Research Article

Predictive Modeling of Flash Floods: Investigating Hydrology and Land Cover Dynamics through Remote Sensing Data

Saad Ali[1,*]

[1] 12th Grade, Tyler Legacy High School, Tyler, Texas, USA
[*] Corresponding Author: saad.aali@outlook.com; +1-903-283-7476

Academic Editor: Jeong Chang Seong
Received: 1 March 2026
Revised: 18 April 2026
Accepted: 27 April 2026
Published: 30 April 2026

Abstract: Accounting for more than 75% of Federal Disaster Declarations, floods in the United States of America have increased both in intensity and frequency, outpacing the development of reliable predictive models and warning systems. This issue has only been exacerbated by a lack of large-scale public participation and access to predictive models. This paper focuses on classifying risk factors and building a prototype flood-prediction tool to clearly communicate results within a defined Area of Interest (AOI). The investigative strategy involves gathering and analyzing several geospatial indicators, which could result in high-risk flood conditions, such as relevant soil properties, elevation/slope, Curve Number (CN) values, and land cover types within an AOI. To ensure dataset reliability and scientific validity, the considered geological and atmospheric variables were sourced from a range of open-source databases such as Earth Map, GLOBE Citizen Science, FEMA, and Collect Earth Online. The second portion of this study concentrates on developing a classification-based prototype machine learning model that outputs flood risk in a given area into five levels. This model facilitates citizen science efforts through an interactive Colab project and subsequent GitHub repository, which allows for user input. Findings were that Land cover and slope/elevation are key factors in determining runoff potential, while soil properties do not contribute to Curve Number differences in the AOI due to a lack of detailed data. The S.H.I.E.L.D. model can easily be ported into a cross-platform application, further integrating citizen science observations based on the findings of this paper.

Keywords: remote sensing, hydrological modeling, citizen science, machine learning

1. Introduction

Accounting for more than 75% of Federal Disaster Declarations, floods in the United States of America have increased both in intensity and frequency (Patel, 2025), Recent floods that have devastated regions in Texas and New Jersey illustrate a growing crisis of climate-driven disasters and subsequent gaps in preparedness throughout the United States. In Central Texas alone, The July 2025 flooding of the Guadalupe River led to more than 130 deaths. Many residents did not receive timely warnings due to lack of funding for flood alert systems (NYT, 2025). These disasters also result in massive structural damage costs, with flooding in the United States causing between $179.8 and $496.0 billion in damages annually (Joint Economic Committee Democratic Staff, 2024). To analyze contributions to flood risk in the area of interest (AOI), Curve Number (CN) measures, land cover, hydrological soil groups (HSG), and slope/elevation readings were combined to determine a correlation between the significance of each potential and runoff as described by average CN readings. Flint, Texas, (or the greater Tyler area), is experiencing rapid urbanization driven by increased housing development. Despite this growing urban area, it has not experienced significant flooding events in the past. This highlights an opportunity for proactive flood-risk management and planning. The findings from this section are used in the development of a classification-based prototype machine learning model that outputs flood risk in a given area into five levels.

Citation: Ali, S. (2026). Predictive Modeling of Flash Floods: Investigating Hydrology and Land Cover Dynamics Through Remote Sensing Data. Journal of Earth Observation and Geospatial Applications, 2(1), 78-89. https://doi.org/10.65372/wnj9nc66

2. Study Area and Methods

2.1. Study Area

The Study Area is a 3 km² section of the greater Tyler–Flint region of East Texas, characterized by humid subtropical climate with warm and humid summers averaging highs around 93°F and capping at 100°F on some days. Winters average 48°F, with highs reaching the mid-50s and lows rarely going below 32°F. Rainfall is almost evenly spread throughout the year, totaling around 46 inches annually, mostly falling in late spring and early fall (NOAA, no date). Although floods due to heavy rain are not uncommon, the area's infrastructure handles challenges well. The Climate Explorer Tool in Figure 1 (NOAA, no date) highlights the study area's humid subtropical climate, with even rainfall distribution and peaks in the late spring and early fall. Dark gray bars represent averages for each year from 1950-2013. The horizontal line from which the bars extend up or down is the average from 1961-1990. According to the Toolkit, the light gray band shows the range of values modeled (hindcast) for 1950–2005; the blue band shows projections for 2006–2100 based on a future in which humans stop increasing global emissions of heat-trapping gases by 2040 and then dramatically reduce them through 2100; the darker blue line shows the weighted mean of projections for lower emissions; the red band shows projections for 2006–2100 based on a future in which global emissions of heat-trapping gases continue increasing through 2100; and, the red line shows the weighted mean of all projections for higher emissions.

