Data in Brief 62 (2025) 111914

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Pavement condition and climatic data in southeast Texas: A dataset for evaluating flood impacts on pavement performance

Hossein Hariri Asli[a], Nicholas Brake[a,*], Mahdi Feizbahr[a], Unique Bakhrel[a], Homayoon Arbabkhah[b], Mohammad Sina Jahangir[c]

[a] Department of Civil and Environmental Engineering, Lamar University, Beaumont, TX 77710, United States
[b] Department of Industrial and Systems Engineering, Lamar University, Beaumont, TX 77710, United States
[c] Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Canada

ARTICLE INFO

Article history:
Received 25 March 2025
Revised 15 June 2025
Accepted 16 July 2025
Available online 22 July 2025

Dataset link: https://zenodo.org/records/15638682
Dataset link: https://data.mendeley.com/datasets/p6vg4v7f9k/2
Dataset link: Comprehensive Dataset for Data-Driven Pavement Performance Prediction and Analysis in Flood-Prone Beaumont, Southeast Texas (Original data)

Keywords:
Pavement performance
Data-driven modeling
Pavement distress
Flood-Prone Infrastructure
Southeast Texas

ABSTRACT

Effective pavement maintenance is essential for economic stability, optimal network performance, and roadway safety. Achieving this requires thorough evaluation of pavement conditions, including structural integrity, surface roughness, and distress characteristics. Pavement performance indicators play a critical role in influencing vehicle safety and ride quality. Recent advances have emphasized the use of data-driven modeling to anticipate pavement behavior, with the goal of optimizing resource allocation and refining Maintenance and Rehabilitation (M&R) strategies through accurate condition assessment. A foundational requirement for these modeling efforts is the availability of standardized, high-quality datasets that can support robust and reproducible infrastructure analysis. This data article presents a comprehensive dataset assembled to facilitate pavement performance prediction, with a geographic focus on Southeast Texas, particularly the flood-vulnerable area of Beaumont. The dataset encompasses pavement and traffic attributes, meteorological records, flood simulation outputs, ground deformation measurements, and topographic indices, enabling detailed

* Corresponding author.
E-mail address: nbrake@lamar.edu (N. Brake).

https://doi.org/10.1016/j.dib.2025.111914

examination of both load-associated and non-load-associated degradation mechanisms. Data preprocessing was performed using ArcGIS Pro, Microsoft Excel, and Python to ensure consistency and usability in data-driven modeling applications, including machine learning workflows. Key contributions of this dataset include its utility in analyzing the climatic and environmental factors affecting pavement conditions, identifying critical predictive features, and enabling in-depth correlation analysis across diverse variables. By filling existing gaps in input variable selection resources, this dataset supports the development of predictive tools for estimating future maintenance demand and enhancing the resilience of pavement networks in flood-impacted areas. The resource highlights the importance of standardized datasets for advancing pavement management practices and provides a robust foundation for ongoing infrastructure performance modeling.

Specifications Table

Subject	Engineering & Materials science
Specific subject area	Integrated datasets for predicting pavement performance in flood-prone areas using pavement condition, climate, flood, and terrain data to enhance management strategies.
Type of data	Table, Geospatial file (Shapefile), Image, Figure Filtered, Processed
Data collection	Pavement condition data were collected using a Laser Road Surface Tester (ASTM D6433–11) [1,2] to record International Roughness Index (IRI), Pavement Condition Index (PCI), and surface distresses. High-resolution imagery and spatial coordinates were also captured. Traffic data and pavement specifications were obtained from the City of Beaumont's Pavement Management System (PMS) dataset [3,4]. Climatic variables came from Parameter-elevation Regressions on Independent Slopes Model (PRISM), synthetic models, and Sentinel-1 radar. Road segments were delineated in ArcGIS Pro using Geographic Information System Identifiers (GISIDs) [5,6]. All datasets were cleaned, processed, and normalized in Python and Microsoft Excel for integration into spatial and predictive modeling workflows.
Data source location	• Institution: Lamar University, Center for Resiliency (CfR) • Agency: Pavement Management System (PMS) data, City of Beaumont • Region: Southeast Texas • County: Jefferson • Country: United States of America
Data accessibility	Zenodo [7] DOI: 10.5281/zenodo.1563868 https://doi.org/10.5281/zenodo.15638682 Data Excel (XLSX): https://doi.org/10.5281/zenodo.15638682 ArcGIS (Spatial features and attribute table): https://doi.org/10.5281/zenodo.15638682 Mendeley Data [8] DOI: 10.17632/p6vg4v7f9k.2 https://data.mendeley.com/datasets/p6vg4v7f9k/2 Data Excel (XLSX): https://data.mendeley.com/datasets/p6vg4v7f9k/2 ArcGIS (Spatial features and attribute table): https://data.mendeley.com/datasets/p6vg4v7f9k/2
Related research article	'none'.

1. Value of the Data

- This dataset integrates pavement condition indices, meteorological data, flood maps, and topographic indices, enabling advanced Machine Learning models to predict pavement degradation and optimize Maintenance and Rehabilitation (M&R) strategies. It also facilitates robust model-independent Input Variable Selection methods to identify critical variables influencing pavement performance, improving model accuracy, efficiency, and interpretability.
- By incorporating nuisance flood maps, historical records, experimental data, synthetic climatic models, and topographic indices, the dataset enables precise assessment of flooding's short- and long-term impacts on pavement. It fills gaps in understanding environmental effects on pavement performance, supporting predictive tools and methodologies to enhance infrastructure resilience in flood-prone regions.
- The data has been preprocessed using ArcGIS Pro, Python, and Microsoft Excel to ensure standardization and readiness for direct application in data-driven modeling, and Machine Learning approaches. The dataset supports comprehensive data analysis, allowing researchers to explore correlations and trends among input variables, enhancing understanding of key factors affecting pavement performance.
- The dataset is valuable for large-scale evaluations, assisting transportation agencies and emergency managers in prioritizing M&R efforts, managing resources, and planning road closures, especially in flood-prone areas like Southeast Texas.
- By leveraging multiple data sources and preprocessing techniques, the dataset enhances the accuracy of pavement performance models, improving decision-making processes for transportation planners and policymakers.
- The integration of diverse data sources in this dataset makes it a powerful tool for studying pavement deterioration under varying environmental and traffic conditions, providing insights that can improve long-term infrastructure sustainability.

2. Background

This dataset was compiled to support data-driven pavement performance modeling with a particular focus on flood-prone urban regions. The motivation emerged from growing infrastructure resilience challenges in Southeast Texas, where intense rainfall and recurring flooding have accelerated pavement degradation. Recognizing the limitations of isolated condition assessments, the dataset was designed to integrate diverse sources of information, including pavement inspection records, traffic characteristics, climate variables, flood simulation outputs, and topographic indicators. A standardized, segment-based framework was developed using GIS techniques to ensure alignment across spatial and tabular data. The theoretical foundation stems from advancements in predictive modeling and input variable selection methods, which require comprehensive and well-structured datasets for effective training and validation. To this end, data were collected, processed, and organized to facilitate reproducibility and direct use in machine learning applications. If used alongside the related research article on flood-related pavement deterioration and variable screening, this dataset offers added transparency and supports extended experimentation, including alternate feature selection pipelines and regional adaptation studies.

3. Data Description

The dataset provides comprehensive information essential for analyzing pavement performance and its interaction with flooding events, particularly within the Beaumont Road network. It is organized into two main categories: Excel files containing tabular datasets and GIS files for spatial analysis in ArcGIS.

Excel Files

The dataset includes ten Excel files, each capturing critical variables related to pavement conditions, specification, climate, maintenance, topography, and flooding:

- Pavement Inspection 2019.xlsx and Pavement Inspection 2023.xlsx: Contain the road segments inspection data for two different years (2019 and 2023). These files include pavement condition indices, distress measurements (such as cracks, rutting, and potholes), structural attributes, M&R activities history, performance indicators, GISIDs, and segment addresses.
- Flood Susceptibility.xlsx: Provides flood risk assessment metrics, including flood susceptibility scores for different road segments.
- Ground Deformation.xlsx: Contains remote sensing-derived ground deformation data, indicating possible pavement displacements and subsurface geotechnical issues.
- Maintenance and Rehabilitation Activities.xlsx: Documents historical M&R activities, including repair dates, methods, and affected pavement sections.
- Segmented Streets Frequency of Flooding.xlsx: Details the frequency of past flood inundation events affecting individual road segments.
- Segmented Streets Precipitation and Temperature.xlsx: Contains climate variables, including precipitation and temperature, to assess their impact on pavement performance.
- Segmented Streets Synthetic Flood Model Depth_Speed_Elevation_Max.xlsx and Segmented Streets Synthetic Flood Model Depth_Speed_Elevation_Mean.xlsx: Provide synthetic flood model outputs, including floodwater depth, velocity, and elevation under different flooding scenarios.
- Segmented Streets Topographic Control Index_Max.xlsx: Includes topographic indices that influence water flow and flood accumulation patterns.

Each file is formatted for integration into data-driven modeling, supporting Machine Learning and statistical approaches to pavement performance prediction.

GIS Files

The GIS component comprises the spatial road network and attribute files, supporting localized and network-wide geospatial analysis:

Segmented Road Network.shp, .shx, .dbf, .prj, .sbn, .sbx, .cpg, .shp.xml: Represent the segment geometries, spatial references, metadata, and attribute fields corresponding to Excel variables via GISID.

The spatial and tabular datasets are cross-compatible, enabling integrated analysis across platforms (e.g., ArcGIS, Python). Fig. 1 summarizes the structure and content of the dataset.

