Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

# Spatial RAG for Urban Crash Hotspot Discovery and Safety Countermeasure Recommendation

*Ziliang Samuel Zhong[1], Long Zhang[2], David Ma[3]*
[1] *New York University, NY, USA*
[2] *Transportation Systems Engineering, Southern Methodist University, TX, USA*
[3] *Software Engineering, UC Irvine, CA, USA*
samuelzhong0702@gmail.com

DOI: 10.69987/JACS.2023.31206

**Keywords**

urban crash analysis; hotspot discovery; STATS19; spatial clustering; BIRCH; DBSCAN; kernel density estimation; retrieval-augmented generation; safety countermeasures; traffic exposure.

**Abstract**

Urban crash screening systems often identify high-burden places but leave safety analysts to translate those places into treatment concepts. This paper develops a Spatial RAG pipeline for urban crash hotspot discovery and countermeasure recommendation using official 2022 Great Britain road safety and traffic data. The study integrates 106,004 STATS19 collision records, 135,480 casualty records, 193,545 vehicle records, 22,240 AADF count-point rows, 206 local-authority traffic records, and 17,840 MRDB major-road links. The primary spatial benchmark uses the 71,763 geocoded urban injury collisions. National clustering compares BIRCH, MiniBatchKMeans, and DBSCAN on British National Grid coordinates. BIRCH and MiniBatchKMeans both produce 204-center partitions aligned with the number of active highway authorities in the urban sample; MiniBatchKMeans reaches the highest sampled silhouette of 0.460 in 2.94 s, while BIRCH reaches 0.447 in 3.07 s and supplies the hierarchical center structure used for downstream hotspot screening. DBSCAN identifies 679 dense components but leaves 29.9% of crashes as noise and forms a 20,192-crash largest component, making it less suitable as the national partition. Within the ten highest-burden BIRCH centers, Spatial RAG captures 14.87% of held-out severity burden at a 5% cell budget and reaches a mean severity-AUC10 of 0.1405, compared with 11.13% and 0.1093 for KDEGrid. The paired center-wise advantage over KDEGrid is significant with a one-sided Wilcoxon p-value of 0.001953. KDEGrid remains the most stable method at the top-5% budget, with a Jaccard overlap of 0.724. The retrieval layer assigns eight of the top ten hotspots to an urban vulnerable-road-user area-wide package and two to a major-road corridor speed-management package. The results show that severity-aware spatial screening and deterministic retrieval can convert an official crash archive into a transparent first-pass safety planning tool.

## Introduction

Crash screening is a routine task in road safety analysis. Agencies collect collision archives, map events, identify dense locations, and rank candidate places for engineering review. A long line of work has studied density surfaces, clustering, and spatial hotspot identification for this purpose [9]-[13]. However, a hotspot map by itself is not enough for a safety program. Analysts also need to know whether the detected pattern is more consistent with an intersection conflict, a corridor speed-management problem, a vulnerable-road-user district, or a broader systemic treatment opportunity.

This paper addresses that gap with a Spatial RAG workflow that keeps the ranking stage quantitative and the recommendation stage traceable. The term RAG is used in the retrieval-augmented-generation sense: candidate outputs are grounded in an external evidence library rather than written freely from a model's internal memory [21]. Instruction-following language-model studies further show that constrained tasks benefit from structured context and explicit output formats [22]-[24]. For transportation safety, the useful form of this idea is not open-ended writing. It is a deterministic retrieval and text-realization layer that attaches a standard package to a measured hotspot footprint.

The data design is intentionally broader than a coordinate-only hotspot exercise. The study uses the Department for Transport's 2022 STATS19 collision, casualty, and vehicle records and links them to 2022 traffic-flow and major-road context from AADF, local-authority traffic, and the Major Roads Database [1], [2]. This makes it possible to evaluate spatial concentration, severity burden, vulnerable-road-user involvement, traffic exposure, and road-context indicators in the same pipeline. The countermeasure side is also constrained by established road safety evidence, including speed-risk literature, the Highway Safety Manual, systemic safety guidance, crash modification factor guidance, empirical Bayes before-after practice, and proven countermeasure compilations [14]-[20].

The contribution is threefold. First, the paper evaluates national and local hotspot screening on official Great Britain data with severity, casualty, vehicle, exposure, and road-context fields. Second, it compares three national clustering methods and four local cell-ranking methods under explicit review budgets. Third, it adds a deterministic Spatial RAG layer that translates the top-ranked hotspot geometry and context into a standardized countermeasure package without changing the measured screening frontier.

## Method

Data integration and spatial preparation

Table 1 summarizes the data components. The collision file provides the base location, severity, date, time, road type, speed limit, junction, lighting, weather, and road-surface variables. The casualty and vehicle files add road-user and vehicle mix. AADF count points provide nearest traffic-flow context, local-authority traffic provides exposure in motor-vehicle miles, and MRDB links provide a major-road network reference. All spatial calculations use British National Grid eastings and northings converted to kilometers. The national clustering and local benchmark are restricted to urban injury collisions with valid coordinates, leaving 71,763 crashes. Figure 1 shows the workflow, and Figure 2 shows the resulting urban collision cloud.