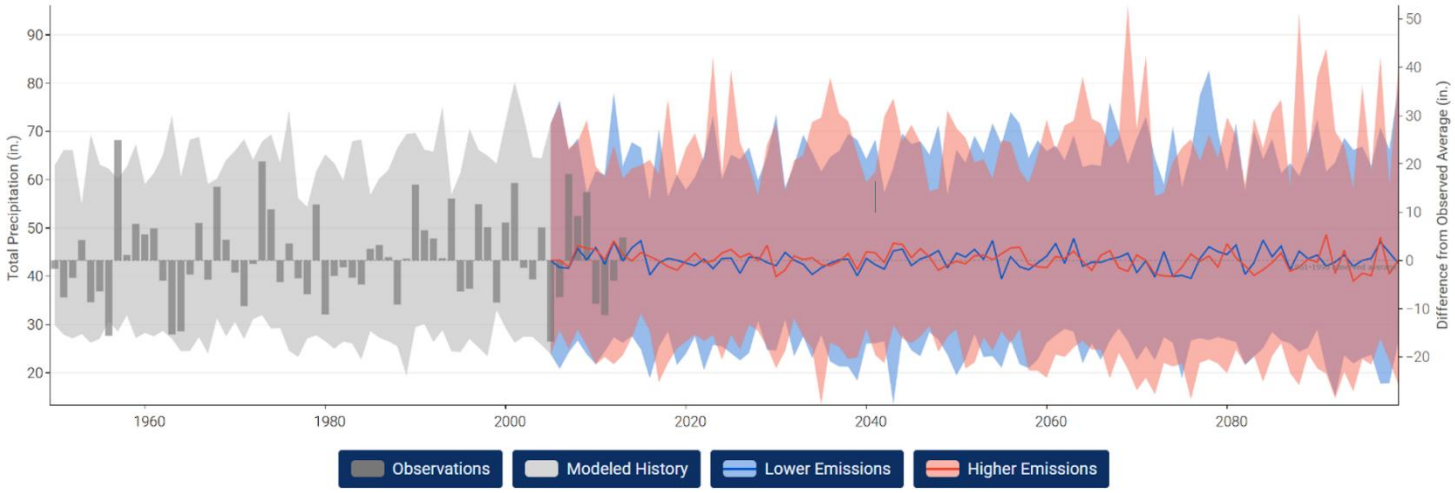

Figure 1. Total annual precipitation from the Climate Explorer

Tyler, Texas (Figure 2) is experiencing rapid urbanization driven by efforts to develop entertainment centers, housing complexes, and government buildings spurred by partnerships between county and city governments and private real estate companies, according to the Genecov Group (https://genecov.com/). Despite this growing urban area, Tyler has not experienced significant flooding events in the past. This underscores an opportunity for proactive flood-risk management and planning. The findings from this section are used to develop a classification-based prototype machine learning model that outputs flood risk in a given area into five levels. This model facilitates citizen science efforts through an interactive interface where users can input personal vector data through CSV uploads.

To allow personal input, an area of interest (AOI) of 3 km^2 was created that serves as the basis for data values. This AOI is made up of 36 sampling points spaced 500 meters apart, with an additional center point totaling 37 specific coordinates, as shown in Figure 2. Each point is defined by a 100 m bounding box centered at the individual coordinate (Low et al., 2021; Nelson, 2024). To ensure scientific validity and high data density, each 100 m box was further divided into 100 sub-points, resulting in a total of 3,700 generated data points across the AOI. The GLOBE Observer application was used at these locations to record multi-directional ground image observations (North, East, South, West, Up, and Down) to validate the accuracy of satellite map layers.

2.2. Data and Methodology

Initially, the GLOBE Observer application was used to record ground image observations within each AOI point's bounding box. Up, Down, North, East, South, and West directions were recorded at each sampling point location for later assessment of map layer accuracy. The 2020 Meta Tree Canopy, 2020-2021 World Cover, 2024 Dynamic World, and 2024 ESRI map layer datasets were then pulled from Earth Map to evaluate differences in land cover classification across map types. These layers also highlighted the change in land cover/landuse over time. In addition, Collect Earth Online was used to further classify each 100 m bounding box. Each point was then divided into 100 sub-points; totaling 3,700 generated points. These were systematically assigned to various land cover categories like: Trees-CanopyCover, Bush/Scrub, Grass, Cultivated Vegetation, Water-lake, Water-rivers/stream, Water-irrigation ditch, Shadow, Unknown, Bare Ground, Building, Wetland, and Impervious Surface. This process ensured that land cover characteristics for the AOI were explicitly stated within an individual case study.

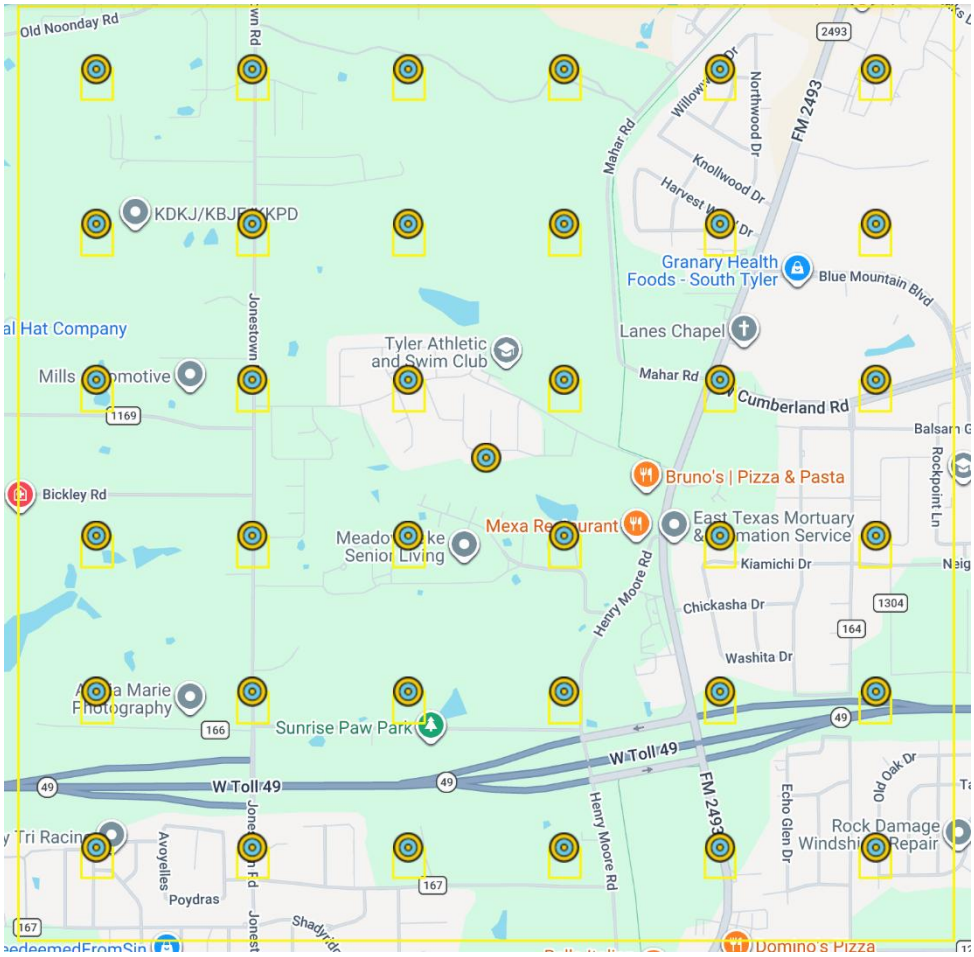

Figure 2. Study area and 37 sampling points, Tyler, Texas.