3.1. General information

This data provides an overview of the key attributes included in the dataset. These attributes are essential for describing the roadway segments, their location, and their unique identifiers within the geographic and pavement management systems.

3.1.1. GISID

Each roadway segment is uniquely identified by a GISID, which links the segment across all spatial and tabular datasets. This identifier enables block-level segmentation, ensuring precise alignment between pavement condition data and geospatial features.

3.1.2. On street

Indicates the name of the street on which the roadway segment is located. This attribute provides critical information for identifying the primary street associated with each segment.

3.1.3. From street

Denotes the cross street or reference point where a roadway segment begins. This attribute helps establish the starting location of each segment within the network.

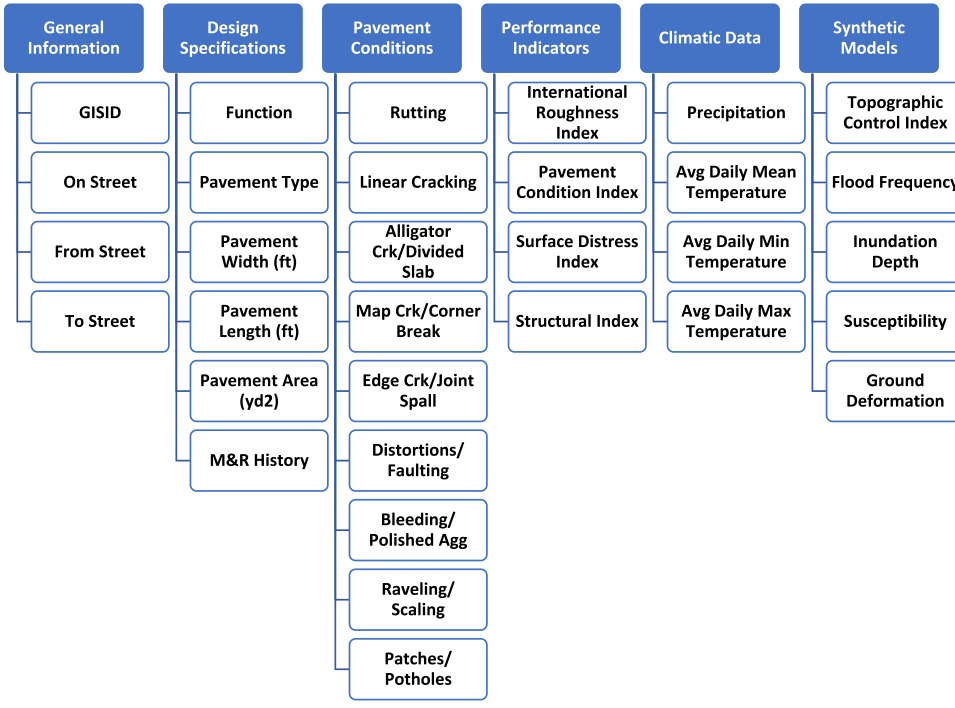

Fig. 1. Overview of the dataset organization.

3.1.4. To street

Marks the intersecting street where the segment ends, establishing the downstream boundary for analysis and management. Fig. 2. demonstrates the segmented road network of Beaumont as displayed in ArcGIS Pro 3, with each segment's attributes linked to a unique GISID.

3.2. Pavement data

This section describes core pavement attributes critical for evaluating structural condition and predicting future performance using machine learning models.

3.2.1. Summary of the pavement design features

3.2.1.1. Function. This attribute categorizes roads based on their functional classification and function codes, such as Arterial (1), Collector (2), or Local (3). The pavement function determines traffic load, frequency of use, and priority for maintenance. Arterials (1) typically carry the highest traffic volume (vehicles/day) and experience faster deterioration, while Collectors (2) serve as intermediate roads connecting arterials to locals, and Locals (3) handle lighter traffic loads, primarily serving residential areas. The units for this attribute are categorical.

3.2.1.2. Pavement type. Identifies the material composition of the pavement, typically Asphalt (pavement code 1), Concrete (pavement code 2), or Composite (pavement code 3). Pavement type influences structural performance, lifecycle costs, and maintenance strategies. Asphalt is more flexible and cost-effective for repairs, while concrete is more durable but costlier to install [9]. The units for this attribute are categorical. Table 1 summarizes the types and functions of the segmented road network in Beaumont.

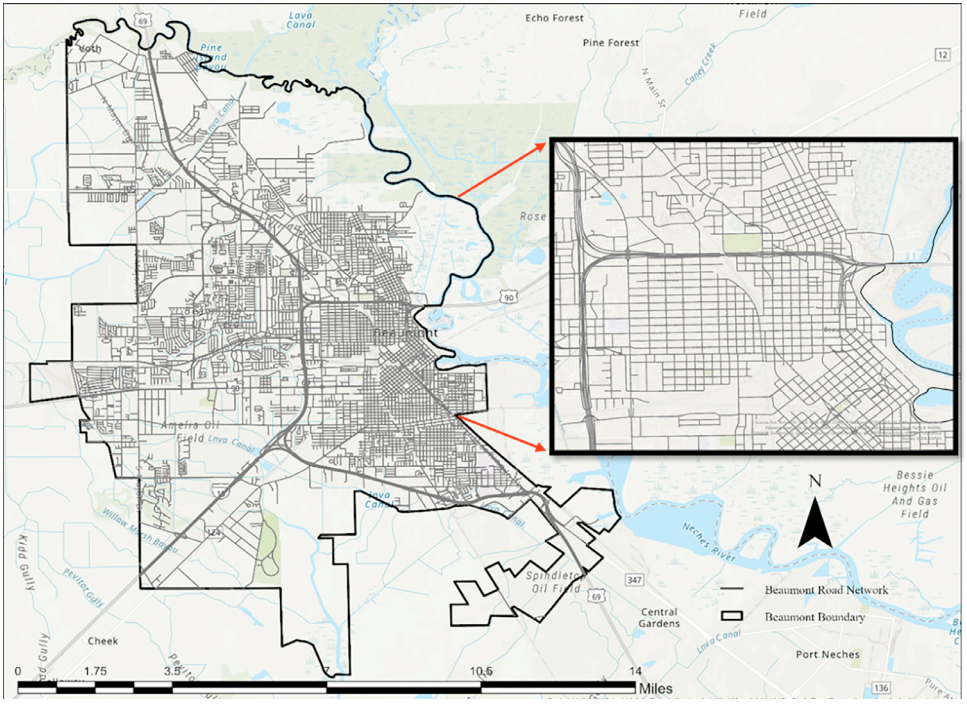

Fig. 2. Segmented road network of Beaumont with unique GISIDs in ArcGIS Pro 3.

Table 1
Network inventory summary by functional class and pavement type.

	Pave type	Network	Arterial	collector	local
Segment Count	All Streets	13,974	1642	1296	11,036
	Asphalt	8752	636	828	7288
	Concrete	5222	1006	468	3748
Network Length (mi)	All Streets	667.1	80.4	90.5	496.2
	Asphalt	408.6	26	55.8	326.7
	Concrete	258.5	54.4	34.7	169.5

3.2.1.3. Pavement width (ft). Represents cross-sectional width, influencing the total area and structural capacity of the road. Wider pavements accommodate higher traffic volumes and larger vehicles, which may require enhanced structural designs. The units for this attribute are in feet (ft).

3.2.1.4. Pavement length (ft). Specifies the longitudinal extent of a road segment from one endpoint to another. Pavement length is used to calculate the total pavement area and plays a role in cost estimation for maintenance and rehabilitation. The units for this attribute are in feet (ft).

3.2.1.5. Pavement area (yd^2). Computed as length × width × adjustment factor, essential for quantifying materials and M&R costs. The units for this attribute are in square yards (yd^2).

3.2.1.6. Pavement thickness (in). Vertical depth of structural layers (surface to sub-base), determining load-bearing capacity and deformation resistance (inch). Fig. 3 shows Beaumont's geocoded pavement coring map on ArcGIS Pro, highlighting the locations of interest for identifying pavement layer thicknesses and types.

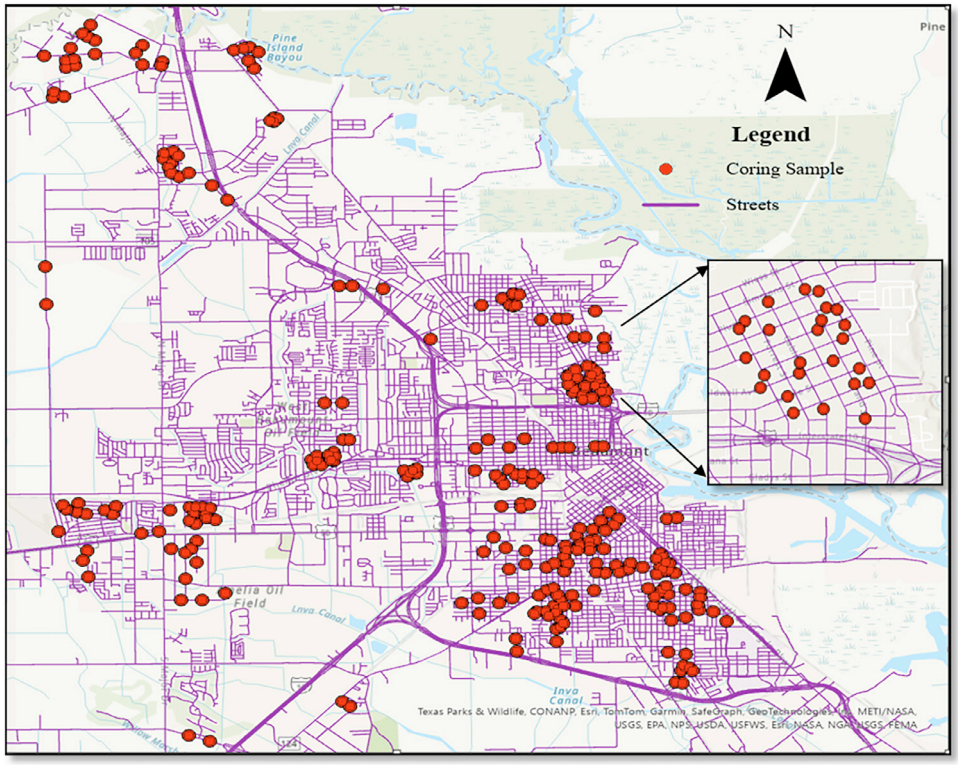

Fig. 3. Beaumont's geo-coded pavement coring map.

