Predicting Road Accident Hotspots Using Street-View and Satellite Imagery

MSc Research Project
Artificial Intelligence

Adam Lysaght
Student ID: x23314869

School of Computing
National College of Ireland

Supervisor: Devanshu Anand

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Adam Lysaght
Student ID:	x23314869
Programme:	Artificial Intelligence
Year:	2025
Module:	MSc Research Project
Supervisor:	Devanshu Anand
Submission Due Date:	11/08/2025
Project Title:	Predicting Road Accident Hotspots Using Street-View and Satellite Imagery
Word Count:	5294
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Adam Lysaght
Date:	11th August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Road Accident Hotspots Using Street-View and Satellite Imagery

Adam Lysaght
x23314869

Abstract

Roads facilitate economic and social connections in every day life, yet recurring traffic accidents at certain locations pose safety risks. This paper presents a multi-modal deep learning framework to identify high-risk road segements, or "hotspots", through the integration and processing of satellite and street-view imagery. A series of machine learning and deep learning models are implemented and evaluated, learning to distinguish visual features associated with historically dangerous road segments. Results demonstrate that combining satellite and street-view data improves classification accuracy by 6% compared to single data source models. The study provides comparative algorithm analysis over a range of machine and deep learning models, offering valuable insights for practitioners aiming to implement targeted interventions.

1 Introduction

Roads are the backbone of modern economies, facilitating trade, employment, and social connectivity. However, road traffic accidents are one of the leading causes of death worldwide, with more than 1 million deaths each year (World Health Organisation, 2023). Variability in road design, weather, and traffic conditions among other factors leads to differences in accident rates across roads and locations (Cheng and Washington; 2005). Certain road segments, often referred to as accident "hotspots" or "blackspots" are characterized by a heightened risk and frequent reoccurrence of traffic accidents (Ghadi and Torok; 2017). Despite these risks, current safety strategies predominantly rely on reactive measures, such as once off post-accident reviews or public awareness campaigns.

Identifying dangerous hotspots presents both a challenge and an opportunity for advancing road safety. The challenge lies in understanding the factors that heighten road risk and what causes an accident, which can be environmental, infrastructural, behavioural, or a combination of these factors. Thus, it can be a challenge for road safety practitioners to implement specific, targeted measures with confidence that it will have a positive safety impact. On the other hand, adopting a comprehensive data-driven view of accidents across the road network can support the identification of recurring patterns that are indicative of accident hotspot locations. With this capability, the identified hotspots can be analysed to determine the underlying factors contributing to the risk. Targeted remediation strategies could then be enacted, aimed at reducing the number of road traffic accidents.

In this regard, identifying road accident hotspots has been a focus for road safety organisations. Historically, approaches have typically relied on statistical measures that are typically linear in nature to correlate accident frequency with observed variables (e.g. traffic volume, road curvature) to infer road risk and determine hotspot locations. These approaches generally rely on static assumptions from a fixed dataset, limiting their utility and flexibility. More recently, advancements in image processing techniques have enabled the ability to infer information about the road network from images. For example, remote sensing techniques enable the extraction of the structure of road networks from satellite imagery (Xu et al.; 2018). Higher resolution, ground level image sources such as street view imagery enable the ability to identify and analyse more nuanced road components, for example, detecting damage to the road network (Arya et al.; 2021). The benefit of these approaches is that they can be applied in an automated manner to support continuous classification and analysis of the road network. These advancements in image processing and geospatial analysis create a foundation for data driven approaches to road safety.

In line with these advancements, this work aims use computer vision and deep learning techniques to assess and perform accident risk assessments of the road network for the benefit of road safety. More specifically, this work aims to develop a binary classification approach to distinguish accident hotspots from non-hotspot locations based on image data of the road network. This is performed by sourcing an accident hotspot dataset, training algorithms to learn features from geospatial image data (satellite, street view) of frequent accident locations, and applying the model on unseen data to predict accident hotspots.

Thus, the proposed research question that this study aims to address are listed as follows:

1. What performance improvements does combining satellite and street view imagery provide compared to single-modal identification of accident hotspots?

2. How do varying machine learning deep learning models perform at identifying road accident hotspots?

3. What recommendations can be provided to road safety practitioners and government agencies to enable unsupervised road accident hotspot identification?

This work investigates the effectiveness of various multi-modal machine & deep learning approaches to predict traffic accident hotspots using street view and satellite imagery. The contribution of this work is two fold - Firstly, it provides a framework for multi-modal accident hotspot prediction and compares the performance of varying machine learning and deep learning algorithms. Secondly, it demonstrates the meaningful performance improvements (5-6% in accuracy) that multi-modal road image data brings over single-modality approaches in predicting accident hotspots locations. This is particularly valuable for traffic safety practitioners to identify and remediate hotspot locations.

The remainder of the paper is structured as follows:

- Section 2, Related Work presents an overview of the existing work to identify road accident hotspots and how this work aligns to the literature.

- Section 3, Methodology details the planned approach to execute this study.

- Section 4, Design & Implementation describes the system architecture and technical details of the implemented solution.
- Section 5, Evaluation reviews experimental results and assesses model performance.
- Section 6, Conclusion summarises the key findings of this work, its limitations, and proposed directions for future work in this area.

2 Related Work

Identifying road accident hotspots has been well studied within the existing literature. Sorensen (2007) presents an overview of approaches used, namely "accident based" methods and "non-accident based" methods. Accident based methods use datasets of previous accident data to identify hotspots and therefore are retrospective in nature. On the other hand, non-accident based methods analyse factors of the surrounding driving environment, irrespective of if an accident took place. These factors include road conditions (e.g., presence of guardrails, road geometry, speed limits), traffic levels, and the driver of the vehicle (Sorensen; 2007). In this regard, non-accident based approaches are proactive in nature and can help to identify roads where there is an increased risk of an accident in the future. Combining accident and non-accident based techniques presents an opportunity to leverage the strengths of each approach, resulting in a more flexible accident hotspot prediction model. This combined approach forms the basis of this work, which will be further examined in subsequent sections.

