Climate Risk Management 46 (2024) 100654

Contents lists available at ScienceDirect

Climate Risk Management

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

# CLIM-SEG: A generalizable segmentation model for heat and flood risk mapping

Anushka Ray [1], Katherine Xu [1], Norhan Bayomi [*], John E. Fernandez

*Massachusetts Institute of Technology, The Urban Metabolism Group, United States*

A R T I C L E I N F O

*Keywords:*
Climate change risk
Climate adaptation
Segmentation
AI

A B S T R A C T

With the escalating impact of climate change coupled with increased urbanization, many cities will experience extreme heat events and intense flooding. Current modeling approaches often fail to incorporate high-resolution, frequently updated data sources, such as aerial imagery from web mapping platforms, limiting their effectiveness in identifying areas at risk. To address this gap, the paper presents CLIM-SEG, a novel framework for high-resolution urban heat and flood risk assessment, addressing critical gaps in current climate risk modeling. This framework integrates semantic segmentation of aerial imagery with a weighted sum approach that integrates environmental, socioeconomic, and building data to provide comprehensive risk evaluations at the census tract level. CLIM-SEG synthesize land cover data with hazard and vulnerability factors, producing risk scores ranging from 0 to 1. This low-cost and efficient framework can enable urban planners to prioritize resources for flood mitigation and heat adaptation, addressing the limitations of current approaches and contributing to the field of urban planning and climate change adaptation. The propoosed methodology incorporates a custom-curated dataset of 545 aerial images, including 145 manually annotated segmentation maps, to fine-tune advanced semantic segmentation models. The optimized Segmenter model achieves a pixel accuracy of 97.85% and an Intersection over Union (IoU) of 0.9578 for key urban features, significantly outperforming baseline models. Boston is selected to represent an ideal representation for both heat and flood risk, as the city experiences severe urban heat islands, and is susceptible to coastal and riverine flooding, with over 11,000 structures expected to be affected by 2070 due to sea level rise and increased precipitation. Results from flood and heat risk models indicate that census tracts in South End have the highest flood risk, with a weighted score value of 0.825, while census tracts in the Fenway-Kenmore neighborhood show the highest heat risk, with a score of 0.991. Both of these results have also been verified with other heat and flood risk mapping sources for Boston. The proposed framework of CLIM-SEG not only addresses the challenges faced by Boston but also has the potential to be scaled to other urban areas dealing with the impacts of climate change, providing a valuable tool for risk assessment and decision-making in the face of a changing climate.

* Corresponding author.
*E-mail address:* nourhan@mit.edu (N. Bayomi).

[1] Equal contribution.

https://doi.org/10.1016/j.crm.2024.100654
Received 28 May 2024; Received in revised form 11 September 2024; Accepted 13 September 2024
Available online 12 October 2024

## 1. Introduction

Climate change presents a pressing global challenge, resulting in elevated temperatures, prolonged heat waves (Meehl and Tebaldi, 2004; Tebaldi et al., 2006), rising sea levels, and intensified precipitation patterns (Mimura, 2013). The rapid urbanization in the United States amplifies these impacts, as cities face the effects of urban heat islands (UHIs), which exacerbate temperature extremes due to the prevalence of dry, impervious surfaces, limited tree canopy for shading, and anthropogenic heat emissions (EPA, 2023; Chapman et al., 2017). Furthermore, the replacement of vegetation and soil with buildings and paved surfaces increases the vulnerability of coastal and riverine cities to severe flooding events, jeopardizing populations, infrastructure, and transportation networks (Konrad, 2016). A climate vulnerability assessment in Boston revealed that without stormwater system enhancements, over 11,000 structures are at risk of stormwater flooding due to sea level rise and increased precipitation (Climate Vulnerability Assessment, 2017). As a result, communities in heat-prone and coastal urban areas face heightened risks of energy insecurity, adverse health outcomes, and flood-related damages (Haines et al., 2006; Harlan and Ruddell, 2011; Luber and McGeehin, 2008; Santamouris, 2020; Climate Vulnerability Assessment, 2022).

To mitigate these risks, urban planners require sophisticated tools to identify areas within cities that necessitate targeted flood mitigation and heat adaptation interventions (Konrad, 2016). Flood and heat risk modeling at the granular census tract level can inform the prioritization of investments in resilience measures such as storm barriers, improved drainage systems, cooling centers, and green infrastructure (Larsen, 2015). Moreover, these models can raise public awareness and guide policy development to enhance preparedness for extreme heat and flooding events (Wolf et al., 2015; Turek-Hankins et al., 2020). However, existing flood and heat risk models have limitations, often relying on socioeconomic and meteorological data while overlooking the influence of the built and natural environments (Landreau et al., 2021). Additionally, their predictions may be difficult to update or generalize to different locations due to dependence on specific datasets or focus on particular cities. Many flood models used for property-level risk assessment are based on historical flood data, which may not accurately represent present and future risk or adapt to the rapidly evolving impacts of climate change (Davis, 2022). There is a lack of cost-effective tools for policymakers to estimate flood and heat risk at the census tract level that integrate land cover, environmental features, socioeconomic, and building data.

To address these limitations, the paper proposes a novel weighted sum model that assesses heat and flood risk for urban areas at the census tract scale. The developed model incorporates hazard and vulnerability factors derived from land surface temperature, elevation, slope, flow accumulation, building, and socioeconomic datasets, as well as features extracted from aerial imagery using semantic segmentation, a cutting-edge computer vision technique that classifies pixels into distinct land cover categories. By integrating these diverse data sources, the proposed approach captures both the severity of potential impacts and the vulnerability of affected populations, providing a more comprehensive risk assessment compared to previous models.

The simultaneous assessment of flood and heat risk in this study is intentional and offers several advantages. Firstly, these risks are often interconnected consequences of climate change in urban environments, with cities experiencing increased heat island effects potentially facing heightened flood risks due to altered precipitation patterns. Secondly, both risks share common vulnerability factors, such as socioeconomic characteristics and built environment features, making a combined analysis more comprehensive. The CLIM-SEG framework, particularly its semantic segmentation approach, is applicable to both heat and flood risk assessments, as land cover characteristics are relevant to both hazards. This methodological synergy allows for efficient resource utilization, maximizing the utility of collected data and developed models. Furthermore, urban planners and policymakers typically need to consider multiple climate risks when developing adaptation strategies; thus, an integrated assessment provides a more holistic view for decision-making. By applying the CLIM-SEG framework to two distinct climate risks, the paper demonstrates its versatility and potential for broader applications in climate risk assessment. This combined approach aligns with the growing recognition in urban climate science of the need for multi-hazard risk assessments to inform effective and efficient adaptation strategies in the face of complex, interrelated climate challenges.

As a proof of concept, heat and flood risk models are developed for Boston, a city that experiences significant UHIs, with temperatures rising up to 10°F in the summer (Urban Heat Islands, 2021), and is projected to face increased coastal and riverine flooding (Climate Vulnerability Assessment, 2022). The proposed models leverage publicly available datasets and aerial imagery to ensure transferability to other urban contexts. To obtain risk factors for the developed heat and flood risk model, the research leverages publicly available datasets, such as those released by national and city governments. For data that may not be readily available, computer vision and machine learning (ML) methods are applied in the form of semantic segmentation to extract features from aerial imagery. Computer vision trains computers to perceive and understand the visual world, and ML automates the process of learning from data to make predictions. An automated approach to select desired features from images minimizes the need for human labor as it enables the training of an ML model to complete the task. The main contributions of this research are:

1. A novel dataset comprising 145 aerial images with manually-annotated segmentation maps, enabling the training of robust semantic segmentation models for feature extraction.
2. The development of high-performing semantic segmentation models that accurately identify built and natural environment features in unseen aerial images.
3. The formulation of a heat risk model that identifies census tracts most susceptible to extreme heat by integrating land surface temperature, building, socioeconomic, and aerial image-derived data, empowering urban planners to prioritize heat adaptation strategies.
4. The development of a flood risk model that pinpoints census tracts most vulnerable to flooding by aggregating elevation, slope, flow accumulation, building, socioeconomic, and aerial image-derived data, facilitating targeted flood mitigation efforts.

5. A rigorous evaluation of the performance and validity of the proposed segmentation models, heat risk, and flood risk models using Boston data, demonstrating their potential for application in urban planning and policymaking.

The proposed approach not only addresses the challenges faced by Boston but also holds promise for scalability to other urban areas confronting the impacts of climate change. By providing a cost-effective, data-driven tool for risk assessment at the census tract level, this research contributes to the advancement of urban planning and climate change adaptation, enabling the development of more targeted and effective strategies for building resilient and sustainable cities.

The paper is structured into section eight main sections. Section two provides a comprehensive background on current heat and flood risk modeling approaches, highlighting existing research gaps. Section three and four details the study area, the curation of datasets. Section five outlines the methodology and the development of the semantic segmentation model, and the integration of various risk factors, specifically the risk features employed in the heat and flood risk models, and the process of curating an aerial image and segmentation dataset for the proposed approach. Moreover, Section six showcases the steps to fine-tune a segmentation model to extract relevant land cover information. Section seven presents the results from the segmentation model, the implications of findings, where we compare results with existing risk assessments, and explores the potential for applying proposed framework to other urban areas, and limitations and generalizability to other cities. Finally, Section eight and nine conclude the paper, summarizing key contributions and suggesting directions for future research. Throughout these sections, we emphasize the novel aspects of CLIM-SEG framework and its potential to enhance urban climate risk assessment and adaptation

## 2. Background

Accurate assessment of heat and flood risk at the urban census tract level is crucial for developing effective mitigation and adaptation strategies in the face of climate change. This section explores the current state of research on heat and flood risk modeling, focusing on the methodologies, data sources, and limitations of existing approaches. The review is organized into three main sections: heat risk modeling, flood risk modeling, and research gaps. The heat risk modeling section discusses studies that quantify heat-related vulnerability, develop indices to identify vulnerable populations, and model heat risk as a function of hazard, exposure, and vulnerability, as illustrated in Fig. 1 below. The flood risk modeling section examines the growing adoption of machine learning methods for flood risk assessment and the use of the Analytical Hierarchy Process (AHP) for determining the relative importance of flood-related features. Finally, the research gaps section highlights the lack of integrated modeling approaches that combine semantically segmented aerial images with environmental, socioeconomic, and building datasets, and discusses the limitations of existing land cover datasets. By identifying these gaps, this review underscores the need for novel approaches that can provide more comprehensive, accurate, and updateable assessments of heat and flood risk at the urban census tract level.

### 2.1. Heat risk modeling

Existing research on heat risk modeling can be categorized into three main themes: vulnerability assessment, risk modeling frameworks, and weighting schemes. Studies focusing on vulnerability assessment aim to quantify heat-related vulnerability and identify populations at risk at scale (Ho et al., 2015). For instance, Zhang et al. (2014) and Morabito et al. (2014) evaluated metrics to

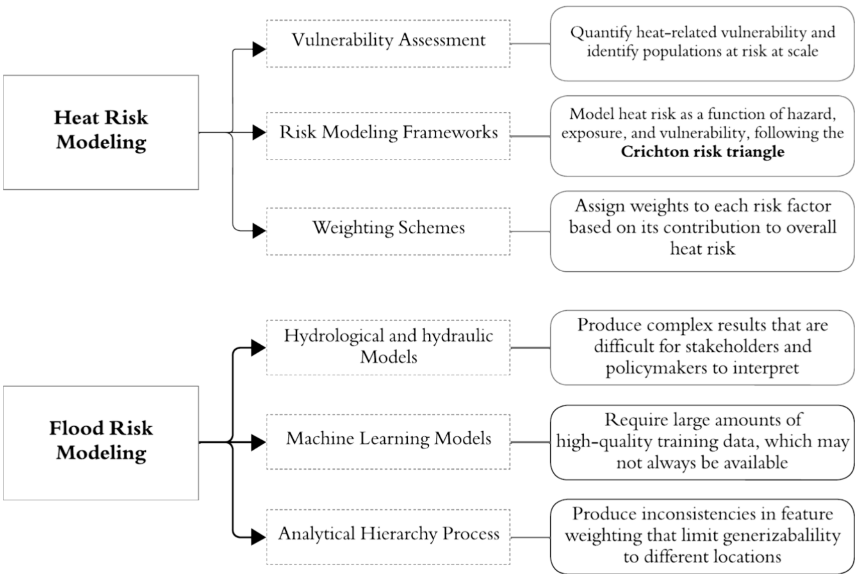


**Fig. 1.** Heat and flood risk literature review overview.

assess heat-related mortality, while Reid et al. (2009, 2012), Vescovi et al. (2005), and Chuang and Gober (2015) developed indices to identify heat-vulnerable populations across various urban areas. These studies provide valuable insights into the factors contributing to heat vulnerability and inform the development of targeted interventions. Building upon this work, Bayomi and Fernandez (2021) proposed a comprehensive framework for heat vulnerability assessment that considers the physical characteristics of the built environment, heat adaptation resources, and individual adaptive capacity, offering a more holistic approach to vulnerability assessment.

Another theme in heat risk modeling research is the development of risk modeling frameworks. A growing body of literature (Buscail et al., 2012; Hu et al., 2017) models heat risk as a function of hazard, exposure, and vulnerability, following the Crichton risk triangle (Crichton, 1999). While these frameworks share a similar overall structure, they vary in the specific risk factors considered and the methods used to aggregate them. For example, Tomlinson et al. (2011) assessed heat health risk in Birmingham, UK by combining high-resolution urban heat island data as the hazard feature, household information to determine exposure, and four vulnerability features such as elderly populations, people with ill health, people living in flats, and people living in high-rise buildings. Specifically, they combine the four vulnerability features into a single "exposed and vulnerable" value by weighting each feature at 25 %, and then they integrate this combined value with hazard features using a weight of 50 % each. Similarly, Buscail et al. (2012) developed a heat-wave health risk index for Rennes, France using land surface temperature as the hazard feature, affected population to assess exposure, and four vulnerability features including socioeconomic status, extreme age, population density, and the presence of old buildings. They follow a similar methodology as Tomlinson et al. (2011) by combining the four vulnerability features using a weight of 25 % each and then incorporating this combined value with hazard features using a weight of 50 % each. These studies demonstrate the importance of integrating diverse data sources to capture the complex nature of heat risk.

The third theme in heat risk modeling research focuses on weighting schemes used to aggregate risk factors. Heat risk models often employ a weighted sum method, assigning weights to each risk factor based on its contribution to overall heat risk. Some models, such as those by Cutter et al. (2003) and Mitchell and Chakraborty (2015), use equal weighting schemes, assuming that all factors contribute equally to heat risk. In contrast, other studies apply statistical analysis to weight features based on their variance (Holand et al., 2011; Wolf and McGregor, 2013) or expert opinion (Alonso and Renard, 2020). Notably, Alonso and Renard (2020) evaluated the vulnerability of various populations based on an expert panel of 35 health professionals. These experts determined that the population category most vulnerable to heat is people aged 75 years and over with a score of 38 %, followed by people with chronic or acute pathologies with a score of 25 %. These studies highlight the importance of carefully considering the relative importance of different risk factors and the potential impact of weighting schemes on model results.

Despite the significant advances in heat risk modeling research, there remain several gaps and limitations. First, many studies rely on a limited set of data sources, often focusing on socioeconomic and meteorological data while overlooking the influence of the built and natural environments. Second, the models developed in these studies may be difficult to update or generalize to different locations due to their reliance on specific datasets or focus on particular cities. Third, there is a lack of research on the integration of high-resolution aerial imagery and advanced computer vision techniques, such as semantic segmentation, for extracting relevant risk factors that can help scale and generalize a modeling approach for climate risks assessment, especially in cities with limited data availability. The paper aims to address these gaps by proposing a novel heat risk modeling approach that integrates diverse data sources, including land surface temperature, building characteristics, socioeconomic data, and features extracted from aerial imagery using semantic segmentation. By incorporating a wider range of risk factors and leveraging advanced computer vision techniques, the proposed model captures the complex nature of heat risk and offers a more comprehensive assessment compared to existing approaches. Furthermore, by using publicly available datasets and aerial imagery, the proposed model democratizes climate risk assessment and can be easily updated and generalized to other urban areas, addressing the limitations of previous studies.

### 2.2. Flood risk modeling

The increasing frequency and intensity of floods due to climate change have heightened the need for accurate and reliable flood risk mapping in urban areas. Traditionally, policymakers and urban planners have relied on hydrological and hydraulic models for flood risk assessment (Amatulli et al., 2018). However, these models often produce complex results that are difficult for stakeholders and policymakers to interpret, limiting their practical application (Maskrey et al., 2022). This limitation has led to the growing adoption of ML methods, which have shown promise in predicting the probability of natural hazards occurrence (Panahi et al., 2020).

ML methods offer several advantages over traditional hydrological and hydraulic models. They can handle large amounts of data from various sources, identify complex patterns and relationships, and provide more accurate predictions (Mosavi et al., 2018). For example, Darabi et al. (2020) trained several ML algorithms, including support vector machines, random forests (Zhu and Zhang, 2022), and artificial neural networks, to predict flood hazard and vulnerability levels in Amol City, Iran. The models were trained on variables such as precipitation, slope, land use, and elevation, demonstrating the ability of ML methods to integrate diverse data sources for flood risk assessment. In addition to predicting the flood hazard and vulnerability levels, the study also determined the importance of each feature with a value between 0 and 1. They found the distance to the channel to be the highest contributing factor with an importance value of 0.92. Conversely, slope had the lowest importance value of approximately 0.39. Although ML methods deliver promising results, they are still limited by requiring large amounts of high-quality training data, which may not always be available (Wang et al., 2015). Moreover, the results of machine learning models can be difficult to interpret, as they often operate as "black boxes," making it challenging to understand the underlying relationships between input variables and predicted outcomes (Rudin, 2019).

Another important aspect of flood risk modeling is determining the relative importance of different flood-related features. The Analytical Hierarchy Process (AHP) has been widely used for this purpose (de Brito and Evers, 2016). AHP is a multi-criteria decision-making method that allows experts to compare and prioritize different factors based on their relative importance (Saaty, 1987). Several studies have applied AHP to flood risk modeling, but the results have been inconsistent. For example, Vojtek and Vojteková (2019) used AHP to create a map of flood-susceptible regions in Slovakia based on seven flood conditioning factors. The study concluded that slope was the most important factor with a normalized weight of 0.39, while elevation had a weight of 0.07. In contrast, Negese et al. (2022) applied AHP to map flood-prone areas in Ethiopia and found that elevation was the most important factor with a weight of 0.17. The next most important features were slope, flow accumulation, annual rainfall, distance to rivers, drainage density, topographic wetness index, land use and land cover, soil types, NDVI, and curvature. These inconsistencies in feature weighting may be attributed to differences in the study areas' geographical and climatic characteristics, as well as the experts' judgments involved in the AHP process. This highlights the need for more research on the relative importance of flood-related features and the development of more robust and generalizable weighting schemes.

Despite the progress made in flood risk modeling research, several gaps and limitations remain. First, many studies focus on a limited set of flood-related features, often overlooking the potential contributions of other relevant factors, such as the built environment and socioeconomic vulnerability (Arrighi et al., 2019). Second, the integration of high-resolution remote sensing data, such as aerial imagery, and advanced computer vision techniques, such as semantic segmentation, has been limited in flood risk modeling research (Ngo et al., 2020; Robertson and Chan, 2009). Third, there is a need for more comprehensive flood risk models that can account for the complex interactions between hazard, exposure, and vulnerability factors (Cutter et al., 2003).