3.2.1.7. Maintenance and rehabilitation (M&R) history. Categorical variable describing past interventions to restore or extend pavement life. Activities vary by pavement type and are selected based on PCI thresholds. For example, asphalt with PCI 85–100 undergoes routine maintenance; PCI <15 may require full-depth reconstruction. For concrete pavements, localized rehabilitation and grinding are applicable for PCI values between 70 and 80, while extensive panel replacement is performed when PCI values fall between 40 and 54. Partial or full-depth reconstruction is necessary for pavements with PCI values below 15 [4]. Table 2 summarizes recommended M&R strategies by pavement type and condition.

These attributes form the analytical foundation for performance modeling and strategy optimization in pavement management systems.

3.2.2. Summary of the pavement condition and distress features

Pavement condition and distress metrics are essential for assessing road performance and developing sustainable maintenance strategies. This dataset incorporates several condition indices, namely, Surface Distress Index (SDI), IRI, PCI, and Structural Index (SI), collected using a Laser Road Surface Tester (Laser RST) in accordance with ASTM D6433 [1].

The Laser RST conducted single-pass surveys for local streets and dual-pass surveys for higher classification roads. Distress data were processed and uploaded into pavement management software using a standardized 0–10 severity scale, aligned with each distress type's impact on SDI. This uniform scale allows consistent classification across severity levels and supports targeted M&R planning.

Both load-associated (e.g., rutting, alligator cracking) and non-load-associated (e.g., raveling, bleeding) distresses were recorded. The following subsections outline the major dis-

Table 2
Categorization of M&R activities for Asphalt, composite, and concrete pavements.

Pavement Type	Rehab Code	Rehab Activity
Asphalt	5	Routine Maintenance
Asphalt	10	Slurry Seal / Seal Coat
Asphalt	20	MicroSurface / Chip Seal
Asphalt	23	MicroSurface / Chip Seal + Structural Patching
Asphalt	26	MicroSurface / Chip Seal + Structural Patching
Asphalt	30	Edge Mill + Thin Overlay (1.5 - 2.0)
Asphalt	33	Edge Mill + Thin Overlay (1.5 - 2.0) + Structural Patching
Asphalt	36	Edge Mill + Thin Overlay (1.5 - 2.0) + Structural Patching
Asphalt	40	EM/FWM + Moderate Overlay (2.0 - 3.0)
Asphalt	43	EM/FWM + Moderate Overlay (2.0 - 3.0) + Structural Patching
Asphalt	46	EM/FWM + Moderate Overlay (2.0 - 3.0) + Structural Patching
Asphalt	50	FWM + Thick Overlay (> 2.0 - 3.0)
Asphalt	53	FWM + Thick Overlay (> 2.0 - 3.0) + Structural Patching
Asphalt	56	FWM + Thick Overlay (> 2.0 - 3.0) + Structural Patching
Asphalt	60	Surf Recon + Base Rehab / FWM + Structural Patching + Overlay
Composite	65	Surf Recon + PCC to Base/FWM + Structural Patching + Overlay
Asphalt	70	ACP Full Depth Reconstruction
Composite	75	Full Depth Recon + PCC to Base
Concrete	510	PCC Joint Rehab & Crack Seal
Concrete	520	PCC Localized Rehab
Concrete	523	PCC Localized Rehab + Grind
Concrete	530	PCC Slight Panel Replacement (<10 %)
Concrete	533	PCC Slight Panel Replacement (<10 %) + Grind
Concrete	540	PCC Moderate Panel Replacement (<20 %)
Concrete	543	PCC Moderate Panel Replacement (<20 %) + Grind
Concrete	550	PCC Extensive Panel Replacement (<33 %)
Concrete	553	PCC Extensive Panel Replacement (<33 %) + Grind
Concrete	560	PCC Partial Reconstruction
Concrete	570	PCC Full Depth Reconstruction

"EM/FWM" refers to Edge Mill/Flexible Wear Maintenance.
"FWM" refers to Flexible Wear Maintenance.
"PCC" refers to Portland Cement Concrete.
"ACP" refers to Asphalt Concrete Pavement.
"Structural Patching" involves localized repair of deteriorated pavement sections before overlay application.

tress types, including measurement criteria and implications for pavement performance modeling.

3.2.2.8. Rutting. Permanent depression along wheel paths due to repeated traffic loading. It is a major contributor to the SDI and significantly affects pavement condition ratings. Rutting negatively affects ride quality and increases hydroplaning risk, often requiring resurfacing or full-depth reconstruction [1].

3.2.2.9. Linear cracking (Longitudinal, transverse, and block cracking). Linear cracking encompasses longitudinal and transverse cracks, often caused by thermal fluctuations, shrinkage, and construction deficiencies. Block cracking, a form of interconnected transverse and longitudinal cracks, typically results from aged or oxidized pavement. These cracks, if left untreated, may propagate into fatigue cracking or allow moisture infiltration, exacerbating subgrade deterioration. Their presence informs early maintenance actions such as sealing or overlays [1].

3.2.2.10. Alligator cracking/divided slab. Alligator cracking, also known as fatigue cracking, appears as interlaced cracks forming a pattern akin to alligator skin and typically signals subgrade or base layer failure in asphalt pavements. In concrete pavements, similar degradation is observed in the form of divided slabs, where panels fracture due to excessive load or weak support. Both distresses serve as red flags for structural failure and are commonly used to trigger high-priority rehabilitation decisions [1].

3.2.2.11. Map cracking/corner break. Map cracking is a network of fine, irregular cracks caused by shrinkage or environmental stress, commonly found in asphalt pavements. Corner breaks, by contrast, are structural failures located at slab corners in concrete pavements, often initiated by curling stresses or inadequate load transfer. Both distresses reflect foundational weaknesses and, if progressive, signal the need for partial or full-depth slab repairs [1].

3.2.2.12. Edge cracking/joint spall. Edge cracking develops along unconfined pavement edges, commonly due to poor drainage, traffic stress, or shoulder drop-off. Joint spalling refers to chipping or fragmentation at slab joints in concrete pavements, often resulting from joint seal failure, freeze-thaw cycles, or inadequate construction. These distresses can compromise load transfer efficiency and water-tightness, increasing vulnerability to further deterioration [1].

3.2.2.13. Distortions/faulting. Distortions include various surface irregularities such as depressions, heaving, and bumps, typically stemming from subgrade movement or moisture-related expansion. Faulting is specific to concrete pavements and involves vertical displacement between adjacent slabs. Both reduce ride quality and safety, and their severity directly influences maintenance urgency and treatment selection [1].

3.2.2.14. Bleeding/polished aggregate. Bleeding manifests as a shiny surface film due to excess asphalt binder, while polished aggregate reflects loss of surface texture from repeated wheel passage. Both conditions significantly diminish skid resistance and are considered safety-critical surface distresses. Their detection plays a central role in friction restoration planning [1].

3.2.2.15. Raveling/scaling. Raveling is characterized by the progressive loss of aggregate on asphalt pavements due to binder degradation, whereas scaling refers to the flaking or peeling of concrete surfaces, often from freeze-thaw cycles or poor finishing practices. These surface distresses accelerate material loss and water infiltration, requiring timely surface sealing or overlays to prevent deeper structural issues [1].

3.2.2.16. Patches/potholes. Patches denote prior repairs on deteriorated surfaces, which may vary in effectiveness depending on the technique and material used. Potholes are bowl-shaped depressions caused by material disintegration under traffic and moisture. Both are key indicators of reactive maintenance and are used to assess the residual serviceability and lifecycle stage of a roadway segment [1].

In support of these detailed descriptions, Table 3 presents a summary of key distress features, including their definitions, measurement units, and severity classification criteria. Additionally, Fig. 4 provides representative annotated photographs of each distress type based on standard visual assessments. These resources aim to help researchers, practitioners, and pavement engineers better interpret the dataset and apply it to predictive modeling and infrastructure management tasks.

This structured classification enables direct incorporation of distress features into machine learning workflows, while supporting infrastructure managers in identifying deterioration mechanisms and tailoring intervention strategies.

3.2.3. Summary of the pavement performance indicators

This dataset includes four key indicators that collectively characterize pavement condition: the IRI, SDI, PCI, and Structural Index (SI). Together, these indices provide a comprehensive understanding of surface smoothness, visible deterioration, structural adequacy, and overall performance, supporting data-driven decision-making for pavement management and rehabilitation planning.

Table 3
Summary of pavement distress features.