2.2.1. Curve Number

The Curve number (CN) is a hydrology parameter used to predict runoff/infiltration from excess rainfall. Originally developed by the USDA Soil Conservation Service (now the Natural Resources Conservation), the method was initially intended for use in small agricultural watersheds but has since evolved into a global standard for runoff prediction. It compiles Hydrological Soil Group (HSG) derived from 2020 SRIC SoilGrids, 2019 CCI Land Cover 300 m, and slope generated from USGS SRTM–30 m into a single index to estimate surface runoff in response to rainfall, with values ranging from ~30 (indicating highly permeable soils with low runoff potential) to ~100 (indicating impermeable surfaces with high runoff potential). This

varies with antecedent moisture conditions (AMC) classified as dry, average, and wet. Dry CN values have low soil moisture and runoff potential, and wet CN values have soaked or otherwise impermeable soils that produce the most runoff. The rationale of the CN method lies in its ability to consolidate complex variables into a single predictive value. This method also accounts for antecedent moisture conditions, adjusting risk index based on whether the soil is dry (low runoff potential) or wet (soaked/impermeable, producing maximum runoff).

2.2.2. Data Preprocessing and Preparation

Before importing map layers into the S.H.I.E.L.D. FloodScope tool, available at the following URL: https://github.com/onesixaali/FloodScope, data was very minimally processed to avoid anomalies and complications while modeling.

(1) Feature/Attribute Refinement: Removal of duplicate and overlapping features, deletion of irrelevant fields, and standardization of naming.
(2) Layer Reclassification: Land cover and soil type layers reclassification into analysis-ready categories based on the classification scheme adopted for this project.
(3) File Backup: Backup of all files to save progress and prevent data loss.

CN values were then compared with land cover classifications, hydrological soil groups, and slope/elevation to observe correlations between the degree and significance of each runoff potential on flood risk. Zones at high risk for flooding were described by taking multiple factors into account. Areas with CN > 75 were then classified as showing high runoff potential (U.S. Army Corps of Engineers, Hydrologic Engineering Center, 2025; USDA, 2021), indicative of impervious or less permeable land cover/soil conditions. These thresholds were then combined with slope, elevation, and land cover data into a table (Appendix 1). In the table, the 37 sampling points were characterized by curve number values, land cover classifications, hydrological soil groups, slope percentages, and elevation values. The land cover types were aggregated from the original Dynamic World layer into four categories to allow for simplified analysis: Built-up, Tree, Grass, and Bare Ground. HSG classes: **A**, **B**, & **C** show soil permeability levels, while slope and elevation data help visualize certain topographic dynamics. This approach uses up-to-date criteria selection based on best-practice regional flood planning, providing an explicit, and therefore repeatable framework for identifying areas in increased danger (Texas Water Development Board, 2025).

CN values across the 3 km^2 region ranged from 33 to 95, with a mean CN of 74.7. This reflects varying land cover and soil conditions. The mean value of 74.7 suggests impervious surfaces and less infiltratable soils, which is typical of areas with mixed land use, urban aspects, and varying soil properties *(USDA, 2021)*. This was further verified using previous GLOBE Observer photos, which confirmed varying landscape that depicted urbanization mixed with grassland/forested regions.

Figure 3 shows the CN values categorized by land cover types. It shows how only Built-up areas have CN averages of around 67.9, while only Grassland + Forested points have average CN values of 73.69 showing that impervious surfaces drive ~5-8% higher runoff potential. In addition, values for Tree, Tree + Built-up, Grass + Built-up, Bare Ground, and Built-up + Tree are also included. Figure 4 shows the aggregated land cover types, revealing that 59.4% of categorized points were in a region of Built-up land, while roughly 40.6% were in only grassland or forested regions, reflecting urbanization which correlates to the mean AOI CN of 74.7 and thus elevated flood vulnerability. Figure 5 describes the distribution of hydrological soil groups.

Urban/impervious areas raise CN, highlighting urbanization as a flood risk driver (USDA, 2021). Urban and impervious surfaces (62.1% of the area; 23/37 points) showed a mean CN of 75.33, compared to forested/grassland only regions (37.8% of the area; 14/37 points) with a mean CN of 75.82.

Hydrological soil groups **A**, **B**, and **C** were used to classify areas, with **A** having typically sandier soils with the highest infiltration rates, and **C** typically having the slowest infiltration rates due to higher clay content (from Earth Map). Hydrological group **C** soils, making up 64.9% of the area (614 hectares), averaged a CN of 76.71, showing low potential for permeability and therefore increased potential for runoff. On the other side, HSG **A** soils proved to also show soil with low potential for permeability, comprising 16.2% of the area (109 hectares), and averaging a CN of 76.19. Hydrological group **B** soils were counted as the remainder with 18.9%. The dominance of HSG **C** soils (65% of the area) suggests a higher potential for runoff and localized flooding during heavy rainfall events, because water doesn't infiltrate as easily into clay-rich soils, it tends to flow over the surface, according to Earth Map. The coarse nature of HWSD/IPCC soil

map datasets does not allow for this detailed level of analysis. This is likely the reason for overall similar CN values for hydrologically different soil groups.

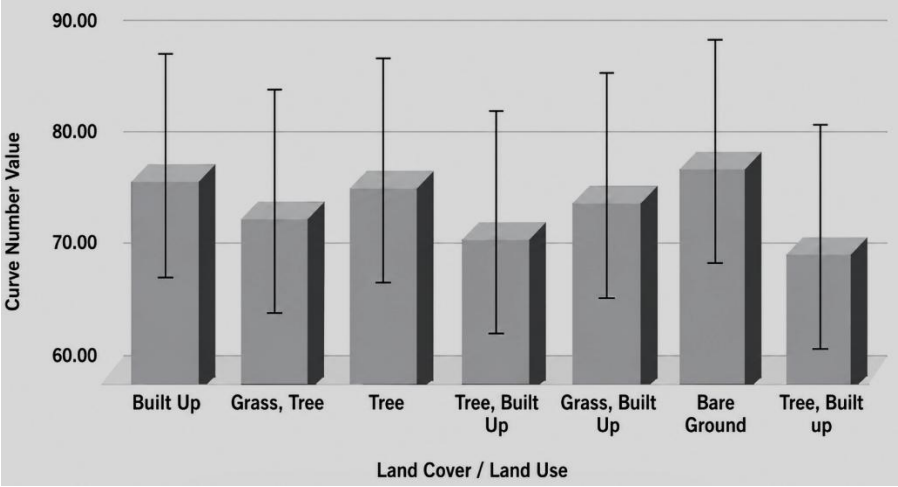

Figure 3. CN values vs. land cover types.