The paper aims to address these gaps by proposing a novel flood risk modeling approach that integrates diverse data sources, including elevation, slope, flow accumulation, building characteristics, socioeconomic data, and features extracted from aerial imagery using semantic segmentation. By incorporating a wider range of flood-related features and leveraging advanced computer vision techniques, the proposed model provides a more comprehensive assessment of flood risk compared to existing approaches. Furthermore, by using publicly available datasets and aerial imagery, the proposed model can be easily adapted to other urban areas, addressing the need for more generalizable flood risk models.

### 2.3. Research gaps

To the best of our knowledge, there has been minimal to no work in modeling the flood and heat risk of census tracts by semantically segmenting aerial images and aggregating relevant hazard and vulnerability features from environmental, socioeconomic, and building datasets. The paper addresses these gaps in heat and flood risk modeling. One major gap is the lack of integrated modeling approaches that combine semantically segmented aerial images with environmental, socioeconomic, and building datasets. Most studies focus on either heat or flood risk modeling, using a limited set of data sources, such as socioeconomic and meteorological data (Cutter et al., 2003; Buscail et al., 2012; Hu et al., 2017). This narrow focus fails to capture the complex interactions between the built environment, natural environment, and social vulnerability factors that contribute to heat and flood risk in urban areas (Cho and Chang, 2017; Koks et al., 2015).


a. Map of Boston in Massachusetts with 2020 Census Tract boundaries. Each grey enclosed area represents a census tract where flood and heat risk metrics were measured and compared.

b. Map of New York City in New York highlighting the boroughs of Manhattan, Queens, and Brooklyn, where the segmentation model was evaluated

**Fig. 2.** Study areas in Boston and New York City with examined census tracts.

Another significant gap is the limitations of existing datasets used for land cover classification and risk modeling. For example, the Open Sentinel Map dataset (Johnson et al., 2022) contains satellite images and segmentations of buildings, roads, water, and 12 land-use categories. However, the annotations are from OpenStreetMap in 2022, while the satellite images span from 2017 to 2020, leading to potential temporal inconsistencies. Moreover, the authors note that the raster format of the OpenStreetMap annotations may introduce geo-registration, rasterization, and temporal errors in the data. Similarly, the National Land Cover Database (NLCD) (Earth Resources Observation and Science Center, 2018) provides satellite imagery with corresponding segmentation maps of land cover classes, but it is updated only every five years, limiting its ability to capture rapid changes in urban environments.

The paper aims to address these gaps by proposing a novel integrated modeling approach that combines semantically segmented aerial images with environmental, socioeconomic, and building datasets to assess heat and flood risk at the census tract level. The paper leverages high-resolution aerial imagery from Apple Maps and Nearmaps data platform, which is updated more frequently than satellite imagery, and contains manually annotated land cover data to ensure high accuracy and consistency. This approach enables the generation of a comprehensive dataset that captures the fine-grained spatial heterogeneity of urban environments and their exposure to heat and flood hazards. The integration of semantically segmented aerial images with environmental, socioeconomic, and building datasets offers several potential benefits for urban risk modeling. First, it enables a more comprehensive and accurate assessment of heat and flood risk by capturing the complex interactions between the built environment, natural environment, and social vulnerability factors. Second, the use of high-resolution aerial imagery allows for the identification of fine-grained spatial patterns of risk within cities, enabling targeted interventions and adaptation strategies. Finally, the ability to update the model with new aerial imagery and data sources ensures that risk assessments remain current and relevant in the face of rapid urban growth and climate change.

## 3. Study area

The primary study area for this research is the city of Boston, Massachusetts, located in the northeastern United States. Boston is the capital and most populous city of the Commonwealth of Massachusetts, with an estimated population of 675,647 as of 2020 (U.S. Census Bureau, 2021a,b; Census Tracts in Boston, 2023). The city covers an area of approximately 89.63 square miles (232.14 km$^2$), of which 48.42 square miles (125.41 km$^2$) is land and 41.21 square miles (106.73 km$^2$) is water (Boston Planning and Development Agency, 2020). Boston has a diverse topography, characterized by a combination of low-lying areas, hills, and a complex coastline. The city's elevation ranges from sea level to 330 feet (100 m) at its highest point in Bellevue Hill (U.S. Geological Survey, 2020). This varied terrain, combined with its coastal location, makes Boston particularly vulnerable to both flooding and heat-related risks. Fig. 2a illustrates the study area, with a map of Boston overlaid with the 2020 Census Tract boundaries. The proposed assessment of vulnerabilities and risks at the granular census-tract level can inform potential mitigation strategies as well. The performance of the segmentation model, which was fine-tuned on Boston census tract images, was also applied and qualitatively evaluated on images from New York City. The goal of this task was to assess the generalizability of an image segmentation model fine-tuned on one urban area to another distinct urban environment.

New York City, particularly the borough of Manhattan, serves as a secondary study area for evaluating the generalizability of our model. New York City is the most populous city in the United States, with an estimated population of 8,804,190 as of 2020 (U.S. Census Bureau, 2021a,b). Manhattan, one of the city's five boroughs, is a densely populated island covering 22.83 square miles (59.13 km$^2$)

**Table 1**
List of all datasets used to extract each feature and their resolution.

| Dataset | Source | Relevant Features | Resolution |
|---|---|---|---|
| Boston and NYC Aerial Image Dataset | Nearmaps and Apple Maps | Land Cover Features:<br>Water<br>Grass<br>Trees<br>Pavement<br>Building<br>Earth | Around 600 × 400 |
| Global Multi-resolution Terrain Elevation Data (GMTED2010) | EarthEnv | Elevation, Slope | 7.5 Arcseconds |
| HydroSHEDS | HydroSHEDS | Flow Accumulation | 3 Arcseconds |
| Boston Buildings Inventory | City of Boston | Building age (year built), height (number of floors), use class (residential, commercial, industrial, public, or vacant), presence of an air conditioning system | N/A |
| Boston Socioeconomic Dataset | 2020 Census Data | Population of:<br>Over 65<br>Limited English proficiency<br>Low to no income<br>With Disabilities<br>People of Color<br>Female Population<br>Non-US Citizens<br>Children under 6 | N/A |
| Surface Urban Heat Islands (SUHI) | Yale Center for Earth Observation | SUHI data from summer 2018 daytime | 300 m |

(NYC Planning, 2020). The borough is characterized by its iconic skyline, with numerous skyscrapers and high-rise buildings, presenting a stark contrast to Boston's urban landscape. Manhattan's average elevation is 33 feet (10 m) above sea level, with its highest natural point reaching 265 feet (81 m) in Washington Heights (U.S. Geological Survey, 2020). Like Boston, New York City experiences a humid subtropical climate (Köppen: Cfa), with hot summers and cool winters (Kottek et al., 2006).

The city faces similar environmental challenges, including vulnerability to sea-level rise, with projections indicating a potential increase of 11–21 in. by the 2050s (New York City Panel on Climate Change, 2019). The urban heat island effect is also significant, with surface temperatures in some areas reaching up to 20°F (11 °C) warmer than surrounding rural areas during summer (NYC Mayor's Office of Resiliency, 2020). These characteristics make New York City, and Manhattan in particular, an excellent counterpoint to Boston for assessing the adaptability of our CLIM-SEG framework to varied urban environments. Fig. 2b shows the specific boroughs of New York City that were included in the evaluation: Manhattan, Queens, and Brooklyn. This cross-application aims to provide insights into the robustness and adaptability of the segmentation approach across different urban settings.

## 4. Materials

This study employs publicly available datasets to develop the heat and flood risk models, and the exact data sources and features contributed by each dataset are discussed in this section and listed in Table 1. One of the main datasets used to extract CLIM-SEG features is the Global Multi-resolution Terrain Elevation Data (GMTED2010). This dataset provides a global digital elevation model (DEM) with a spatial resolution of 7.5 arc-seconds (approximately 250 m at the equator), derived from multiple data sources. Elevation is a relevant feature to measure for flood risk, as areas closer to sea level face higher inundation levels during rainfall events. Slope is the change in elevation between two points, and it is another key element in measuring flood susceptibility. Low slope areas store runoff accumulated from higher slope areas, and steeper slopes help remove runoff water.

## 5. Methodology

This study aims to develop a comprehensive methodology for assessing heat and flood risk at the census tract level in urban areas. The proposed approach integrates a wide range of risk factors, including socioeconomic, built environment, and natural environment characteristics, as well as features extracted from high-resolution aerial imagery using semantic segmentation techniques. The methodology section is divided into two main parts: (1) Risk Features, which describes the various factors considered in the heat and flood risk models, and (2) Dataset curation, which details the process of curating aerial image datasets for fine-tuning the segmentation model and analyzing Boston census tracts. The Risk Features section discusses the features relevant to both heat and flood risk, such as land cover, socioeconomic vulnerability, and building characteristics, as well as features specific to each risk type, such as land surface temperature for heat risk and elevation, slope, and flow accumulation for flood risk. The dataset curation section elaborates on the data collection and preprocessing steps involved in creating the aerial image datasets, including manual segmentation, georeferencing, and masking techniques. By integrating these diverse data sources and employing advanced computer vision and geospatial analysis techniques, this methodology provides a robust framework for assessing and mapping heat and flood risk in urban areas, enabling targeted adaptation and mitigation strategies. Fig. 3 illustrates the overall workflow of the study.

### 5.1. Risk features

To develop comprehensive heat and flood risk models for urban areas, the research considered a wide range of factors encompassing the socioeconomic, built environment, and natural environment characteristics of a city. These factors were selected based on their established relevance in previous studies (Cutter et al., 2003; Weill, 2023; Fatemi et al., 2020) and their potential to provide a holistic understanding of heat and flood risk at the census tract level. While some features, such as land cover, socioeconomic, and

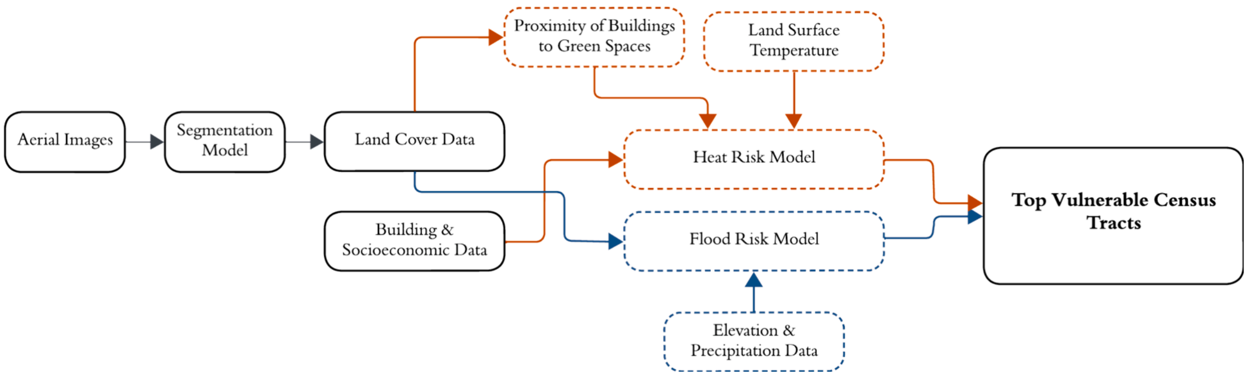


**Fig. 3.** Overview of how features are extracted from various datasets and utilized in the heat and flood risk models, which determine the top vulnerable census tracts.

building data, are relevant to both heat and flood risk, others, such as land surface temperature, are specific to heat risk, and features like elevation and precipitation are more pertinent to flood risk. This section describes the features used in the development of heat and flood risk models, their data sources, and the preprocessing steps applied.

#### 5.1.1. Features relevant to both heat and flood risk

Features contributing to flood and heat risk are classified into two categories: hazards and vulnerabilities. Hazards encompass environmental and geographical factors that influence an area's exposure to extreme heat and flooding, while vulnerabilities capture the characteristics of the population and infrastructure that affect their susceptibility to these hazards. Hazard features include the percentage of a census tract area containing buildings, pavement, trees, grass, earth, and water, which are derived from land cover data. These features provide insights into the physical characteristics of an area that can amplify or mitigate heat and flood risk (Maragno et al., 2020). However, it is important to note that land cover data may have limitations, such as misclassification errors or temporal inconsistencies, which can introduce uncertainties in the risk assessment.

Vulnerability features, on the other hand, measure how the same flood or heat event would impact different social groups or areas. To capture socioeconomic vulnerability, the most recent 2020 census data was used for Boston, which was preprocessed to handle missing data and aggregated at the census tract level. The research focused on specific demographic groups that have been identified as more vulnerable to heat and flood risk, including:

- **Total population:** A common indicator for heat and flood exposure because it enumerates the vulnerable individuals over an urban area (Li et al., 2022).
- **People of color:** Defined as individuals belonging to racial or ethnic minority groups, such as African Americans, Hispanics, Asians, and Native Americans. These groups are more vulnerable to flood and heat risk due to historical disparities in income, housing, and access to resources (Johnson, 2022). In addition, a higher population density of people of color has been linked with heat vulnerability according to a study in New York City (Rosenthal et al., 2014).
- **People with disabilities:** Individuals with physical or mental impairments that limit their ability to prepare for and respond to natural disasters. They are also more likely to face discrimination in times of crisis such as a flood or heat event (Climate Ready Boston Social Vulnerability, 2017).
- **People with limited English proficiency:** Individuals who struggles to speak or understand English, might not be aware of news regarding natural disasters or communicate with emergency personnel (Climate Ready Boston Social Vulnerability, 2017).
- **Women**: Due to physiological differences and social inequities, women are more susceptible to heat-related illnesses and injuries from flood events (Foster et al., 1976; Mehnert et al., 2002). Additionally, a study found that women often suffer more physical injuries from flood related damage, and they may experience harassment in flood shelters (Azad et al., 2013).
- **People who are foreign-born and non-US citizens:** These individuals may face additional barriers in accessing assistance during natural disasters, particularly if they are undocumented (Climate Ready Boston Social Vulnerability, 2017).
- **Children under six years old:** Families who have young children under six years old require more assistance and support during disasters such as flooding. Children are also more vulnerable from stress after a natural disaster (Climate Ready Boston Social Vulnerability, 2017).
- **People in poverty:** Economic constraints can limit individuals' ability to prepare for and recover from floods and extreme heat events *(*Climate Ready Boston Social Vulnerability, 2017*)*.
- **Men and women over 65 years old**: Elderly populations are more likely to suffer from heat-related mortality due to increased physiological limitations (Kenny et al., 2010). It is also more difficult for them to evacuate during flood events (Climate Ready Boston Social Vulnerability, 2017).

While census data provides valuable information on socioeconomic vulnerability, it is important to acknowledge potential limitations, such as undercounting of marginalized populations or temporal lags in data collection (Luetz and Merson, 2019).

In addition to socioeconomic vulnerability, building characteristics that can influence heat and flood risk are also considered. Using the Boston Buildings Dataset, the vulnerability of each census tract is evaluated based on building age, height, type, and the presence of air conditioning units. The dataset was preprocessed to classify buildings into relevant categories, such as old (built before 1980) and tall (six or more stories) buildings. Building age is a significant factor in heat and flood risk, as older buildings may have poorer insulation and be more susceptible to damage (Clarke, 1972; Naughton et al., 2002). The cutoff year of 1980 was chosen based on previous studies that identified buildings constructed before this time as more vulnerable (Buscail et al., 2012). Tall buildings can contribute to the urban heat island effect by impeding natural wind flow and heat dissipation (Urban Heat Islands, 2023; Learn About Heat Islands | US EPA, 2023), while also providing vertical evacuation options during flood events (Aerts et al., 2020).

Building type is another important consideration, with residential buildings being particularly relevant to heat risk due to their high occupancy rates and energy consumption (Williams et al., 2020). Moreover, people spend nearly 90 % of their time indoors, predominantly in residential settings (Klepeis et al., 2001). Additionally, Boston's climate vulnerability assessment predicts that the majority of damages due to flooding will be to residential and mixed-use properties (Climate Resilience Initiatives, 2022). The Boston Buildings Dataset categorizes buildings into residential, commercial, industrial, public, and vacant types, allowing the calculation of the percentage of residential buildings in each census tract. It is worth noting that there may be interactions and dependencies between different risk features. For example, building age and type may be correlated, as older buildings are more likely to be of certain types (Uejio et al., 2011).

Similarly, socioeconomic vulnerability may be linked to building characteristics, as low-income neighborhoods often have older and poorly maintained housing stock (Chuang and Gober, 2015). While the proposed models do not explicitly account for these interactions, they are an important consideration when interpreting the results. Finally, it is crucial to address data privacy and ethics when working with census data and sensitive information about vulnerable populations. In the paper, all data used in the study was aggregated at the census tract level and did not contain any personally identifiable information. Moreover, the study adhered to the ethical guidelines set forth by institutional review board and the American Psychological Association (2017) to protect the privacy and well-being of the individuals represented in the data.

