# Road Danger Prediction - Classic models, AI models and Data Challenges

Presentation by Richard Owen, Craig Smith and George Ursachi, Agilysis, UK

Roundtable on Artificial Intelligence in Road Traffic Crash Prevention
10-12 February 2021

# ROAD DANGER PREDICTION

## CLASSIC MODELS, AI MODELS AND DATA CHALLENGES

RICHARD OWEN, CRAIG SMITH & GEORGE URSACHI

# AGENDA

- Classic models
- AI models
- Data requirements
- New developments in AI – Explainable AI
- New datasets and missing datasets
- Pilot RAPIER

agilysis


agilysis

# AT A GLANCE

GEORGE URSACHI

# CLASSIC MODELS

"PAIR" VARIANCE OF THE DEPENDENT VARIABLE WITH VARIANCE OF THE INPUT VARIABLES

## Advantages

- Very strong and straightforward with "perfect" data
- Expose coefficients for each selected variable
- Expose the decision process for variable inclusion/exclusion
- Allow for interrogation

## Disadvantages

- What can't be "paired" is assigned to constant
- Does not always fit best to the actual situation
- Work with user defined limitations and assumptions
- Sometimes assign (pair) effects "wrongly" to other variables (collinear or co-dependent with missing data or information)

# AI MODELS

ALSO "PAIR" VARIANCE OF THE DEPENDENT VARIABLE WITH VARIANCE OF THE INPUT VARIABLES

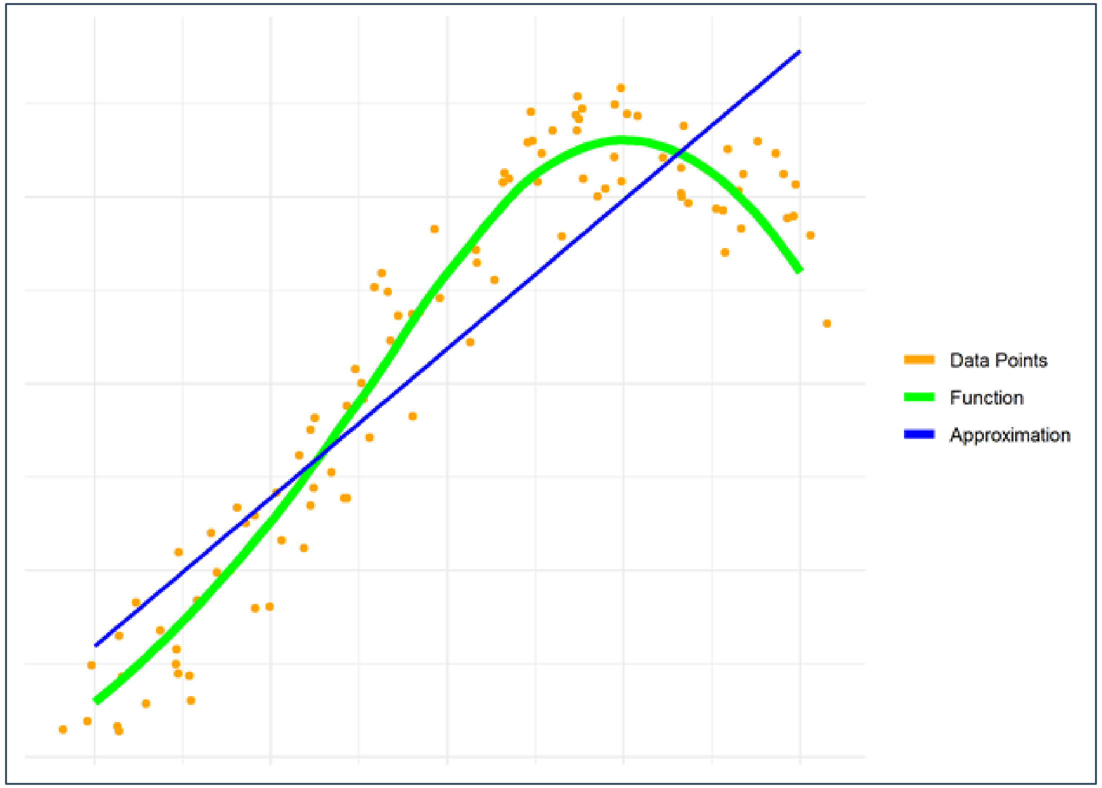
**Advantages**

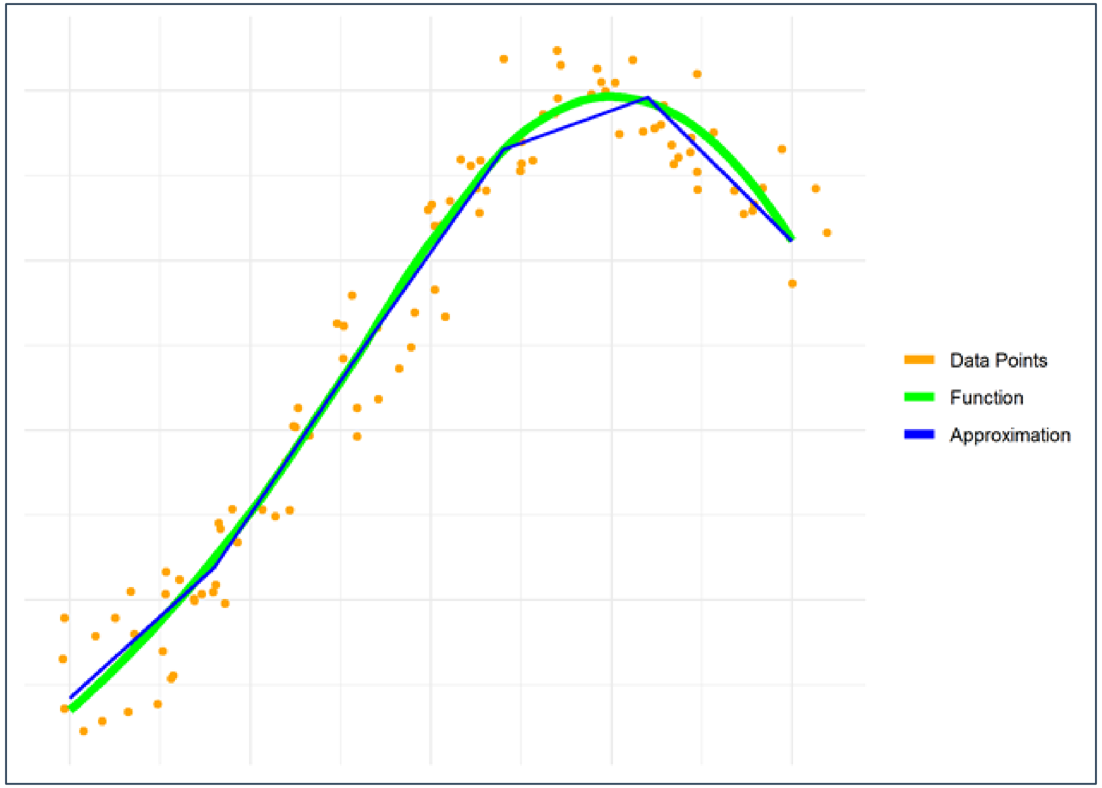
- Usually exhibit better results than classic models
- Find the best fit, without pre-set limitations
- Allow for more interaction between input variables

**Disadvantages**

- Do not expose the coefficients or the decisions processes
- More difficult to interrogate and therefore spot warnings
- Also sometimes assign effects "wrongly" to other variables (collinear or co/inter-dependent)

# CLASSIC MODELS AND AI MODELS



# CLASSIC MODELS AND AI MODELS

agilysis

| Variables | Density | Rate | Density (PC) | Rate (PC) | Density (Ped) |
|---|---|---|---|---|---|
| Dual Carriageway | -0.049 | -0.041 | -0.081 | -0.045 | -0.115 |