2.1 Statistical methods to identifying road accident hotspots

Earlier work in the field employed statistical techniques such as regression (poisson, binomial), and empirical bayes among others to quantify the relationship between road accident risk factors and accident frequency. A number of countries developed accident prediction models based on these approaches, such as Denmark and Portugal (Sorensen; 2007), and more recently Ireland, where a negative binomial distribution was found to be the best model statistical model in predicting accident hotspots (Wallbank et al.; 2023). While these models are simple to implement, the disadvantages are widely acknowledged in that each model has its own predefined relationships between dependent and independent variables (Silva et al.; 2020), meaning that these prediction models are have limited flexibility and are static in nature.

2.2 Spatial clustering algorithms to identify road accident hotspots

The application of spatial algorithms has proven useful in identifying road accident hotspots. Wang et al. (2023) propose a clustering approach to identify accident hotspots using a variant of DBSCAN, a density-based spatial clustering algorithm. The authors group collections of accidents into clusters based on their spatial proximity, identifying 39 accident hotspots from a dataset of 3,536 traffic accidents (Wang et al.; 2023). The authors evaluation method involved visiting the location of identified hotspots and investigating the likely causes of accidents, for example, heavy traffic volume. While this manual approach limits the repeatability and thus the verifiability of the effectiveness of the algorithm, the work presents a useful approach to systematically construct accident hotspots from a historic road accident dataset, a method which will be applied within the context of this study.

2.3 Machine learning algorithms to identify road accident hotspots

The emergence of machine learning algorithms have enabled advanced algorithms to identify accident hotspots more accurately. Strong results were obtained by Amorim et al. (2023) who applied SVM, Random Forest, and a multi-layer perceptron neural network algorithms on an accident dataset of 24 features (e.g. road layout, weather conditions, vehicles) on Brazilian highways. The authors predicted whether a highway section had a risk of severe or nonsevere accidents, achieve a recall score of 83%, a strong result given the rarity of hotspot locations compared to non-hotspot locations. Overall, the work demonstrated strong performance using supervised machine learning and highlighted the prominent road features that impact the likelihood of serious road accidents. Despite this, the approach proposed relies heavily on structured data (for example, weather, road type) and predefined severity labels based on if there was a serious injury for each accident historically recorded. This limits the ability to discover novel or emerging hotspots that may not align with historical patterns, a gap that could be addressed by incorporating unsupervised approaches (e.g. clustering of accident coordinates) to identify patterns in the road network.

2.4 Image segmentation and deep learning techniques to identify road accident hotspots

In contrast to the accident-based hotspot detection methods discussed so far, the rapid advancement in deep learning algorithms has enabled the ability to model road features from images of the road network, for example, satellite or street view images.

An initial step in processing satellite imagery involves using remote sensing techniques to identify and extract road networks from images. This process is typically regarded as a semantic segmentation task, where road and non-road labels are assigned to all pixels in an image, achieving binary semantic segmentation (Mo et al.; 2024). Expanding beyond simply extracting roads from images, existing work has employed computer vision algorithms such as CNNs to gather information about roads for safety purposes. Najjar et al. (2017) utilised a CNN to create maps of New York and Denver that indicate the safety of roads in three different levels, achieving an accuracy of 78%. The authors utilised a road accident dataset and the corresponding satellite images of accident locations to label roads as high, neutral, and low risk of accidents, and subsequently trained a CNN model to predict based on satellite imagery of other cities. The work demonstrated that visual features contained in satellite imagery can be effectively used as a proxy indicator of road safety (Najjar et al.; 2017). However, augmenting the road data with additional data such as street view imagery or tabular road attributes could have the potential enhance the model's understanding of relationships between road characteristics and accident likelihood.

He et al. (2021) builds on this work and combines GPS data with the satellite imagery. The author models the risk maps on a 5m x 5m basis in comparison to a 150m x 150m used in previous work, providing more granular and detailed risk assessment of road segments.

The approach used a Kernel density model, achieving stronger performance than previous work. In a similar manner, Wijnands et al. (2021) apply a deep auto-encoder architecture with convolution layers to images of road intersections taken from satellite imagery. The authors augment the model using telematics dataset to capture driving behaviour. While the authors did not specifically apply the work to detect accident hotspots, they highlight the importance of extracting domain specific features such as road type, size, shape, and lane markings using a thorough feature extraction process, and how this can be used to identify unsafe road designs.

The viability of extracting features from road images to model road risk is further emphasised by Tanprasert et al. (2020) who segment street view images to capture the presence of 32 classes (e.g. traffic light, fence) in each pixel of the image. Applying a fully connected neural network to the images of the roads, the authors achieved an accuracy of 70% at identifying known accident hotspots. The authors conclude that not only the condition and layout of the road is a contributor to a road's accident-proneness, but also objects around the road, specifically within the driver's eye level, are key factors. While the work primarily focused on the latter, integrating both street view and satellite imagery has the potential to provide a more comprehensive approach to testing these hypotheses, an objective which this study aims to address.

2.5 Combining street view and satellite imagery to identify accident hotpots

Guo et al. (2024) advanced this methodology by designing a novel two branch deep learning model to extract global structural features from satellite imagery and dynamic continuous features from street view imagery. The authors apply CNNs to learn structural patterns such as lane and intersection configurations, transformers to capture longer range contextual dependencies like urban sprawl and proximity to commercial zones, and RNNs with an attention mechanism to prioritise features (e.g. road signs) within the driver's field of view. The work tested a variety of architectures and algorithms, achieving a peak accuracy of 93.7%. The results highlight the effectiveness of deep learning algorithms and the combination of satellite and street view imagery to identify road accident hotspots.