No.	Distress Type	Description	Measurement Unit	Severity Classification
1	Rutting	Depressions in wheel paths due to traffic loading	Depth (mm)	Minor (≤6), Moderate (6–12), Severe (>12)
2	Linear Cracking	Longitudinal, transverse, or block cracks from shrinkage or stress	Width (mm), Length (m)	Low (<3), Moderate (3–10), High (>10)
3	Alligator Cracking / Divided Slab	Interconnected fatigue cracking / fractured concrete slabs	Area (m^2)	Based on extent and displacement
4	Map Cracking / Corner Break	Shrinkage or load-induced cracking in patterns or corners	Area (m^2)	Based on crack density and depth
5	Edge Cracking / Joint Spall	Cracks at edges or slab joints due to stress or poor drainage	Length (m)	Based on length and width
6	Distortion / Faulting	Surface unevenness or joint height differences	Height/Depth (mm)	Based on severity of distortion
7	Bleeding / Polished Aggregate	Binder overflow or smoothed aggregate surface	Area (m^2)	Based on extent of affected area
8	Raveling / Scaling	Disintegration of asphalt/concrete surface layers	Area (m^2)	Based on area and depth of material loss
9	Patches / Potholes	Repaired areas or open holes from material loss	Area (m^2), Depth (mm)	Based on patch quality and pothole size

2.2.3.17. International roughness index (IRI). The IRI is a standard measure of pavement smoothness that quantifies elevation changes along the pavement profile. It is reported in millimeters per meter (mm/m) and indicates ride quality. Lower IRI values (e.g., <2 mm/m) represent smoother surfaces, while higher values (e.g., >4 mm/m) indicate rougher conditions. IRI is crucial for assessing user satisfaction and identifying sections requiring maintenance to enhance ride comfort [10].

3.2.3.18. Surface distress index (SDI). The SDI is derived from distress observations loaded, including both load-associated distresses (e.g., rutting, alligator cracking) and non-load-associated distresses (e.g., raveling, patching) into the pavement management software, aggregating severity scores on a 0 to 100 scale. Load-associated distresses have a higher impact on the SDI. This metric reflects the severity and extent of surface distress, making it a vital input for maintenance prioritization [11].

3.2.3.19. Pavement condition index (PCI). The PCI measures the overall condition of a pavement segment. In this dataset, the PCI was computed using the City of Beaumont's official pavement assessment methodology. Following field data collection using the Laser Road Surface Tester (RST), two sub-indices were derived for each road segment: the Surface Distress Index, based on visual assessments of distresses such as cracking, rutting, and patching, and the Roughness Index, derived from longitudinal profile measurements. The final PCI score was calculated using a weighted average:

$$\text{PCI} = 0.67 \times \text{Surface Distress Index} + 0.33 \times \text{Roughness Index}$$

This weighting reflects the city's emphasis on surface condition in driving maintenance priorities while still incorporating ride quality. PCI scores range from 0 (failed) to 100 (excellent), with thresholds used to categorize pavement segments for management purposes (e.g., PCI 85–100 = Excellent; PCI 0–40 = Poor/Very Poor or backlog). These scores were used to support

Fig. 4. Representative images of common pavement distresses observed in the dataset. Labels indicate the distress type. Photographs were captured by the research team during field surveys in Southeast Texas and illustrate typical manifestations of each distress type as classified in accordance with ASTM D6433 [1].

budget allocation, rehabilitation prioritization, and long-term pavement performance planning [12].

3.2.3.20. Structural index (SI). The SI offers a simplified classification of pavement strength, based on the severity and type of structural distresses. Unlike PCI, the SI does not directly influence overall condition scores but supports decision-making related to structural adequacy. Pavements are categorized as weak (SI = 30), moderate (SI = 60), or strong (SI = 80). This classification informs the selection of appropriate rehabilitation strategies based on underlying load-bearing capacity [13].

Table 4 summarizes the definitions, scales, and formulas associated with these performance indices, including the conversion of IRI into a normalized Roughness Index (RI) and the classification schemes for load- and non-load-associated distresses (Fig. 5, Fig. 6).

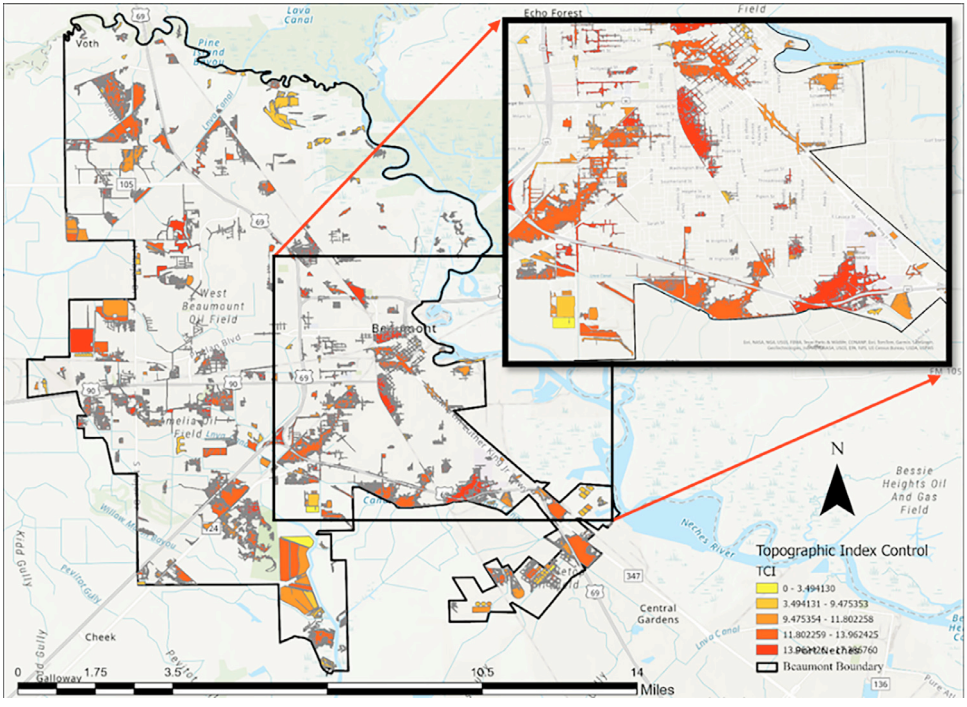

Fig. 5. The TCI map of Beaumont, TX, generated using ArcGIS, shows topographic variations across the area.

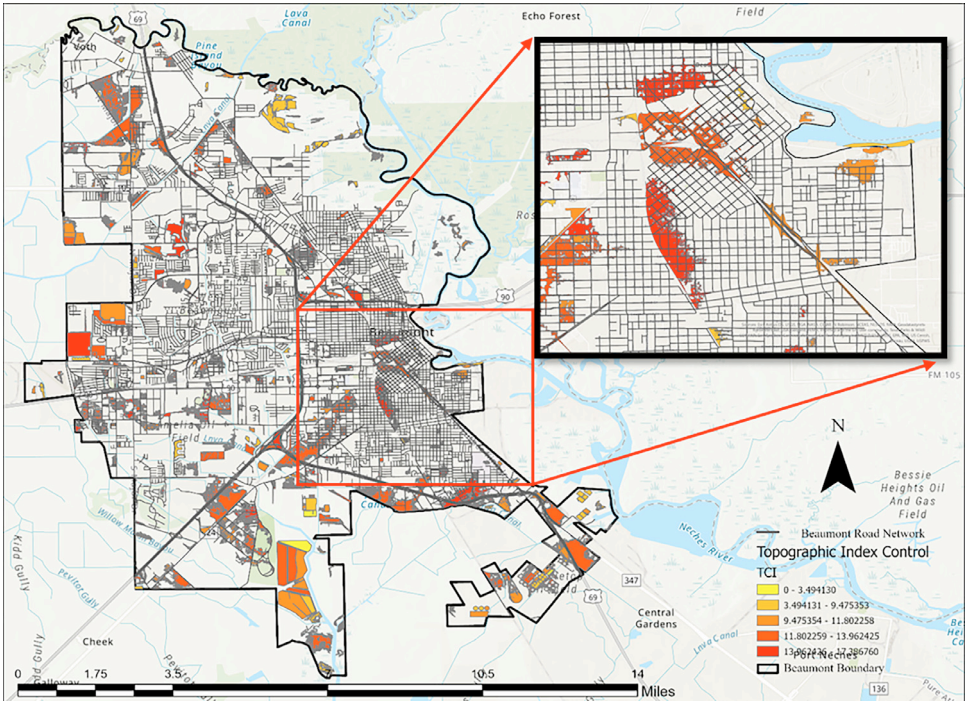

Fig. 6. Localized TCI values for each road segment in Beaumont, TX, assigned using ArcGIS spatial tools.

Table 4
Summary of pavement condition indices, distress features, and corresponding scales used for evaluating pavement performance.

Feature		Description	Scale/Formula
Surface Distress Index (SDI)		Aggregates observed pavement defects.	0 to 100, with distress severity rated on a scale from 0 to 10, depending on type and severity.
	Load Associated Distresses (LADD): Heavily influence the SDI	Flexible pavement: alligator cracking, map cracking, and wheel path rutting	Rated 0 to 10 based on severity and extent.
		Rigid pavement: divided slabs, distortions, faulting, and corner breaks	Rated 0 to 10 based on severity and extent.
	Non-Load Associated Distresses (NLAD): Less impact on SDI	Flexible pavement: raveling, scaling, patches, potholes, bleeding and patching	Rated 0 to 10 based on severity and extent.
		Rigid pavement: joint spalling, edge cracking, and polished aggregate	Rated 0 to 10 based on severity and extent.
Roughness Index (RI)		A measure of pavement roughness based on the IRI to assess ride quality and surface smoothness.	RI = (11 - 3.5 × ln(IRI)) × 10, scale from 0 to 100.
Structural Index (SI)		Assesses pavement strength by relating distress severity to overall condition, classifying pavements into weak, moderate, and strong categories. Does not contribute to PCI.	Weak: 30, Moderate: 60, Strong: 80
Pavement Condition Index (PCI)		Overall health and performance of a pavement segment. PCI combines SDI, RI, and Structural Index. Newer or well-maintained pavements score higher, while pavements in poor condition score lower.	0 to 100 (PCI > 65 = good; PCI < 60 = poor).