| Shared Use Carriageway | -0.081 | -0.102 | -0.178 | -0.195 | -0.163 |
| Carriageway Width | +0.200 | +0.194 | +0.260 | +0.158 | +0.167 |
| Garden | -0.014 | -0.009 | -0.027 | -0.026 | |
| 20 mph Limit | +0.014 | +0.014 | +0.018 | | +0.021 |
| Average Speed | +0.160 | +0.248 | +0.394 | +0.285 | +0.415 |

→ The unpleasant part of using ANN is that you can get more accurate predictions, but you can't understand why, or which variable is influencing the model in what way.

→ With GLM on the other hand, individual variables contribution can be assessed from the model.

***ANN = Artificial Neural Network

***GLM = Generalised Linear Model

# DATA REQUIREMENTS

- Relevant
- Reliable
- To cover the complete picture of variables that can exhibit influence on the dependent variable (driver characteristics?)
- Appropriate granularity
- Clean, no systematic errors

- For both classic models and AI the following is valid:

Garbage in → Garbage out

- Classic models allow sometimes to point out that there is a problem, AI might cover it through fitting better to whatever information it has been fed with

# DATA REQUIREMENTS

- Until we are sure the data fulfils the mentioned requirements at the minimum acceptable at least, AI should not be seen as a better solution, but better as to be used together with classic models
- Data is still the Holy Grail

- First use of AI → to improve data:
  - Quality
  - Quantity (granularity)
  - Availability
- Then use AI together with classic models to improve predictions


# NEW DEVELOPMENTS IN AI

CRAIG SMITH

# EXPLAINABLE AI

- The limitations of AI and its inherent “black box” nature, paired with the increasing use of AI to make significant real-world decisions, leads to a demand for explainability
- Model-Agnostic Explainable AI (XAI) is becoming more popular and more accessible
- Bridges the gap between the more accurate/flexible AI models and the classical models that are more easily interrogated