2.6 Research Niche

Overall, there is promising research in leveraging deep learning and geospatial data for accident hotspot identification. However, existing frameworks are predominantly validated in regions with specific regions (e.g. China, US), leaving gaps in understanding how these models can be applied to contexts like the UK. Furthermore, there remains the opportunity to test a variety of machine and deep learning algorithms on both street-view and satellite, assessing the predictive performance of each data set and algorithm.

3 Methodology

This section presents the research methodology used to undertake this work, from data collection to analysis and evaluation.

3.1 Research Approach & Problem Formulation

This work uses supervised machine learning and deep learning algorithms for binary classification of geographic accident locations as either accident hotspots or non-hotspots. The basis of this approach lies in the enhanced strength and flexibility of deep learning architectures to capture relevant patterns in images compared to traditional statistical approaches. Applied to road accident hotspot prediction, identifying visual and spatial characteristics of the road network consistent with high-risk locations using deep learning provides an accurate, scalable, and flexible method to detect road accident hotspot locations.

The problem can be framed as a classification task where models are trained to classify geographic locations as either accident hotspots or non-hotspots based on their visual characteristics. For each accident location (defined by latitude and longitude coordinates), the models process two types of imagery: street view images of the surrounding area and a satellite image above the location. This multi-modal approach enables the models to learn from both ground level details and aerial patterns. Formally, this involves learning a function $f^* : \mathcal{X} \to \mathcal{Y}$ that maps the input images $\mathcal{X} = \{\mathbf{x}_{sv}, \mathbf{x}_{sat}\}$ to binary predictions $\mathcal{Y} = \{0, 1\}$, where $\mathbf{x}_{sv}$ represents the street view images, $\mathbf{x}_{sat}$ represents the satellite image, and the output indicates whether a location is a hotspot $(y = 1)$ or not $(y = 0)$. The trained models can then be applied to new locations to identify potential accident hotspots based solely on their visual and spatial characteristics, providing a proactive approach to road safety assessment.

3.2 Methodology Overview

This work is structured around four interconnected phases: Data Collection and Preprocessing, Feature Engineering, Model Development, and Evaluation; as described in Figure 1 below.

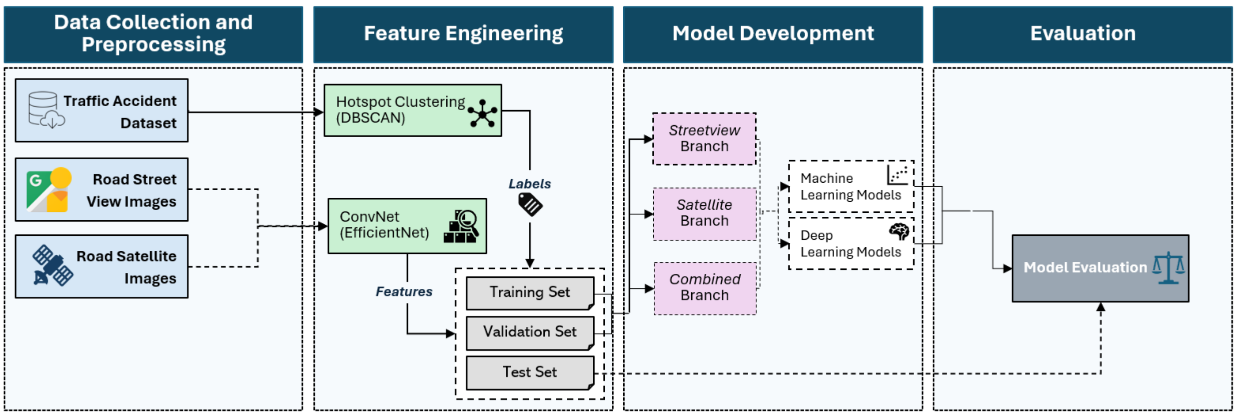


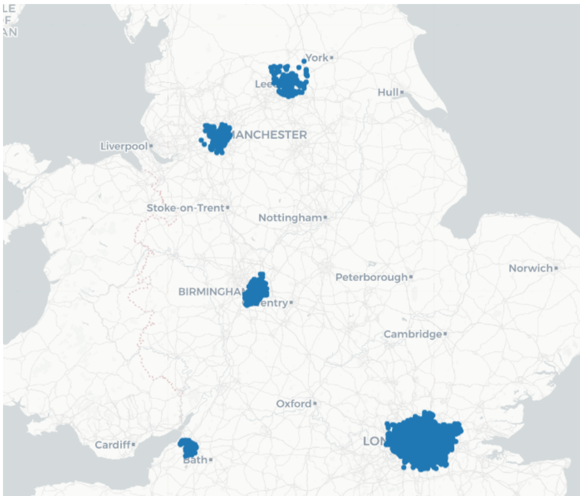
Figure 1: Methodology Overview

3.2.1 Data Collection and Preprocessing

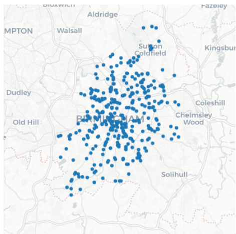
3.2.1.1 Road Accident Data Acquisition:

To reliably train deep learning algorithms, a verified set of accident locations was required. As such, a Traffic Accident dataset was sourced that contained historic road traffic accidents in the UK. It was sourced from the UK Department for Transport [1]. It contains approximately 100,000 police-reported road traffic collisions from 2023, using accident coordinates (latitude, longitude) as primary location identifiers. This dataset serves as the ground truth for model validation in this study.