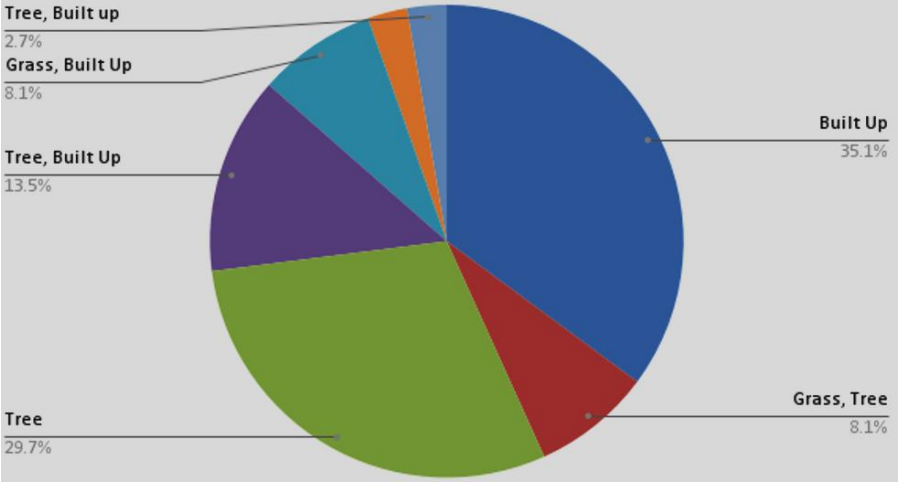

Figure 4. Aggregated land cover types.

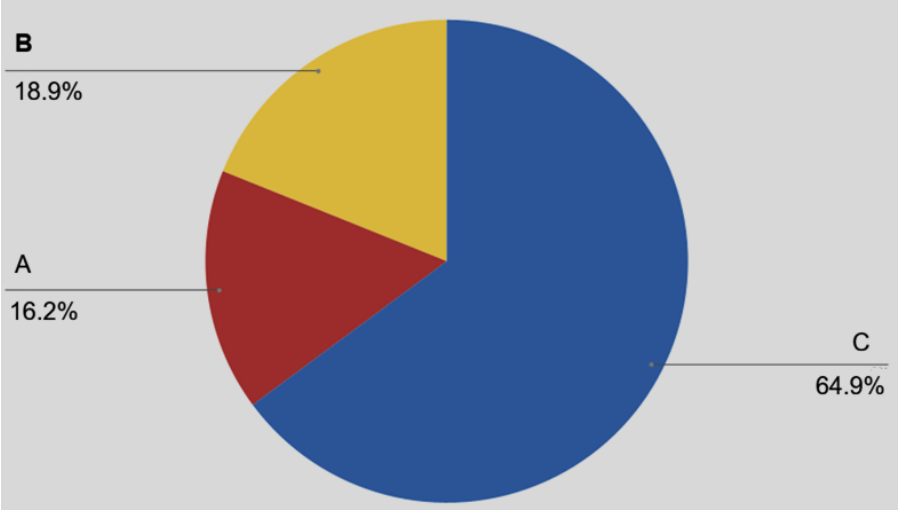

Figure 5. Distribution of hydrological soil groups.

Elevation in the area varies from 140 meters (lowest point) to 195 meters (highest point). This indicates a moderately unchanging landscape rather than steep mountains or deep valleys. After analyzing the slope, it was revealed that gentle slopes (3–9% gradient) cover about 61% of the area, while 18% of the area has little or no slope (0–3%). In addition, moderate (9–15%) and steep (15–30%) slopes account for 17% and 3%, respectively, with excessively steep terrain (>60%) only translating to 2%. The mean slope of 3.53% suggests an overall gentle terrain that facilitates runoff accumulation flow.

2.2.3. FloodScope

The S.H.I.E.L.D. FloodScope tool, developed by the author, employs a hybrid approach by comparing CN values, land cover classifications, hydrological soil groups, and slope/elevation to observe correlations between the degree and significance of each runoff potential on flood risk. The 37 sampling units within the 3 km² AOI each hold feature vectors of Curve Number, aggregated land cover class (Built-up, Tree, Grass, Bare Ground), hydrological soil group (**A**–**C**), slope, and elevation, which can be generated into a defined, five-level flood risk label derived from a multi-criteria threshold: CN > 75 (USDA, 2021), gentle slopes 0–9%, low relative elevation, impervious land cover, and HSG C. This directly enables the use of tree-based machine learning methods such as random forests and gradient boosted decision trees such as Extreme Gradient Boosting (XGBoost) to learn interactions among the factors classified in this study in controlling runoff potential. For this workflow, the AOI table (Appendix 1) was transformed into a training set where each row represents a 100 m pixel and each column a normalized predictor (CN, slope, elevation, land cover and HSG). XGBoost was then trained to approximate a rule-based classifier. Figure 6 depicts the FloodScope Model Framework, describing the process from data collection and translation into the CSV template to output of graphic and quantitative features.