The proposed heat and flood risk models incorporate a comprehensive set of risk features, including land cover, socioeconomic vulnerability, and building characteristics. These features were selected based on their established relevance in literature and their potential to provide a holistic understanding of heat and flood risk at the census tract level. While the paper acknowledges the limitations and uncertainties associated with these data sources, steps were taken to preprocess the data, ensure consistency in terminology, and address data privacy and ethics concerns. By considering a wide range of risk features and their potential interactions, the proposed models aim to provide a robust and nuanced assessment of heat and flood risk in urban areas

*5.1.2. Heat features*

In addition to the shared features, factors specific to increasing resilience to extreme heat are considered (Li et al., 2022), such as land surface temperature and the presence of air conditioning in buildings. To acquire land surface temperature data, the paper leverages the Surface Urban Heat Islands (SUHI) dataset (Chakraborty and Lee, 2019) from the Yale Center for Earth Observation. This dataset contains annual, summer, and winter SUHI intensities in degrees Celsius for day and nighttime, with a spatial resolution of 1 km, for years between 2003 and 2018, and for over 10,000 urban areas worldwide. SUHI intensities are calculated as the difference between the mean land surface temperature of an urban area and its surrounding non-urban buffer, using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on NASA's Terra and Aqua satellites (Chakraborty and Lee, 2019). As hotter temperatures that contribute to heat risk tend to occur during summer days (feeling hot, hot, hot: How the summer heat can get you down, 2016), and to ensure the use of the most recent data available, the proposed heat risk model incorporates SUHI data from summer 2018 daytime. While land surface temperature is a key factor in assessing heat risk, it is important to acknowledge its limitations. Land surface temperature may not fully capture the impact of other factors, such as humidity or wind speed, on human

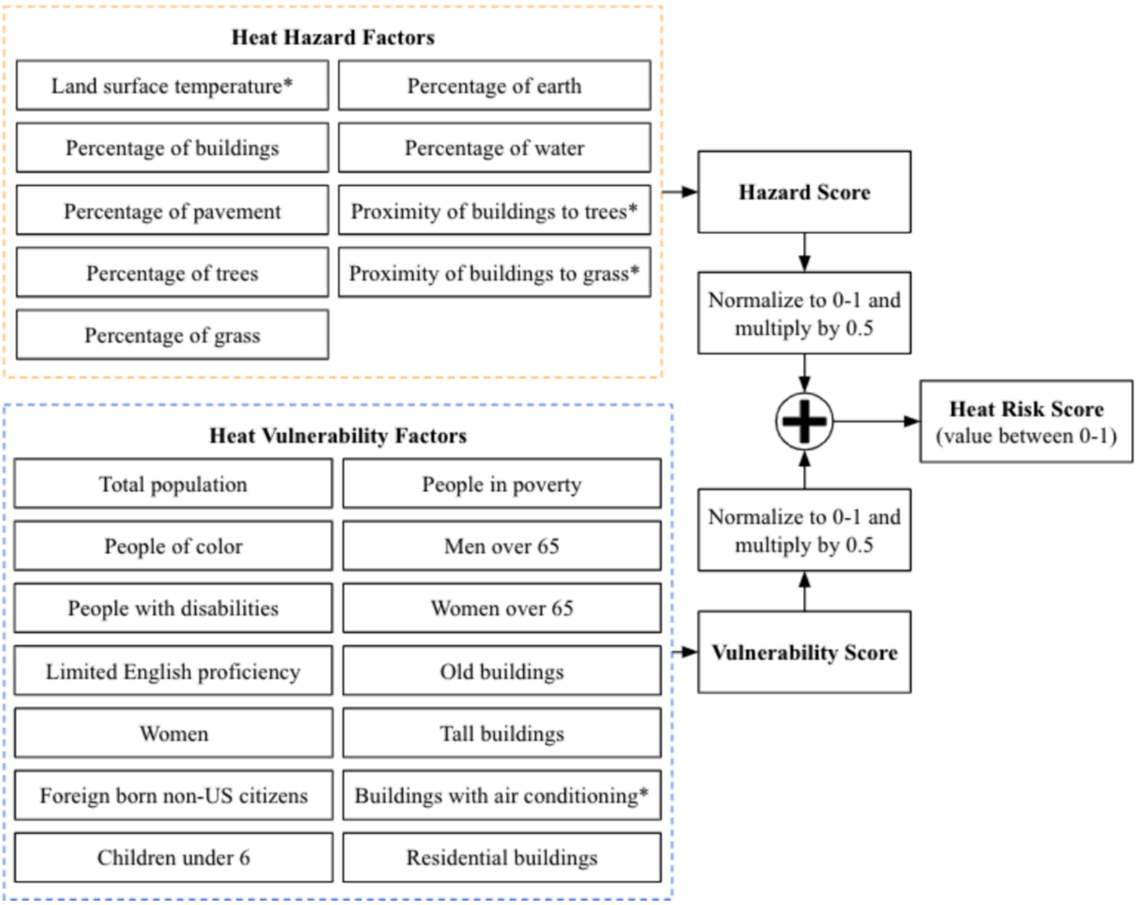


**Fig. 4.** Framework for heat vulnerability combined score calculations. Factors with an asterisk (*) are unique to the heat risk model.

thermal comfort (Zhao et al., 2021). Future studies could consider incorporating additional meteorological variables to provide a more comprehensive assessment of heat exposure.

The presence of air conditioning in buildings is another significant factor to cope with heat exposure. Air conditioning allows buildings to maintain cooler indoor temperatures, providing a refuge from extreme heat (Ostro et al., 2010). In Boston, approximately 82 % of households have air conditioning. Studies have shown that the prevalence of air conditioning is associated with reduced heat-related morbidity and mortality (Skarha et al., 2022). To incorporate this factor into the heat risk model, the percentage of buildings with air conditioning systems in each census tract is computed using data from the Boston Buildings Dataset. Furthermore, when a building lacks air conditioning, proximity to green spaces, such as trees and grass, can help reduce heat risk. Green spaces mitigate urban heat through evapotranspiration and shading (Ziter et al., 2019). The trained segmentation model classifies areas of buildings and green spaces, enabling the calculation of the minimum distance from buildings to trees and grass per census tract. This measure of proximity to green spaces provides insights into the potential cooling effects of vegetation on buildings without air conditioning. However, it is important to recognize the limitations of using minimum distance to green spaces as a measure of proximity. This approach does not account for the size, quality, or accessibility of the green spaces (Nesbitt et al., 2019). Future research could explore more comprehensive metrics, such as the percentage of green space within a certain radius of buildings or the connectivity of green spaces, to better capture the cooling potential of urban vegetation.

The approach of a weighted sum model was used to estimate extreme heat risk at the census tract level by aggregating the heat hazard and heat vulnerability factors, as shown in Fig. 4. After obtaining a single combined risk score between 0 and 1 for each census tract, the top five census tracts with the highest risk scores are identified, indicating the areas that are considered the most vulnerable to extreme heat.

It is also crucial to consider the potential interactions between heat risk factors. Land surface temperature, air conditioning prevalence, and proximity to green spaces may interact with each other and with other vulnerability factors, such as socioeconomic status and building characteristics (Jenerette et al., 2011). For example, low-income neighborhoods may have lower air conditioning prevalence and less access to high-quality green spaces, exacerbating their vulnerability to extreme heat. Future studies could employ more advanced statistical techniques, such as multilevel modeling or structural equation modeling, to investigate these complex interactions and their implications for urban heat resilience.

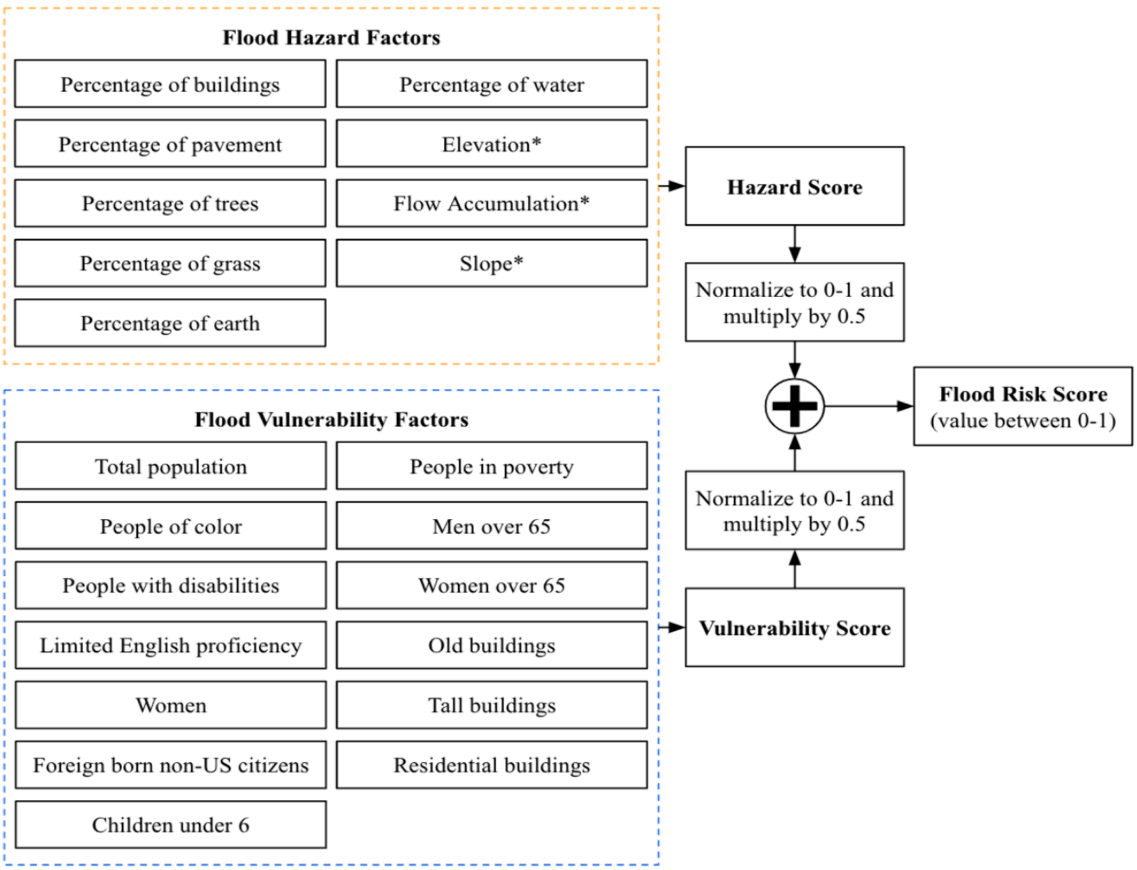


**Fig. 5.** Framework for flood vulnerability combined score calculations. Factors with an asterisk (*) are unique to the flood risk model.

### 5.1.3. Flood features

The proposed flood risk model incorporates hazard features specific to flood risk, such as elevation, slope, and flow accumulation. These factors play a crucial role in determining the susceptibility of an area to flooding (Manfreda et al., 2015). Elevation and slope data are obtained from the Global Multi-resolution Terrain Elevation Data (GMTED2010) dataset (Danielson and Gesch, 2011). GMTED2010 provides a global digital elevation model (DEM) with a spatial resolution of 7.5 arc-seconds (approximately 250 m at the equator), derived from multiple data sources, including the Shuttle Radar Topography Mission (SRTM) and the ASTER Global Digital Elevation Model (GDEM) (Danielson and Gesch, 2011). The dataset has a vertical accuracy of ±5 m at the 95 % confidence level, making it suitable for regional-scale flood risk assessment.

Elevation is a critical factor in flood risk assessment, as areas closer to sea level face higher inundation levels during rainfall events, and lower elevation areas are more susceptible to flooding due to the natural flow of water from higher to lower regions (Hitouri et al., 2024; Roopnarine et al., 2022). The average, minimum, maximum, and median elevation values for each census tract to capture the range and central tendency of elevation within the tract, providing a comprehensive characterization of the tract's topography. Slope, defined as the change in elevation between two points, is another key factor in flood susceptibility (Swain et al., 2020; Ramesh and Iqbal, 2020). Low slope areas tend to store runoff accumulated from higher slope areas, while steeper slopes facilitate the rapid disposal of runoff water. The average, maximum, minimum, and median slope values are calculated in degrees for each census tract, using the same statistical measures as for elevation to ensure consistency and comparability between the two factors.

Flow accumulation, a measure of the amount of water that drains through and accumulates at a specific point, is a valuable indicator of flood risk at the census tract level (Vojtek and Vojteková, 2019). The flow accumulation values are calculated for each census tract using the D8 flow direction algorithm (Tarboton, 2003) applied to the GMTED2010 DEM. The D8 algorithm determines the direction of flow from each cell to its steepest downslope neighbor, allowing for the delineation of drainage networks and the calculation of flow accumulation. Next, the average, maximum, minimum, and median flow accumulation values are computed for each census tract to characterize the spatial distribution of water accumulation within the tract. While flow accumulation provides valuable insights into the potential for water accumulation and flooding, it is important to acknowledge its limitations. Flow accumulation does not account for the influence of soil type, land cover, or drainage infrastructure on water flow and accumulation (Mahmoud and Gan, 2018). These factors can significantly modify the actual water flow patterns and the resulting flood risk (Mahmoud and Gan, 2018). Future research could integrate additional data layers, such as soil maps and land cover classifications, to refine the flow accumulation estimates and provide a more accurate representation of flood risk.

To quantify flood risk, the sum of flood vulnerability (which includes socioeconomic and building features) scores and flood hazard scores was used to obtain a singular flood risk score for each census tract as shown in Fig. 5. The hazard and vulnerability scores are normalized and combined with a weight of 0.5 each to obtain flood risk scores between 0 and 1.

It is also crucial to consider the potential interactions between flood risk factors and their temporal variability. Elevation, slope, and flow accumulation may interact with each other and with other vulnerability factors, such as building characteristics and socioeconomic status, to create complex flood risk patterns (Koks et al., 2015). Moreover, flood risk factors, particularly those related to rainfall and flow accumulation, can vary significantly over time (Muis et al., 2015). Future studies could employ dynamic modeling approaches, such as coupled hydrologic-hydraulic models, to capture the temporal evolution of flood risk and its interactions with other factors.

## 5.2. Dataset curation

To extract features from aerial images for the proposed heat and flood risk models, two image datasets were curated: one for fine-tuning a semantic segmentation model and another for analyzing Boston census tracts. The dataset for fine-tuning the segmentation model consists of 545 aerial images of Boston with manually created segmentation masks, while the dataset for census tract analysis covers all 207 Boston census tracts using georeferenced images.

**Table 2**
Comparison of the proposed Boston and NYC Aerial Image Dataset and other relevant image segmentation datasets.

| Dataset | Types of Images | Aerial Imagery | Image Resolution | Semantic Segmentation | All Required Classes |
|---|---|---|---|---|---|
| ADE20K | Real-world Images | No | 2048 × 1536 | Yes | No |
| SpaceNet + DeepGlobe | Pre and Post Disaster Images | No | Varies | Yes | Yes |
| OpenSentinelMap | Real-world Images | No | 192 × 192 | Yes | Yes |
| Chesapeake Land Cover Dataset | Real-world Images | No | Low | Yes | Yes |
| AIDER | Post Disaster Images | Yes | Varies | No | N/A |
| UC Merced Land Use Dataset | Real-world Images | Yes | 256 × 256 | No | N/A |
| FloodNet | Post Disaster Images | Yes | 4000 × 3000 | Yes | No |
| LandCover.ai | Real-world Images | Yes | 512 × 512 | Yes | No |
| Proposed Boston and NYC Aerial Image Dataset | Real-world Images | Yes | Around 600 × 400 | Yes | Yes |

#### 5.2.1. Dataset for fine-tuning the segmentation model

Aerial images provide valuable information about the built and natural environments of an urban area with minimal manual effort. However, access to high-quality datasets of urban aerial images with segmentation maps for semantic classes purposes is limited, as shown in Table 2. To address this issue, a novel dataset is curated by collecting 82 images from Apple Maps (Maps, 2023) and 63 images from the Nearmap MapBrowser (MapBrowser, 2023) captured in August 2022. The locations were selected to represent a diverse range of urban landscapes across Boston, covering residential, commercial, and industrial areas.

The paper focused on classifying image pixels into six semantic classes: building, pavement, tree, grass, earth, and water. These classes were chosen based on their relevance to heat and flood risk factors, such as the presence of impervious surfaces, vegetation, and water bodies. Pixels that did not fit into these six classes were labeled as “clutter.” Manual segmentation was performed using Adobe Photoshop (Adobe Inc., 2019) to create high-quality ground truth data. However, it is important to acknowledge the limitations of manual segmentation, such as human error and subjectivity (Ulku and Akagunduz, 2022). To mitigate these issues, a single annotator performed all segmentations following a consistent protocol and received regular feedback from the research team.

To supplement the manually segmented dataset and increase its size and diversity, a dataset of 400 images and segmentation maps from the public Drone Dataset was incorporated (Aerial Semantic Segmentation Drone Dataset, 2021). Although this dataset contains 24 semantic classes, only the six classes relevant to the study (building, pavement, tree, grass, earth, and water) and the “clutter” class were considered. The final dataset of 545 images used for fine-tuning the segmentation model consists of 82 images from Apple Maps, 63 images from Nearmap MapBrowser, and 400 images from the Drone Dataset. Fig. 6 shows an example images from Apple Maps and Nearmap MapBrowser with manually created segmentation maps.

While a dataset of 545 images is reasonable for fine-tuning a segmentation model, it may still be considered relatively small compared to some large-scale benchmarks (Cordts et al., 2016). To ensure the robustness and generalizability of the trained model, data augmentation techniques, such as random rotations, flips, and color jittering, were utilized during the training process (Shorten and Khoshgoftaar, 2019). Additionally, the model’s performance was carefully monitored on a separate validation set to detect potential overfitting.

#### 5.2.2. Dataset for analyzing on Boston census tracts

To analyze the land cover composition of each Boston census tract, a comprehensive dataset using georeferenced aerial images from Nearmap. These images were obtained through the Nearmap API and cover the entire spatial extent of all 207 census tracts in Boston. The georeferenced images have a spatial resolution of 7.5 cm per pixel, providing a high level of detail for land cover classification. To associate image pixels with their corresponding census tracts, the georeferencing information provided by Nearmap was utilized. Each image is accompanied by a coordinate file containing the resolution in meters per pixel and the world coordinates of the top-left image pixel. Using this information, the pixel coordinates of each image was converted to world coordinates, allowing the aggregation of land cover data for each census tract.

One challenge encountered during the creation of the census tract dataset was the presence of multiple census tracts within a single aerial image. To address this issue, masks were generated for each image-census tract pair, where pixels corresponding to a particular census tract were marked as white and all other pixels were marked as black. This masking process ensures that pixels are correctly assigned to their respective census tracts, avoiding overcounting or undercounting issues. For example, the aerial image in Fig. 7 has two census tracts in it: 2.01 and 3.01. Therefore, there are two masks that are inverses of each other. The masks are overlaid on the segmentation to only consider the pixels in the white areas of the mask to be added to the land cover count for that census tract.

It is important to consider the potential impact of image quality and seasonality on the accuracy of segmentation results. While the Nearmap images used in this study were captured during the same season (summer) to minimize the effects of seasonal variations on vegetation, some variations in image quality and lighting conditions were still present. To mitigate these issues, image normalization techniques, such as histogram equalization, were applied to enhance the consistency of the images (Singh and Kapoor, 2014).

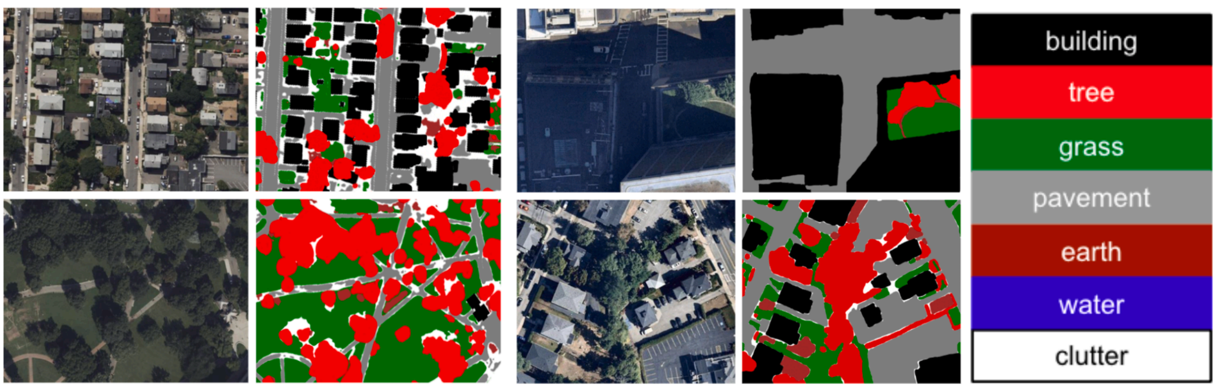


**Fig. 6.** Example aerial images from Apple Maps (Maps, 2023) and Nearmap MapBrowser (MapBrowser, 2023) and the manually created segmentations for fine-tuning the segmentation model. A legend displaying the color for each class is displayed next to the ground-truth images and segmentations.