# EXPLAINABLE AI

- One of the leading techniques is to build local approximations to complex AI models out of simpler, more explainable models
- Can be done by looking at how the AI model reacts to perturbations to the data around specific datapoints, and fitting a local linear model
- Or using game-theoretic Shapley values to assess the local contributions of input variables
- These techniques provide the model “coefficients” that are so valuable in classical models

# EXPLAINABLE AI

- There are also techniques for global explanations of models
- Like building simpler global "surrogate" models
- Or measuring feature importance (how does accuracy change if input variables are dropped)
- Or feature interactions (Friedman's H-statistics)

- XAI is a fast-growing field, so new techniques are constantly being developed.


# NEW DATASETS

RICHARD OWEN

# NEW DATASETS

More data!

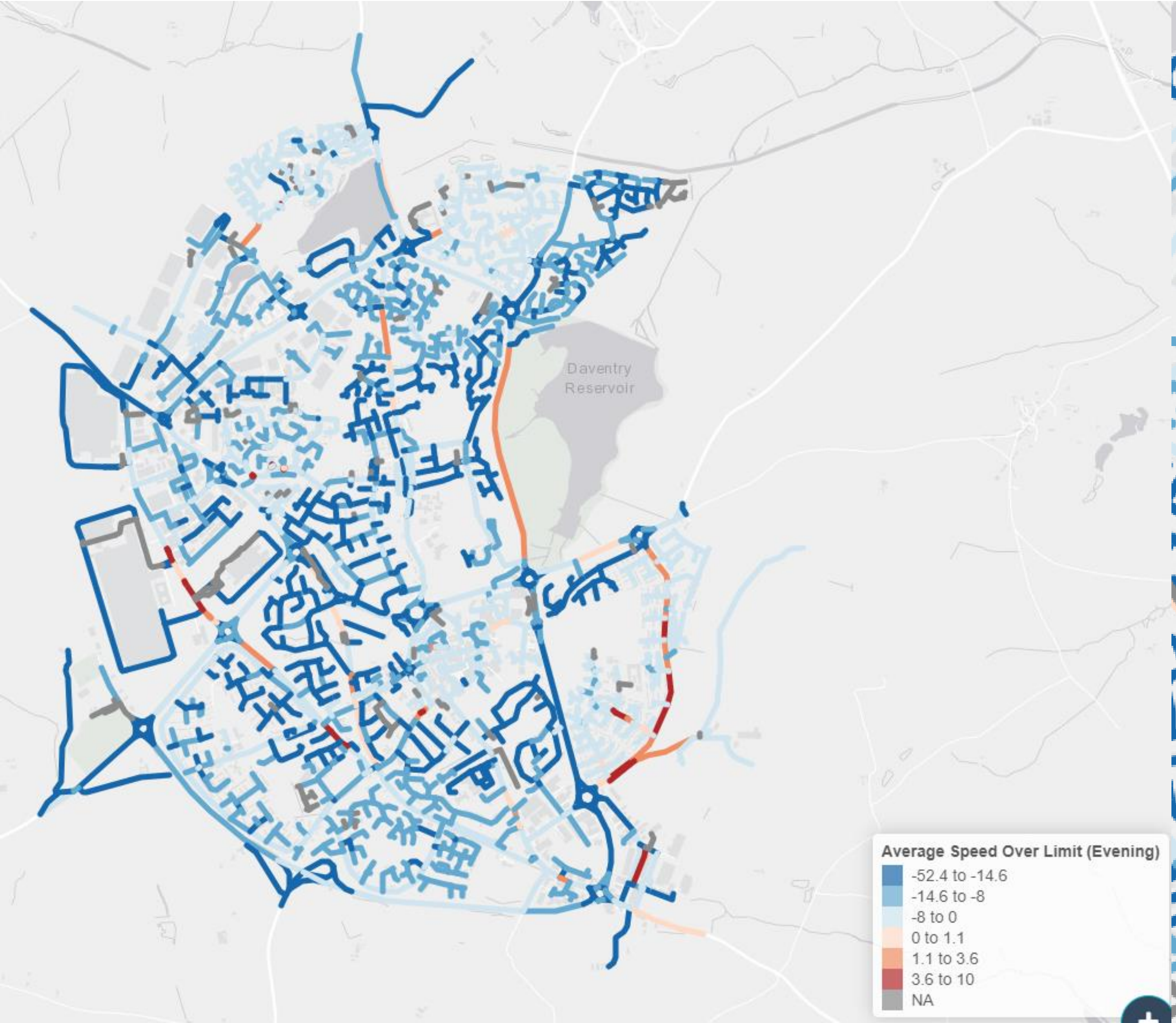
agilysis.co.uk

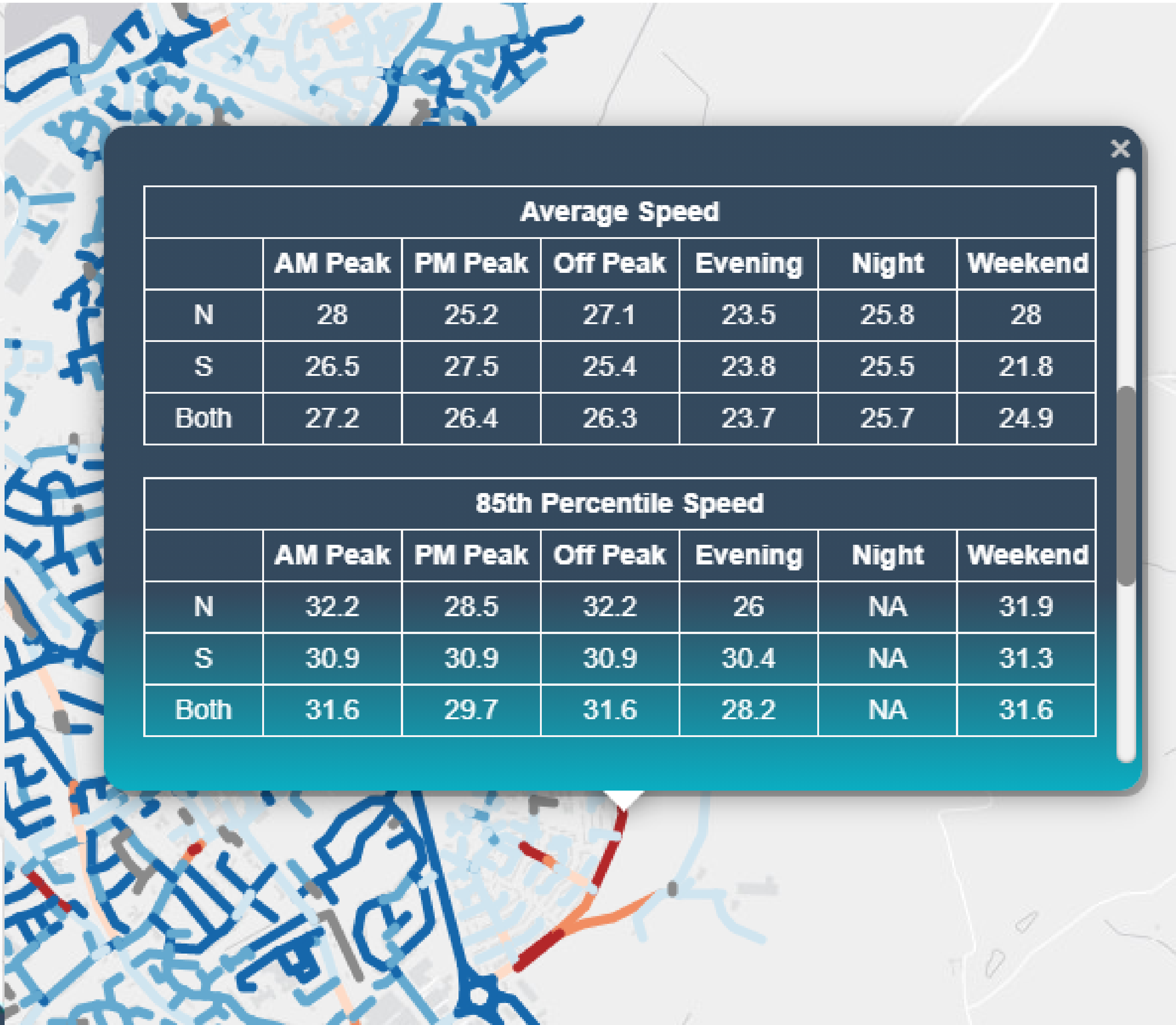
# NEW DATASETS - TELEMATICS

- Global Positioning System (GPS)
  - Vehicle location, speed and movement
- Sensors
  - Capture of data on driver activity, including acceleration, braking and cornering, driver monitoring, vehicle features
- Vehicle diagnostics
  - Post-collision information from sensors