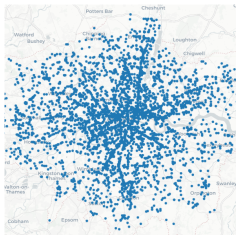
The dataset was preprocessed to address quality and computational constraints. Entries without valid latitude/longitude coordinates were excluded. To reduce computational demands while maintaining a spread of geographies, data was filtered to five major UK cities: London, Birmingham, Manchester, Leeds, and Bristol, resulting in approximately 28,000 collisions. The study area is visualised in Figure 2 below.



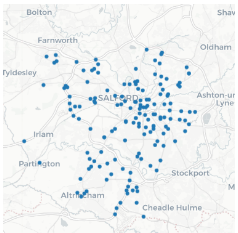
(a) Overall UK Study Area



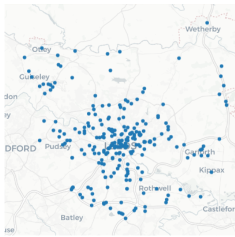
(a) Birmingham



(b) London



(c) Manchester



(d) Leeds

(b) Detailed View of Cities Selected

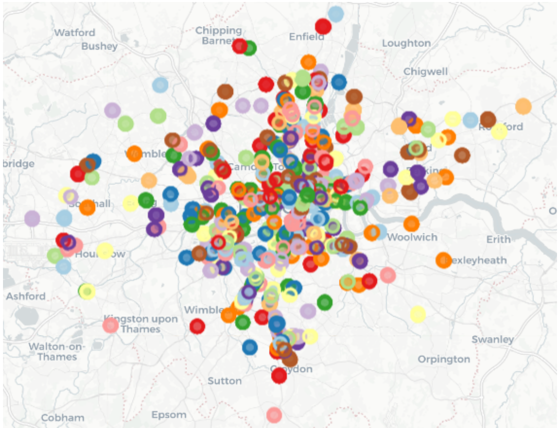
Figure 2: Visualisation of historic accident locations across the study area: (left) Overall UK study area, (right) Detailed views of selected cities (a) Birmingham, (b) London, (c) Manchester, and (d) Leeds.

3.2.1.2 Geographic Clustering Methodology (DBSCAN)

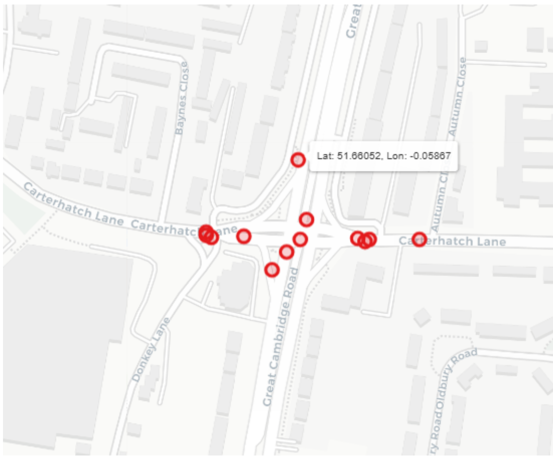
To classify the set of accident data into *hotspot* and *non-hotspot* groups, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm was used. As employed by Guo et al. (2024) and Wang et al. (2023) for similar use cases, DBSCAN identifies dense clusters of geographic data while separating noise points. This makes it well suited to identify accident hotspot areas versus non hotspot areas based on historic accident data. Using parameters of E=60 meters (maximum point separation) and min-samples=6 (minimum number of accidents per cluster), locations with a high density of accidents in close proximity were labelled as accident hotspots, while isolated points were labelled non-hotspots. This process generated binary labels for accident locations to support model training phase at a later stage. In total, 461 accident hotspots were identified totalling 4,044 accident points, with an average of 9 accident points per cluster. The remainder of the data set (24,400) was labelled as not part of an accident

[1] Road Safety Data - Collisions 2023: `https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-accidents-safety-data`

hotspot; a sample of these points (4,044) were selected for inclusion in the training process to ensure balance between the two classes. An example of the DBSCAN clustering process is visualised in Figure 3 below.



(a) Classification of accident hotspots in London, identified through DBSCAN. Each coloured point represents an accident hotspot zone based on a high concentration of historic accident locations.

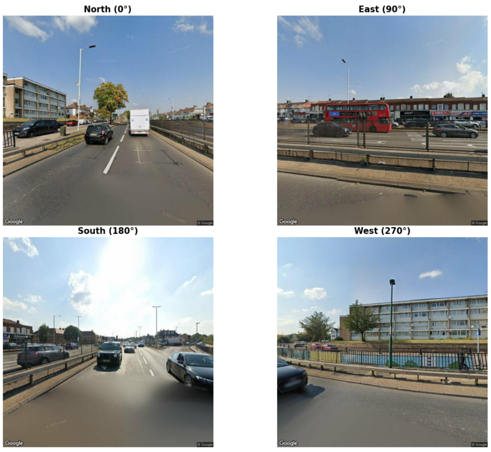


(b) An example of an individual hotspot zone from figure (a). Each point represents a historic accident location.

Figure 3: Geospatial clustering (DBSCAN) approach used to identify and label accident hotspot zones.

3.2.1.3 Street View and Satellite Data Collection

For each accident point, both street view and satellite imagery were collected from the Google Maps API. Street view images were collected at 640x640 pixel resolution across varying angles (0°, 90°, 180°, 270°) to provide a full field of view of the road network at each location. Satellite images were collected at 640x640 pixel resolution with zoom level 20 (0.30 m/px), providing an aerial view of road layout and infrastructure. An example of the street view and satellite imagery sourced for each accident location is visualised in Figure 4 below. This imagery is the input to the feature extraction process in the next stage.



(a) Street view images (4) for a specific location

(b) Satellite image for a specific location

Figure 4: An example of (a) street view imagery and (b) satellite imagery sourced for each accident location.