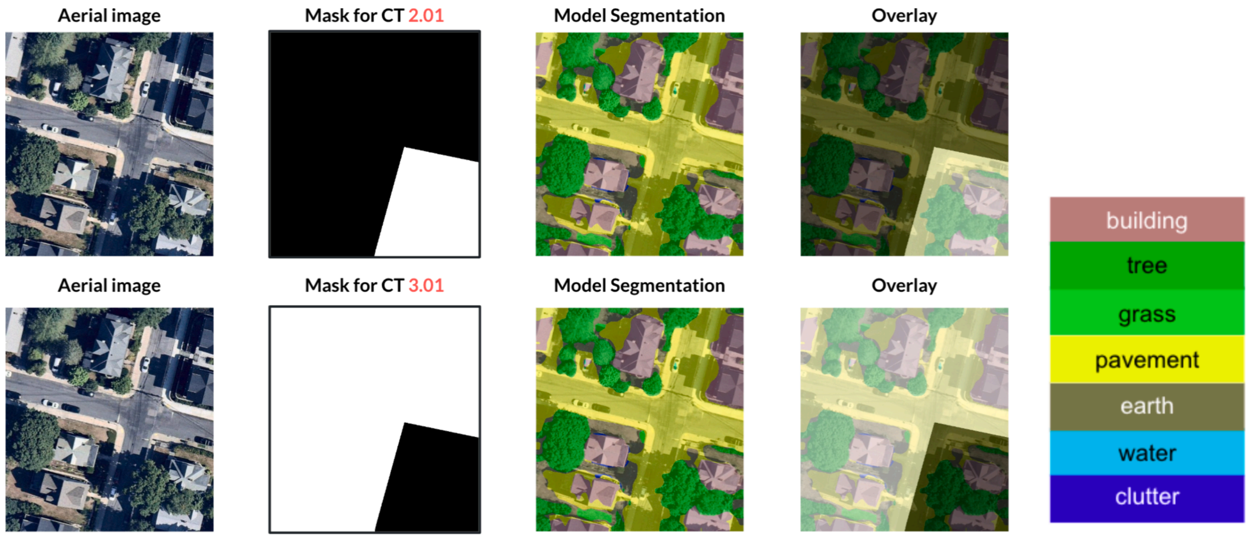


**Fig. 7.** Aerial image and segmentation masking process.

Finally, the use of aerial imagery for urban analysis raises ethical concerns regarding privacy. To address these concerns, all aerial images used in this study were obtained from licensed providers (Apple Maps and Nearmap) and did not contain any personally identifiable information. Furthermore, all land cover data was aggregated at the census tract level, preventing the identification of individual properties or households.

## 6. Segmentation model fine-tuning

Several machine learning architectures have been employed for semantic segmentation on scene images, including convolutional neural networks (CNNs) and vision transformers. CNNs, such as PSPNet (Zhao et al., 2017), utilize convolutional, pooling, and fully-connected layers to capture spatial and temporal features in images, making them well-suited for pattern recognition tasks (Saha, 2018). However, CNNs tend to focus on local information in an image, potentially missing more global features (Dosovitskiy et al., 2020). On the other hand, vision transformer models, such as Segmenter (Strudel et al., 2021) and KNet (Zhang et al., 2020), divide images into patches, encode them into vectors, and use these vectors along with positional information as inputs to the model (Siddiqui, 2022). This approach allows vision transformers to capture both local and global context in images. In this paper, the performance of three state-of-the-art segmentation models, PSPNet (CNN), Segmenter, and KNet (vision transformers) is compared on the task of semantic segmentation of aerial images. PSPNet employs a pyramid pooling module to capture multi-scale context information, while Segmenter and KNet leverage self-attention mechanisms to model long-range dependencies in images. By including

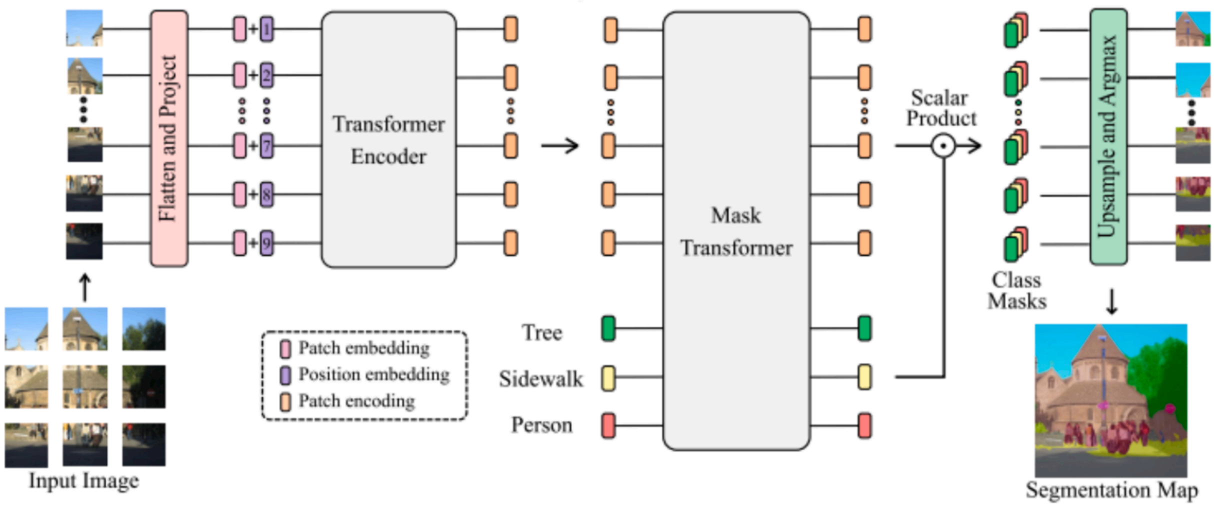


**Fig. 8.** Overview of Segmenter architecture.

both CNN and vision transformer architectures in the comparison, the paper aims to assess their relative strengths and weaknesses for aerial image segmentation.

Training these segmentation models from scratch requires large amounts of labeled data, which can be time-consuming and costly to acquire. Therefore, the paper leverages pre-trained models that have been trained on the ADE20K dataset (Zhou et al., 2016), a large-scale scene parsing benchmark containing 20,210 images with 150 semantic categories. ADE20K was chosen as the pre-training dataset due to its diverse set of scene images and high-quality annotations, which have been shown to facilitate transfer learning for various computer vision tasks (Xie et al., 2021). However, the performance of models pre-trained on ADE20K for aerial image segmentation has not been extensively studied.

### 6.1. Segmenter

For the task of identifying features in aerial images, Segmenter (Strudel et al., 2021) emerged as the best-performing semantic segmentation model that was fine-tuned. Segmenter is a transformer-based encoder-decoder model that predicts a semantic class for each image pixel, as shown in Fig. 8. The encoder learns to map the image embeddings to patch encodings, and the decoder maps the patch encodings to patch-level class scores, which are up sampled to pixel-level scores.

An input image is divided into patches, which are flattened and projected into patch embeddings. The patch embeddings are fed with position embeddings into the transformer encoder to obtain patch encodings. The patch encodings are then decoded by the mask transformer to obtain segmentation masks for each class.

#### 6.1.1. Encoder

Each input image with dimensions H (height) * W (width) * C (number of channels) is divided into N patches of size $P \times P$, where N equals $(H * W)/P^2$. The choice of patch size P is a crucial hyperparameter that affects the model's performance and computational efficiency. Smaller patch sizes lead to a higher number of patches and, consequently, a larger number of computations, while larger patch sizes may result in a loss of fine-grained details. In the study's experiments, a patch size of $16 \times 16$ pixels was found to provide a good balance between performance and efficiency.

Next, a linear projection layer is applied to each patch, mapping it to a D-dimensional embedding space. The resulting patch embeddings are denoted as $x\_0 = [E_{x1}, E_{x2},\ldots,E_{xN}]$, where $E\_xi \in R\hat{}D$. To incorporate positional information, learnable positional embeddings $pos = [pos_1, pos_2,\ldots pos_N]$, where $pos\_i \in R\hat{}D$, are added to the patch embeddings. The final input to the transformer encoder is a sequence of tokens $z\_0 = x\_0 + pos$.

The transformer encoder consists of L layers, each comprising a multi-headed self-attention (MSA) block and a multilayer perceptron (MLP). The MSA block allows the model to attend to different parts of the input sequence, capturing long-range dependencies and contextual information. The MLP block applies non-linear transformations to the attended features, enabling the model to learn complex representations. Layer normalization (LN) is applied before each block to normalize the activations and stabilize the training process, as follow:

$$a_{i-1} = MSA(LN(z_{i-1})) + z_{i-1}$$

$$z_i = MLP(LN(a_{i-1})) + a_{i-1}$$

where $i = \{1, \ldots.., L\}$.

Residual connections are used to facilitate the flow of information and prevent the vanishing gradient problem. The output of the L-th layer, z_L, represents the contextualized patch encodings.

#### 6.1.2. Decoder

The decoder takes the patch encodings z_L and generates a segmentation map with dimensions H*W*K, where K is the number of

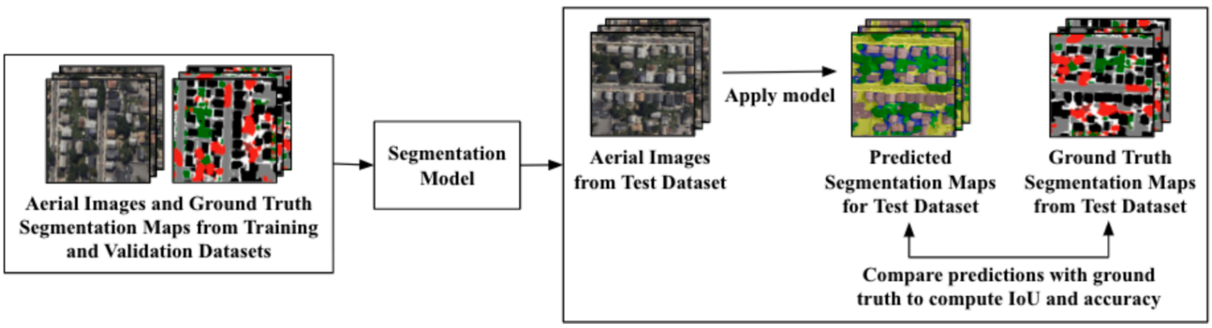


**Fig. 9.** Overview of the fine-tuning process. The aerial images from the curated datasets were used and segmentations to fine-tune the segmentation model. Then, the fine-tuned segmentation model is applied on images from the test dataset, and the model is evaluated by comparing the model predictions with ground truth segmentations.

semantic classes. The decoder consists of a mask transformer that processes the patch encodings along with a set of K learnable class embeddings cls = [cls_1, …, cls_K] of size (K*D), where cls_i ∈ R^D. These class embeddings are randomly initialized and learned during training to represent the semantic classes. The mask transformer computes the scalar product between the L2-normalized patch embeddings and the class embeddings. The L2-normalization ensures that the patch embeddings have unit norm, which helps to stabilize the training process and improves the model's convergence. The resulting scores are then reshaped and upsampled to the original image size H*W using bilinear interpolation. Finally, a softmax function is applied along the class dimension to obtain pixel-wise class probabilities, generating the final segmentation map.

### 6.2. Fine-tuning process

The paper hypothesizes that, since these pre-trained models were originally designed for street view images, they may not generalize well to aerial images without fine-tuning. To test this hypothesis, a set of experiments were conducted to compare the performance of the pre-trained models on curated aerial image dataset before and after fine-tuning, as shown in Fig. 9. The fine-tuning process involves adapting the pre-trained models to the specific characteristics of aerial images by updating the model parameters using a smaller learning rate and a limited number of training iterations. Specifically, all layers of the pre-trained models are fine-tuned using a learning rate of 0.0001 and a batch size of 8 for 50 epochs. The models are optimized using the AdamW algorithm (Loshchilov and Hutter, 2017) with a weight decay of 0.01.

To evaluate the performance of the fine-tuned models, two widely adopted metrics in semantic segmentation are used: intersection over union (IoU) and pixel accuracy (Long et al., 2015). IoU measures the overlap between the predicted and ground truth segmentation masks. Pixel accuracy represents the percentage of correctly classified pixels in the segmentation map, while IoU measures the overlap between the predicted and ground truth segmentation masks. IoU is calculated as the area of intersection between the predicted and ground truth masks divided by their area of union. These metrics provide a comprehensive assessment of the model's performance in correctly classifying pixels and accurately delineating object boundaries.

Training machine learning models, especially larger transformer architectures, requires parallel computing power not capable of CPUs alone. Therefore, cloud-based processing with a large computational capacity is used to train the semantic segmentation models using 1 GPU and 20 CPUs (Reuther et al., 2018). This approach allowed to efficiently train the models while maintaining cost-effectiveness compared to using multiple GPUs or a larger cluster of CPUs.

## 7. Results

### 7.1. Socioeconomic data

To assess the socioeconomic vulnerability of Boston's census tracts, demographic data from the 2020 American Community Survey (ACS) 5-year estimates (U.S. Census Bureau, 2021a,b) are analyzed. The paper focused on eight vulnerable groups that are more susceptible to the impacts of extreme heat and flooding events: people of color, people with disabilities, people with limited English proficiency, women, foreign-born non-US citizens, children under 6 years, people living below the poverty line, and elderly populations (65 years and older). These groups are considered vulnerable due to various factors, such as limited access to resources, reduced mobility, and pre-existing health conditions, which can exacerbate the adverse effects of heat and flood hazards (Cutter et al., 2003; Flanagan et al., 2011).

The population density of each vulnerable group is calculated by dividing the group's population by the total area of the census tract. The overall socioeconomic vulnerability score is calculated as the sum of the population densities of the eight vulnerable groups, with a maximum possible score of eight. A higher score indicates a greater concentration of vulnerable populations within a census tract, suggesting increased susceptibility to heat and flood risks. Table 3 presents the top five census tracts with the highest socio-economic vulnerability scores. Two of these census tracts (705.02 and 709.01) are located in the Shawmut neighborhood, which is known for its diverse population and high proportion of low-income households (Boston Planning and Development Agency, 2020). The Fenway-Kenmore (104.03) and Chinatown (702.02) neighborhoods also contain census tracts with high concentrations of vulnerable populations, likely due to the presence of student communities and immigrant enclaves, respectively (Boston Planning and Development Agency, 2020).

**Table 3**
Top five census tracts with the highest socioeconomic vulnerability scores in Boston.

| Census Tract | Approximate Location | Score |
|---|---|---|
| 705.02 | Shawmut | 5.949 |
| 104.03 | Fenway–Kenmore | 5.900 |
| 702.02 | Chinatown | 5.375 |
| 607 | D Street / West Broadway | 5.179 |
| 709.01 | Shawmut | 5.134 |

### 7.2. Buildings

To assess the building vulnerability of Boston's census tracts, the data from the Boston Buildings Dataset is analyzed (Boston Planning and Development Agency, 2020). The paper considers three main factors that contribute to a building's vulnerability to heat and flood risks: age, height, and type. Older buildings (constructed before 1950) are more likely to have poor insulation and outdated infrastructure, making them less resilient to extreme weather events (Hinojosa and Meléndez, 2017). Low-rise buildings (under six floors) are more susceptible to heat exposure and lack vertical evacuation options during flooding events (Aerts et al., 2014). Residential buildings are of particular concern due to their high occupancy rates and the potential for prolonged exposure to hazardous conditions (Klinenberg, 2015). Google Street View images confirmed that these census tracts are densely populated with residential buildings that are lower in height. This aligned with the data analysis showing 98 % of buildings are residential and below 6 floors, and 92 % of buildings were built before 1950.

Next, the percentage of buildings in each census tract that met these vulnerability criteria is calculated (pre-1950 construction, under six floors, and residential type). The building vulnerability score was then computed by taking the average of these three percentages and normalizing the scores to a range of 0–1. A higher score indicates a greater proportion of vulnerable buildings within a census tract. Table 4 presents the census tracts with the highest and lowest building vulnerability scores. The top five most vulnerable tracts are located in neighborhoods with a high concentration of older, low-rise residential buildings, such as Dorchester Heights (602 and 603.01), Ashmont (1004), and West Roxbury (1105.01). Conversely, the least vulnerable tracts are found in areas with a mix of newer, taller, and non-residential buildings, such as the South End (704.02), West End (203.01 and 203.04), and Seaport (9812.02). This area includes several non-residential buildings such as the Massachusetts General Hospital, The Liberty Hotel, and Shriner's Children's Hospital, and only 10 % of buildings in this area are residential.

### 7.3. Models assessment

The research compared the performance of PSPNet, Segmenter, and KNet on the curated aerial image dataset to determine the most suitable model for the segmentation task. The models were fine-tuned using the same hyperparameters and evaluated using pixel accuracy and intersection over union (IoU) metrics. The segmentation models were evaluated in three scenarios: pre-trained on ADE20K, pre-trained and then fine-tuned on curated dataset, and trained only on curated dataset without prior fine-tuning. The three scenarios, as shown in Table 5, reveal that the fine-tuned and trained segmentation models show significant improvement in identifying heat and flood risk features in aerial images over the pre-trained models. The fine-tuned Segmenter model achieves the highest performance in mean IoU and pixel accuracy scores for building, tree, pavement, and water classes, with a pixel accuracy of 97.85 % and an IoU of 0.9578, compared to PSPNet (pixel accuracy: 93.76 %, IoU: 0.875) and KNet (pixel accuracy: 79.27 %, IoU: 0.719). For KNet and Segmenter, the fine-tuned versions outperform the pre-trained models and those trained only on aerial image dataset. Furthermore, these models can segment a batch of 1000 images in under 2 s, significantly reducing human labor costs. The predicted segmentations were qualitatively compared with ground truth segmentations from human annotators in addition to quantitatively assessing them using class-level IoU and pixel accuracy scores.

Next, the segmentation models are analyzed using test dataset for fine-tuning, which contains ground truth segmentations. Then, the best-performing model is identified to use in subsequent experiments, such as predicting the segmentation maps for the heat and flood risk models. Analysis findings show that the fine-tuned Segmenter model outperforms PSPNet and KNet in qualitative and quantitative evaluations. In the qualitative assessment, all three fine-tuned models can effectively identify tree and grass areas. Segmenter and KNet outperformed PSPNet in detecting building areas, and PSPNet is more likely to misclassify buildings as pavement. In addition, PSPNet and KNet are more likely than Segmenter to incorrectly classify water as buildings or pavement. Fig. 10 shows the segmentation maps predicted by the fine-tuned PSPNet, Segmenter, and KNet models for a subset of the aerial images in the test dataset.