**Table 1.** Data components used in the integrated 2022 Great Britain benchmark.

| Data component | Records | Role in analysis |
|---|---|---|
| STATS19 collision records 2022 | 106,004 | collision coordinates, severity, road type, junction, light, weather, road surface |
| Geocoded collision records used | 105,982 | records with valid longitude/latitude and British National Grid coordinates |
| Urban geocoded collisions used in clustering | 71,763 | urban_or_rural_area = 1 |
| STATS19 casualty records 2022 | 135,480 | casualty severity and road-user class |

| | | |
|---|---|---|
| STATS19 vehicle records 2022 | 193,545 | vehicle type and vehicle manoeuvre fields |
| AADF count-point rows 2022 | 22,240 | nearest traffic-flow and road-category context |
| Local-authority traffic rows 2022 | 206 | area-level motor-vehicle miles exposure |
| MRDB major-road links 2022 | 17,840 | major-road link identifiers joined through count points |

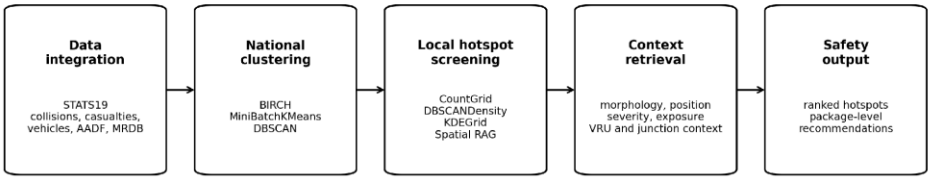


**Figure 1.** Spatial RAG workflow for clustering, hotspot ranking, exposure/context linkage, and countermeasure retrieval.

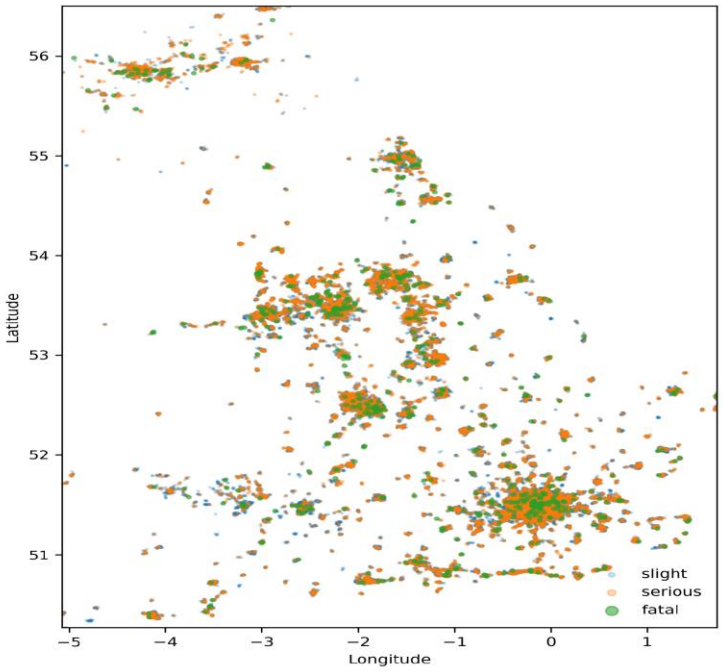


**Figure 2.** Geocoded 2022 urban STATS19 collision point cloud by recorded collision severity.

Nearest-neighbor percentiles in Table 2 characterize the spatial scale of the urban collision cloud before clustering. The density baseline uses the 75th percentile of the 5-nearest-neighbor distance, 0.594 km, as DBSCAN eps. The 95th percentile values are reported to show the upper tail of the spacing distribution but are not used for DBSCAN because that scale merges too many nearby urban components at the national level. Table 3 lists the final parameter settings.

**Table 2.** Full urban-data nearest-neighbor distance percentiles in projected kilometers.

| Percentile | 5-NN distance (km) | 10-NN distance (km) |
|---|---|---|
| 10 | 0.100 | 0.193 |
| 25 | 0.190 | 0.322 |
| 50 | 0.349 | 0.550 |
| 75 | 0.594 | 0.906 |
| 90 | 0.960 | 1.461 |
| 95 | 1.285 | 2.017 |

**Table 3.** Hyperparameters and deterministic settings used in the experiments.

| Component | Setting |
|---|---|
| Projection | British National Grid eastings/northings divided by 1,000 to obtain kilometers |
| Urban filter | urban_or_rural_area = 1 and valid coordinates |
| Severity weights | fatal = 3, serious = 2, slight = 1 |
| BIRCH | threshold = 3.0 km; branching_factor = 50; n_clusters = active 2022 highway authorities |
| MiniBatchKMeans | n_clusters = active 2022 highway authorities; batch_size = 10,000; n_init = 3; random_state = 42 |
| DBSCAN | eps = 75th percentile of 5-NN distance; min_samples = 10; ball-tree search |
| Local grid | 0.5 km x 0.5 km square cells |
| Local split | chronological 80/20 split within each selected BIRCH center |

| | |
|---|---|
| Spatial RAG score | count, severity burden, local density, AADF context, VRU, junction, darkness and wet-surface context |
| Stability | 20 bootstrap resamples per center at top-5% budget |

National clustering

The national stage compares BIRCH, MiniBatchKMeans, and DBSCAN. BIRCH incrementally compresses dense regions into a clustering-feature tree and is well suited to large spatial point clouds [4]. MiniBatchKMeans represents scalable centroid partitioning [5]. DBSCAN represents density-connected clustering with noise labeling [3]. BIRCH and MiniBatchKMeans are both set to 204 clusters, matching the number of active highway authorities represented in the urban sample, so their partitions are evaluated at the same administrative scale. DBSCAN is left unconstrained in cluster count. Runtime, cluster count, noise ratio, largest-cluster size, median cluster size, silhouette, Calinski-Harabasz, and Davies-Bouldin are reported. Silhouette, Calinski-Harabasz, and Davies-Bouldin are computed on the same deterministic 3,000-point sample for comparability [6]-[8].