# NEW DATASETS - TELEMATICS

## Speed



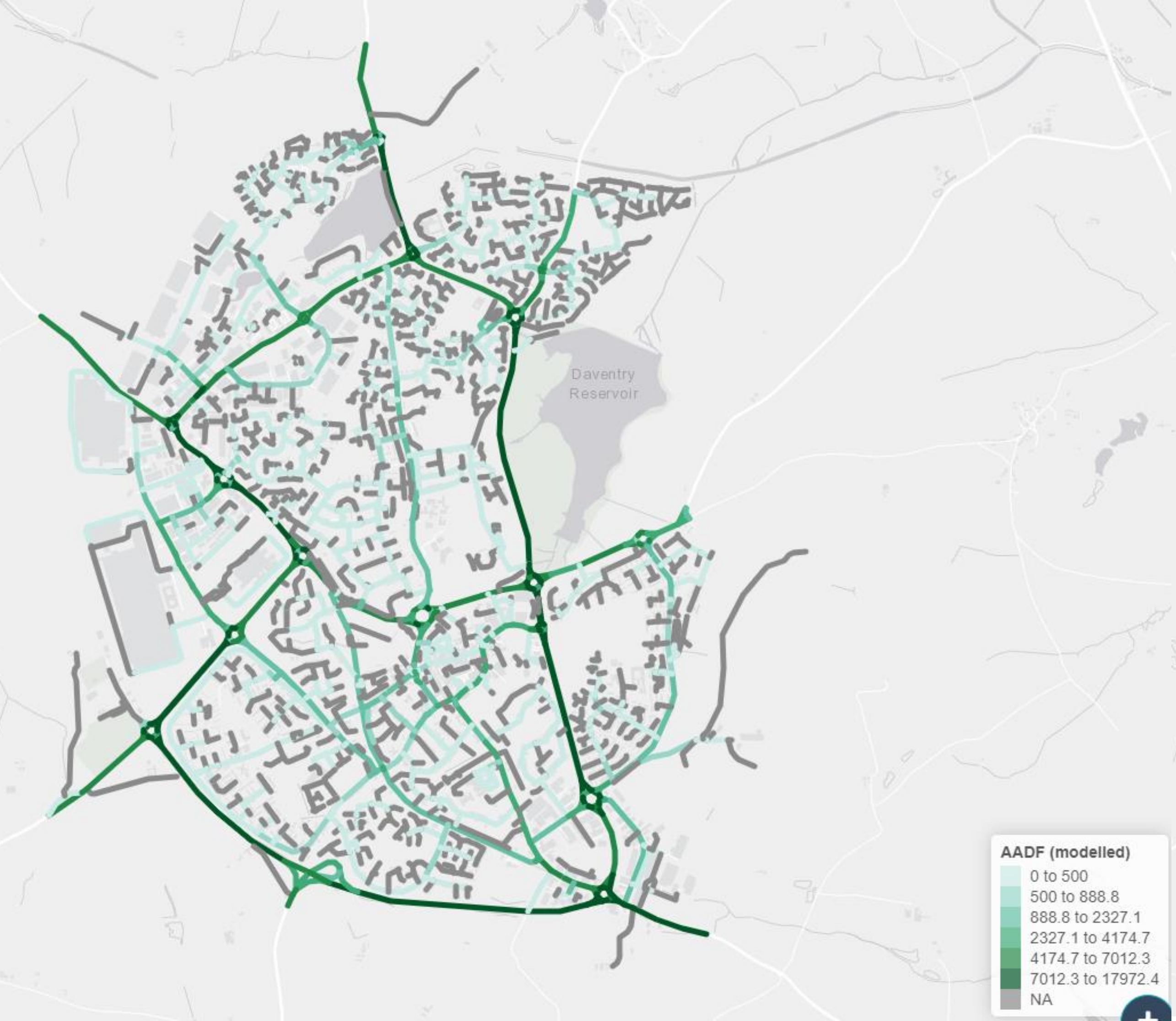
| Average Speed | | | | | | |
|---|---|---|---|---|---|---|
| | AM Peak | PM Peak | Off Peak | Evening | Night | Weekend |
| N | 28 | 25.2 | 27.1 | 23.5 | 25.8 | 28 |
| S | 26.5 | 27.5 | 25.4 | 23.8 | 25.5 | 21.8 |
| Both | 27.2 | 26.4 | 26.3 | 23.7 | 25.7 | 24.9 |