3.2.2 Feature Engineering

The feature engineering phase uses convolutional neural networks (CNN) to extract relevant visual features from both street view and satellite images prior to these features being used as inputs by prediction models in the final stage. This approach was chosen to decouple feature extraction from classification, enabling the flexibility of testing of various downstream models without reprocessing the original imagery. The CNN backbones (EfficientNet, ResNet) were pre-trained on ImageNet, with the backbone weights frozen to maintain generalisability.

Additionally, $Pooling$ was set to $None$ to preserve spatial and directional relationships rather than flattening features into 1D vectors. The backbone output outputs 3D vectors, (H, W, C), where H and W indicate the spatial positioning (height and width), and C represents the feature channels in each position. For accident hotspot prediction, not only the presence of objective but the spatial relationships between objects and geographic context of an object may be a key determinant of the decision to predict an area as a hotspot. For example, objects such as road intersections, traffic infrastructure, and road patterns across the area may be a factor in road risk. This is illustrated in Figure 5 below where the pooling/flattening of features (right) is compared against higher dimensionality feature extraction (left) which preserves the relationships between objects.

3.2.3 Model Development

3.2.3.1 Approach

Using the extracted features, three separate branches are set up and tested to evaluate the performance of the varying modalities (Street view, Satellite, Combined) and algorithms in predicting accident hotspots.

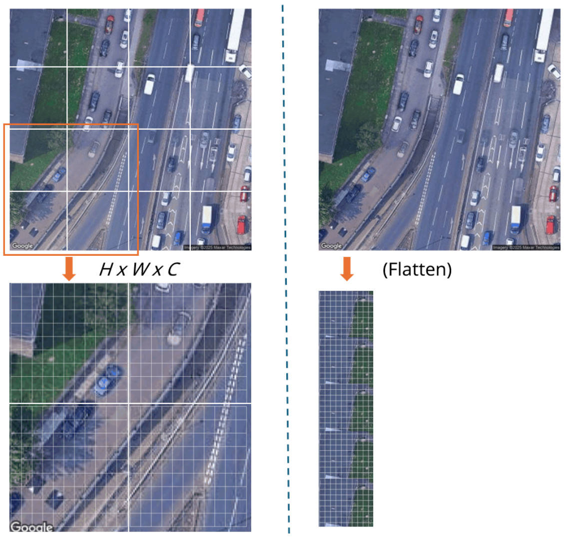


Figure 5: Example of spatial relationship preservation (left) versus pooling/flattening (right); the model could potentially learn there is a curve in the road onto a busier road.

3.2.3.2 Machine Learning Approaches

Various Machine Learning algorithms were implemented, trained, and tested on both the street view branch and the satellite branch. While evaluating the performance of these algorithms was not the primary focus of this work, they provide a useful baseline of a minimum performance level of the which the deep learning approaches could be evaluated against. As traditional machine learning algorithms operate on 1 dimension feature sets, global and mean pooling was applied to reduce the 3-dimensional spatial-temporal tensors to 1-dimensional feature vectors. As such, it is expected that these algorithms would perform sub-optimally relative to deep learning approaches which preserve spatial relationships.

3.2.3.3 Deep Learning Approaches

Deep learning was used to leverage its capacity for processing multi-dimensional data while preserving the inherent spatial relationships within the feature sets. The approach used prioritised using proven architecture layers such as BiGRU, Attention, and 3D Convolutions. Further information on the models and architecture used is provided within Section 4, Design.

3.2.4 Evaluation

In the evaluation phase, models were trained by minimising binary cross-entropy loss of model predictions. Subsequently, the trained models were evaluated on the test sets, with their performance metrics evaluated.

3.2.4.1 Performance Metrics

The models implemented are assessed and compared using a variety of performance metrics, namely Accuracy, Recall, F1-score, and Precision. Accuracy measures overall classification correctness across both classes, while precision calculates the proportion of correctly identified hotspots among all predicted hotspots. Recall determines the proportion of actual hotspots correctly identified by the model, and the F1-score provides a harmonic mean balancing precision and recall trade-offs. In addition, computational efficiency metrics were evaluated, including model training time, inference time, and number of parameters. These metrics provide insight into practical deployment considerations beyond predictive performance.

3.2.4.2 Ablation Studies

Ablation studies evaluate the contribution of individual components, for example, assessing the contribution of solely processing street view sequences without satellite data. This was achieved naturally given the three branch approach which is discussed in further detail in Section 4 Design.

4 Design & Implementation

This section details the system architecture and design decisions underlying the system, and the technical implementation of final solution.

4.1 Architecture Overview

The proposed architecture processes satellite and street view images of the road network into binary accident risk predictions through three interconnected subsystems, as illustrated in Figure 6 below. The architecture comprises of (1) **Collision Data Processing & Spatial Clustering**, generating ground truth labels through accident clustering, (2) a **Multi-Modal Feature Extraction** framework extracting features from street view sequences and satellite imagery, and (3) a **Machine & Deep Learning Pipeline** to create binary predictions and evaluate the strongest performing model at identifying accident hotspot locations.