Local hotspot ranking and evaluation

The local benchmark uses the ten highest-burden BIRCH centers. Each center is split chronologically into an 80% training period and a 20% held-out period. Training crashes define occupied 0.5 km by 0.5 km cells. The held-out period is used only for evaluation. Four methods rank the occupied training cells: CountGrid ranks by collision count; KDEGrid ranks by a Gaussian-smoothed count surface, following kernel-density hotspot practice [9]-[11]; DBSCANDensity ranks by a count-dominated local-neighborhood density score; and Spatial RAG combines count, severity burden, local density, AADF context, vulnerable-road-user share, junction share, major-road proximity, darkness, and wet-surface context. Capture@b is the share of held-out crashes captured by the top b percent of occupied cells, and SeverityCapture@b uses fatal=3, serious=2, and slight=1 weights. AUC10 summarizes the capture curve from 0% to 10%. Top-5% stability is measured through 20 bootstrap resamples, and paired center-level severity-AUC10 differences are tested with one-sided Wilcoxon signed-rank tests.

Countermeasure retrieval

The retrieval layer is deterministic. The top Spatial RAG footprint in each evaluated center is classified by morphology, position, and context. Compact cell groups are treated as point hotspots, elongated groups as corridor hotspots, and broader cell groups as area hotspots. The position label is central or peripheral relative to the BIRCH center. Additional context indicators include vulnerable-road-user share, junction share, major-road share, darkness share, and wet-surface share. The selected trigger returns one package from a closed safety library built from standard safety references [14]-[20]. The output is intended for first-pass scoping, not final design.

## Results and discussion

National clustering and center selection

Table 4 and Figure 3 summarize the national comparison. MiniBatchKMeans reaches the highest sampled silhouette, 0.460, in 2.94 s. BIRCH is nearly as fast at 3.07 s and reaches a sampled silhouette of 0.447. DBSCAN finishes in 3.79 s but produces 679 clusters, labels 29.9% of crashes as noise, and creates a 20,192-crash largest component. For downstream screening, BIRCH is used because it provides a hierarchical spatial partition that produces interpretable urban-center units while retaining comparable runtime and internal quality.

**Table 4.** National clustering comparison on the 2022 urban STATS19 benchmark.

| Method | Runtime (s) | Clusters | Noise ratio | Largest cluster | Median cluster size | Silhouette | Calinski-Harabasz | Davies-Bouldin |
|---|---|---|---|---|---|---|---|---|
| BIRCH | 3.07 | 204 | 0.000 | 10,139 | 153 | 0.447 | 10,116.7 | 0.545 |
| MiniBatch KMeans | 2.94 | 204 | 0.000 | 1,785 | 228 | 0.460 | 12,847.0 | 0.629 |
| DBSCAN | 3.79 | 679 | 0.299 | 20,192 | 17 | 0.071 | 2,093.5 | 0.349 |

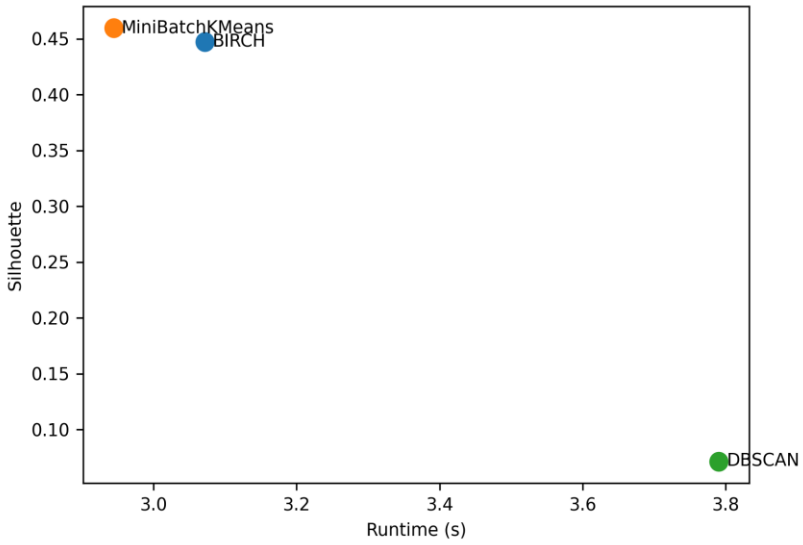


**Figure 3.** Runtime-quality trade-off of the national clustering baselines.

Table 5 and Figure 4 show the twelve highest-burden BIRCH centers. The largest center is a central-London cluster dominated by Lambeth-area collisions, with 10,139 crashes and a severity burden of 11,945. Other high-burden centers include Brent, Redbridge, Lewisham, Birmingham, Croydon, Sandwell, Leeds, Liverpool, and Enfield. The top ten centers form the local benchmark described in Table 6.

