speed-log (density) relationship and (ii) a kinetic energy criterion based on density times squared speed, together with speed and acceleration rules to identify breakdown and recovery transitions. Using one week of data (1 to 7 April 2022) covering 2, 598 links, we compute each link's TP frequency and convert it into a categorical severity level using K-Means clustering. Thirteen severity classes ('a' to 'm', low to high TP occurrence) are then predicted from eight non-traffic link attributes via a J48 (C4.5) decision tree classifier with 10-fold cross validation in Weka. J48 (C4.5) is a statistical classification algorithm based on decision trees and Weka is a machine learning and data analysis software package. Results show an overall classification accuracy of 76.14% but highly uneven performance across severity levels. The classifier performs strongly for low severity classes ('a' to 'h') yet struggles to discriminate high severity classes ('i' to 'm'), where misclassification becomes close to random. The findings suggest that basic link descriptors can support network wide screening of generally stable links but are insufficient for reliably identifying the most TP prone links without richer features capturing network topology, bottleneck geometry, and operational context.

Keywords: tipping points, traffic congestion, severity classification, decision trees, J48, K-Means clustering

1 Introduction

Recurring congestion on strategic road networks is costly and disruptive, motivating methods that can detect the onset of breakdown and support proactive intervention. National Highways datasets provide high frequency link level measures of traffic flow and speed, creating opportunities for network scale analysis but also requiring automation to be practical. A key concept in this context is the tipping point (TP), the moment when traffic transitions from free flow to congestion, and the corresponding recovery transition back to free flow. Manual identification is feasible for individual road links, but not for thousands of road links over long periods.

DOI: https://doi.org/10.5592/CO/CETRA.2026.1887

This paper concentrates on classifying road links based on TP frequency, using only non-traffic link attributes. The aim is to assess whether readily available descriptors, such as link category, carriageway type, hard shoulder presence, and speed limit, can predict which links frequently experience TPs. Such capability would support prioritization for monitoring and targeted mitigation, even when richer operational features are unavailable.

2 Background: tipping points and automated detection

A tipping point (TP) is treated as a threshold at which a link's operating state changes sharply. In speed time profiles, breakdown TPs typically occur immediately before a sustained drop in speed, while recovery TPs occur when speeds return to a stable free flow level. Figure 1 gives an example of such TPs. Classical fundamental diagrams [6] provide relationships between flow, speed, and density, but in practice these diagrams do not always fit observed data and their theoretical critical points do not necessarily coincide with breakdown and recovery moments.

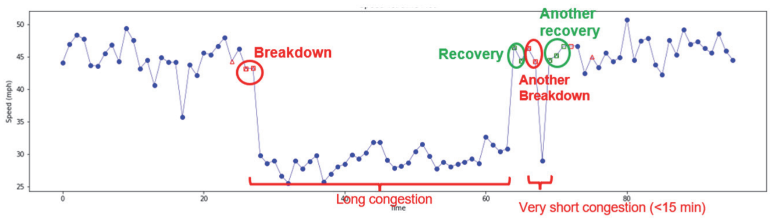

Figure 1 An example of breakdown and recovery tipping points based on a speed-time plot (time axis gives 15-min interval numbers)

To generate labels for the classification task, TPs were detected automatically using a cross-checked rule-based model. Two complementary methods were used. First, a maximum curvature method fitted a sigmoid to the speed-log(density) relationship and identified candidate transitions near the point of highest curvature (such an example is given in figure 2). Second, a kinetic energy method computed and identified candidate transitions associated with high kinetic energy conditions, where and are the traffic density and speed at timestamp , respectively.