Quantitatively, the fine-tuned models show comparable performance on the test dataset. However, the fine-tuned Segmenter and KNet models show better results for segmenting trees, pavement, earth, and water than the fine-tuned PSPNet. Segmenter also outperforms PSPNet and KNet in segmenting buildings. Overall, it was determined that the fine-tuned Segmenter model performs the best

**Table 4**
Census tracts with the highest and lowest building vulnerability scores in Boston.

| High Building Vulnerability Areas | | | Low Building Vulnerability Areas | | |
|---|---|---|---|---|---|
| Census Tract | Location | Vulnerability Score | Census Tract | Location | Vulnerability Score |
| 602 | Dorchester Heights | 0.961 | 704.02 | South End | 0.250 |
| 603.01 | Dorchester Heights | 0.952 | 203.01 | West End | 0.301 |
| 1004 | Ashmont | 0.949 | 909.01 | Dorchester | 0.314 |
| 1105.01 | West Roxbury | 0.937 | 9811 | Forest Hills | 0.314 |
| 2.01 | Brighton | 0.935 | 9813 | Boston Logan Airport | 0.336 |
| 911 | Uphams Corner | 0.933 | 203.04 | West End | 0.354 |
| 1006.01 | Dorchester Center | 0.927 | 9807 | Stony Brook Reservation | 0.370 |
| 2.02 | Brighton | 0.926 | 9812.02 | Seaport (Cruise Terminal) | 0.372 |
| 501.01 | Eagle Hill | 0.924 | 808.01 | Mission Hill | 0.378 |
| 502 | Eagle Hill | 0.922 | 9815.02 | Beachmont | 0.389 |

**Table 5**
IoU and accuracy values for each model type across the six land cover classes.

| Segmentation Model Results (Baseline and Fine-tuned) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Model | Metric | Building | Tree | Grass | Pavement | Earth | Water |
| Pre-trained PSPNet | IoU | 21.37 | 24.52 | 22.48 | 0.91 | 9.18 | 0.11 |
| | Acc | 42.09 | 64.31 | 26.37 | 0.91 | 20.86 | 0.12 |
| Fine-tuned PSPNet | IoU | 65.55 | 61.16 | 86.35 | 72.72 | 66.66 | 81.49 |
| | Acc | 81.12 | 78.93 | 93.13 | 89.12 | 76.78 | 81.98 |
| Trained PSPNet | IoU | 66.68 | 61.58 | **87.56** | 77.31 | 70.16 | 95.12 |
| | Acc | 86.47 | 83.31 | **93.76** | 88.34 | 76.46 | 95.69 |
| Pre-trained KNet | IoU | 25.4 | 35.45 | 26.85 | 2.65 | 9.41 | 0.0 |
| | Acc | 55.96 | 72.25 | 31.53 | 2.66 | 24.6 | 0.0 |
| Fine-tuned KNet | IoU | 67.12 | 65.28 | 87.27 | 78.09 | **71.93** | 95.12 |
| | Acc | 82.16 | 79.81 | 92.09 | 92.45 | **79.27** | 95.84 |
| Trained KNet | IoU | 53.74 | 59.37 | 44.91 | 58.09 | 56.85 | 95.33 |
| | Acc | 65.95 | 78.97 | 47.63 | 91.15 | 76.89 | 96.09 |
| Pre-trained Segmenter | IoU | 36.36 | 42.39 | 36.66 | 13.47 | 10.99 | 12.37 |
| | Acc | 81.86 | 75.19 | 41.37 | 13.93 | 32.61 | 12.42 |
| Fine-tuned Segmenter | IoU | **71.88** | **65.73** | 86.69 | **78.24** | 70.2 | **95.78** |
| | Acc | **85.17** | **82.49** | 92.81 | **89.94** | 76.35 | **97.85** |
| Trained Segmenter | IoU | 55.57 | 50.09 | 80.79 | 62.64 | 64.59 | 92.99 |
| | Acc | 76.91 | 70.21 | 89.56 | 82.66 | 71.57 | 93.79 |

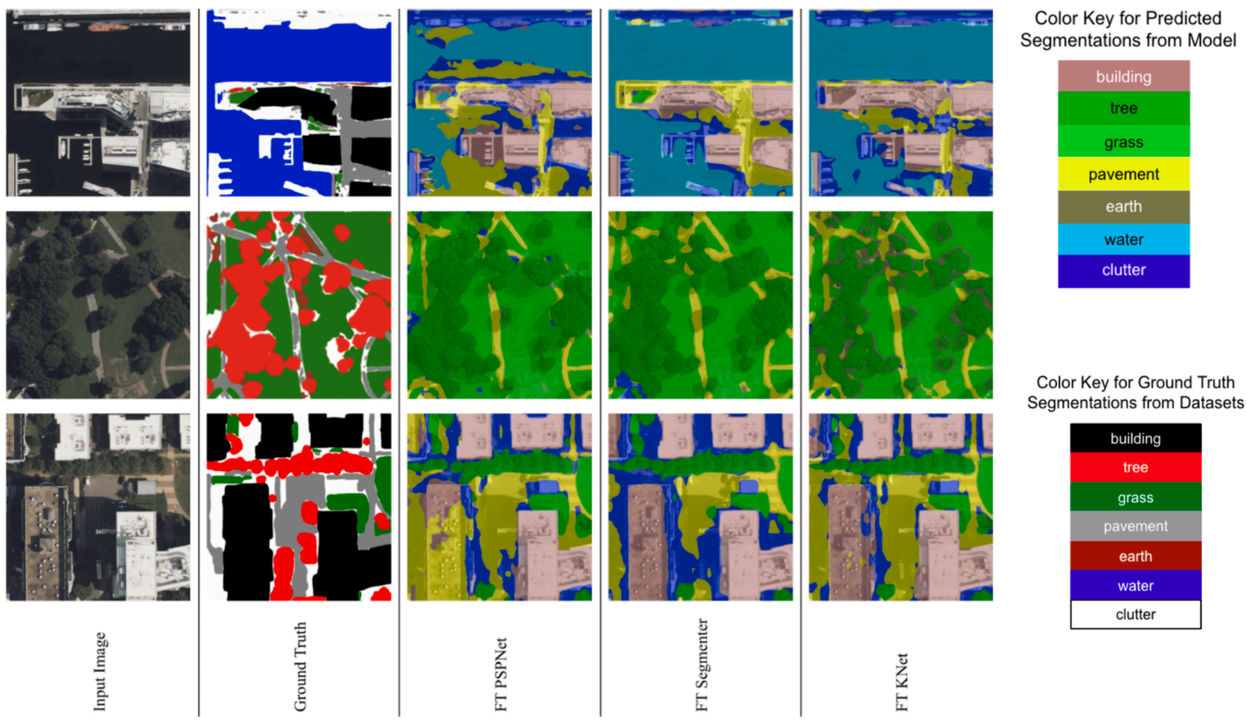


**Fig. 10.** Predicted segmentations for Boston images using the fine-tuned PSPNet, Segmenter, and KNet.

among these three models. As a result, Segmenter was selected for the remainder of the segmentation experiments.

## 7.4. Segmentation results for Boston census tract images

To assess the land cover composition and extract relevant features for the heat and flood risk models, the fine-tuned Segmenter model is applied to aerial images of all Boston census tracts. The Segmenter model automatically generates semantic segmentation maps, significantly reducing the need for manual labor in analyzing land cover patterns. Fig. 11 presents examples of Boston aerial images alongside their corresponding segmentation maps predicted by the fine-tuned Segmenter model. The model accurately delineates buildings, pavement, trees, grass, earth, and water, providing a detailed characterization of the urban landscape.

Using the predicted segmentation maps, the percentage of each census tract covered by the six land cover classes is calculated. Table 6 highlights the top five census tracts with the highest percentage of each land cover class, revealing the spatial heterogeneity of urban form across Boston. For instance, census tract 701.02 in Downtown Boston has the highest percentage of building coverage (31.43 %), while census tract 9807, which encompasses Stony Brook Reservation, has the highest percentage of tree coverage (54.32 %).

To quantify the overall heat and flood risk contributed by land cover, the individual land cover percentages are aggregated using a weighted sum approach. An equal weight of +1 or −1 to each risk factor were applied to compute a heat and flood risk score per census tract between 0 and 1. Positive weights were assigned to classes that increase risk, such as buildings and pavement, while negative weights were assigned to classes that mitigate risk, such as trees and grass. The weighted sum hazard score for each census tract was calculated using the following formula:

$$\text{Hazard Score} = \Sigma(\text{w\_i} \times \text{p\_i})$$

where w_i is the weight assigned to land cover class i, and p_i is the percentage of census tract area covered by class i. Results showed that the heat risk scores had a mean of 0.495 with a standard deviation of 0.162, as shown in Fig. 12. Also, the median of 0.510 is slightly above the mean, indicating that the distribution of scores is slightly left skewed. The skew suggests that more census tracts have a heat risk score above 0.5, but there are a few census tracts that have very low heat risk scores. This observation is reasonable because there are some census tracts in Boston that consist of only parks and have minimal heat risk, such as census tract 9817 for Boston Common and Public Garden.

### 7.4.1. Risk models

Results from the risk quantification indicated that census tract 705.02 in the South End of Boston is most susceptible to flood risk.

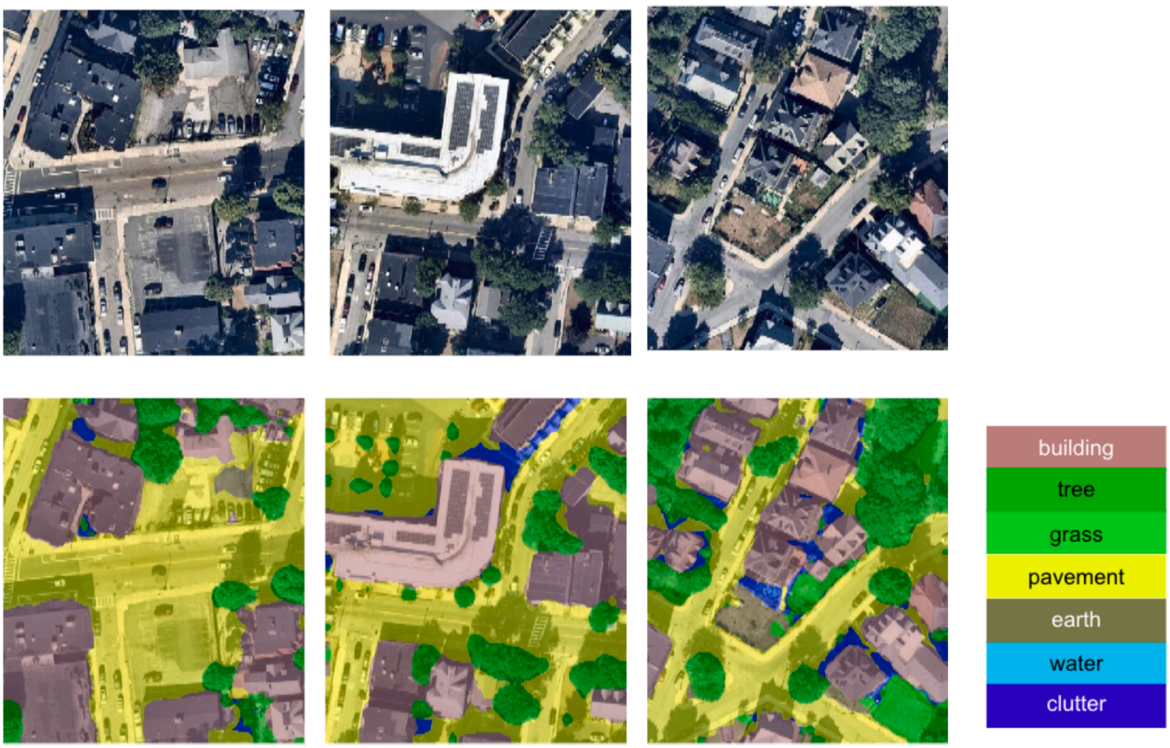


**Fig. 11.** Segmentations of Boston Aerial Images Produced by Each Fine-tuned Segmenter Model.

**Table 6**
Top 5 Boston Census Tracts with the highest percentage of each class.

| Top 5 Grass | | Top 5 Tree | | Top 5 Earth | | Top 5 Paved | | Top 5 Water | | Top 5 Building | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CT | % | CT | % | CT | % | CT | % | CT | % | CT | % |
| 9816.0 | 22.41 | **9807** | **54.32** | 9815.02 | 21.52 | 9813 | 43.80 | 9812.01 | 43.91 | **701.02** | **31.43** |
| 9809 | 22.22 | 9803 | 45.45 | 9816 | 16.37 | 907 | 42.08 | 9801.01 | 41.05 | 106 | 30.71 |
| 9811 | 22.09 | 1106.01 | 45.30 | 9813 | 13.50 | 612.02 | 41.79 | 606.04 | 40.95 | 701.04 | 29.71 |
| 9817 | 18.54 | 9810 | 43.37 | 9812.01 | 13.14 | 606.02 | 40.63 | 305 | 39.07 | 403 | 28.68 |
| 9810 | 17.32 | 1201.05 | 40.99 | 909.01 | 12.30 | 606.03 | 39.40 | 503 | 36.03 | 303.02 | 27.79 |

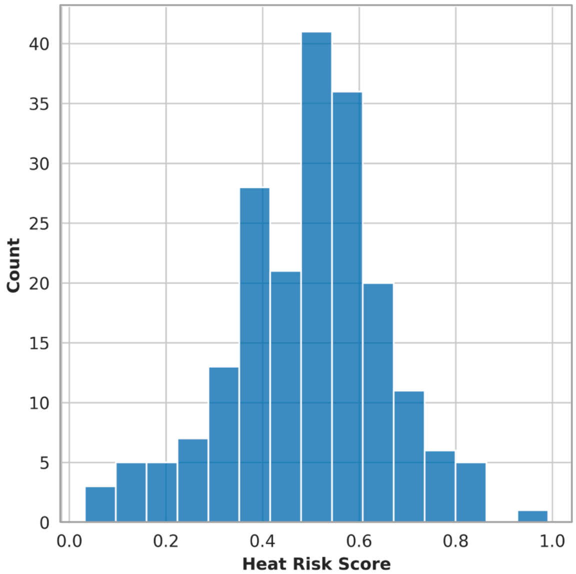


**Fig. 12.** Distribution of heat risk scores for 207 Boston census tracts using equal weights to combine the individual risk factors.

Based on Census Reporter, 42.5 % of people in this census tract are below the poverty line, more than twice the poverty rate of Boston. Additionally, based on the processed flood hazard data, census tract 705.02 has a low slope and is close to sea level. East Boston also has high flood risk census tracts like 507 near Mystic River, which has a large POC population and is characterized by higher percentages of building and sidewalk. According to Boston's Climate Report, both South End and East Boston have a significant population of low to no income residents, and these areas will experience the greatest increase in land exposed to stormwater flooding as sea levels rise and rainfall becomes more extreme. Census tract 9807 exhibits the lowest susceptibility to flood risk. This is the Stony Brook Reservation, which is a highly vegetated area. These areas consist of grass, trees, earth, and minimal buildings and limited paved surfaces. Moreover, these locations do house a large vulnerable population.

Tables 7 and 8 present the weights flood and heat risk scores calculated each census tract, respectively, along with the top five census tracts with the highest hazard scores. The identification of high-risk areas based on land cover composition can inform urban planning decisions and help prioritize targeted mitigation strategies, such as increasing green space in heat-prone neighborhoods or improving stormwater management in flood-prone areas. For heat risk, two census tracts (104.03 and 104.04) are located in Fenway–Kenmore, and another two census tracts (705.02 and 709.01) are in Shawmut within the South End. These areas are known for

**Table 7**
The 10 Boston census tracts with the highest flood risk scores based on a weighted sum of risk factors using equal weights.

| Census Tract | Approximate Location | Hazard Score | Vulnerability Score | Flood Risk Score |
|---|---|---|---|---|
| 705.02 | South End | 0.650 | 1.000 | 0.825 |
| 104.03 | Fenway | 0.694 | 0.952 | 0.823 |
| 709.01 | South End | 0.696 | 0.926 | 0.811 |
| 507 | East Boston | 0.631 | 0.947 | 0.789 |
| 502 | East Boston | 0.625 | 0.953 | 0.789 |
| 501.01 | East Boston | 0.646 | 0.897 | 0.772 |
| 104.04 | Northeastern University | 0.712 | 0.818 | 0.765 |
| 504 | East Boston | 0.744 | 0.761 | 0.752 |
| 610 | South Boston | 0.634 | 0.827 | 0.731 |
| 506 | East Boston | 0.736 | 0.719 | 0.727 |

**Table 8**
The 10 Boston census tracts with the highest heat risk scores based on a weighted sum of risk factors using equal weights.

| Census Tract | Approximate Location | Hazard Score | Vulnerability Score | Heat Risk Score |
|---|---|---|---|---|
| 104.03 | Fenway-Kenmore | 0.983 | 1 | 0.991 |
| 104.04 | Fenway-Kenmore | 0.929 | 0.783 | 0.856 |
| 705.02 | Shawmut | 0.694 | 0.991 | 0.842 |
| 702.02 | Chinatown | 0.8 | 0.874 | 0.837 |
| 701.02 | Downtown Boston | 0.885 | 0.772 | 0.828 |
| 709.01 | Shawmut | 0.76 | 0.882 | 0.821 |
| 607 | D Street/West Broadway | 0.783 | 0.795 | 0.789 |
| 502 | East Boston | 0.646 | 0.866 | 0.756 |
| 202 | Beacon Hill | 0.8 | 0.708 | 0.754 |
| 301 | North End | 0.781 | 0.711 | 0.746 |

their highly diverse communities based on US Census data. Additionally, there are two census tracts (702.02 and 701.02) from Downtown Boston and Chinatown that have many tall buildings and roads, resulting in higher land surface temperatures.

In addition to the weighted sum hazard scores, building proximity features are extracted for the heat risk model. Previous studies have shown that the proximity of buildings to green spaces, such as trees and grass, can provide localized cooling effects and mitigate heat stress. Using the predicted segmentation maps, the minimum Euclidean distance from the centroid of each building pixel to the nearest tree or grass pixel is calculated for each census tract. Table 9 present the top five census tracts with the shortest minimum distance from buildings to trees. These proximity features provide valuable information on the accessibility of green spaces for building occupants during extreme heat events.

While the segmentation-based approach offers a comprehensive assessment of land cover patterns and their contributions to heat and flood risk, it is important to acknowledge potential limitations. The accuracy of the Segmenter model, although high, may be affected by factors such as image resolution, shadow effects, and the presence of obscured or ambiguous land cover features. Future improvements could include the incorporation of additional land cover classes, such as water bodies and impervious surfaces, and the use of higher-resolution aerial or satellite imagery to capture finer-scale variations in urban form.

#### *7.4.2. Risk model evaluation*

There are few studies combining flood hazard features from aerial images with vulnerability features from socioeconomic data and building datasets to rank susceptible urban areas as granular as census tracts. As a result, it is difficult to quantitatively validate the accuracy of these rankings. However, the city of Boston has developed an interactive tool incorporating results from various flood models to predict the areas affected by different types of floods across time frames in the future. Fig. 13 illustrates flooding visualizations from the Boston Climate Ready Map for the top 10 high-risk census tracts identified by the Flood Risk model (shown in Table 7). The teal regions are the predicted long-term storm water flooding areas. The light blue color represents areas impacted by coastal flooding as a result of a 36 in. rise in sea-level in the long term (2070 s). For a majority of the top 10 high risk census tracts, the Boston Climate Ready models align with the findings.

In contrast, Fig. 14 shows visualizations from the Boston Climate Ready Map of areas impacted by flooding for the 10 low-risk census tracts. It is evident that Fig. 14 depicts a higher concentration of open-spaces in the low risk areas compared to the images in Fig. 13. Aside from the Belle Isle Marsh Reservation (census tract 9816) there are significantly smaller areas predicted to be affected by storm-water flooding, with virtually no areas expected to be impacted by coastal flooding. Census tract 9816 has a minimal population which contributes to its lower flood risk rank. Lastly, a smaller proportion of the low risk census tracts in the Boston Climate Ready Map is covered by vulnerable population layers. As a result, these areas are both low risk in terms of flood hazard and flood vulnerability. These observations suggest qualitative consistency between the outputs from the flood risk model and the predictions made by the Boston Climate Ready map.