3.3. Climatic data

Climatic conditions significantly influence pavement degradation, particularly in regions like Southeast Texas where variations in precipitation and temperature can exacerbate surface and structural deterioration. This study incorporates four key meteorological variables, Total Monthly Precipitation, Average Daily Minimum Temperature, Average Daily Mean Temperature, and Average Daily Maximum Temperature, as critical input features for Machine Learning-based pavement performance prediction. These attributes were identified during the initial phase of the research and acquired from the PRISM Climate Group and NOAA/National Weather Service datasets [6,14].

All climatic data were spatially localized to Beaumont's segmented road network using ArcGIS Pro. This process involved geospatial operations such as raster reclassification, Inverse Distance Weighting (IDW) interpolation, and zonal statistics to assign accurate climate values to each roadway segment. The resulting 72-month dataset (36 months preceding each inspection year: 2019 and 2023) was structured to support time-aware predictive modeling (Table 5).

Table 5
Necessary meteorological input features of modeling.

Climatic Input Feature	Data Source
Annual Average Precipitation	NOAA/ National Weather Service
Total Monthly Precipitation	Oregon State University/ PRISM Climate Data
Average Daily Minimum Temperature	Oregon State University/ PRISM Climate Data
Average Daily Mean Temperature	Oregon State University/ PRISM Climate Data
Average Daily Maximum Temperature	Oregon State University/ PRISM Climate Data

3.3.1. Total monthly precipitation

Total Monthly Precipitation reflects the cumulative precipitation, including rainfall, snow, sleet, and hail measured in millimeter over a calendar month. Sourced from PRISM [6], daily values were aggregated to monthly totals and spatially mapped to the road network using IDW interpolation and zonal statistics. Precipitation plays a critical role in subgrade saturation, freeze-thaw activity, and moisture-related distresses. Given Beaumont's annual rainfall range of 40–50 inches, high-resolution precipitation data enhances model accuracy in forecasting water-induced pavement degradation [6].

3.3.2. Average daily min temperature

This metric represents the average of daily minimum temperatures recorded within each month, expressed in degrees Celsius. Using PRISM datasets and geospatial techniques in ArcGIS Pro, minimum temperatures were interpolated and linked to individual segments. Although Southeast Texas typically experiences mild winters, episodic low temperatures can initiate thermal cracking and damage, especially in aged or moisture-sensitive pavements. This attribute contributes to modeling freeze-thaw susceptibility and seasonal effects on pavement layers [6].

3.3.3. Average daily mean temperature

The Average Daily Mean Temperature is the average of all daily temperatures within a given month. It provides a thermal baseline that influences asphalt performance, particularly in terms of binder behavior and material fatigue. Data was sourced from PRISM, interpolated via IDW, and localized using zonal statistics. Beaumont's climatic profile requires this input to capture sustained thermal exposure effects in performance models [6].

3.3.4. Average daily max temperature

This variable denotes the monthly average of daily maximum temperatures, also expressed in degrees Celsius. Derived from PRISM and processed similarly through geospatial methods, it captures peak heat exposure, which contributes to surface softening, rutting, and bleeding. Maximum temperature data is essential for modeling high-temperature stress on asphalt pavements and informing heat-resilient design and maintenance strategies [6].

3.4. Summary of the synthetic models' data

This section introduces the synthetic geospatial and climatic model outputs developed by our research group and partners to enhance pavement performance modeling. The dataset incorporates five derived attributes: Topographic Control Index (TCI), Frequency of Flooding, Inundation Depth, Flood Susceptibility Index, and Ground Deformation. These variables were generated using hydrologic and geospatial models tailored to Southeast Texas conditions, specifically a two-year return period and 15-minute precipitation intervals and were aligned to Beaumont's segmented road network using ArcGIS Pro.

Each attribute captures distinct environmental and terrain-related stressors influencing pavement degradation. Their integration supports the development of resilient infrastructure management strategies and provides high-resolution input for predictive modeling. The subsections below describe the derivation, spatial processing, and importance of each synthetic feature.

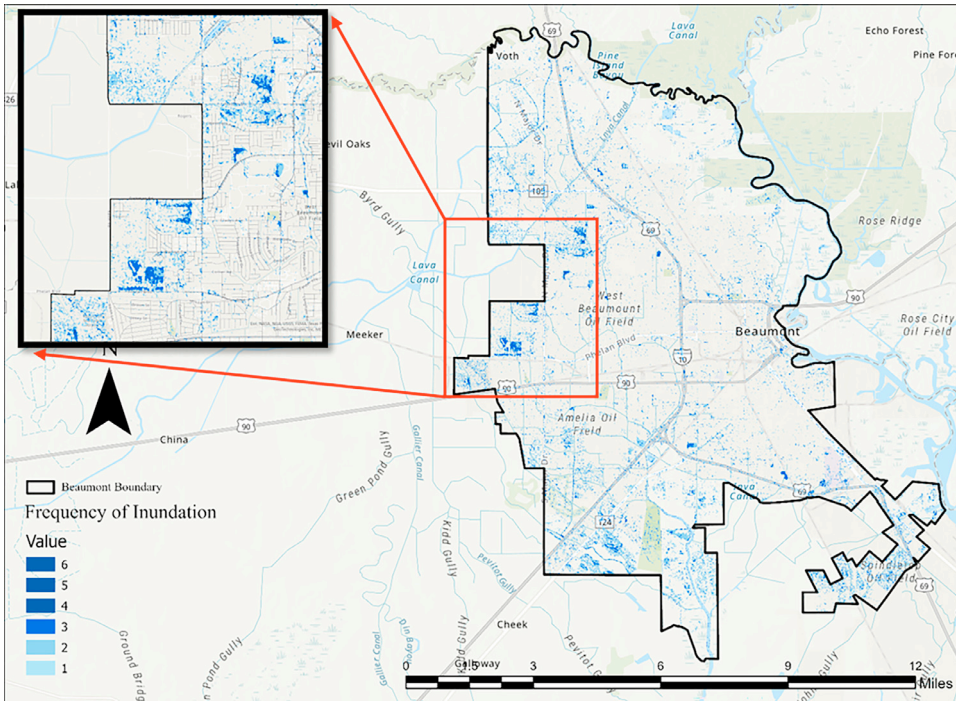

Fig. 7. Frequency of flooding map for Beaumont, TX, indicating areas with varying levels of flood recurrence.

3.4.1. Topographic control index (TCI)

The Topographic Control Index (TCI) is a dimensionless metric used to estimate the tendency of water to accumulate in terrain depressions. It highlights areas where runoff is likely to concentrate, thereby indicating potential surface water retention and flood susceptibility. TCI values were derived by analyzing digital elevation models in ArcGIS Pro using geoprocessing tools such as IDW interpolation and zonal statistics. Each TCI value was localized to individual road segments using spatial joins and statistical overlays. This integration ensures that each road segment has a specific TCI value, capturing the localized risk of water retention [15].

To enable integration into the segmented road dataset, localized TCI values were assigned to GISID-linked segments, as shown in the map below.

TCI plays a key role in identifying areas where topography amplifies water retention risk. High TCI values correlate with higher flood potential, making this attribute instrumental in proactive maintenance planning.

3.4.2. Frequency of flooding

This attribute quantifies how often each road segment experiences flooding, encompassing both nuisance and severe flood events. Expressed as the number of inundation occurrences, flood frequency reflects the historical recurrence of flood exposure at the segment level. Fig. 7 shows the frequency of flooding map for the study area, Beaumont, TX, generated through spatial analysis to highlight areas prone to recurrent flooding events.

The dataset was compiled using Sentinel-1 radar imagery, localized ground sensor data, and rainfall estimates provided by Oregon State University [6]. Spatial operations such as intersect, join, and aggregation were performed in ArcGIS Pro to assign segment-specific flood frequency values. Fig. 8 illustrates the distribution of flooding frequency across Beaumont, TX, based on

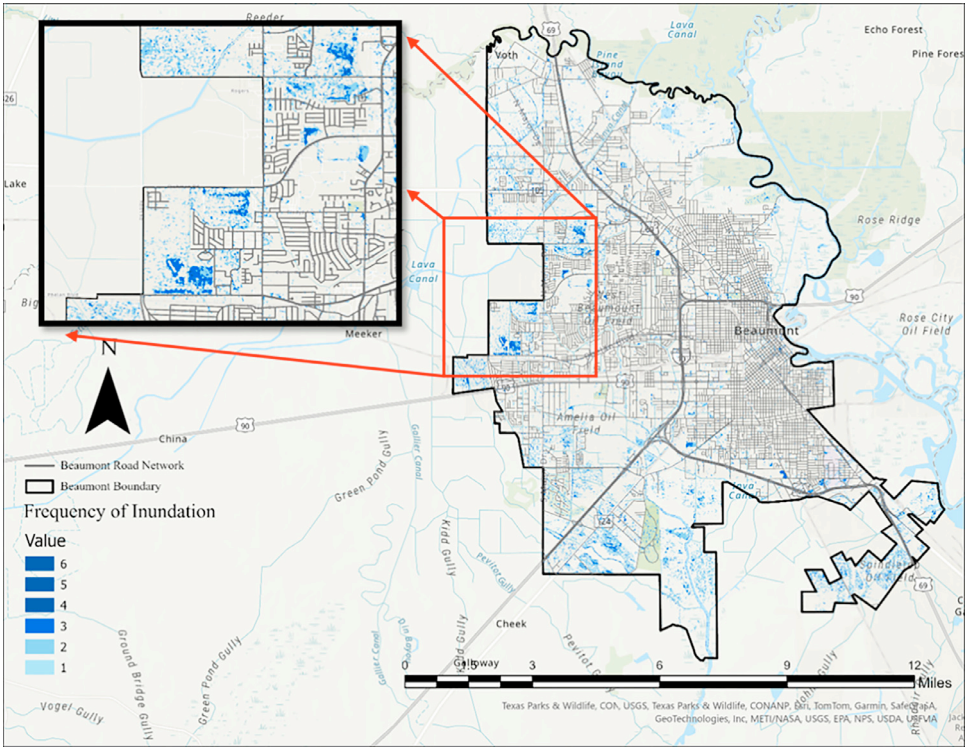

Fig. 8. Map depicting the frequency of flooding in Beaumont, TX, derived from historical data and spatial analysis.

historical data and geoprocessing tools, overlaying with the Beaumont segmented road network [15].