**Table 5.** Top twelve BIRCH centers ranked by severity-weighted burden.

| Cluster ID | Approx. local authority | Crash count | Severity burden | Fatal collisions | Serious collisions | Centroid lon | Centroid lat | BBox density | Mean nearest AADF | VRU collision share |
|---|---|---|---|---|---|---|---|---|---|---|
| 83 | Lambeth | 10,139 | 11,945 | 32 | 1,742 | -0.113 | 51.513 | 19.06 | 21,375 | 0.762 |
| 69 | Brent | 4,041 | 4,625 | 19 | 546 | -0.306 | 51.545 | 6.30 | 27,681 | 0.596 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 78 | Redbridge | 2,587 | 2,965 | 13 | 352 | 0.085 | 51.562 | 4.70 | 29,020 | 0.504 |
| 57 | Lewisham | 2,252 | 2,610 | 11 | 336 | 0.043 | 51.439 | 3.85 | 20,156 | 0.567 |
| 128 | Birmingham | 2,171 | 2,572 | 22 | 357 | -1.859 | 52.483 | 5.01 | 21,335 | 0.381 |
| 89 | Croydon | 2,069 | 2,458 | 9 | 371 | -0.150 | 51.380 | 6.15 | 19,920 | 0.579 |
| 60 | Sandwell | 1,892 | 2,256 | 10 | 344 | -2.048 | 52.537 | 3.07 | 17,842 | 0.409 |
| 76 | Leeds | 1,639 | 2,131 | 16 | 460 | -1.566 | 53.769 | 3.36 | 21,600 | 0.477 |
| 51 | Liverpool | 1,648 | 1,975 | 13 | 301 | -2.963 | 53.414 | 2.91 | 16,420 | 0.579 |
| 119 | Enfield | 1,371 | 1,577 | 6 | 194 | -0.122 | 51.633 | 4.23 | 25,414 | 0.557 |
| 31 | Manchester | 1,196 | 1,528 | 18 | 296 | -2.252 | 53.455 | 1.80 | 25,669 | 0.510 |
| 16 | Nottingham | 1,178 | 1,443 | 8 | 249 | -1.209 | 52.972 | 1.70 | 18,102 | 0.544 |

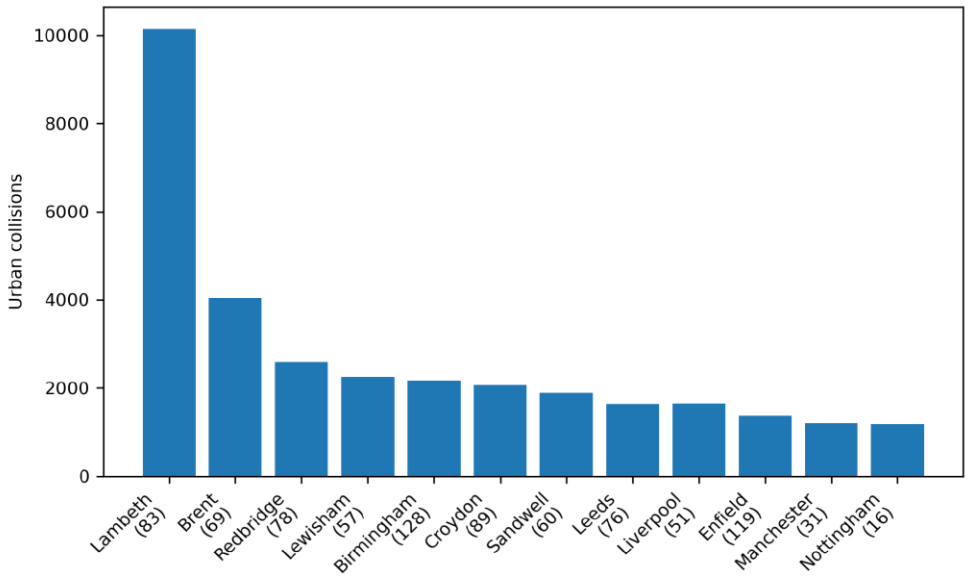


**Figure 4.** Top BIRCH centers by severity-weighted burden.

**Table 6.** Center-level benchmark geometry for the ten evaluated BIRCH centers.

| Center rank | Center ID | Center name | Total crashes | Train crashes | Test crashes | Train cells | Test cells | Mean nearest AADF | VRU collision share |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 83 | Lambeth | 10,139 | 8,111 | 2,028 | 999 | 739 | 21,375 | 0.762 |
| 2 | 69 | Brent | 4,041 | 3,232 | 809 | 842 | 460 | 27,681 | 0.596 |
| 3 | 78 | Redbridge | 2,587 | 2,069 | 518 | 553 | 298 | 29,020 | 0.504 |
| 4 | 57 | Lewisham | 2,252 | 1,801 | 451 | 647 | 305 | 20,156 | 0.567 |
| 5 | 128 | Birmingham | 2,171 | 1,736 | 435 | 608 | 298 | 21,335 | 0.381 |
| 6 | 89 | Croydon | 2,069 | 1,655 | 414 | 503 | 257 | 19,920 | 0.579 |
| 7 | 60 | Sandwell | 1,892 | 1,513 | 379 | 730 | 302 | 17,842 | 0.409 |
| 8 | 76 | Leeds | 1,639 | 1,311 | 328 | 544 | 231 | 21,600 | 0.477 |
| 9 | 51 | Liverpool | 1,648 | 1,318 | 330 | 580 | 237 | 16,420 | 0.579 |
| 10 | 119 | Enfield | 1,371 | 1,096 | 275 | 335 | 166 | 25,414 | 0.557 |

Local hotspot ranking

Table 7 and Figure 5 show the prospective screening results. At the 5% budget, DBSCANDensity captures 15.44% of held-out crashes and 15.03% of held-out severity burden. Spatial RAG captures 15.19% of held-out crashes and 14.87% of severity burden. CountGrid is similar, with 15.18% crash capture and 14.83% severity capture. KDEGrid is lower at 11.30% crash capture and 11.13% severity capture. The same pattern appears in AUC10: Spatial RAG and DBSCANDensity both reach about 0.1405 in severity-AUC10, compared with 0.1093 for KDEGrid.