Furthermore, heat risk model results are evaluated by comparing the census tracts with the highest heat risk to existing neighborhood-level heat risk estimations (Fig. 15) from the Risk Factor website (Risk Factor, 2023). This comparison provides evidence of the validity of the results. Also, the proximity of high heat risk regions to heat adaptation amenities, such as official cooling centers from the City of Boston, (Preparing for Heat, 2023) are computed to investigate where additional resources can be supplied. Since ground truth heat risk values are not readily available to our knowledge, the heat risk model is primarily validated by comparing

**Table 9**
The five Boston census tracts with the farthest minimum distance (in miles) from any building to a tree. The predicted segmentations from the fine-tuned segmentation model are used to determine the locations of buildings and trees, and only census tracts with buildings in their segmentation maps are considered.

| Census Tract | Approximate Location | Building-Tree Distance |
|---|---|---|
| 302 | North End | 0.0021 |
| 9816 | Belle Isle Marsh | 0.0020 |
| 5.05 | Brighton | 0.0015 |
| 301 | North End | 0.0012 |
| 9818 | Emerald Necklace | 0.0010 |

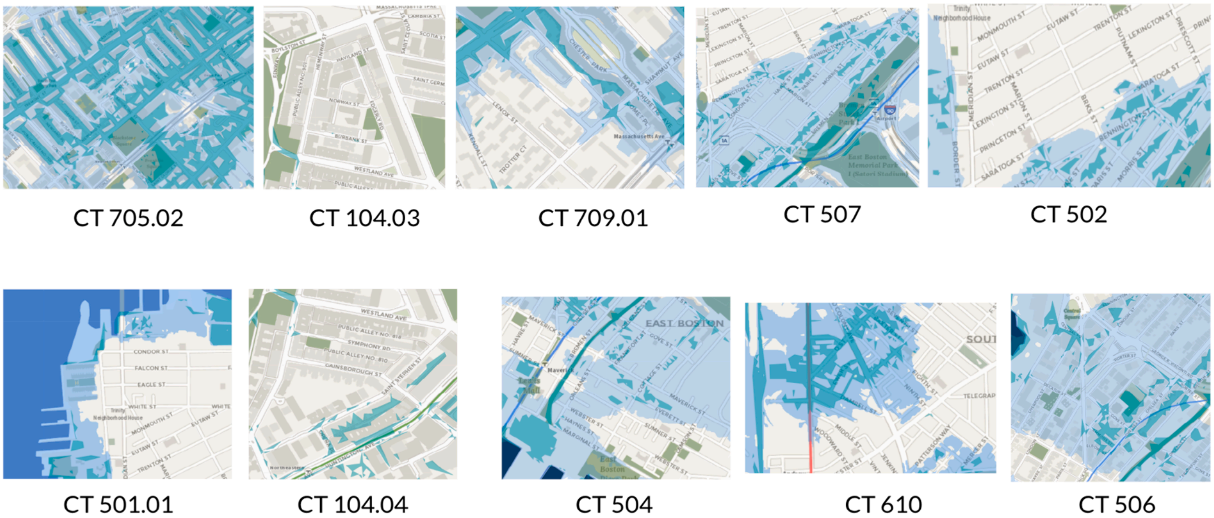


**Fig. 13.** Census tracts with high flood risk shown on Boston Climate Vulnerability Map.

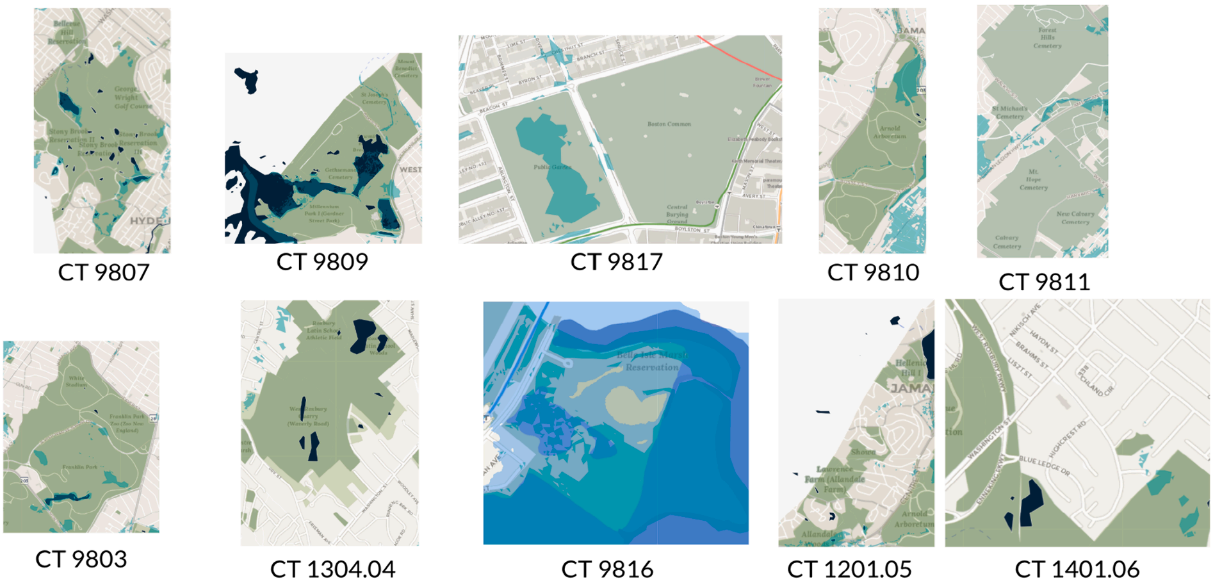


**Fig. 14.** Census tracts with low flood risk shown on Boston Climate Vulnerability Map.

its predictions for high heat risk census tracts with risk estimations from Risk Factor (Risk Factor, 2023). If the high heat risk areas identified by the Segmenter model correspond with the regions identified by Risk Factor as having major heat risk, then this alignment offers evidence supporting the validity of the model's predictions.

Table 10 presents the neighborhoods of Boston containing the highest risk census tracts from the heat risk model with a risk rating and the number of properties at risk sourced from the Risk Factor website (Risk Factor, 2023). Specifically, the high heat risk census tracts from the developed model are based on using both equal and different weights, so there are 12 top census tracts in total. The heat risk rating is at the neighborhood-level, and includes the number of properties with severe, major, and moderate heat risk from the website. All these neighborhoods are labeled as having a major heat risk, with 1–203 severe and 182–8700 major risk properties, including homes. Notably, the Fenway–Kenmore neighborhood contains the highest risk census tracts and the highest risk census tract from the developed model (104.03). It is also the neighborhood with the largest number of severe and major risk properties. Fig. 15 shows a screenshot from Risk Factor with heat risk details for the Fenway–Kenmore neighborhood. These observations provide compelling evidence supporting the validity of the proposed heat risk model. The regions that the model has identified as having higher heat risk are located in the neighborhoods that are described as having major heat risk by Risk Factor (Risk Factor, 2023).

The census tracts identified as having the highest heat risk by the model are also compared with the locations of official cooling centers in Boston. If a region exhibits high heat risk but lacks heat adaptation amenities, then it can be recommended that additional

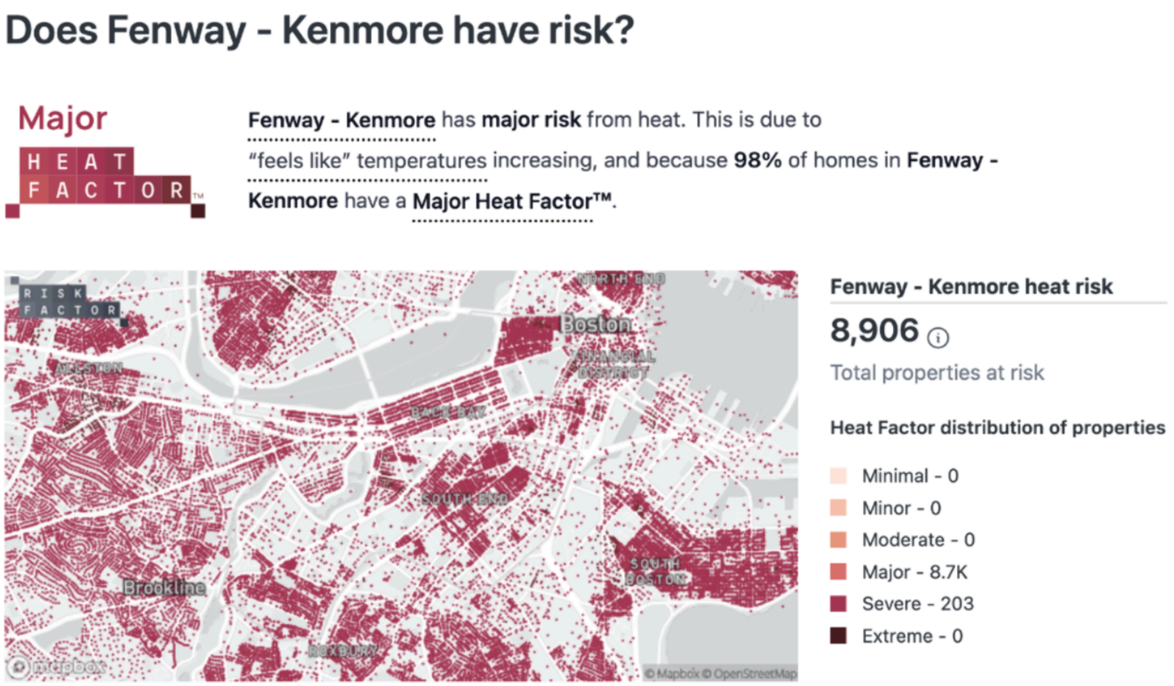


**Fig. 15.** Risk Factor (Risk Factor, 2023) classifies the Fenway–Kenmore neighborhood of Boston as having major risk from heat due to its increased "feels like" temperatures and high proportion of homes with major heat risk.

**Table 10**
Heat risk estimations from the Risk Factor website (Risk Factor, 2023) for the neighborhoods containing the top census tracts (included below neighborhood name) from the heat risk model.

| Neighborhood with Census Tracts | Heat Risk | Severe Risk Properties | Major Risk Properties | Moderate Risk Properties |
|---|---|---|---|---|
| 104.03, 104.04, 102.05 Fenway-Kenmore | Major | 203 | 8700 | 0 |
| 705.02, 709.01 – Shawmut | Major | 1 | 930 | 0 |
| 702.02 – Chinatown | Major | 19 | 182 | 0 |
| 701.02 – Downtown Boston | Major | 46 | 842 | 0 |
| 607 – D Street/West Broadway | Major | 103 | 2100 | 1 |
| 502 – East Boston | Major | 102 | 6600 | 0 |
| 202 – Beacon Hill | Major | 56 | 1300 | 0 |
| 301 – North End | Major | 29 | 1200 | 1 |
| 5.05 – Brighton | Major | 185 | 5400 | 0 |

resources be provided to these regions to urban planners and policymakers. Fig. 16 illustrates the regions of Boston that are farther than a 10-minute walk to one of the 21 designated cooling centers as of 2020 (Preparing for Heat, 2023). The map shows that the Fenway–Kenmore, South End, South Boston, and Chinatown neighborhoods of Boston face longer heat events, and some parts of these neighborhoods are also far from an official cooling center. More recently in 2022, there were 16 official cooling centers in various neighborhoods of Boston that could be activated when the mayor declares a heat emergency (Keeping Cool in the Heat, 2023). When comparing the top census tracts for heat risk from the model with the regions of Boston that are not within a 10-minute walk from a cooling center, it was found that all of the top 10 census tracts from the model with equal weights are in close proximity to a cooling center. However, census tracts 5.05 located in Brighton and 102.05 located in Fenway–Kenmore from the model with different weights are not near a cooling center despite having high heat risk scores. This finding, along with the fact that roughly only 31 % of the total population of Boston is located within 0.5 miles of a cooling center (Boston University School of Public Health, 2023), indicates a need for more resources to help residents in census tracts 5.05 and 102.05 stay cool during extreme heat.

### *7.5. Segmentation results for NYC*

Although the model was trained on aerial images from Boston, its performance was tested on images from New York City to assess its generalizability to other urban areas. Fig. 17 shows the qualitative performance of the fine-tuned model on Manhattan, Queens, and Brooklyn. Sub Fig. 17a and c shows Manhattan, which is more urbanized than Boston, Queens, and Brooklyn. Additionally, Manhattan consists of a higher density of taller buildings than in Boston, resulting in many areas in the Manhattan aerial images being occluded by

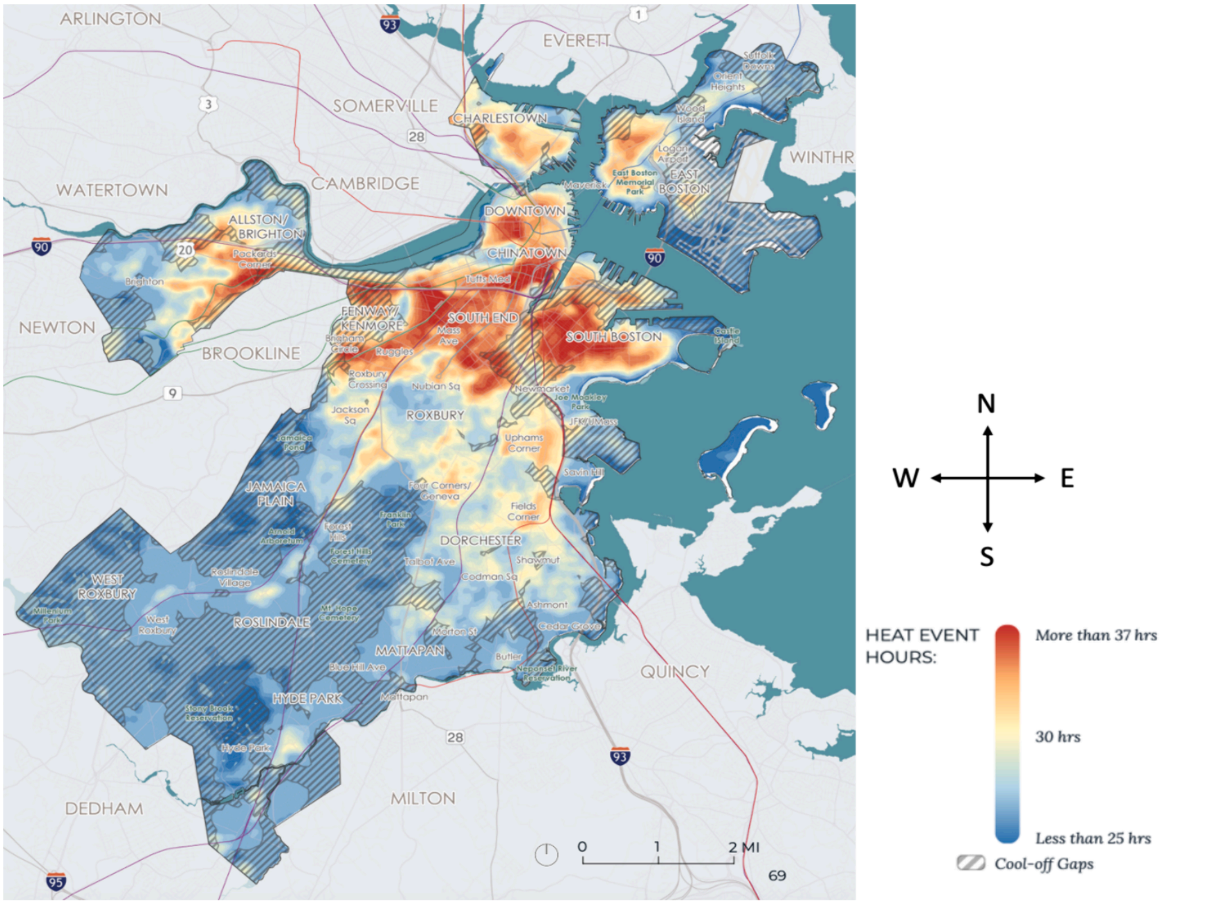


**Fig. 16.** Map of regions in Boston that require more than a 10-minute walk from an official cooling center (dashed area), layered on a map of heat event duration (red represents a duration of more than 37 h, and blue represents a duration of less than 25 h). Image taken from the City of Boston (Preparing for Heat, 2023), which used summer 2020 cooling centers in the visualization.

shadows. The rooftops of buildings in Manhattan are also more complex than those of buildings in Boston, Queens, and Brooklyn. Despite these differences, the Segmenter model fine-tuned on Boston performed qualitatively well on images from Manhattan. Subfigures 17a and 17c show that the model correctly segments heavily shadowed areas. Images 17b and 17d are from Queens, and 17e and 17f are from Brooklyn. In contrast to Manhattan, these areas are less urban and similar to many Boston areas, so the model performs well for these locations.

## 8. Discussion

The paper maps heat and flood risk at the granular census tract level, offering valuable insights for urban planners and policymakers. This detailed mapping helps identify specific regions within urban areas that need improved heat and flood adaptation amenities. To develop the proposed models, data on land cover, socioeconomic factors, building characteristics, and environmental conditions are collected in addition to aerial imagery. Results from the paper include a ranking of census tracts in Boston most vulnerable to extreme heat and flood events. From this ranking, city planners can target the top census tracts that will suffer from these events to build amenities for mitigating flood risk and reducing the effects of extreme heat. Additionally, the ranking of census tracts is consistent with results by other heat and flood models. For example, the top 5 census tracts most vulnerable to flood risk are all shown to experience coastal and stormwater flooding in the Boston Climate Vulnerability map. Additionally, these locations have a high percentage of vulnerable populations, such as people of color, low income households, and people with limited English proficiency. Similarly, the top five census tracts in Boston that are most vulnerable to heat risk correspond to neighborhoods with major heat risk according to Risk Factor, a free tool developed by the nonprofit First Street Foundation. The work presented in the paper demonstrates the effectiveness of using image segmentation techniques to gather land cover values of areas, and then integrate those in a larger flood and heat risk model. Additionally, the fine-tuned Segmenter model was found to excel in accurately segmenting earth, water, buildings, trees, grass, and pavement. The fine-tuned model also yields promising qualitative results for unseen aerial images of New York City.