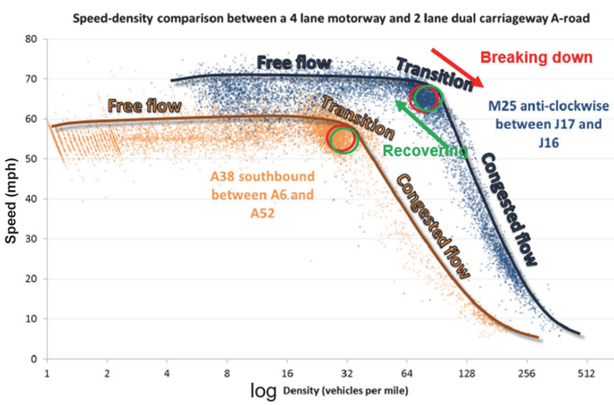

Figure 2 A speed-log(density) diagram. The values on the x-axis still represent the original densities rather than their logarithms (Source: National Highways)

Speed and acceleration rules were applied to align candidates with breakdown and recovery dynamics, and a crosscheck rule retained only events supported by both methods. Detected TPs were then aggregated to compute TP frequency for each link over the study week.

3 Data and severity labelling

The dataset comprised 2, 598 road links observed over one week (1 to 7 April 2022). For each link, the crosschecked TP detection model produced a count of detected TPs over the week. This TP frequency formed the basis for downstream severity labelling and classification. Predictor variables were limited to eight non-traffic link attributes, listed in table 1. Rather than predicting raw TP counts, links were grouped into categorical TP severity levels using K-Means clustering applied to TP [2, 3]. The clustering yielded 13 severity classes, labelled ‘a’ to ‘m’ from low to high TP occurrence. These classes formed training labels for supervised classification.

Table 1 Non-traffic link attributes used for TP severity prediction

Field	Description
Road_Name	Name of the road
Link_Type	Type of road link (e.g. primary, secondary)
Link_Category	Category of the road link (e.g. urban, rural)
Link_Length	Length of the road link
Carriageway	Description of the carriageway
Hard_Shoulder	Presence of a hard shoulder
Speed_Limit	Speed limit of the road link

4 Methods

A J48 classifier, Weka’s implementation of Quinlan’s C4.5 decision tree algorithm, was used because of its interpretability and its ability to handle mixed categorical and numerical predictors [4, 5]. Weka is a machine learning and data analysis software package, and J48 (C4.5) is a statistical classification algorithm based on decision trees. The classifier learnt to predict the probability of each link belonging to each TP severity class using only the non-traffic attributes in table 1. Model training and evaluation used 10-fold cross validation. Each fold trained on nine tenths of the data and tests on the remaining tenth, repeating until all folds have served as test data. Table 2 summarizes the classifier configuration used to support robust performance and to mitigate overfitting through pruning.

Table 2 Parameter setting for J48 tree classifier

Parameter	Value
Test mode	10-fold cross-validation
Batch size	100 instances
Collapse tree	True, to simplify the decision tree by merging nodes where feasible
Confidence factor	0.25, to control the pruning of the decision tree and mitigate overfitting
Minimum number of objects	2, specifying the minimum instances per leaf

Performance was evaluated using overall accuracy and probability based error measures (MAE and RMSE) computed from the classifier's probabilistic outputs. A link was counted as correctly classified when its true class was the class with the highest predicted probability. Class wise precision and recall were also reported to identify where performance varied across severity levels.

5 Results

5.1 Overall predictive performance

Table 3 reports overall performance across all 2, 598 links. The classifier achieved 76.14% accuracy. However, aggregate accuracy masked substantial differences between low and high severity classes.

Table 3 Overall performance metrics

Metric	Value
Correctly classified instances	1978 (76.14%)
Incorrectly classified instances	620 (23.86%)
Mean absolute error (MAE)	0.04, indicating the average magnitude of prediction errors
Root mean squared error (RMSE)	0.1652, measuring the square root of the average squared prediction error
Relative absolute error (RAE)	29.00%, comparing the MAE to the mean absolute deviation
Root Relative squared error (RRSE)	62.92%, comparing the RMSE to the standard deviation of the actual values
Total Number of Instances	2598

5.2 Class wise performance

To diagnose performance across severity levels, table 4 summarizes the primary per class metrics and table 5 reports class wise results. The classifier performed strongly for low severity classes ('a' to 'h') but struggled for high severity classes ('i' to 'm').