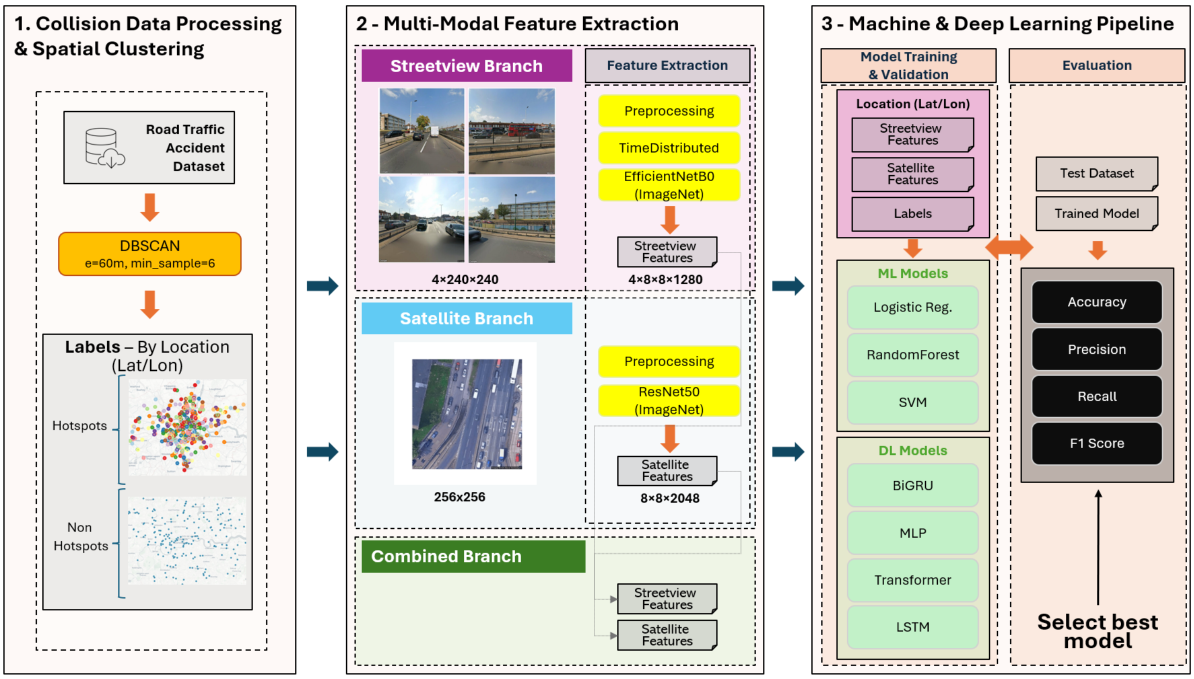


Figure 6: System Architecture Overview

4.2 Collision Data Processing & Spatial Clustering

4.2.1 DBSCAN Configuration

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was used to cluster groups of accident locations together, providing the ground truth labels for "hotspot" versus "non-hotspot" accident locations. Accordingly, accident locations grouped into a cluster by the algorithm were classified as hotspots (label=1), whereas the remaining data points, designated as noise, were considered non-hotspots (label=0).

The algorithm was configured two parameters: **epsilon (ε)**: the maximum distance between any two points in the cluster, and **MinPts**: the minimum number of points required to form a cluster (hotspot). These parameters were configured as 60 meters and 6 points respectively. The selection of these values was guided by exploratory and visual analysis using a $kNearestNeighbors$ distance graph, plotting the distance of each point to its 5th nearest neighbor (where k=MinPts-1). It was found that 60 meters and 6 points were appropriate parameters to capture meaningful clusters in areas with a higher density of accidents compared to other areas.

4.3 Multi-Modal Feature Extraction

4.3.1 Street View Branch

Street View image features are extracted using an EfficientNetB0 backbone pre-trained on ImageNet. EfficientNetB0 (5m parameters) was chosen due its computation efficiency against other backbones, e.g. ResNet50 (25m parameters). Input images are resized from 640×640 to 240×240 pixels to reduce computational requirements. Four street view images (0°, 90°, 180°, 270°) for each accident location are processed as a batch through a TimeDistributed layer and the EfficientNet layers, generating a 5D tensor with dimensions (batch, sequence=4, height=8, width=8, channels=1280). This architecture ensures spatial relationships are captured within each image, and between the between the 4 street view image sequence, for example, recognising a 4 way road intersection across the images.

4.3.2 Satellite Branch

For the satellite branch, a similar design was adopted. Given there is only 1 satellite image for each location, the ResNet50 backback was used as the primary option as it outputs 2,080 feature channels in comparison to 1,280 of ResNet50 at 240x240 image resolution. This configuration outputs a 4D tensor with dimensions (batch, height=8, width=8, channels=2080).

4.3.3 Combined Branch

A central Dataframe acts as a dictionary between accident locations, the street view features, and the satellite features. This dataframe serves as the combined branch which later passes both the street view and satellite features as inputs into the deep learning models.

4.4 Machine Learning & Deep Learning Pipeline

4.4.1 Dataset Split

Firstly, feature and label data are divided into training and validation sets (80%), and test sets (20%) with equal label distribution across each set. To prevent data leakage, all accident locations as part of a hotspot are assigned exclusively to training or testing sets.

4.4.2 Pipeline Summary & Model Implementation

The framework implements and evaluates three distinct algorithm categories:

Machine Learning: Seven standard machine learning algorithms are implemented using scikit-learn. These algorithms process engineered features through multiple feature extraction strategies (temporal aggregation, spatial pooling, combined features) with hyperparameter optimization. A list of the machine learning algorithms implemented are shown in Table 1 below.

Table 1: Machine Learning Algorithms

No.	Algorithm
1	Logistic Regression
2	Support Vector Machine
3	Gradient Boosting
4	Random Forest
5	Decision Tree
6	Gaussian Naive Bayes
7	K-Nearest Neighbors

Deep Learning - Street View/Satellite (Separate): Deep learning models were set up to process both street view and satellite data (separately). These models process 3D spatial tensors through TensorFlow/Keras architectures.

Deep Learning - Street View/Satellite (Combined): Finally, deep learning models that support multi-modal data were set up. These combine street view and satellite inputs through concatenation and attention mechanisms.

The selection, design, and implementation of the Deep Learning models and architectures was informed by testing, hyperparameter optimization, and through the review of related work. A detailed list and summary architecture of the deep learning algorithms implemented are shown in Table 2 below.