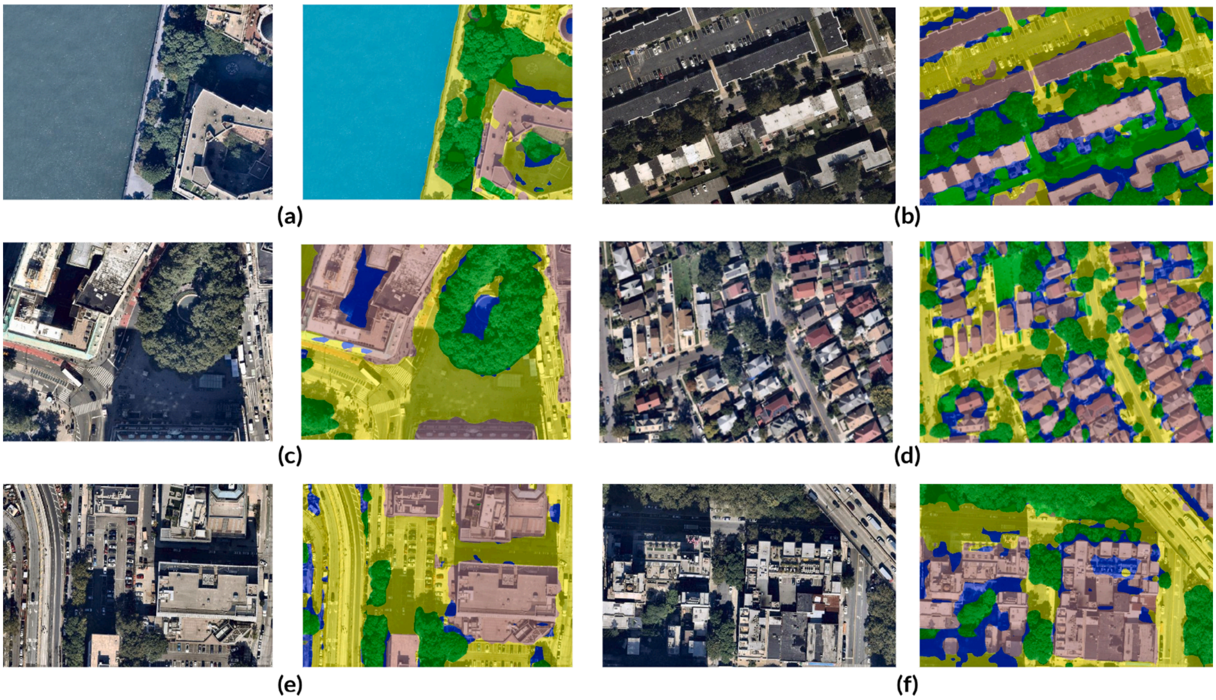

**Fig. 17.** Aerial images and the model's predicted segmentations for New York City.

The diverse array of features incorporated into the CLIM-SEG framework contributes significantly to the overall risk assessment produced. Land cover features extracted through semantic segmentation, particularly the proportion of impervious surfaces and vegetation, exert a substantial influence on both heat and flood risk predictions. Census tracts with higher percentages of buildings and pavement consistently demonstrate elevated risk scores, underscoring the urban heat island effect and increased runoff potential. Conversely, areas with greater tree canopy and grass coverage exhibit lower risk profiles, highlighting the mitigating effects of urban greenery. Socioeconomic variables, such as the proportion of elderly residents and low-income households, markedly amplify vulnerability scores, especially in densely populated urban cores. Topographic features, including elevation and slope, prove critical in delineating flood-prone areas, with low-lying, flat regions showing heightened susceptibility. The integration of building characteristics, particularly age and height, refines the risk assessment by capturing the varying resilience of urban infrastructure. Notably, the inclusion of air conditioning prevalence in the heat risk model significantly modulates vulnerability scores, revealing stark disparities in adaptive capacity across different neighborhoods. This multifaceted approach enables a more comprehensive understanding of urban climate risks, facilitating targeted intervention strategies that address both physical and social vulnerabilities.

The comprehensive set of features employed in the proposed CLIM-SEG framework, while extensive, raises important considerations regarding data redundancy and feature relevance. Analysis reveals that while most features contribute unique and valuable information to the risk assessment, some exhibit potential collinearity. For instance, the proportion of impervious surfaces derived from semantic segmentation correlates strongly with population density from census data, both indicative of urbanization intensity. Similarly, building age and the presence of air conditioning systems show a moderate inverse relationship, reflecting the modernization of urban infrastructure. However, these apparent redundancies often provide complementary rather than duplicate information. The impervious surface data, for example, offers higher spatial resolution and more frequent updates compared to census-derived urbanization metrics. Future iterations of the CLIM-SEG framework could benefit from more sophisticated feature selection techniques, such as principal component analysis or LASSO regression, to further optimize the balance between comprehensiveness and non-redundancy in the feature set.

### *8.1. Model limitation*

One limitation of the flood and heat risk models is the assumption that each feature contributes a weight of either $-1$ or $+1$ to the final flood and heat risk values. This positive or negative weighting is based on whether the feature would mitigate or increase risk as determined by prior literature. However, there are other works that utilize unequal weights based on the Analytical Hierarchy Process.

Another limitation of the work presented in this paper arises from the use of mutually exclusive labels in the semantic segmentation approach to obtain land cover features. In this methodology, each pixel in an image is assigned to a single class, implying that a particular area of land contains only one land cover type. However, this assumption may not always hold true in real-world scenarios, where overlapping features are common (Cheng et al., 2024). For instance, a tree canopy might cover a patch of pavement, or a building might cast a shadow over a grassy area. The use of mutually exclusive labels can introduce potential errors in the land cover classification process. When a pixel is assigned to a single class, the model fails to capture the multi-layered nature of the land cover (Qin and Liu, 2022). This can lead to an oversimplification of the urban landscape and may result in inaccuracies in the derived risk factors (Rao et al., 2023).

Moreover, the presence of shadows in aerial imagery can further complicate the classification process. Shadows can obscure the underlying land cover, making it challenging for the model to assign the correct label (Zhu and Woodcock, 2012). This issue is particularly relevant in urban areas with tall buildings, where shadows are prevalent. Misclassification of shadowed regions can introduce errors in the land cover percentages and subsequently affect the accuracy of the heat and flood risk assessment. To address these limitations, future work could explore the use of multi-label classification techniques that allow each pixel to be associated with multiple classes (Sumbul et al., 2020). This approach would enable the model to capture the co-occurrence of different land cover types within a single pixel, providing a more accurate representation of the urban landscape. Additionally, incorporating shadow detection and removal techniques could help mitigate the impact of shadows on the classification process (Luo et al., 2019).

### *8.2. Assumptions on land cover risk contributions*

The relationship between land cover and flood risk is complex and depends on various factors beyond just the type of land cover. For example, a well-designed urban area with adequate drainage systems and green infrastructure can effectively manage stormwater runoff, reducing the flood risk (Sörensen et al., 2016). Conversely, a poorly maintained grassland or park with compacted soil and limited infiltration capacity may contribute to increased runoff and flood risk (Ferreira et al., 2013).

Another limitation of the work presented in this paper lies in the assumptions made regarding the risk contribution of different land cover types. In the flood risk model, grass is assumed to have an overall mitigating effect on flood risk, while built-up areas have an overall increasing effect. However, this assumption may not always be accurate, as factors such as drainage capacity and the presence of green infrastructure can significantly alter the flood risk of built-up areas (Zhang et al., 2023). Moreover, the assumption that all built-up areas have an equal contribution to flood risk overlooks the heterogeneity within urban landscapes. Different types of urban development, such as high-density residential areas, commercial districts, and industrial zones, can have varying levels of impervious surface cover and drainage infrastructure (Li et al., 2023). These differences can significantly influence the flood risk contribution of each urban land cover type.

To address these limitations, future research could incorporate more detailed information on the characteristics of each land cover type, such as the presence and capacity of drainage systems, the type and density of urban development, and the implementation of green infrastructure. This information could be obtained through high-resolution remote sensing data, ground surveys, and collaboration with local authorities and urban planners (Leitão et al., 2016). Furthermore, the use of hydrological models that simulate surface runoff and infiltration processes could provide a more accurate assessment of the flood risk contribution of different land cover types (Salvadore and Bronders, 2015). These models can consider factors such as soil properties, topography, and rainfall characteristics, enabling a more comprehensive understanding of the relationship between land cover and flood risk. By incorporating these additional factors and using more advanced modeling techniques, future work can refine the assumptions made in the current flood risk model and provide a more nuanced assessment of the risk contribution of different land cover types.

### *8.3. Generalizability of the framework to other cities*

While the paper focuses on Boston as a case study, the proposed framework for heat and flood risk mapping has the potential to be generalized to other cities. The example of applying the fine-tuned segmentation model to aerial images of New York City demonstrates the transferability of the proposed approach. However, it is essential to discuss the city-specific data challenges and opportunities that may arise when extending the proposed framework to other urban areas.

One of the key advantages of the proposed framework is the use of publicly available datasets, such as census data, environmental data, and building information, which are often available for many cities across the United States, such as US Census Bureau, and Environmental Protection Agency (EPA), for example. This allows for the replication of the methodology in other urban contexts without the need for extensive data collection efforts. However, the availability and quality of city-specific data may vary, presenting challenges in generalizing the proposed framework. For instance, some cities may have limited or outdated building information, which could affect the accuracy of the vulnerability assessment. Similarly, the resolution and coverage of environmental data, such as land surface temperature or precipitation, may differ across cities, impacting the precision of the hazard assessment.

Another challenge lies in the transferability of the fine-tuned segmentation model to cities with distinct urban morphologies and land cover patterns. While the model performed well on aerial images of New York City, it may require further fine-tuning or adaptation when applied to cities with significantly different urban landscapes, such as those with extensive informal settlements or unique architectural styles, similar to those found in developing countries and the global south. Transfer learning techniques and domain adaptation methods could be employed to improve the model's performance in these cases (Shi et al., 2022).

Despite these challenges, the increasing availability of high-resolution aerial and satellite imagery, along with advancements in computer vision and ML techniques, presents opportunities for generalizing the framework to other cities. The growing adoption of open data policies and the development of urban data platforms can further facilitate the access to city-specific datasets. Moreover, the framework's modular nature allows for the incorporation of additional city-specific risk factors or the adaptation of the weighting scheme to reflect the local context and priorities. This flexibility enables urban planners and decision-makers to tailor the risk assessment to their specific needs and integrate local knowledge into the process.

## 9. Conclusion

The main contributions of this paper include three fine-tuned image segmentation models that achieve high accuracy and IoU scores in identifying land cover. To fine-tune these models, a dataset of 82 aerial images is curated with manually-created segmentation maps. Additionally, the land cover data from the top-performing image segmentation model is integrated with other relevant features, allowing to identify individual Boston census tracts most susceptible to the harms of extreme heat and flood events. The presented research closes gaps in heat and flood risk modeling by segmenting land cover features from high-resolution aerial images, along with hazard and vulnerability features from environmental, socioeconomic, and building datasets. Aerial images from web mapping platforms are utilized to minimize the potential temporal, geo-registration, and rasterization errors from prior work. The results show that the fine-tuned Segmenter model performs the best for segmenting land cover classes in the aerial image dataset. Additionally, the fine-tuned model generalizes well to other unseen urban areas such as New York City. In the heat and flood risk models, the land cover values are combined from applying the segmentation model to Boston images with other pertinent features affecting heat and flood risk. These features are combined with equal weighting, and the resulting Boston census tracts identified as highly vulnerable to heat and flood risk by the model align with results from other established models.

The proposed CLIM-SEG model serves as a critical component within a broader framework for urban heat and flood risk assessment. The segmentation model enables the extraction of fine-grained land cover data from aerial imagery, which is then integrated with other environmental, socioeconomic, and building data to perform comprehensive risk analyses. This integration of advanced computer vision techniques with risk assessment methodologies represents a novel approach in urban climate risk modeling. Although a significant portion of the paper discusses risk analysis methods, this reflects the end-to-end nature of the proposed framework, where the segmentation model plays a pivotal role in enhancing the accuracy and granularity of risk assessments. Results from the segmentation model outputs highlight the performance and applications of the segmentation model, demonstrating its crucial contribution to the overall risk assessment process. By presenting both the segmentation model and its application in risk analysis, the paper aims to provide a complete picture of the innovative approach to urban climate risk assessment. Future research can build on heat and flood risk models by refining the image segmentation model to account for overlapping land cover features in a given location, such as a tree canopy covering a street. Moreover, more relevant features can be also incorporated, such as the drainage capacity for the flood risk model, and improve the weighting scheme by analyzing the importance of features used in heat and flood risk models.

## CRediT authorship contribution statement

**Anushka Ray:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Katherine Xu:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Norhan Bayomi:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Formal analysis, Conceptualization. **John E. Fernandez:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation.

## 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.

## Data availability

Data will be made available on request.

## References


Adobe Inc., 2019. Adobe Photoshop. Retrieved February, 2023, from https://www.adobe.com/products/photoshop.html.

Aerial Semantic Segmentation Drone Dataset, 2021. Institute of Computer Graphics and Vision. Retrieved February, 2022, from http://dronedataset.icg.tugraz.at.

Aerts, J.P.M., Uhlemann-Elmer, S., Eilander, D., Ward, P.J., 2020. Comparison of estimates of global flood models for flood hazard and exposed gross domestic product: a China case study. Nat. Hazards Earth Syst. Sci. 20, 3245–3260. https://doi.org/10.5194/nhess-20-3245-2020.

Alonso, L., Renard, F., 2020. A comparative study of the physiological and socio-economic vulnerabilities to heat waves of the population of the metropolis of Lyon (France) in a climate change context. Int. J. Environ. Res. Public Health 17 (3), 1004. https://doi.org/10.3390/ijerph17031004.

Amatulli, G., Domisch, S., Tuanmu, M., Parmentier, B., Ranipeta, A., Malczyk, J.T., Jetz, W., 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci. Data 5. https://www.nature.com/articles/sdata201840.

Arrighi, C., Pregnolato, M., Dawson, R.J., Castelli, F., 2019. Preparedness against mobility disruption by floods. Sci. Total Environ. 654, 1010–1022.

Azad, A.K., Hossain, K.M., Nasreen, M., 2013. Flood-induced vulnerabilities and problems encountered by women in northern Bangladesh. Int. J. Disaster Risk Sci. 4, 190–199. https://doi.org/10.1007/s13753-013-0020-z.

Bayomi, N.M., Fernandez, J.E., 2021. Heat vulnerability and risk analytics for the built environment. Massachusetts Institute of Technology. https://hdl.handle.net/1721.1/14006.

Boston Planning & Development Agency, 2020. Boston by the Numbers 2020.

Boston University School of Public Health, 2023. Where to beat the heat in Boston. Available at: https://arcg.is/1j58m.

Buscail, C., Upegui, E., Viel, J.-F., 2012. Mapping heatwave health risk at the community level for public health action. Int. J. Health Geogr. 11, 38. https://doi.org/10.1186/1476-072X-11-38.

Census Tracts in Boston, 2023. Analyze Boston. Retrieved June 21, 2023, from: https://data.boston.gov/dataset/2020-census-tracts-in-boston.

Chakraborty, T., Lee, X., 2019. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. Int. J. Appl. Earth Observ. Geoinf. 74, 269–280.

Chapman, S., Watson, J.E.M., Salazar, A., Thatcher, M., McAlpine, C.A., 2017. The impact of urbanization and climate change on urban temperatures: a systematic review. Landsc. Ecol. 32 (6), 1921–1935. https://doi.org/10.1007/s10980-017-0561-4.

Cheng, X., Li, B., Deng, Y., Tang, J., Shi, Y., Zhao, J., 2024. MMDL-Net: Multi-band multi-label remote sensing image classification model. Appl. Sci. 14 (6), 2226. https://doi.org/10.3390/app14062226.

Cho, S., Chang, H., 2017. Recent research approaches to urban flood vulnerability, 2006–2016. Nat. Hazards 88 (1), 633–649.

Chuang, W.-C., Gober, P., 2015. Predicting Hospitalization for Heat-Related Illness at the Census-Tract Level: Accuracy of a Generic Heat Vulnerability Index in Phoenix, Arizona (USA). Environ. Health Perspect. 123 (6), 606–612. https://doi.org/10.1289/ehp.1307868.

Clarke, J., 1972. Some effects of the urban structure on heat mortality. Environ. Res. 5 (1), 93–104. https://doi.org/10.1016/0013-9351(72)90023-0.

Climate Ready Boston Social Vulnerability, 2017. Analyze Boston. Available at: https://data.boston.gov/dataset/climate-ready-boston-social-vulnerability.

Climate Resilience Initiatives, 2022. Boston.gov. Retrieved February, 2022, from https://www.boston.gov/sites/default/files/imce-uploads/2017-01/crb_-_focus_area_ri.pdf.

Climate Vulnerability Assessment, 2017. Boston.gov. Retrieved October 15, 2023, from https://www.boston.gov/sites/default/files/imce-uploads/2017-01/crb_-_focus_area_va.pdf.

Climate Vulnerability Assessment, 2022. Boston.gov. Retrieved February, 2022, from https://www.boston.gov/sites/default/files/imce-uploads/2017-01/crb_-_focus_area_va.pdf.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.

Crichton, D., 1999. The Risk Triangle. In: Natural Disaster Management. Tudor Rose, pp. 102–103.

Cutter, S.L., Boruff, B.J., Shirley, W.L., 2003. Social vulnerability to environmental hazards. Soc. Sci. Q. 84, 242–261. https://doi.org/10.1111/1540-6237.8402002.

Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010) (No. 2011-1073). US Geological Survey.

Darabi, H., Haghighi, A.T., Mohamadi, M.A., Rashidpour, M., Ziegler, A., Hekmatzadeh, A.A., Kløve, B., 2020. Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. Hydrol. Res. 51 (1), 127–142. https://doi.org/10.2166/nh.2019.090.

Davis, M., 2022. Flood Risk Models Vary Widely: Here's What You Need to Know (and How to Mitigate Threats). Available at: https://www.valuepenguin.com/flood-risk-study.

de Brito, M.M., Evers, M., 2016. Multi-criteria decision-making for flood risk management: a survey of the current state of the art. Nat. Hazards Earth Syst. Sci. 16 (4), 1019–1033.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv, abs/2010.11929. https://arxiv.org/abs/2010.11929.

Earth Resources Observation and Science Center, 2018. National Land Cover Database. U.S. Geological Survey. https://www.usgs.gov/centers/eros/science/national-land-cover-database.

Fatemi, M.N., Okyere, S.A., Diko, S.K., Kita, M., Shimoda, M., Matsubara, S., 2020. Physical vulnerability and local responses to flood damage in peri-urban areas of Dhaka, Bangladesh. Sustainability 12 (10), 3957. https://doi.org/10.3390/su12103957.

Feeling hot, hot, hot: How the summer heat can get you down, 2016. Climate Impacts Group. Available at: https://cig.uw.edu/news-and-events/feeling-hot-hot-hot-how-the-summer-heat-can-get-you-down.

Ferreira, S., Hamilton, K., Vincent, J.R., 2013. Does development reduce fatalities from natural disasters? New evidence for floods. Environ. Dev. Econ. 18 (6), 649–679. https://doi.org/10.1017/S1355770X13000387.

Flanagan, B.E., Gregory, E.W., Hallisey, E.J., Heitgerd, J.L., Lewis, B., 2011. A social vulnerability index for disaster management. J. Homeland Security Emergency Manage. 8 (1), 3. https://doi.org/10.2202/1547-7355.1792.

Foster, K.G., Ellis, F.P., Doré, C., Exton-Smith, A.N., Weiner, J.S., 1976. Sweat responses in the aged. Age Ageing 5 (2), 91–101. https://doi.org/10.1093/ageing/5.2.91.

Haines, A., Kovats, R.S., Campbell-Lendrum, D., Corvalan, C., 2006. Climate change and human health: Impacts, vulnerability and public health. Public Health 120 (7), 585–596. https://pubmed.ncbi.nlm.nih.gov/16542689/.

Harlan, S.L., Ruddell, D.M., 2011. Climate change and health in cities: impacts of heat and air pollution and potential co-benefits from mitigation and adaptation. Curr. Opin. Environ. Sustain. 3 (3), 126–134. https://www.sciencedirect.com/science/article/abs/pii/S1877343511000029.

Hinojosa, J., Meléndez, E., 2017. The housing crisis in Puerto Rico and the impact of Hurricane Maria. Center for Puerto Rican Studies.