Recurrent flooding accelerates moisture-induced deterioration of pavements. By incorporating this metric into modeling efforts, planners can identify high-risk segments and prioritize interventions accordingly.

3.4.3. *Inundation depth*

Inundation Depth refers to the estimated depth of floodwater accumulating over each road segment during a synthetic flood scenario. It was modeled for a two-year return period with 15-minute rainfall intensity and is expressed in feet (Fig. 9).

Using ArcGIS Pro, depth rasters were generated and spatially assigned to segments through operations including raster conversion, intersection, and zonal merging. The final dataset links each GISID segment with a localized inundation depth. Fig. 10 illustrates the predicted inundation depths across Beaumont, TX, overlaid on the segmented road network. Greater inundation depths signify increased stress and structural risk to pavement systems. This attribute provides critical input for modeling the severity of flood exposure and its impact on long-term pavement performance.

3.4.4. *Flood susceptibility index*

Mapping flood susceptibility is essential for mitigating flood risks by identifying vulnerable areas using remote sensing (RS) and geographic information system (GIS) technologies. The Jefferson County flood susceptibility map was generated using the XGBoost machine learning algorithm at a 30 m resolution. Thirteen geospatial input factors were utilized, including elevation (DEM), slope, topographic wetness index (TWI), normalized difference vegetation index (NDVI),

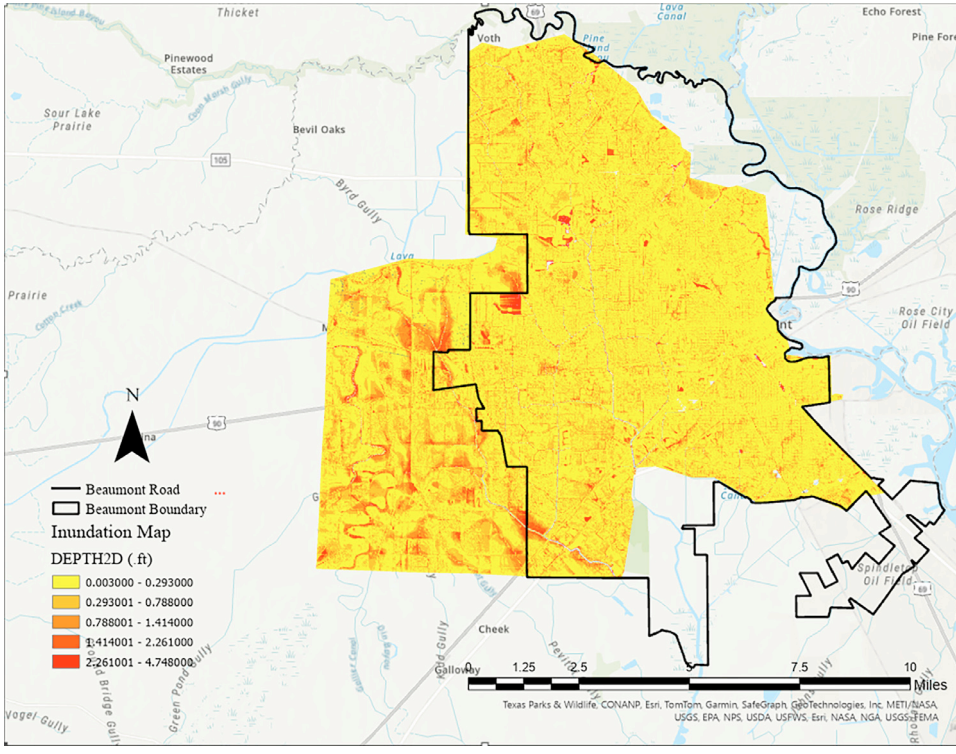

Fig. 9. Inundation depth map for Beaumont, TX, based on a two-year return period and 15-minute precipitation.

soil type, rock unit, land use/land cover (LULC), depression areas, soil hydrologic group, average yearly precipitation, and the distances to both major and minor streams and roads. In this dataset, the map was clipped to the boundary of Beaumont, resulting in a dimensionless flood susceptibility index ranging from 0 to 0.47, where higher values indicate a greater vulnerability to flooding. Fig. 11 illustrates the resulting flood susceptibility map for Beaumont, highlighting areas with varying levels of flood risk.

3.4.5. Ground deformation

The ground deformation data in this dataset stems from a parallel research effort focused on monitoring land subsidence and uplift in Beaumont, Texas. Persistent Scatterer Interferometric Synthetic Aperture Radar (PS-InSAR) techniques were used to process 28 Sentinel-1 SAR images acquired between January and December 2023. This method detects millimeter-level vertical surface displacements by identifying stable radar reflectors across the urban landscape. Persistent scatterers were selected based on coherence and amplitude stability thresholds. The resulting line-of-sight (LOS) deformation values, ranging from +20.4 mm (uplift) to –21.9 mm (subsidence), were converted to vertical displacements assuming a near-vertical incidence angle, then spatially projected and assigned to road segments using GISID.

Areas characterized by high road network density and extensive built-up infrastructure exhibited the most pronounced subsidence, indicating a correlation between urban development and ground settlement. Seasonal fluctuations in precipitation and temperature influenced deformation patterns, rainfall-induced soil saturation led to temporary uplift, while extended dry periods resulted in compaction and settlement. A geo-detector statistical analysis quantified both environmental and anthropogenic influences, providing insight into the factors driving ground stability in the region.

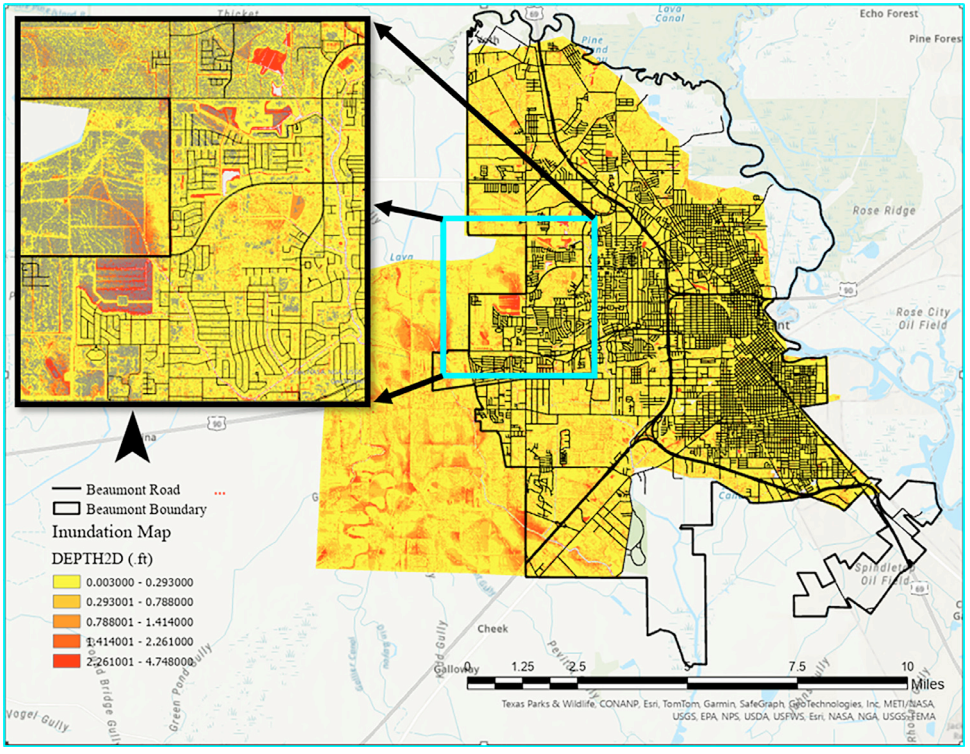

Fig. 10. Inundation depths in Beaumont, TX, overlaying with the Beaumont segmented road network.

Further methodological details, including image preprocessing, LOS-to-vertical projection, and validation protocols, are documented in a related graduate thesis and a peer-reviewed article [16,17]. These sources provide guidance for those wishing to replicate the PS-InSAR processing workflow or extend it to other urban pavement environments.

The spatial distribution of these deformation patterns is visualized in an ArcGIS Pro map (Fig. 12), which highlights areas of significant uplift and subsidence.

In summary, these attributes (TCI, Frequency of Flooding, Inundation Depth, Flood Susceptibility Index, and ground deformation) were derived through synthetic models and their integration with Beaumont's segmented road network. Together, they form a robust dataset that enables advanced machine learning applications for pavement performance modeling, ensuring informed decision-making for flood-resilient infrastructure management.