**Table 7.** Local hotspot ranking performance under cell-review budgets.

| Method | Capture@1% | Capture@3% | Capture@5% | Capture@10% | SeverityCapture@1% | SeverityCapture@3% | SeverityCapture@5% | SeverityCapture@10% | AUC10 | SeverityAUC10 |
|---|---|---|---|---|---|---|---|---|---|---|
| CountGrid | 0.047 | 0.106 | 0.152 | 0.250 | 0.044 | 0.103 | 0.148 | 0.244 | 0.144 | 0.140 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| DBSCANDensity | 0.047 | 0.104 | 0.154 | 0.251 | 0.045 | 0.101 | 0.150 | 0.244 | 0.145 | 0.141 |
| KDEGrid | 0.033 | 0.076 | 0.113 | 0.205 | 0.033 | 0.075 | 0.111 | 0.201 | 0.111 | 0.109 |
| Spatial RAG | 0.047 | 0.107 | 0.152 | 0.251 | 0.045 | 0.103 | 0.149 | 0.245 | 0.144 | 0.141 |

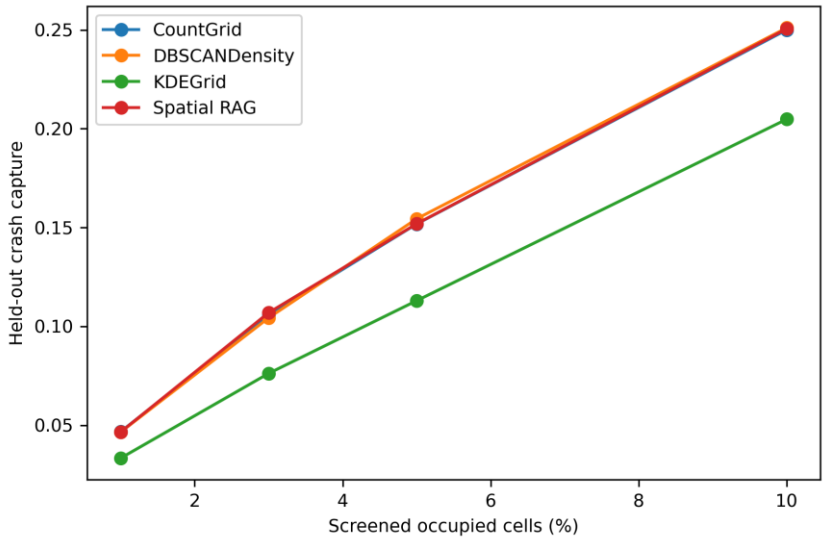


**Figure 5.** Severity-weighted capture curves for the local hotspot ranking methods.

The stability result in Table 8 explains the main trade-off. KDEGrid has the highest top-5% bootstrap Jaccard overlap, 0.724, because the smoothed surface changes less under resampling. That stability does not translate into better prospective capture. The paired Wilcoxon results in Table 9 show that CountGrid, DBSCANDensity, and Spatial RAG each exceed KDEGrid in center-wise severity-AUC10 with p=0.001953. Table 10 confirms that the advantage is broad across centers, although Liverpool is the one center where KDEGrid is close to or above the count-dominated methods.

**Table 8.** Bootstrap stability of the top-5% screened cells.

| Method | Top-5% Jaccard |
|---|---|
| CountGrid | 0.471 |
| DBSCANDensity | 0.543 |
| KDEGrid | 0.724 |
| Spatial RAG | 0.496 |

**Table 9.** One-sided Wilcoxon signed-rank tests on center-level severity-AUC10.

| Method A | Method B | Mean Δseverity-AUC10 | Wilcoxon statistic | One-sided p-value |
|---|---|---|---|---|
| CountGrid | KDEGrid | 0.031 | 54.0 | 0.0020 |
| DBSCANDensity | KDEGrid | 0.031 | 54.0 | 0.0020 |
| Spatial RAG | KDEGrid | 0.031 | 54.0 | 0.0020 |

**Table 10.** Center-wise severity-AUC10 values for the four local methods.

| Center rank | Center | CountGrid | DBSCANDensity | KDEGrid | Spatial RAG |
|---|---|---|---|---|---|
| 1 | Lambeth (83) | 0.122 | 0.124 | 0.101 | 0.123 |
| 2 | Brent (69) | 0.162 | 0.163 | 0.115 | 0.163 |
| 3 | Redbridge (78) | 0.138 | 0.141 | 0.111 | 0.135 |
| 4 | Lewisham (57) | 0.146 | 0.142 | 0.098 | 0.148 |
| 5 | Birmingham (128) | 0.150 | 0.137 | 0.118 | 0.147 |
| 6 | Croydon (89) | 0.154 | 0.157 | 0.114 | 0.161 |
| 7 | Sandwell (60) | 0.102 | 0.103 | 0.083 | 0.102 |
| 8 | Leeds (76) | 0.155 | 0.159 | 0.140 | 0.154 |
| 9 | Liverpool (51) | 0.120 | 0.131 | 0.134 | 0.121 |
| 10 | Enfield (119) | 0.153 | 0.148 | 0.078 | 0.151 |

Table 11 shows the local-score ablation. Adding severity and density context changes the frontier only modestly. The best mean severity-AUC10 is obtained by Count + Severity + Density, 0.1406, and the full Spatial RAG score reaches 0.1405. This indicates that the leading cells are still strongly determined by count and local density, while the extra Spatial RAG variables are most valuable for describing and retrieving a treatment package after the high-burden cells have been identified.