Hitouri, S., Mohajane, M., Lahsaini, M., Ali, S.A., Setargie, T.A., Tripathi, G., D'Antonio, P., Singh, S.K., Varasano, A., 2024. Flood susceptibility mapping using SAR data and machine learning algorithms in a small watershed in Northwestern Morocco. Remote Sens. (Basel) 16 (5), 858.

Ho, H.C., Knudby, A., Huang, W., 2015. A spatial framework to map heat health risks at multiple scales. Int. J. Environ. Res. Public Health 12 (12), 16110–16123. https://pubmed.ncbi.nlm.nih.gov/26694445/.

Holand, I., Lujala, P., Rød, J.K., 2011. Social vulnerability assessment for Norway: A quantitative approach. Norsk Geografisk Tidsskrift – Norwegian J. Geogr. 65 (1), 1–17. https://doi.org/10.1080/00291951.2010.550167.

Hu, K., Yang, X., Zhong, J., Fei, F., Qi, J., 2017. Spatially explicit mapping of heat health risk utilizing environmental and socioeconomic data. Environ. Sci. Tech. 51 (3), 1498–1507. https://doi.org/10.1021/acs.est.6b04355.

Jenerette, G.D., Harlan, S.L., Stefanov, W.L., Martin, C.A., 2011. Ecosystem services and urban heat riskscape moderation: water, green spaces, and social inequality in Phoenix, USA. Ecol. Appl. 21 (7), 2637–2651.

Johnson, N., Treible, W., Crispell, D., 2022. Open Sentinel Map: A Large-Scale Land Use Dataset using OpenStreetMap and Sentinel-2 Imagery. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1333–1341.

Johnson, R., 2022. Disaster Disparities: Natural Hazards, Climate Change Threaten Underserved Communities. USNews.com. Available at: https://www.usnews.com/news/health-news/articles/2022-06-22/disaster-disparities-natural-hazards-climate-change-threaten-underserved-communities.

Keeping cool in the heat, 2023. Boston.gov. Available at: https://www.boston.gov/departments/emergency-management/keeping-cool-heat.

Kenny, G., Yardley, J., Brown, C., Sigal, R., Jay, O., 2010. Heat stress in older individuals and patients with common chronic diseases. Can. Med. Assoc. J. 182 (10), 1053–1060. https://doi.org/10.1503/cmaj.081050.

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.R., Behar, J.V., Hern, S.C., Engelmann, W.H., 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J. Expo. Anal. Environ. Epidemiol. 11, 231–252. https://www.nature.com/articles/7500165.

Klinenberg, E., 2015. Heat Wave: A Social Autopsy of Disaster in Chicago, second ed. University of Chicago Press.

Koks, E.E., Jongman, B., Husby, T.G., Botzen, W.J., 2015. Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. Environ. Sci. Policy 47, 42–52.

Konrad, C.P., 2016. Effects of Urban Development on Floods. Effects of Urban Development on Floods. Retrieved October 15, 2023, from https://pubs.usgs.gov/fs/fs07603/.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15 (3), 259–263.

Landreau, A., Juhola, S., Jurgilevich, A., Rasanen, A., 2021. Combining socio-economic and climate projections to assess heat risk. Clim. Change 167, 1–20. https://doi.org/10.1007/s10584-021-03148-3.

Larsen, L., 2015. Urban climate and adaptation strategies. Front. Ecol. Environ. 13 (9), 486–492. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/150103.

Learn About Heat Islands | US EPA. (2023). Environmental Protection Agency. Retrieved 2022, from https://www.epa.gov/heatislands/learn-about-heat-islands.

Leitão, J.P., Moy de Vitry, M., Scheidegger, A., Rieckermann, J., 2016. Assessing the quality of digital elevation models obtained from mini unmanned aerial vehicles for overland flow modelling in urban areas. Hydrol. Earth Syst. Sci. 20, 1637–1653. https://doi.org/10.5194/hess-20-1637-2016.

Li, C., Sun, N., Lu, Y., Guo, B., Wang, Y., Sun, X., Yao, Y., 2023. Review on urban flood risk assessment. Sustainability 15 (1), 765. https://doi.org/10.3390/su15010765.

Li, F., Yigitcanlar, T., Nepal, M., Thanh, K.N., Dur, F., 2022. Understanding urban heat vulnerability assessment methods: A PRISMA review. Energies 15 (19), 6998. https://doi.org/10.3390/en15196998.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Luber, G., McGeehin, M., 2008. Climate change and extreme heat events. Am. J. Prev. Med. 35 (5), 429–435.

Luetz, J.M., Merson, J., 2019. Climate change and human migration as adaptation: conceptual and practical challenges and opportunities. In: Leal Filho, W., Azul, AM, Brandli, L., Özuyar, PG, Wall, T. (Eds.), Climate Action. Encyclopedia of the UN Sustainable Development Goals, pp. 1–13.

Luo, Y., Guan, K., Peng, J., 2019. STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sens. Environ. 214, 87–99.

Mahmoud, S.H., Gan, T.Y., 2018. Urbanization and climate change implications in flood risk management: Developing an efficient decision support system for flood susceptibility mapping. Sci. Total Environ. 636, 152–167.

Manfreda, S., Samela, C., Gioia, A., Consoli, G.G., Iacobellis, V., Giuzio, L., Sole, A., 2015. Flood-prone areas assessment using linear binary classifiers based on flood maps obtained from 1D and 2D hydraulic models. Nat. Hazards 79 (2), 735–754.

MapBrowser, 2023. Nearmap. Retrieved November 29, 2023, from https://apps.nearmap.com/maps.

Maps, 2023. Apple. Available at https://www.apple.com/maps.

Maragno, D., Dalla Fontana, M., Musco, F., 2020. Mapping heat stress vulnerability and risk assessment at the neighborhood scale to drive urban adaptation planning. Sustainability. 12 (3), 1056. https://doi.org/10.3390/su12031056.

Maskrey, S., Mount, N., Thorne, C., 2022. Doing flood risk modelling differently: Evaluating the potential for participatory techniques to broaden flood risk management decision-making. J. Flood Risk Manage. 15 (1), e12757.

Meehl, G.A., Tebaldi, C., 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. Science. Science 305 (5686), 994–997. https://doi.org/10.1126/science.1098704.

Mehnert, P., Bröde, P., Griefahn, B., 2002. Gender-related difference in sweat loss and its impact on exposure limits to heat stress. Int. J. Ind. Ergon. 29 (6), 343–351. https://doi.org/10.1016/S0169-8141(02)00073-2.

Mimura, N., 2013. Sea-level rise caused by climate change and its implications for society. Proc. Jpn. Acad. 89 (7), 281–301. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758961/.

Mitchell, B., Chakraborty, J., 2015. Landscapes of thermal inequity: disproportionate exposure to urban heat in the three largest US cities. Environ. Res. Lett. 10 (11). https://doi.org/10.1088/1748-9326/10/11/115005.

Morabito, M., Crisci, A., Messeri, A., Capecchi, V., Modesti, P.A., Gensini, G.F., Orlandini, S., 2014. Environmental temperature and thermal indices: what is the most effective predictor of heat-related mortality in different geographical contexts? Sci. World J. 2014. https://pubmed.ncbi.nlm.nih.gov/24523657/.

Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. Water 10 (11), 1536.

Muis, S., Güneralp, B., Jongman, B., Aerts, J.C., Ward, P.J., 2015. Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. Sci. Total Environ. 538, 445–457.

Naughton, M., Henderson, A., Mirabelli, M., Kaiser, R., Wilhelm, J., Kieszak, S., Rubin, C., McGeehin, M., 2002. Heat-related mortality during a 1999 heat wave in Chicago. Am. J. Prev. Med. 22 (4), 221–227. https://doi.org/10.1016/s0749-3797(02)00421-x.

Negese, A., Worku, D., Shitaye, A., Getnet, H., 2022. Potential flood-prone area identification and mapping using GIS-based multi-criteria decision-making and analytical hierarchy process in Dega Damot district, northwestern Ethiopia. Appl Water Sci 12, 255. https://doi.org/10.1007/s13201-022-01772-7.

Nesbitt, L., Meitner, M.J., Girling, C., Sheppard, S.R., Lu, Y., 2019. Who has access to urban vegetation? A spatial analysis of distributional green equity in 10 US cities. Landsc. Urban Plan. 181, 51–79.

New York City Panel on Climate Change, 2019. New York City Panel on Climate Change 2019 Report. Ann. N. Y. Acad. Sci. 1439 (1), 1–112.

Ngo, P.T.T., Hoang, N.D., Pradhan, B., Nguyen, Q.K., Tran, X.T., Nguyen, Q.M., Tien Bui, D., 2020. A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using sentinel-1 SAR imagery and geospatial data. Sensors 20 (18), 5190.

NYC Mayor's Office of Resiliency, 2020. Cool Neighborhoods NYC: A Comprehensive Approach to Keep Communities Safe in Extreme Heat.

NYC Planning, 2020. New York City Borough Profile: Manhattan.

Ostro, B., Rauch, S., Green, R.S., Malig, B.J., Basu, R., 2010. The effects of temperature and use of air conditioning on hospitalizations. Am. J. Epidemiol. 172 (9), 1053–1061.

Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., Lee, S., Bui, D.T., 2020. Deep learning neural networks for spatially explicit prediction of flash flood probability. Geosci. Front. 12 (3). https://doi.org/10.1016/j.gsf.2020.09.007.

Preparing for heat, 2023. Boston.gov. Available at: https://www.boston.gov/departments/environment/preparing-heat.

Qin, R., Liu, T., 2022. A review of landcover classification with very-high resolution remotely sensed optical images analysis unit, model scalability and transferability. Remote Sens. 14 (3), 646. https://doi.org/10.3390/rs14030646.

Ramesh, V., Iqbal, S.S., 2020. Urban flood susceptibility zonation mapping using evidential belief function, frequency ratio and fuzzy gamma operator models in GIS: a case study of Greater Mumbai, Maharashtra, India. Geocarto Int. 37, 581–606.

Rao, P., Tassinari, P., Torreggiani, D., 2023. Exploring the land-use urban heat island nexus under climate change conditions using machine learning approach: A spatio-temporal analysis of remotely sensed data. Heliyon 209 (8). https://doi.org/10.1016/j.heliyon.2023.e18423.

Reid, C.E., O'Neill, M.S., Gronlund, C.J., Brines, S.J., Brown, D.G., Diez-Roux, A.V., Schwartz, J., 2009. Mapping community determinants of heat vulnerability. Environ. Health Perspect. 117 (11), 1730–1736. https://doi.org/10.1289/ehp.0900683.

Reid, C.E., Mann, J.K., Alfasso, R., English, P.B., King, G.C., Lincoln, R.A., Margolis, H.G., Rubado, D.J., Sabato, J.E., West, N.L., Woods, B., Navarro, K.M., Balmes, J. R., 2012. Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study. Environ. Health Perspect. 120 (5), 715–720. https://doi.org/10.1289/ehp.1103766.

Reuther, A., Kepner, J., Byun, C., Samsi, S., Arcand, W., Bestor, D., Bergeron, B., Gadepally, V., Houle, M., Hubbell, M., Jones, M., Klein, A., Milechin, L., Mullen, J., Prout, A., Rosa, A., Yee, C., Michaleas, P., 2018. Interactive supercomputing on 40,000 cores for machine learning and data analysis. In: IEEE High Performance Extreme Computing Conference (HPEC), pp. 1–6. https://arxiv.org/abs/1807.07814.

Risk Factor, 2023. Find Your Property's Climate Risks. Risk Factor. Available at: https://riskfactor.com/.

Robertson, N.M., Chan, T., 2009. Aerial image segmentation for flood risk analysis. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 597–600.

Roopnarine, C., Ramlal, B., Roopnarine, R., 2022. A comparative analysis of weighting methods in geospatial flood risk assessment: A Trinidad case study. Land. https://www.mdpi.com/2073-445X/11/10/1649.

Rosenthal, J.K., Kinney, P.L., Metzger, K.B., 2014. Intra-urban vulnerability to heat-related mortality in New York City, 1997–2006. Health Place 30, 45–60.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1 (5), 206–215.

Saaty, T.L., 1987. The analytic hierarchy process—what it is and how it is used. Math. Modell. 9 (3–5), 161–176.

Saha, S., 2018. A Guide to Convolutional Neural Networks: the ELI5 way. Saturn Cloud. Available at: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

Salvadore, E., Bronders, J., 2015. Hydrological modelling of urbanized catchments: A review and future directions. J. Hydrol. 529 (Part 1), 62–81, ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2015.06.028.

Santamouris, M., 2020. Recent progress on urban overheating and heat island research. Integrated assessment of the energy, environmental, vulnerability and health impact. Synergies with the global climate change. Energ. Build. 207.
Shi, Y., Ying, X., Zha, H., 2022. Unsupervised domain adaptation for semantic segmentation of urban street scenes reflected by convex mirrors. IEEE Trans. Intell. Transp. Syst. 23 (12), 24276–24289. https://doi.org/10.1109/TITS.2022.3208334.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.
Siddiqui, R., 2022. What Are Vision Transformers and How Are They Important for General Purpose Learning? Towards Data Science. Available at: https://towardsdatascience.com/what-are-vision-transformers-and-how-are-they-important-for-general-purpose-learning-edd008545e9e.
Singh, S., Kapoor, N., 2014. Health implications of electromagnetic fields, mechanisms of action, and research needs. Adv. Biol. 24. https://doi.org/10.1155/2014/198609.
Skarha, J., Dominick, A., Spangler, K., 2022. Provision of air conditioning and heat-related mortality in Texas Prisons. JAMA Netw. Open 5 (11). https://doi.org/10.1001/jamanetworkopen.2022.39849.
Sörensen, J., Persson, A., Sternudd, C., Aspegren, H., Nilsson, J., Nordström, J., Jönsson, K., Mottaghi, M., Becker, P., Pilesjö, P., Larsson, R., Berndtsson, R., Mobini, S., 2016. Re-thinking urban flood management–time for a regime shift. Water 8 (8), 332, 1–15.
Strudel, R., Garcia, R., Laptev, I., Schmid, C., 2021. Segmenter: Transformer for semantic segmentation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 7242–7252.
Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2020. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. Proceedings of the IEEE/CVF International Conference on Computer Vision 2266–2275.
Swain, K.C., Singha, C., Nayak, L., 2020. Flood susceptibility mapping through the GIS-AHP technique using the cloud. ISPRS Int. J. Geo Inf. 9 (12), 720.
Tarboton, D.G., 2003. Terrain analysis using digital elevation models in hydrology. In: 23rd ESRI International USERS Conference, San Diego, California, vol. 14.
Tebaldi, C., Hayhoe, K., Arblaster, J.M., Meehl, G.A., 2006. Going to the extremes. Clim. Change 79 (3–4), 185–211. https://opensky.ucar.edu/islandora/object/articles:7249.
Tomlinson, C.J., Chapman, L., Thornes, J.E., Baker, C.J., 2011. Including the urban heat island in spatial heat health risk assessment strategies: a case study for Birmingham, UK. Int. J. Health Geogr. 10, 42. https://doi.org/10.1186/1476-072X-10-42.
Turek-Hankins, L.L., Hino, M., Mach, K.J., 2020. Risk screening methods for extreme heat: Implications for equity-oriented adaptation. PLoS One 15 (11).
U.S. Census Bureau, 2021. QuickFacts: Boston city, Massachusetts.
U.S. Census Bureau, 2021. QuickFacts: New York City, New York.
U.S. Geological Survey, 2020. The National Map – Elevation.
Uejio, C.K., Wilhelmi, O.V., Golden, J.S., Mills, D.M., Gulino, S.P., Samenow, J.P., 2011. Intra-urban societal vulnerability to extreme heat: the role of heat exposure and the built environment, socioeconomics, and neighborhood stability. Health Place 17 (2), 498–507.
Ulku, I., Akagunduz, E., 2022. A survey on deep learning-based architectures for semantic segmentation on 2D images. Appl. Artif. Intell. 36 (1).
Urban Heat Islands, 2021. Climate Central. Retrieved November 29, 2023, from https://www.climatecentral.org/climate-matters/urban-heat-islands.
Urban Heat Islands, 2023. National Integrated Heat Health Information System. Available at: https://www.heat.gov/pages/urban-heat-islands.
Vescovi, L., Rebetez, M., Rong, F., 2005. Assessing public health risk due to extremely high temperature events: climate and social parameters. Climate Res. 30 (1), 71–78.
Vojtek, M., Vojteková, J., 2019. Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. Water. https://www.mdpi.com/2073-4441/11/2/364.
Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. J. Hydrol. 527, 1130–1141.
Weill, J.A., 2023. Flood Risk Mapping and the Distributional Impacts of Climate Information,Finance and Economics Discussion Series 2023–066. Board of Governors of the Federal Reserve System, Washington https://doi.org/10.17016/FEDS.2023.066.
Williams, A.A., Baniassadi, A., Igaza González, P., Buonocore, J.J., Cedeño-Laurent, J.G., Samuelson, H.W., 2020. Health and climate benefits of heat adaptation strategies in single-family residential buildings. Front. Sustain. Cities.
Wolf, T., McGregor, G., 2013. The development of a heat wave vulnerability index for London, United Kingdom. Weather Clim. Extremes 1, 59–68. https://doi.org/10.1016/j.wace.2013.07.004.
Wolf, T., Mehnert, W.-C., McGregor, G., 2015. On the science-policy bridge: Do spatial heat vulnerability assessment studies influence policy? Int. J. Environ. Res. Public Health 12 (10), 13321–13349. https://doi.org/10.3390/ijerph121013321.
Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. Adv. Neural Inf. Proces. Syst. 34, 12077–12090.
Zhang, L., Xu, D., Arnab, A., Torr, P.H., 2020. Dynamic graph message passing networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3726–3735.
Zhang, W., Hu, B., Liu, Y., Zhang, X., Li, Z., 2023. Urban flood risk assessment through the integration of natural and human resilience based on machine learning models. Remote Sens. (Basel) 15 (14), 3678. https://doi.org/10.3390/rs15143678.
Zhang, K., Li, Y., Schwartz, J.D., O'Neill, M.S., 2014. What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. Environ. Res. 132, 350–359.
Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890.
Zhao, L., Oleson, K., Bou-Zeid, E., Krayenhoff, E.S., Bray, A., Zhu, Q., Oppenheimer, M., 2021. Global multi-model projections of local urban climates. Nat. Clim. Chang. 11 (2), 152–157.
Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2016. Semantic Understanding of Scenes through the ADE20K Dataset. Int. J. Comput. Vis. 127, 302–321. https://arxiv.org/abs/1608.05442.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imager. Remote Sens. Environ. 118, 83–94. https://doi.org/10.1016/j.rse.2011.10.028.
Zhu, Z., Zhang, Y., 2022. Flood disaster risk assessment based on random forest algorithm. Neural Comput. Appl. 34, 3443–3455. https://doi.org/10.1007/s00521-021-05757-6.
Ziter, C.D., Pedersen, E.J., Kucharik, C.J., Turner, M.G., 2019. Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. Proc. Natl. Acad. Sci. 116 (15), 7575–7580.

## *Further reading*

Flash Flooding Definition, 2022. National Weather Service. Retrieved February, 2022, from https://www.weather.gov/phi/FlashFloodingDefinition.
Jia, P., Cao, X., Yang, H., Dai, S., He, P., Huang, G., Wang, Y., 2021. Green space access in the neighbourhood and childhood obesity. Obes. Rev. 22, e13100.