4. Experimental Design, Materials and Methods

4.1. Study area

The study focuses on Beaumont, a city in Southeast Texas that is highly susceptible to flooding and extreme weather events. Located in the Texas Gulf Coast region, Beaumont experiences a subtropical climate with substantial annual rainfall, contributing to frequent pavement deterioration. The dataset concentrates on the city-managed roadway network, including arterial, collector, and local streets, rather than interstate highways or major state roads. These city-owned roadways are vital in local transportation and emergency response during flood events. However, Beaumont's low-lying topography and urban development patterns make them especially susceptible to water-related damage [18]. Historic flood events, such as Hurricane Harvey in 2017

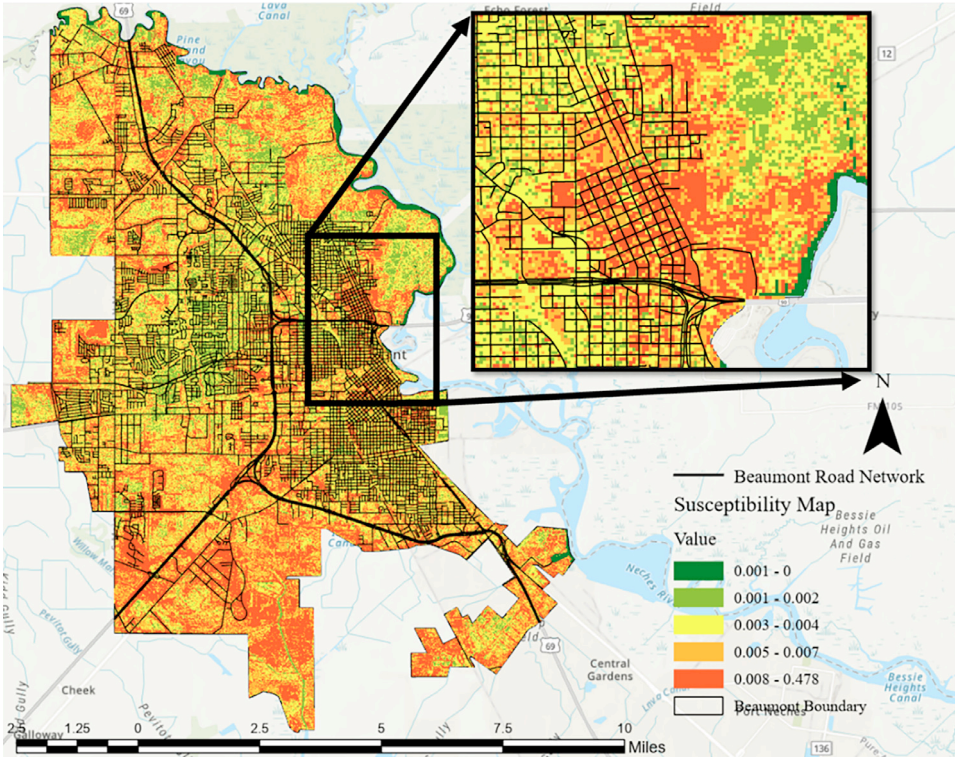

Fig. 11. Flood susceptibility map of Beaumont, Texas, generated using the XGBoost method.

and Tropical Storm Imelda in 2019, have highlighted the need for effective pavement management in this area. Beaumont's roadway network provides an ideal case study for assessing the impacts of flooding on pavement performance, optimizing M&R strategies, and improving infrastructure resilience against future climatic challenges. Fig. 13 illustrates the study area's location in southeast Texas, Beaumont, along with the segmented road network within the city's boundaries, as displayed in ArcGIS Pro 3.

4.2. Field operation

In 2019 and 2023, high-resolution pavement images were collected on 667 miles of asphalt and concrete roads in the City of Beaumont, TX, using an Integrated Road Information System (IRISpro Pave) (Fig. 14). This data facilitated the assessment of pavement cracking, rutting, and roughness. The analysis, conducted according to the American Society for Testing and Materials (ASTM) D6433 standard, provided a PCI for each roadway segment [1]. The ASTM D6433 method ensures a standardized, objective, and repeatable evaluation of pavement conditions, offering a more reliable alternative to methods dependent on subjective human ratings [19].

4.3. ArcGISPro3

ArcGIS Pro was employed as a geospatial analysis tool in this study for spatial data management, preprocessing, and analysis, enabling the integration, visualization, and detailed as-

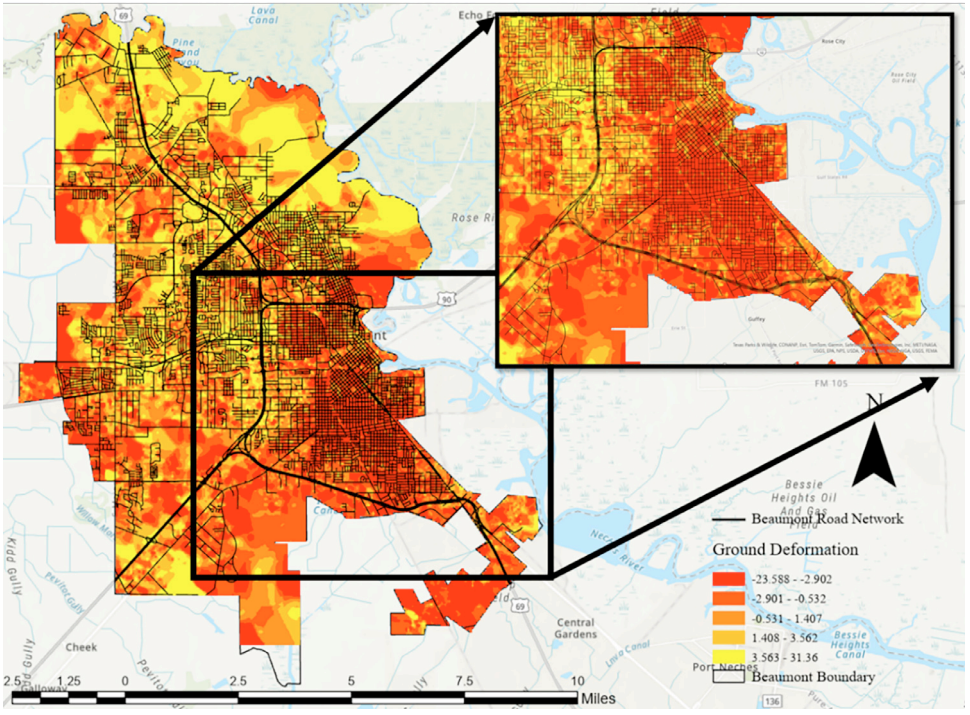

Fig. 12. ArcGIS Pro visualization of ground deformation in Beaumont, Texas.

sessment of diverse datasets. It facilitated the overlaying and merging of pavement condition indices, traffic data, meteorological records, flood maps, DEMs, and topographic indices, which were spatially joined to road network segments. Spatial data cleaning was performed to ensure accuracy and alignment, using tools for coordinate system adjustments and spatial clipping. This integration allowed for localized analyses of pavement conditions and flood impacts, including identifying road segments within flood-prone areas. Geoprocessing workflows were used to overlay flood maps with pavement performance data, assessing short-term and long-term impacts on specific roadway segments. Zonal statistics were also generated to analyze the correlation between flood intensity and pavement degradation. ArcGIS Pro also supported the creation of visualizations of segmented road networks and the export of attribute tables for further preprocessing, ensuring accurate integration with other analysis software.

4.4. Microsoft Excel

In this work, Microsoft Excel served as a vital intermediate tool for data preprocessing and formatting before further analysis in Python and Machine Learning models. Following data extraction from ArcGIS Pro, Excel cleaned and prepared datasets for model training, validation, and testing. This included resolving inconsistencies, addressing missing values, and formatting data for subsequent analysis. Excel's functionalities updated the organization and sorting of large datasets, ensuring compatibility with Machine Learning algorithms. Its user-friendly interface and efficient data management capabilities made it indispensable for preparing high-quality, standardized datasets, which are critical for ensuring the accuracy and reliability of predictive models.