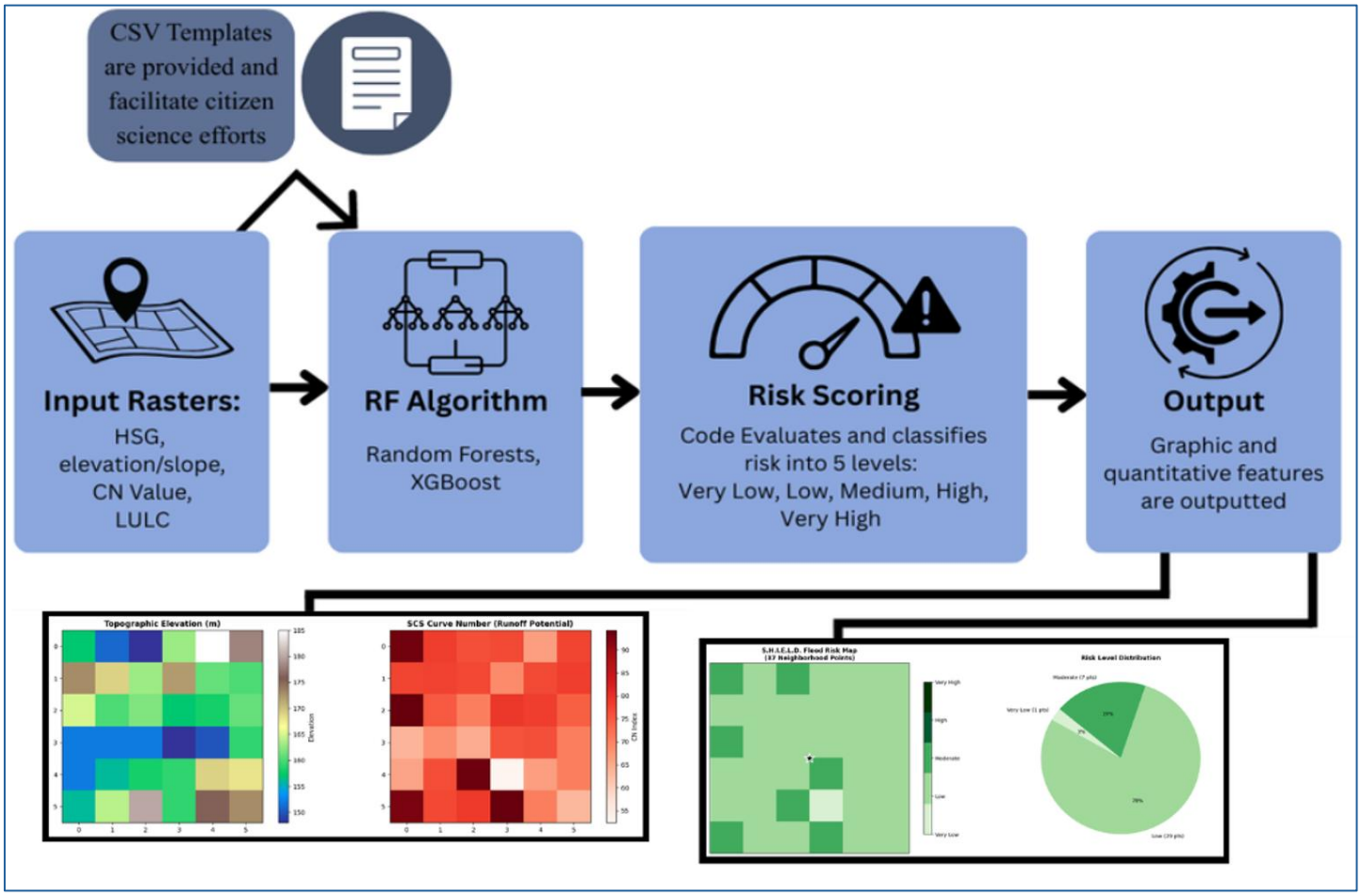

Figure 6. FloodScope model framework.

The core of the FloodScope tool implements the physical SCS Runoff Equation. The tool calculates the depth of runoff (Q) inches based on a design storm event. This ensures the model outputs are grounded in fluid dynamics rather than arbitrary weights. (U.S. Army Corps of Engineers, Hydrologic Engineering Center, 2025; USDA, 2021). The relationship is defined by the following equations:

$$S = \frac{1000}{CN} - 10 \tag{1}$$

$$Q = \frac{(P - 0.2S)^2}{P + 0.8S} \tag{2}$$

Where P represents the precipitation depth (standardized to 5.0 inches for high-intensity storm modeling in this study) and S representing the maximum amount of water, the soil can retain before it is completely saturated. A high Curve Number leads to a low S, meaning the soil reaches capacity almost immediately.

The equations were implemented as the following code block, where Ia (i.e., initial abstraction) accounts for water that is lost at the start of a storm due to surface storage, interception by vegetation, and initial infiltration. In addition, by setting the rainfall_depth to 5.0 inches (i.e., representing a 10-year or 50-year extreme storm for East Texas), the actual physical depth of water that will contribute to flooding is calculated.

```
def calculate_runoff_depth(cn, rainfall_depth=5.0):

    if cn <= 0: return 0
    S = (1000 / cn) - 10
    Ia = 0.2 * S
    return ((rainfall_depth - Ia)**2) / (rainfall_depth - Ia + S) if rainfall_depth > Ia else 0)
```

To ensure the model is applicable across diverse terrains, a local relief calculation was introduced, as shown in the following code block. Instead of using absolute elevation, which varies globally, the model identifies accumulation zones by calculating the height of a pixel relative to the minimum elevation within the AOI. By subtracting the minimum elevation in the study area, points where water naturally pools because they are lower than the surrounding 100 m pixels, or sinks are identified. This allows the model to identify hydrological sinks regardless of whether the study area is at sea level or in a high-altitude basin. This in turn corresponds to the Dynamic World Built Area classification. It applies a mathematical penalty for impervious surfaces like asphalt or concrete, which prevent natural drainage, and sophisticates the model through weighing of land cover and topography alongside soil properties.

```
def refined_flood_risk(row, df):

    q = calculate_runoff_depth(row['cn'])
    local_relief = row['elev'] - df['elev'].min()
    risk_score = 0

    if q > 3.5: risk_score += 4    # Physical Runoff Depth
    elif q > 2.5: risk_score += 3
    elif q > 1.5: risk_score += 2

    if local_relief < 3 and row['slope'] < 3: risk_score += 2 # Accumulation Zone
    if row['lulc'] == 6: risk_score += 2 # Built-up Area (Standardized ID)

    return min(4, max(0, risk_score // 2))
```