**Table 11.** Ablation of the Spatial RAG local scoring function.

| Variant | AUC10 | SeverityAUC10 | Capture@5% | SeverityCapture@5% |
|---|---|---|---|---|
| Count | 0.144 | 0.140 | 0.152 | 0.148 |
| Count + Severity | 0.143 | 0.139 | 0.150 | 0.146 |
| Count + Severity + Density | 0.144 | 0.141 | 0.152 | 0.149 |
| Full Spatial RAG | 0.144 | 0.141 | 0.152 | 0.149 |

Exposure and roadway context

Exposure-normalized results in Table 12 and Figure 6 add a safety-planning perspective that raw hotspot maps do not provide. Hackney, Westminster, Islington, Haringey, Camden, the City of London, Lambeth, and Southwark have the highest severity burden per 100 million vehicle-miles among the authorities shown. These results should not replace site-level diagnosis, but they help separate high crash counts caused by high traffic volume from high crash burden relative to local exposure.

**Table 12.** Highest local-authority severity burden rates after traffic-exposure normalization.

| Local authority | Collisions | Severity burden | Fatal | Serious | VRU collisions | Crashes / 100M vehicle-miles | Severity burden / 100M vehicle-miles |
|---|---|---|---|---|---|---|---|
| Hackney | 804 | 949 | 1 | 143 | 635 | 299.6 | 353.6 |
| Westminster | 1,287 | 1,550 | 6 | 251 | 1,007 | 289.4 | 348.5 |
| Islington | 626 | 734 | 2 | 104 | 514 | 277.5 | 325.4 |
| Haringey | 906 | 1,021 | 2 | 111 | 666 | 285.3 | 321.5 |
| Camden | 743 | 860 | 3 | 111 | 599 | 268.7 | 311.0 |
| City of London | 173 | 232 | 1 | 57 | 141 | 231.0 | 309.7 |
| Lambeth | 1,131 | 1,353 | 5 | 212 | 864 | 248.8 | 297.6 |
| Southwark | 1,062 | 1,243 | 3 | 175 | 822 | 248.8 | 291.2 |
| Wandsworth | 969 | 1,165 | 3 | 190 | 758 | 213.4 | 256.6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Kensington and Chelsea | 611 | 727 | 4 | 108 | 479 | 201.1 | 239.2 |
| Tower Hamlets | 1,038 | 1,210 | 4 | 164 | 743 | 203.1 | 236.8 |
| Hammersmith and Fulham | 559 | 670 | 2 | 107 | 442 | 184.4 | 221.0 |

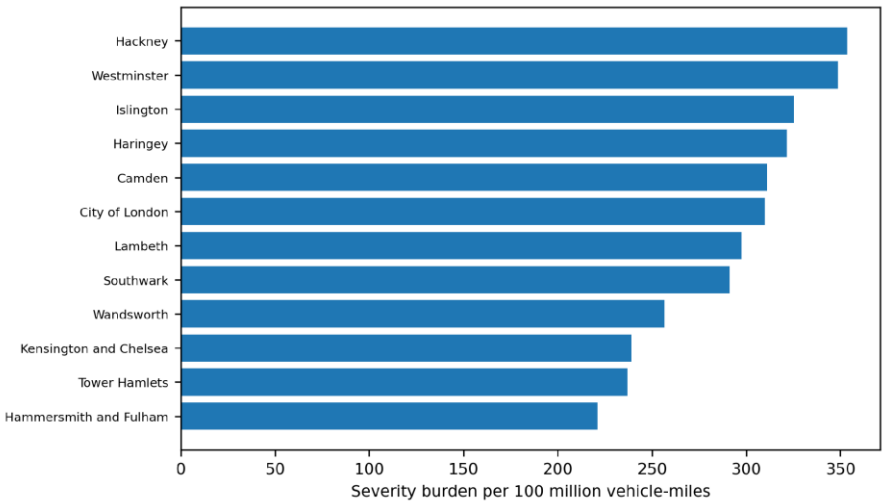


**Figure 6.** Highest local-authority severity burden rates after exposure normalization.

Recommendation retrieval

Table 13 reports the highest-ranked Spatial RAG hotspot from each evaluated center, and Figure 7 maps the top cells in the highest-burden center. The selected Lambeth-area footprint is an area-central hotspot with high vulnerable-road-user and junction shares, so it retrieves the urban vulnerable-road-user area-wide package. Across the ten hotspots, Figure 8 shows that eight retrieve the urban vulnerable-road-user area-wide package and two retrieve the major-road corridor speed-management package. Table 14 lists the trigger classes and representative measures used by the retrieval layer.

**Table 13.** Highest-ranked Spatial RAG hotspots and their retrieved recommendation packages.

| Center rank | Approx. local authority | Hotspot crash count | Severity burden | Spatial RAG score | Morphology | Position | VRU share | Junction share | Major-road share | Recommendation |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Lambeth | 1,344 | 1,589 | 0.624 | area | central | 0.792 | 0.714 | 0.987 | Urban VRU area-wide |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | package |
| 2 | Brent | 659 | 760 | 0.566 | area | peripheral | 0.693 | 0.634 | 0.979 | Major-road corridor speed-management package |
| 3 | Redbridge | 405 | 463 | 0.585 | area | central | 0.543 | 0.531 | 0.928 | Urban VRU area-wide package |
| 4 | Lewisham | 350 | 403 | 0.599 | area | peripheral | 0.666 | 0.680 | 0.931 | Major-road corridor speed-management package |
| 6 | Croydon | 331 | 394 | 0.599 | area | central | 0.665 | 0.619 | 0.961 | Urban VRU area-wide package |
| 5 | Birmingham | 312 | 381 | 0.535 | area | central | 0.417 | 0.519 | 0.766 | Urban VRU area-wide package |
| 8 | Leeds | 258 | 335 | 0.472 | area | central | 0.554 | 0.682 | 0.841 | Urban VRU area-wide package |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 7 | Sandwell | 246 | 292 | 0.512 | area | central | 0.423 | 0.602 | 0.841 | Urban VRU area-wide package |
| 9 | Liverpool | 232 | 280 | 0.538 | area | central | 0.668 | 0.556 | 0.935 | Urban VRU area-wide package |
| 10 | Enfield | 188 | 211 | 0.739 | area | central | 0.622 | 0.644 | 0.888 | Urban VRU area-wide package |