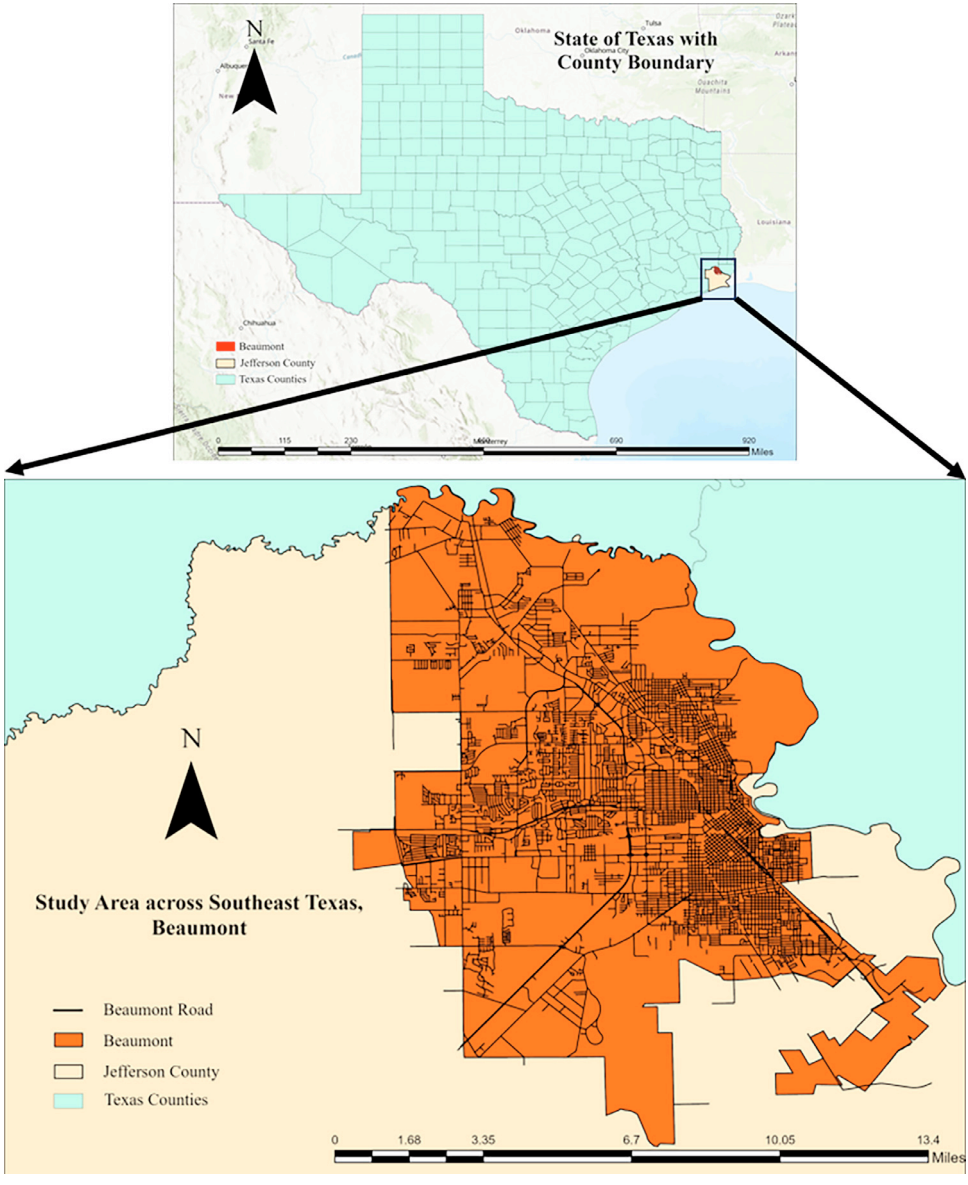

Fig. 13. Study area across southeast Texas.

4.5. Python

Python, executed within Jupyter Notebooks, served as the primary programming language for advanced data preprocessing and analysis. Python libraries were utilized to manage and analyze large datasets efficiently. The tool enabled the implementation of Machine Learning models to predict pavement performance. Python was also critical for data visualization, error analysis, and developing hyperparameter optimization techniques to improve model performance. The ability

Fig. 14. Example of pavement assessment using IRISpro Pave, similar to the method applied in Beaumont, TX, under ASTM D6433 standards [19].

to handle complex, large-scale datasets and integrate various data sources made Python an essential component of this work.

Limitations

The dataset is limited to two inspection years (2019 and 2023), which may constrain long-term trend analyses. Additionally, some M&R records contain missing entries.

Ethics Statement

The authors confirm that they have read and complied with the ethical requirements for publication in Data in Brief. This work does not involve human participants, animal experiments, or data collected from social media platforms.

Disclaimer

This manuscript reflects the perspectives and/or findings of the author(s). It is not to be interpreted as an official policy by the U.S. Department of Homeland Security (DHS) or the U.S. Department of Energy (DOE), whether stated directly or implied. The content herein does not imply the DHS or DOE approval of any equipment tested or assessed. The opinions, results, conclusions, or suggestions presented are solely those of the author(s) and may not coincide with the position of Lamar University or the CfR.

Data Availability

https://zenodo.org/records/15638682 (Zenodo).
https://data.mendeley.com/datasets/p6vg4v7f9k/2 (Mendeley Data).
Comprehensive Dataset for Data-Driven Pavement Performance Prediction and Analysis in Flood-Prone Beaumont, Southeast Texas (Original data) (Mendeley Data).

CRediT Author Statement

Hossein Hariri Asli: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Nicholas Brake:** Conceptualization, Data curation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing; **Mahdi Feizbahr:** Data curation, Software, Writing – review & editing; **Unique Bakhrel:** Data curation, Writing – review & editing; **Homayoon Arbabkhah:** Data curation, Software, Writing – review & editing; **Mohammad Sina Jahangir:** Conceptualization, Writing – review & editing.

Acknowledgements

This material is based upon work funded and supported by the U.S. Department of Homeland Security Cooperative Research and Development Agreement No. 21-TCSP-004, the U.S. Department of Energy, Office of Science, Biological, and Environmental Research Program under Award Number DE-SC0023216, the Center for Resiliency (CfR) at Lamar University. The authors gratefully acknowledge the City of Beaumont for providing the indispensable pavement management system data that has been integral to the advancement of this project. The authors also sincerely thank the Department of Civil and Environmental Engineering at Lamar University for their expert guidance and robust support, and they appreciate the valuable contributions of the dedicated students, faculty, and staff at Lamar University, whose collaborative efforts have significantly enriched this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] ASTM D6433-20Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys, ASTM International, 2020.
[2] ASTM D6433-11, Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys, ASTM International, 2011. Accessed: Feb. 23, 2025. [Online]. Available: https://www.astm.org/d6433-11.html.
[3] "Traffic count database System (TCDS)." Accessed: Feb. 23, 2025. [Online]. Available: https://txdot.public.ms2soft.com/tcds/tsearch.asp?loc=Txdot&mod=TCDS.
[4] "Street Rehabilitation Program | Beaumont, TX." Accessed: Feb. 23, 2025. [Online]. Available: https://www.beaumonttexas.gov/390/Street-Rehabilitation-Program.
[5] "Sentinel-1." Accessed: Feb. 23, 2025. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1.
[6] "PRISM Climate Group at Oregon State University." Accessed: Feb. 23, 2025. [Online]. Available: https://prism.oregonstate.edu/.
[7] H. Hariri Asli, N. Brake, M. Feizbahr, U. Bakhrel, H. Arbabkhahand M. S. Jahangir, "Comprehensive dataset for data-driven pavement performance prediction in flood-prone Beaumont, Southeast Texas". Zenodo, Mar. 19, 2025. doi: 10.5281/zenodo.15638682.
[8] H. Hariri Asli, N. Brake, M. Feizbahr, U. Bakhrel, H. Arbabkhah, and M.S. Jahangir, "Comprehensive dataset for data-driven pavement performance prediction and analysis in flood-prone Beaumont, Southeast Texas," vol. 2, Mar. 2025, doi: 10.17632/p6vg4v7f9k.2.
[9] J. Styer, L. Tunstall, A. Landis, J. Grenfell, Innovations in pavement design and engineering: a 2023 sustainability review, Heliyon 10 (13) (Jul. 2024) e33602, doi:10.1016/j.heliyon.2024.e33602.
[10] D. Chen, J. Hildreth, N. Mastin, Determination of IRI limits and thresholds for flexible pavements, J. Transp. Eng. Part B Pavement. 145 (2) (Jun. 2019) 04019013, doi:10.1061/JPEODX.0000113.
[11] T. Arianto, M. Suprapto, Syafi'i, Pavement condition assessment using IRI from Roadroid and Surface Distress Index method on National road in Sumenep Regency, IOP Conf. Ser. Mater. Sci. Eng. 333 (1) (Mar. 2018) 012091, doi:10.1088/1757-899X/333/1/012091.
[12] P. Marcelino, M. de Lurdes Antunes, E. Fortunato, Comprehensive performance indicators for road pavement condition assessment, Struct. Infrastruct. Eng. 14 (11) (Nov. 2018) 1433–1445, doi:10.1080/15732479.2018.1446179.

[13] J. Bryce, G. Flintsch, S. Katicha, B. Diefenderfer, Developing a network-level structural capacity index for asphalt pavements, J. Transp. Eng. 139 (2) (Feb. 2013) 123–129, doi:10.1061/(ASCE)TE.1943-5436.0000494.
[14] N. US Department of Commerce, "Climate Information." Accessed: Feb. 24, 2025. [Online]. Available: https://www.weather.gov/lch/bptclimategraphs.
[15] "Identifying urban pluvial nuisance flooding hotspots using the topographic control index and remote sensing radar images - ProQuest." Accessed: Feb. 24, 2025. [Online]. Available: https://www.proquest.com/openview/e87a1414d4e5dea4b9d794fdcde2f16e/1?pq-origsite=gscholar&cbl=18750&diss=y.
[16] Md.S. Chowdhury, A. Nur, Y.J. Kim, Comprehensive analysis of ground deformation in Beaumont, Texas: integrating PS-InSAR and GeoDetector techniques to evaluate impact factors and their interactions, Geotechnic. Front. (2025) GSP 365, doi:10.1061/9780784485989.030.
[17] "Comprehensive analysis of ground deformation in Beaumont, Texas: integrating PS-InSAR, and geodetector techniques to evaluate impact factors and their interactions - ProQuest." Accessed: Feb. 24, 2025. [Online]. Available: https://www.proquest.com/openview/125a271593e372ace25050b6106713d2/1?pq-origsite=gscholar&cbl=18750&diss=y.
[18] H. Hariri Asli, N. Brake, J. Kruger, L. Haselbach, M. Adesina, Field surveying data of low-cost networked flood sensors in southeast Texas, Data Brief. 50 (Oct. 2023) 109504, doi:10.1016/j.dib.2023.109504.
[19] "IrisPRO Pave," ICC-IMS. Accessed: Feb. 24, 2025. [Online]. Available: https://icc-ims.com/equipment/pavement-survey/irispro-pave/.