Table 4 Primary class wise metrics

Metric	Description
P+ Rate (Recall)	The proportion of actual positives that are predicted positive by the model
P- Rate	The proportion of actual negatives that are incorrectly classified as positives
Precision	The proportion of positive predictions which are correctly classified as positives

Table 5 Class wise accuracy for TP severity levels

Class	P+ Rate (Recall)	P- Rate	Precision
a	0.938	0.042	0.739
b	1.000	0.002	0.982
c	0.989	0.038	0.878
d	1.000	0.002	0.977
e	0.856	0.023	0.714
f	1.000	0.003	0.969
g	1.000	0.000	1.000
h	1.000	0.001	0.960
i	0.310	0.045	0.361
j	0.153	0.034	0.217
k	0.482	0.035	0.488
l	0.230	0.020	0.435
m	0.270	0.021	0.420

5.3 Confusion matrix

The confusion matrix highlighted that a substantial fraction of misclassifications occurred within higher severity levels (‘i’ to ‘m’), where predictions were dispersed across neighboring classes.

```
=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   j   k   l   m   <-- classified as
 272   0   3   0   0   0   0   0   5   2   7   1   0 |   a
   0 218   0   0   0   0   0   0   0   0   0   0   0 |   b
   1   0 560   0   0   0   0   0   3   0   0   2   0 |   c
   0   0   0 167   0   0   0   0   0   0   0   0   0 |   d
   7   0   4   0 137   0   0   0   5   1   3   1   2 |   e
   0   0   0   0   0 220   0   0   0   0   0   0   0 |   f
   0   0   0   0   0   0 118   0   0   0   0   0   0 |   g
   0   0   0   0   0   0   0  48   0   0   0   0   0 |   h
  28   1  24   0  12   1   0   1  61  17  18  17  17 |   i
  17   1  15   0  10   3   0   0  31  23  24  13  13 |   j
  13   2  12   2  17   2   0   0  10  15  80   3  10 |   k
  20   0  11   0  10   0   0   1  34  20  19  37   9 |   l
  10   0   9   2   6   1   0   0  20  28  13  11  37 |   m
```

Figure 3 Confusion matrix for 13 severity classes

6 Discussion

The decision tree achieved reasonable overall accuracy, but performance was highly uneven across severity levels. The high recall observed for classes ‘a’ to ‘h’ indicated that basic link descriptors contained useful information for identifying links that rarely reached tipping conditions. In contrast, the poor recall and precision for classes ‘i’ to ‘m’ suggested that severe TP frequency was driven by factors not captured in the available non-traffic attributes. From a practical perspective, the model was best viewed as a first pass screening tool.

It could help prioritize links that were likely to be low severity, but it should not be relied upon to identify the most TP prone links without additional features. Likely improvements included adding topology and capacity related descriptors (e.g. proximity to junctions, lane drops, merge and diverge structure, and recurrent demand generators) and using modelling approaches that better handled nonlinear interactions and class imbalance. Exploratory experiments were also conducted using Simple Expectation Maximization clustering [1] on non-traffic attributes, which did not yield meaningful clusters, reinforcing the need for richer features and stronger representations of link context.

7 Conclusion

This paper examined whether non-traffic link attributes could predict how frequently road links experienced congestion tipping points. TPs were labelled using a crosschecked automated detection model combining maximum curvature and kinetic energy criteria with speed and acceleration rules. Link TP frequencies over one week were clustered into 13 severity classes and predicted using a J48 decision tree. Overall accuracy reached 76.14%, with strong performance for low severity classes but weak discrimination among high severity classes. The results indicated that readily available link descriptors were informative for broad screening, but insufficient for reliable prediction of severe TP frequency without richer contextual and operational features. Future work should focus on enhanced feature sets and alternative models aimed specifically at improving performance for the most TP prone links.

References

[1] Bilmes, J.A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, International Computer Science Institute, Berkeley, 1998.

[2] Lloyd, S.P.: Least squares quantization in PCM, Bell Labs Technical Report RR 5497, 1957.

[3] MacQueen, J.B.: Some methods for classification and analysis of multivariate observations, Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297, 1967.

[4] Quinlan, R.: C4.5: Programs for machine learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[5] Weka official documentation, https://ml.cms.waikato.ac.nz/weka/documentation.html, 7.3.2025.

[6] Gerlough, D.L., Huber, M.J.: Traffic flow theory: a monograph, 1975.