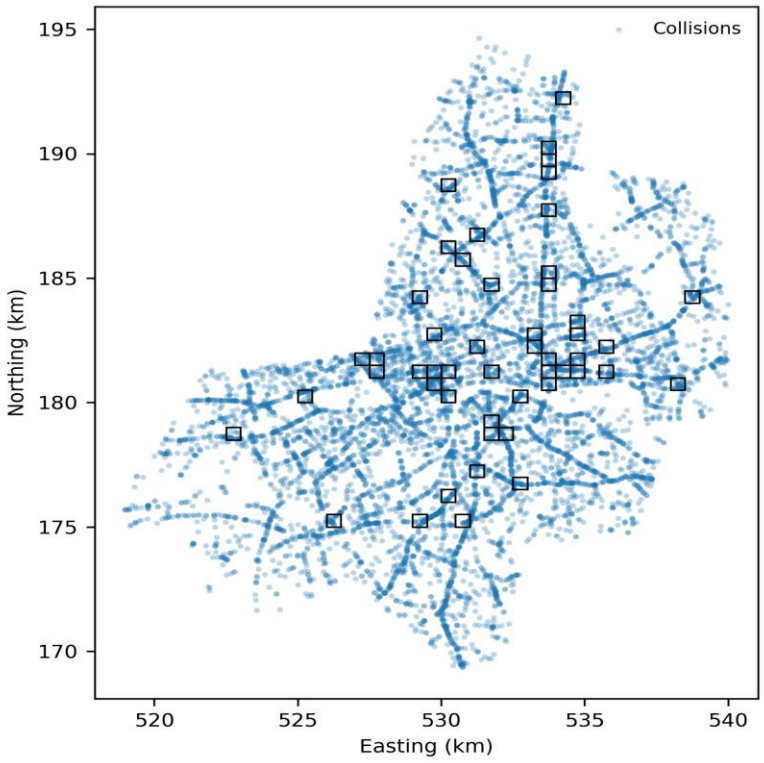


**Figure 7.** Top-5% Spatial RAG cells in the highest-burden BIRCH center.

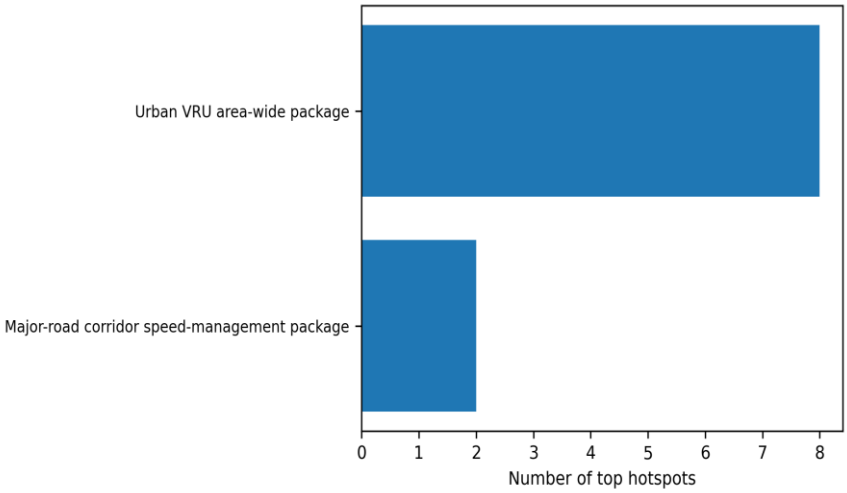


**Figure 8.** Distribution of retrieved countermeasure packages among the top ten hotspots.

**Table 14.** Deterministic countermeasure library used by the retrieval layer.

| Trigger class | Retrieved package | Representative measures |
|---|---|---|
| point or high-junction hotspot | Intersection diagnostic package | signal timing and protected-turn review; high-visibility markings; skid resistance and sight-distance checks |
| corridor or high-major-road-share hotspot | Major-road corridor speed-management package | speed management; crossing/refuge spacing; access management; signing and delineation review |
| central area with high vulnerable-road-user involvement | Urban VRU area-wide package | area speed management; pedestrian-priority timing; curb extensions; cycle and crossing visibility upgrades |
| darkness or wet-surface concentration | Visibility and surface-condition package | lighting audit; drainage and surface-friction review; reflective signing and marking refresh |
| central area without a single point pattern | Central area-wide safety package | network-level speed management; conflict reduction at frequent junctions; pedestrian visibility improvements |
| peripheral area or dispersed district pattern | District-scale systemic package | screen similar sites; low-cost signing and marking refresh; maintenance and lighting audit |

The recommendation frequencies indicate that the leading 2022 urban burden in this benchmark is mostly an area-level vulnerable-road-user problem rather than a set of isolated point conflicts. This is consistent with the high VRU shares in Table 13 and with the known relationship between speed, exposure, and injury risk [14]. The two major-road corridor packages also show why roadway context is useful: a broad area footprint can still require corridor-oriented treatments when the selected cells are closely tied to high-flow major-road links.