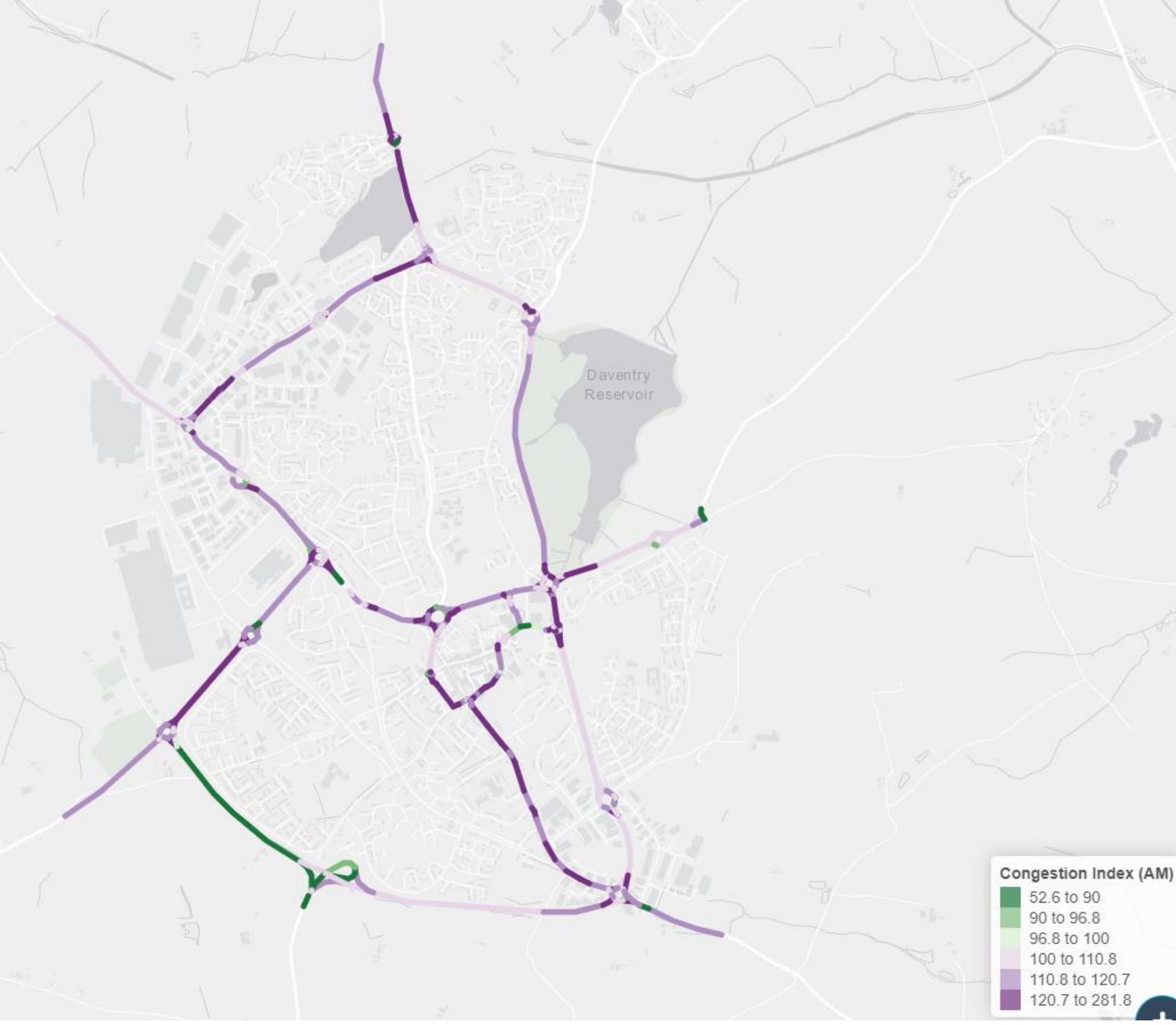
| 85th Percentile Speed | | | | | | |
|---|---|---|---|---|---|---|
| | AM Peak | PM Peak | Off Peak | Evening | Night | Weekend |
| N | 32.2 | 28.5 | 32.2 | 26 | NA | 31.9 |
| S | 30.9 | 30.9 | 30.9 | 30.4 | NA | 31.3 |
| Both | 31.6 | 29.7 | 31.6 | 28.2 | NA | 31.6 |