Table 2: Deep Learning Architectures

Algorithm	Key Layer Type	Parameters	Hidden Units	Layers	Data Support
2D-CNN	Conv2D + BatchNorm	50-80K	32	4	Satellite
3D-CNN	Conv3D + MaxPool3D	350K	64	4	Street View
Multilayer Perceptron	Dense + Dropout	270K	128	5	Both
BiGRU	Bidirectional GRU	50-70K	16	3	Both
Transformer	Multi-Head Attention	50-100K	32, 2 heads	4	Both
ConvNeXt	Depthwise Conv2D	60-100K	32	3	Both
LSTM	Conv2D + LSTM	400K	64	4	Street View
Combined (Concat)	Dual encoder + Dense	100-150K	64	6	Multi-modal
Combined (Attention)	Dense + Multi-Head Attention	150-200K	256	8	Multi-modal

4.4.3 Development Environment

To execute this study, the following development components were used:

Tooling: The core framework components include Python as the development environment, TensorFlow as the primary deep learning framework, NumPy for numerical computing, Pandas for data manipulation, and Scikit-learn for machine learning utilities.

Storage: The image dataset comprising approximately 40,000 samples (8,000 locations with 4 street view angles each (32,000 images) and 8,000 corresponding satellite images) were stored in a Google Cloud Storage Bucket. Images were downloaded into the virtual machine's cache during processing pipelines.

Computation: Models were executed on a Google Cloud virtual machine instance configured as n1-standard-8 (8 vCPUs, 30 GB RAM) and accelerated by an NVIDIA T4 GPU.

5 Evaluation

The section provides a comprehensive summary and analysis of the results of the models implemented. The main findings and implications of the study are presented.

5.1 Street View Branch

The results of the street view branch are presented in Tables 3 and 4 below. The models performed reasonably, with an average accuracy of 63% across all algorithms tested. Overall, the deep learning models performed with greater consistency than the machine learning models, with a range of 61% to 65%. Surprisingly, the model with the highest accuracy was Logistic Regression, a relatively simple machine learning model.

Table 3: Performance comparison of machine learning algorithms on the street view branch

No.	Algorithm	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
1	K-Nearest Neighbors	57.6	58.5	56.0	57.2
2	Gaussian Naive Bayes	58.3	61.4	52.2	56.4
3	Decision Tree	58.6	61.6	52.5	56.7
4	Random Forest	63.5	65.3	60.7	62.9
5	Gradient Boosting	65.2	66.8	62.9	64.8
6	Support Vector Machine	65.5	66.8	63.7	65.2
7	Logistic Regression	66.4	67.0	65.6	66.3

5.2 Satellite Branch

The results of the satellite branch are presented in Tables 5 and 6 below. The satellite branch performed underwhelmingly, with an average accuracy of 54% across all models. BiGRU was the strongest performer with an accuracy of 60%. Despite trialing a number of varying parameters, several of the deep learning models (ConvNeXt, Transformer, 2D CNN) failed to learn any patterns in the feature set and model weights likely were not updating.

Table 4: Performance comparison of deep learning architectures on the street view branch

No.	Algorithm	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Params
1	LSTM	60.8	60.8	60.6	60.7	393 569
2	ConvNeXt	62.5	62.9	61.9	62.4	60 001
3	3D CNN	62.9	63.5	61.9	62.7	344 225
4	Transformer	65.0	65.5	64.3	64.9	50 625
5	BiGRU	65.2	65.8	64.4	65.1	46 337

Table 5: Performance comparison of machine learning algorithms on the satellite branch

No.	Algorithm	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
1	K-Nearest Neighbors	51.5	52.7	56.0	49.7
2	Decision Tree	55.2	55.7	52.5	54.9
3	Gradient Boosting	55.2	55.5	62.9	55.2
4	Random Forest	55.3	56.1	60.7	54.9
5	Gaussian Naive Bayes	57.8	57.8	52.2	57.8
6	Logistic Regression	58.0	58.1	65.6	58.0
7	Support Vector Machine	58.3	58.6	63.7	58.3

Table 6: Performance comparison of deep learning architectures on the satellite view branch

No.	Algorithm	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Params
1	ConvNeXt	48.0	23.1	48.1	31.2	95 329
2	Transformer	48.0	23.1	48.1	31.2	559 937
3	2D CNN	48.0	23.1	48.1	31.2	617 057
4	Multilayer Perceptron	57.9	57.9	57.9	57.9	273 153
5	BiGRU	59.8	60.0	59.6	59.8	70 913

5.3 Combined Branch

The results of the satellite branch are presented in Table 7 below. This branch performed strongest with a highest accuracy of 71%. In addition the Attention model had 70% across all Evaluation metrics. These results indicate that the model was successful in learning patterns across both feature sets and generating more reliable predictions as a result.

Table 7: Performance comparison of deep learning architectures on the combined branch

No.	Algorithm	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Params
1	Concat	66.3	69.1	76.5	66.3	221 506
2	Attention	71.3	72.9	75.8	71.3	1 181 570

5.4 Discussion

In general, the performance of the implemented models was varied. The combined branch proved to be the most effective, followed by the street view branch and then the satellite branch.

The attention-based combined model reached an accuracy of 71.3%, a 5-6% improvement over single modality approaches. This finding answers the original the research question, that combining satellite and street view imagery increases the performance of accident hotspot identification compared to single modal prediction. The integration of both street view and satellite layers into an attention based architecture had a measurable impact, with model able to develop and learn relationships between the modalities.

Evaluating the individual branches, the street view branch consistently outperformed the satellite branch by 6-18% (60-66% vs 48-60% accuracy) across all models. This indicates that ground level image sequences capture more predictive information for accident risk than overhead perspectives. A key factor that likely contributed to this performance improvement was the 360° image collection process, in which the streetview branch likely could extract more detailed features from infrastructure elements such as road markings, signage, and intersection complexity. In contrast, the satellite branch processes a single overhead image of a location, which may be lacking in sufficient detail to extract similar features.