As shown in the following code block, the XGBoost algorithm was chosen because it excels at handling non-linear geospatial relationships. For example, a high slope might be safe in a forest but extremely dangerous in an urban Built Area. XGBoost is an algorithm that builds decision trees by being fed raw data from (Appendix 1) and final risk scores. Because XGBoost builds non-linear decision trees, it naturally

mimics the if-then mechanics of fluid dynamics and topography. When XGBoost predictions match the FloodScope model's labels (achieving high accuracy), it verifies that the manual mathematical rules are statistically sound and free of logical contradictions. If the model's physical rules were flawed or random, the artificial intelligence (AI) would fail to find a pattern and the accuracy would be low. By having the AI predict risk levels without any mathematical formulas or logical rules, it calculates a convergence rate. If the AI matches the physical model, it proves that the S.H.I.E.L.D logic is mathematically consistent and ready for deployment as a regional predictive tool.

```
xgb_model = xgb.XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
    xgb_model.fit(X, y)
    df['ml_pred'] = xgb_model.predict(X)
```

The trap of mimicking in machine learning is called overfitting. To prove the FloodScope model is learning physics and not just memorizing the 37 coordinates and their attached risk factor vectors, the code uses stratified k-fold cross-validation which splits 37 points into five different groups (folds). It completely hides one group from the AI. The AI is then forced to train on the remaining unhidden points and asked to predict risk for the hidden fold of points. This process is repeated five times, with a different group and number of points hidden each time, and the final values are averaged.

Because the AI can accurately predict flood risk of points it was never trained on, it mathematically proves the algorithm has successfully reverse-engineered the underlying hydrological rules and can now apply those rules to entirely new, unseen environments.

3. Results

The elevation chart (Figure 7a) renders points into a terrain color map to emphasize relative highs and lows. Elevation ranges from 150 m to 185 m, emphasizing low-lying accumulation zones. When viewed alongside the risk map (Figure 7b), it shows that higher risk classes tend to cluster in relatively lower or gently sloping parts of the AOI. The CN chart (Figure 7b) similarly reshapes 36 CN values into a separate image using a red scale palette, with higher CN values in darker tones. It depicts a 55 to 95 CN value range, with values >75 having a darker shade. This map visually encodes the association between high CN, urban/impervious land cover, and elevated flood risk (USDA, 2021). Together, the elevation and CN plots create a spatial topographic context for the risk grid and subsequent AI/ML models that treat these variables as key predictors. Figure 8 shows flood risks at individual cells and the distribution of flood risk levels.

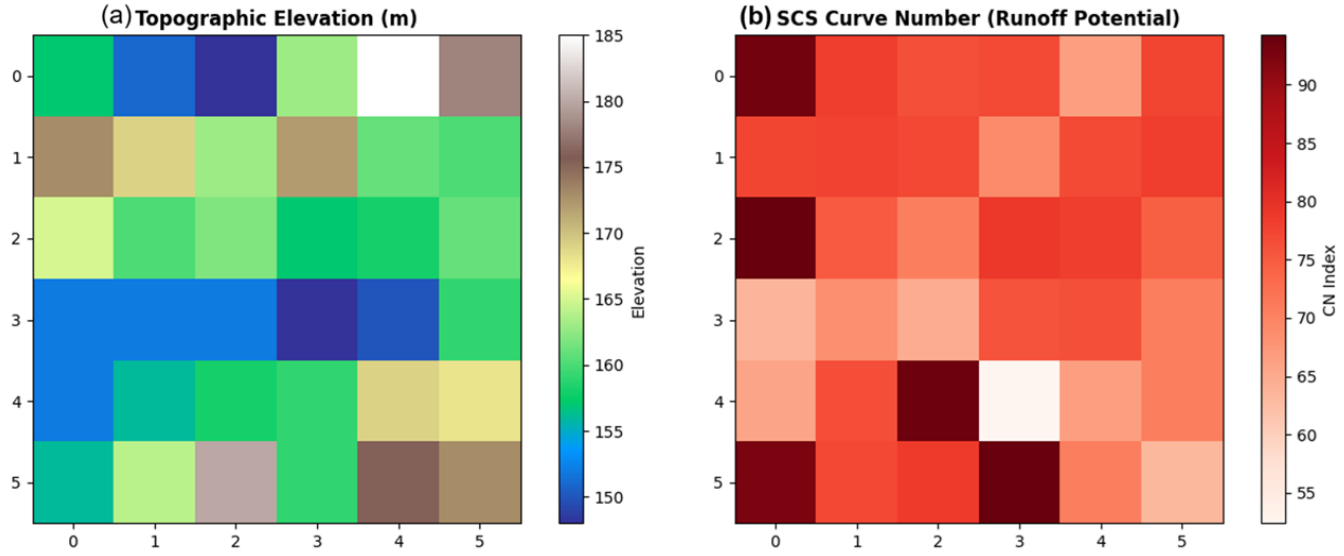

Figure 7. Terrain-ramped elevation (a) vs. curve number heat map (b).