# NEW DATASETS - TELEMATICS

agilysis

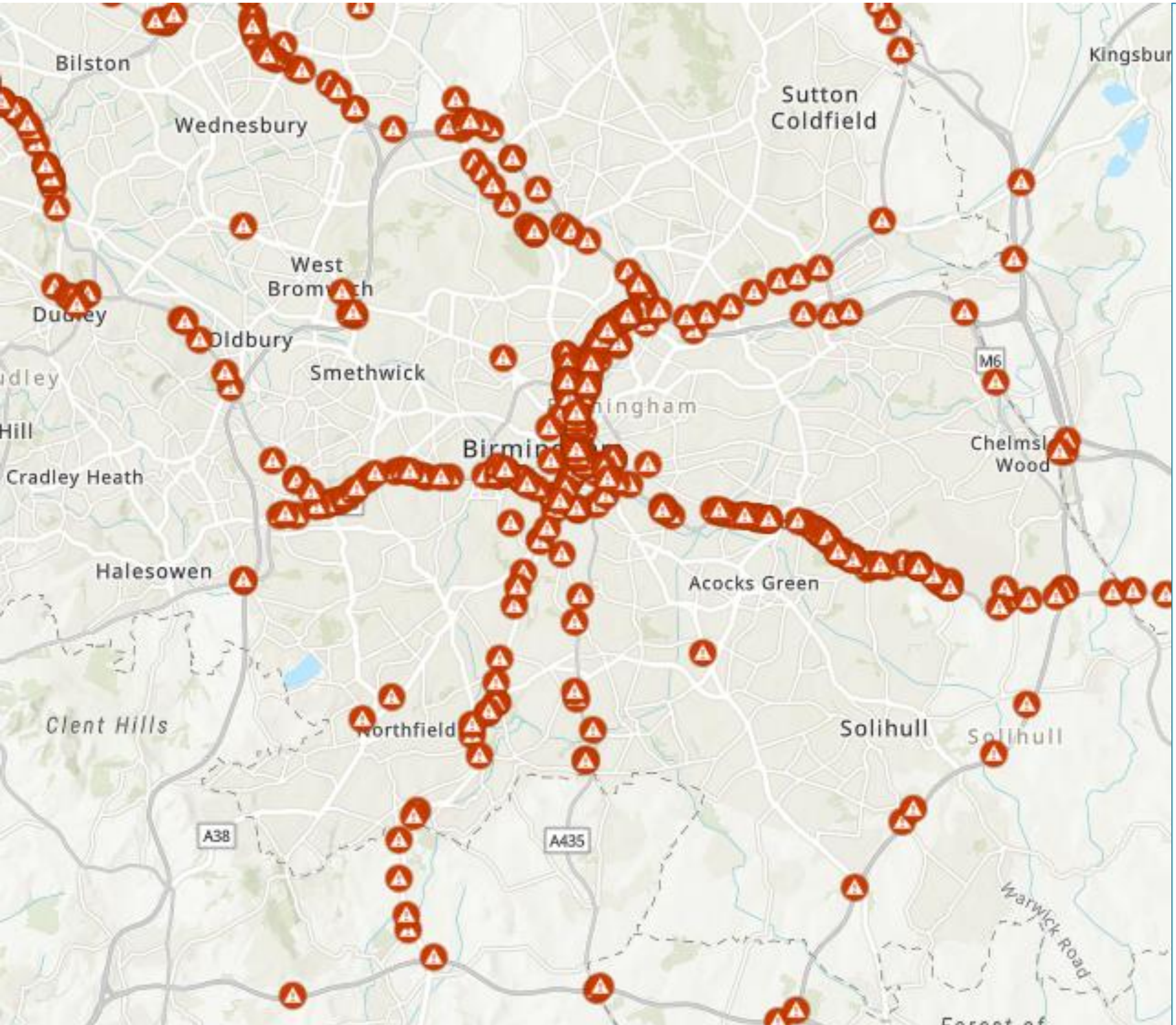
## Flow (modelled)