## Limitations

The analysis uses reported personal-injury collisions and therefore does not include damage-only crashes or unreported injury events. The exposure linkage is also approximate: nearest AADF count points and local-authority motor-vehicle miles are useful for screening, but they do not replace turning counts, pedestrian volumes, signal timing, lane configuration, or field-observed conflicts. The benchmark covers one calendar year, so it supports prospective held-out screening within 2022 but not a multi-year before-after safety evaluation. Finally, the retrieval layer recommends package-level treatment concepts. It does not estimate crash modification factors for a specific design or replace a site audit, benefit-cost analysis, or engineering judgment.

## Conclusion

This paper presented a Spatial RAG framework for urban crash hotspot discovery and safety countermeasure recommendation using integrated 2022 Great Britain road safety and traffic data. The national comparison shows that BIRCH and MiniBatchKMeans both provide fast 204-center partitions, while DBSCAN is less useful as a national partition because of its high noise share and very large largest component. The local benchmark shows that count-dominated methods and Spatial RAG outperform KDEGrid in severity-weighted prospective capture under constrained review budgets. KDEGrid is more stable under bootstrap resampling, but its smoothing lowers held-out capture. The added value of Spatial RAG is therefore not to discard simple hotspot ranking. Its value is that it keeps the strong screening frontier and then attaches a transparent, evidence-grounded countermeasure package to each selected hotspot.

## References

[1] Department for Transport, Road Safety Data: Collisions, Vehicles and Casualties, 2022. London, U.K.: Department for Transport, 2023.

[2] Department for Transport, GB Road Traffic Counts: AADF, Local Authority Traffic and Major Road Link Network 2022. London, U.K.: Department for Transport, 2023.

[3] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, 'A density-based algorithm for discovering clusters in large spatial databases with noise,' in Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining, 1996, pp. 226-231.

[4] T. Zhang, R. Ramakrishnan, and M. Livny, 'BIRCH: An efficient data clustering method for very large databases,' in Proc. ACM SIGMOD Int. Conf. Management of Data, 1996, pp. 103-114.

[5] D. Sculley, 'Web-scale k-means clustering,' in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 1177-1178.

[6] Shilu He, Haowei Tu, and Isa Liu, “Safe PD Capacity Forecasting with Time-Series Foundation Models and Calibrated Uncertainty for Heterogeneous GPU Clusters”, JACS, vol. 3, no. 4, pp. 48–66, Apr. 2023, doi: 10.69987/JACS.2023.30404.

[7] T. Calinski and J. Harabasz, 'A dendrite method for cluster analysis,' Commun. Stat., vol. 3, no. 1, pp. 1-27, 1974.

[8] D. L. Davies and D. W. Bouldin, 'A cluster separation measure,' IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-1, no. 2, pp. 224-227, 1979.

[9] T. K. Anderson, 'Kernel density estimation and K-means clustering to profile road accident hotspots,' Accid. Anal. Prev., vol. 41, no. 3, pp. 359-364, 2009.

[10] Z. Xie and J. Yan, 'Kernel density estimation of traffic accidents in a network space,' Comput. Environ. Urban Syst., vol. 32, no. 5, pp. 396-406, 2008.

[11] Z. Xie and J. Yan, 'Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated

approach,' J. Transp. Geogr., vol. 31, pp. 64-71, 2013.
[12] A. Montella, 'A comparative analysis of hotspot identification methods,' Accid. Anal. Prev., vol. 42, no. 2, pp. 571-581, 2010.
[13] Z. Cheng, Z. Zu, and J. Lu, 'Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification,' Sustainability, vol. 11, no. 1, Art. no. 160, 2019.
[14] L. Aarts and I. van Schagen, 'Driving speed and the risk of road crashes: A review,' Accid. Anal. Prev., vol. 38, no. 2, pp. 215-224, 2006.
[15] R. Elvik, A. Hoye, T. Vaa, and M. Sorensen, The Handbook of Road Safety Measures, 2nd ed. Bingley, U.K.: Emerald, 2009.
[16] American Association of State Highway and Transportation Officials, Highway Safety Manual, 1st ed. Washington, DC, USA: AASHTO, 2010.
[17] Federal Highway Administration, Systemic Safety Project Selection Tool. Washington, DC, USA: U.S. Department of Transportation, 2013.
[18] Federal Highway Administration, A Guide to Developing Quality Crash Modification Factors, Report FHWA-SA-10-032. Washington, DC, USA: U.S. Department of Transportation, 2010.
[19] Daren Zheng, Chenyu Li, and Harvey Davidson, "Continual Red-Teaming for In-the-Wild Jailbreaks via Online Guardrail Updates and Guardrail Distillation", JACS, vol. 3, no. 2, pp. 35–49, Feb. 2023, doi: 10.69987/JACS.2023.30203.
[20] Federal Highway Administration, Proven Safety Countermeasures. Washington, DC, USA: U.S. Department of Transportation, 2021.
[21] Kai Zhang, Siquan Meng, and Eric Zhou, "Evidence-Grounded Trading Desk Risk Memos over SEC Filings: Retrieval-Augmented Generation with XBRL Numeric Verification", JACS, vol. 3, no. 2, pp. 60–76, Feb. 2023, doi: 10.69987/JACS.2023.30205.
[22] T. B. Brown et al., 'Language models are few-shot learners,' in Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 1877-1901.
[23] L. Ouyang et al., 'Training language models to follow instructions with human feedback,' in Adv. Neural Inf. Process. Syst., vol. 35, 2022, pp. 27730-27744.
[24] Xiaohan Chang, Tong Ye, and Sophia Luo , "LLM-as-Reranker for Personalized Recommendation: Popularity Bias Mitigation and Faithful Natural-Language Explanations on MovieLens 100K", JACS, vol. 3, no. 8, pp. 61–78, Aug. 2023, doi: 10.69987/JACS.2023.30806.