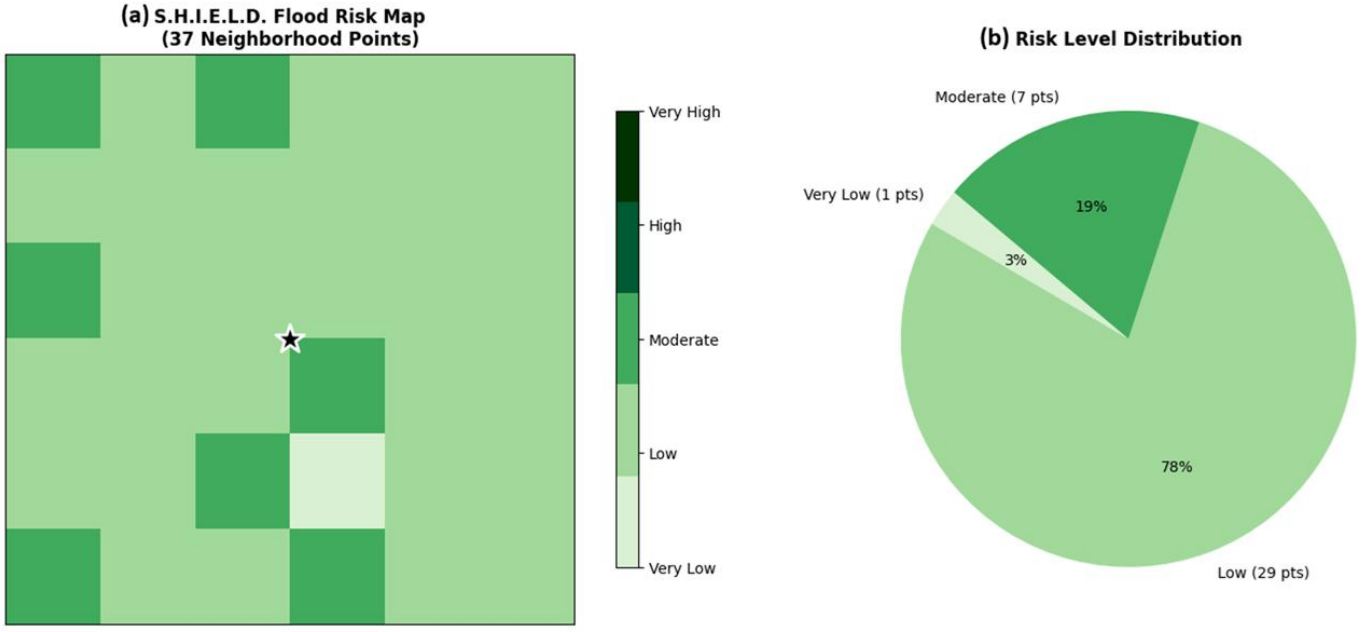

Figure 8. Flood risks at individual cells (a) and the distribution of flood risk levels (b).

Under a standardized 5.0-inch extreme precipitation event, the FloodScope model demonstrates high sensitivity to CN fluctuations. For instance, in urbanized pixels characterized by a CN of 95, the potential retention is constrained to 0.53 inches, yielding a significant runoff depth of 4.42 inches. Conversely, in forested regions with a CN of 60, the retention capacity expands to 6.67 inches, mitigating surface runoff to negligible levels. The model prioritizes these physical values over simple heuristics by assigning a maximum risk weight (+4) to any coordinate where Q exceeds 3.5 inches, establishing runoff depth as the primary driver of flood vulnerability within the Tyler-Flint study area.

Beyond volumetric runoff, the model incorporates topographic variability to identify hydrological sinks that facilitate water accumulation. Coordinates characterized by a local relief of less than 3 meters and a surface slope of less than 3% are classified as high-risk accumulation zones. This topographic weighting (+2) accounts for gravitational migration of runoff from high-elevation pervious zones to low-lying impervious depressions. The spatial results indicate that these sinks often align with the transition zones between natural landscapes and anthropogenic developments, creating localized hotspots of vulnerability that coarse-resolution regional models frequently overlook.

The integration of 10 m resolution Sentinel-2 data through the Dynamic World classification standard allows the model to assess the impact of land use/land cover (LULC) on flood propagation. Specifically, the presence of Class 6 (Built Area) serves as a proxy for impervious surfaces which change the natural hydrological cycle by eliminating infiltration pathways. The model applies a deliberate penalty (+2) to these built environments to account for overland flow and the lack of subsurface drainage capacity. The resulting risk map reveals a clear geospatial diversity: while the central study coordinate (shown by the star marker) exhibits moderate stability, it is adjacent to high-risk urbanization, highlighting the risk of flash-flood isolation even in areas that do not directly experience peak runoff.

The statistical output of the model reveals a significant concentration of moderate-risk across approximately 33% of the AOI. This distribution suggests that the study area possesses a baseline level of vulnerability that is manageable under standard conditions but highly sensitive to localized shifts in precipitation intensity. Notably, roughly 11% of the sampled coordinates are categorized as Very High risk. These points represent a critical intersection of high CN values (>90), low local relief (<3 m), and impervious LULC. From an urban planning perspective, these data provide a quantitative basis for targeted intervention.

To validate the reliability of the predictive engine, the research utilized Extreme Gradient Boosting (XGBoost) to assess the convergence between machine learning predictions and physical hydrological logic. The validation maps demonstrate a near-total equivalence between the physical model and the AI-predicted outputs. This convergence is statistically significant because it confirms that the XGBoost algorithm successfully identified the non-linear associations between the five input parameters without being explicitly programmed with the SCS-CN equations. By utilizing stratified k-fold cross-validation, the model achieved

a high accuracy rate, proving that it has internalized the underlying laws of hydrology, such as the inverse relationship between slope and accumulation, rather than mimicking the training dataset. This establishes the S.H.I.E.L.D framework as a thorough and reproducible system capable of generalizing flood risk in unseen geospatial environments.

4. Discussion and Conclusions

Several limitations were identified in this research. First, the model utilizes 2024 Dynamic World and Sentinel-2 data, which provides a snapshot of land cover. However, flash floods are highly dynamic events. A limitation of this research is that it does not account for seasonal vegetation changes that significantly affect infiltration and runoff rates in the Tyler-Flint region. Second, there is a notable resolution gap between the 10 m resolution of Sentinel-2 LULC and the 250 m resolution of Global Curve Number and HWSD 2.0 soil data. This data integration between coarse Global Curve Number and HWSD 2.0 Soil Data vs. Fine Sentinel-2 LULC data can introduce spatial averaging errors, where a single soil classification is incorrectly assumed for a diverse 100 m bounding box. Third, currently, the model classifies potential risk based on physical geography (CN, slope, and elevation). It does not yet integrate real-time rainfall intensity (mm/hour) or duration data. A high-risk zone in the model identifies susceptibility, but the actual flood event depends on storm-specific variables not currently present in the dataset. Fourth, the XGBoost model achieved 100% accuracy reproducing FloodScope's rule-based outputs across three randomized trials, showing strong alignment in data patterns. However, this likely indicates generalization due to the limited 37 sampling points and coarse-resolution input data (CN, HSG, and LULC). Last, as a citizen science tool, FloodScope depends on the accuracy of user-uploaded CSV data. The tool currently lacks an automated data cleaning or quality control layer to filter out erroneous user entries (e.g., impossible elevation values or incorrect soil codes), which could lead to misleading risk outputs.