# NEW DATASETS - TELEMATICS

## Harsh Braking




Collisions
0

Near Collisions
0

Following Distance
3

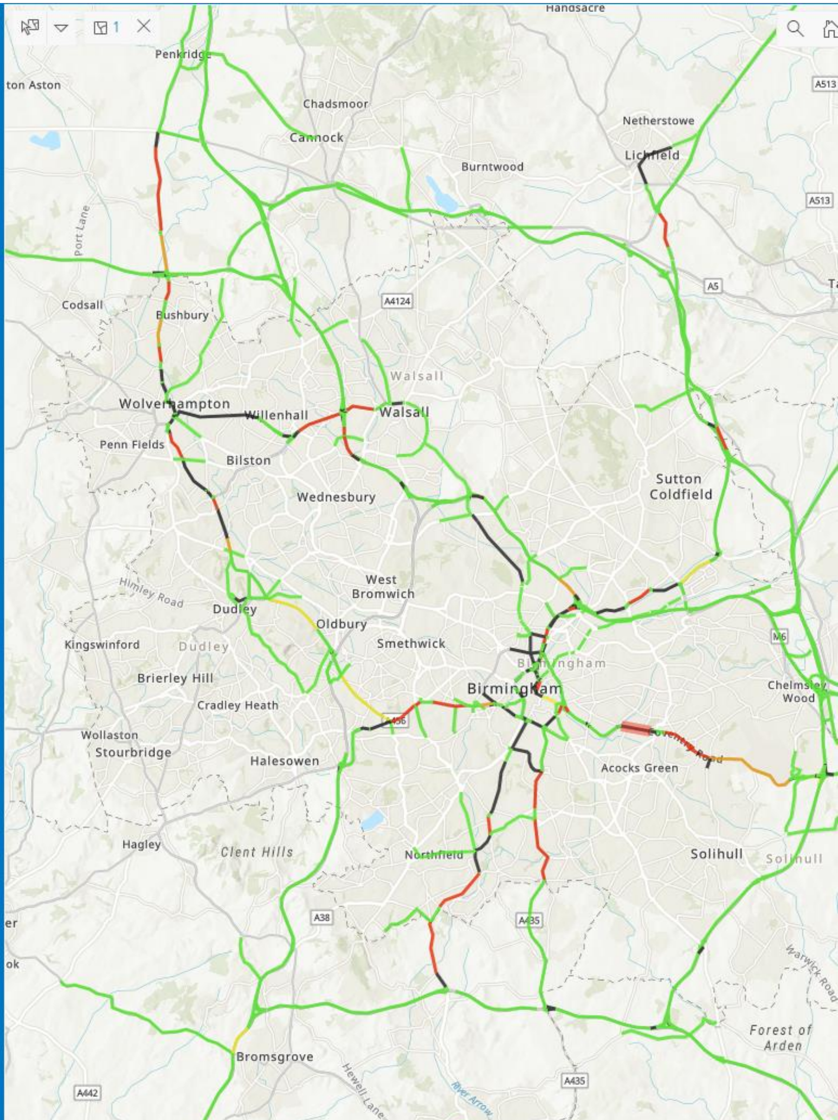
Late Response
2

Red Light
0

Braking
18

Drowsy
0

Failed to Keep an Out
0



# NEW DATASETS - TELEMATICS

What metrics do we KNOW cause higher risk?

Are these road risks or user risks?

Harsh Braking?

Swift Acceleration?

Unusual Cornering Speed?

Following Distance?

Distraction?

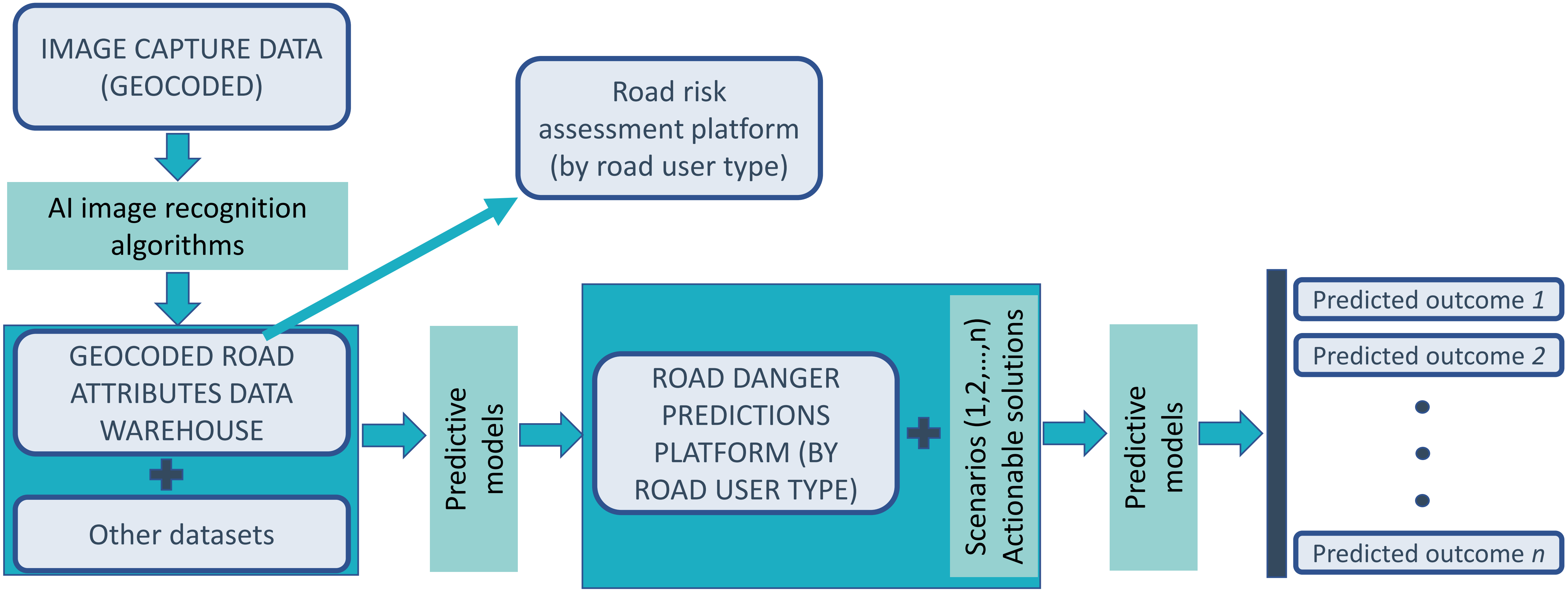
Drowsiness?

# NEW DATASETS - TELEMATICS

Full access to manufacturer vehicle data!

- Speed
- Flow
- Braking
- Acceleration
- Steering input
- Following distance
- Activated safety measures (AEB)
- Post-collision investigation

# PROJECT RAPIER



# PROJECT RAPIER – PILOT 1

https://s3-eu-west-1.amazonaws.com/agilysis.media/video/Rapier/VIRB0066_annotated.mp4 agilysis

RICHARD OWEN

+44 (0) 1295 731810

Richard.Owen@agilysis.co.uk