Focusing on the performance of the individual implemented models, there was mixed performance. On some occasions, the simpler, lower parameter models outperformed larger models. As an example, within the street view branch the Transformer (50K parameters) achieved 65% accuracy while the 3D CNN (340K parameters) achieved 62.9%. Additionally, the machine learning models performed similarly to some of the deep learning models, and in some cases better. In line with Related Work, it was expected that the deep learning models would consistently outperform the machine learning algorithms. Further work with increased sample sizes and computational capacity may reveal that deep learning approaches do indeed outperform machine learning models when testing on a larger scale.

Overall, model performance did not reach levels reported in comparable literature, partly due to computational capacity and model implementation challenges. However, the performance improvements demonstrated as feature complexity increased, progressing from single satellite images to multi-perspective street views, and finally combined modalities highlight the practical utility of image processing algorithms for road accident hotspot identification.

6 Conclusion and Future Work

6.1 Conclusion

The data used in this work identified over 100,000 accidents on the roads in the UK in one year alone. Accidents cannot be completely eliminated on the roads, however, providing road safety practitioners with the capability to identify where they are most likely to happen may enable informed decision making to reduce the number of accidents on the roads.

This work developed a binary classification approach to identify accident hotspots locations based on image data of the road network. This was performed by sourcing an accident hotspot dataset, training algorithms to learn features from geospatial image data (satellite, street view) of frequent accident locations, and applying the model on unseen data to predict accident hotspots.

The work compared the performance of varying machine learning and deep learning algorithms in road accident hotspot prediction. Although the performance of the models was often suboptimal, the study successfully demonstrated the viability of multi-modal fusion of street view and satellite data, achieving a 5-6% accuracy improvement over single-modality approaches.

6.2 Limitations and Future Work

Computational capacity remained a challenge throughout this study. Loading approximately 8,000 street view feature tensors with dimensions (8000, 4, 8, 8, 1280) required approximately 5GB memory, which scales higher during training. Satellite imagery was not as memory intensive; however, it proved challenging to train and test models in the combined branch due to the collective file size of their feature tensors. Google Cloud Platform deployment partly remediated this issue but the challenge remained due to the cost of larger VM instances. As a result, reducing model complexity was often required, for example, decreasing the number of model layers, neurons, and batch sizes. These compromises likely reduced the model's predictive performance. Re-engineering and evaluating more advanced models with greater computational to execute them could be a valuable area for further research.

The binary hotspot classification of accident locations based on DBSCAN clustering may oversimplify accident risk. This approach was chosen as it identifies locations with a high density of historic accidents. However, this labelling approach potentially discards valuable information about accident frequency, severity, and other location specific factors and patterns that could inform more nuanced risk models. Evaluating risk of an accident location on a continuous scale, rather than on a binary one, could be a valuable area for further research.

Finally, this work used visual features alone for accident prediction. The approach captures visual characteristics but doesn't incorporate additional factors that influence accident risk probability such as traffic, weather conditions, road topology, speed limits etc. Incorporating additional data modalities could enhance the predictive performance, supporting road management authorities.

Overall, this work demonstrates the viability of applying deep learning in the field, with promising areas for future work.

References

Amorim, B. D. S. P., Firmino, A. A., Baptista, C. D. S., Junior, G. B., Paiva, A. C. D. and Junior, F. E. D. A. (2023). A machine learning approach for classifying road accident hotspots, *ISPRS International Journal of Geo-Information* **12**(6): 227.

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T. and Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries, *Automation in Construction* **132**: 103935.

Cheng, W. and Washington, S. (2005). Experimental evaluation of hotspot identification methods, *Accident Analysis and Prevention* **37**: 870–881.

Ghadi, M. and Torok, A. (2017). Comparison different black spot identification methods, *Transportation Research Procedia* **27**: 1105–1112.

Guo, W., Xu, C. and Jin, S. (2024). Fusion of satellite and street view data for urban traffic accident hotspot identification, *International Journal of Applied Earth Observation and Geoinformation* **130**: 103853.

He, S., Sadeghi, M. A., Chawla, S., Alizadeh, M., Balakrishnan, H. and Madden, S. (2021). Inferring high-resolution traffic accident risk maps based on satellite imagery and gps trajectories, *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11957–11965.

Mo, S., Shi, Y., Yuan, Q. and Li, M. (2024). A survey of deep learning road extraction algorithms using high-resolution remote sensing images, *Sensors* **24**(5).

Najjar, A., Kaneko, S. and Miyanaga, Y. (2017). Combining satellite imagery and open data to map road safety, *Proceedings of the AAAI Conference on Artificial Intelligence* **31**(1).

Silva, P. B., Andrade, M. and Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review, *Journal of Traffic and Transportation Engineering (English Edition)* **7**(6): 775–790.

Sorensen, M. (2007). *Best practice guidelines on black spot management and safety analysis of road networks*, Transportokonomisk Institut Oslo.

Tanprasert, T., Siripanpornchana, C., Surasvadi, N. and Thajchayapong, S. (2020). Recognizing traffic black spots from street view images using environment-aware image processing and neural network, *IEEE Access* **8**: 121469–121478.

Wallbank, C., Harpham, N. and Fletcher, J. (2023). Collision prediction model for the irish national road network, *Technical report*, Transport Infrastructure Ireland.

Wang, D., Huang, Y. and Cai, Z. (2023). A two-phase clustering approach for traffic accident black spots identification: integrated gis-based processing and hdbscan model, *International Journal of Injury Control and Safety Promotion* **30**: 1–12.

Wijnands, J. S., Zhao, H., Nice, K. A., Thompson, J., Scully, K., Guo, J. and Stevenson, M. (2021). Identifying safe intersection design through unsupervised feature extraction from satellite imagery, *Computer-Aided Civil and Infrastructure Engineering* **36**(3): 346–361.

Xu, Y., Xie, Z., Feng, Y. and Chen, Z. (2018). Road extraction from high-resolution remote sensing imagery using deep learning, *Remote Sensing* **10**(9): 1461.