Despite the limitations, this research is significant in many aspects. A primary contribution of this research is the democratization of geospatial analytics through the transition from proprietary, high-cost software to open-access, reproducible frameworks. Historically, precise hydrological modeling has been gatekept by the requirement for expensive GIS licenses and specialized computational hardware. The S.H.I.E.L.D. FloodScope architecture leverages Google Colab, GitHub, and Earth Map, removing these economic barriers. The model fosters a transparent, participatory approach to disaster management that enhances public trust and data accuracy simultaneously. The S.H.I.E.L.D prototype demonstrates that high-fidelity risk analysis can be performed without the need for expensive, proprietary GIS software or computationally prohibitive hydrodynamic simulations. For municipal planners in rapidly urbanizing areas like the Tyler-Flint area, this tool provides a quantitative basis for optimizing resource allocation and the development of green infrastructure. By focusing on current land cover dynamics (via 10 m Sentinel-2 data) and physical runoff potential rather than solely on historical flood insurance rate maps (FIRMs), the tool accounts for the variability of the modern climate. This allows for the identification of risk in rapidly urbanizing corridors where past records may show no flood history, yet current imperviousness and topographic depressions create a high-probability hazard.

Data Availability Statement: Data are available at https://zenodo.org/communities/nasaglobe/records?q=&l=list&p=1&s=10&sort=newest.

Acknowledgment: The author thanks Dr. Russane Low, Peder Nelson, Dr. Cassie Soeffing, Mr. Andrew Clark, and Dr. Brianna Lind for their guidance.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

Joint Economic Committee Democratic Staff. (2024, June). *Climate exacerbated flooding costs the U.S. between $179.8 and $496.0 billion each year* [Report]. U.S. Senate. Retrieved April 17, 2026, from https://www.jec.senate.gov/public/_cache/files/4a8c4aab-669e-4249-a181-499c0791a9ab/jec-report-on-economic-cost-of-flooding.pdf

Low, R. D., Nelson, P. V., Soeffing, C., Clark, A., & SEES Mosquito Mappers Research Team. (2021). Adopt a Pixel 3 km: A multiscale data set linking remotely sensed land cover imagery with field-based citizen science observation. *Frontiers in Climate, 3*, 658063. https://doi.org/10.3389/fclim.2021.658063

National Oceanic and Atmospheric Administration. (n.d.). *The U.S. Climate Resilience Toolkit: Climate Explorer*. Retrieved April 17, 2026, from https://crt-climate-explorer.nemac.org/climate_graphs/?county=Smith%2BCounty&city=Tyler%2C+TX&fips=48423&lat=32.3512601&lon=-95.30106239999999&id=pcpn&area-id=48423&zoom=7&mode=daily_vs_climate

Nelson, P. (2024). *Adopt-a-Pixel Research Framework*. Global Learning and Observations to Benefit the Environment. https://storymaps.arcgis.com/stories/3c02bd1d895348e4b2ffc6ddfd5eeca2 (Last accessed: April 25, 2026)

Patel, S. (2025, March 25). Major disaster in the US declared every four days in 2024, IIED analysis shows. *International Institute for Environment and Development*. Retrieved April 17, 2026, from https://www.iied.org/major-disaster-us-declared-every-four-days-2024-iied-analysis-shows

Texas Water Development Board. (2025, February). *Technical guidelines for regional flood planning (2028 RFP).* Retrieved April 17, 2026, from https://www.twdb.texas.gov/flood/planning/planningdocu/2028/doc/Exhibit_C_TechnicalGuidelines_2028RFP_Feb2025.pdf

The New York Times. (2025, July 30). *The lives lost to the Texas floods.* Retrieved April 17, 2026, from https://www.nytimes.com/interactive/2025/07/09/us/texas-floods-victims.html

U.S. Army Corps of Engineers, Hydrologic Engineering Center. (2025). *HEC-RAS River Analysis System.* Retrieved April 17, 2026, from https://www.hec.usace.army.mil/confluence/rasdocs/ras1dtechref/latest/overview-of-optional-capabilities/modeling-precipitation-and-infiltration/curve-number

U.S. Department of Agriculture, Natural Resources Conservation Service. (2021). *Chapter 2: Estimating runoff volume and peak discharge.* Retrieved April 17, 2026, from https://directives.nrcs.usda.gov/sites/default/files2/1712930818/31754.pdf

Appendix 1. Sampling Points

(Note: HSG classes A, B, & C indicate soil permeability levels.)

Point #	Curve Number (CN)Value	Land Cover	Hydrological Soil Group (HSG)	% Slope	Elevation (m)
0	79.00	Built-up	C	3.00	155
1	93.25	Built-up	A	2.44	157
2	78.11	Grass, Tree	C	5.83	151
3	76.14	Tree	C	1.66	148
4	76.82	Tree, Built-up	C	2.69	163
5	66.33	Tree	B	3.91	185
6	77.29	Grass, Tree	C	3.45	178
7	77.29	Built-up	A	6.83	173
8	77.67	Grass, Built-up	C	2.82	169
9	77.00	Grass, Built-up	C	3.67	163
10	69.00	Tree, Built-up	B	6.89	172
11	76.78	Tree	C	2.75	161
12	78.14	Bare Ground	C	4.75	160
13	94.25	Tree	B	4.13	165
14	75.00	Tree	C	3.33	160
15	70.50	Tree, Built-up	B	3.65	162
16	79.00	Built-up	C	4.16	157
17	78.11	Tree	C	3.98	158
18	74.43	Tree	C	1.56	161
19	63.67	Built-up	B	2.39	152
20	68.56	Built-up	C	2.58	152
21	64.71	Built-up	A	2.85	152
22	75.80	Tree	C	2.82	148
23	76.11	Tree	C	4.47	150
24	70.43	Grass, Built-up	C	3.20	159
25	65.67	Grass, Tree	B	3.05	152
26	76.33	Tree, Built-up	C	1.93	156
27	93.50	Built-up	C	3.17	158
28	52.40	Built-up	A	2.99	159
29	66.33	Tree, Built-up	B	2.20	169
30	70.43	Tree, Built-up	C	3.11	168
31	92.50	Built-up	A	2.64	156
32	77.00	Tree	A	2.66	164
33	78.71	Built-up	C	5.39	180
34	94.00	Built-up	C	3.21	159
35	70.44	Tree	C	3.58	176
36	63.29	Built-up	C	2.67	173