WILEY

Author Query Form

Journal: TGIS

Article: 13236

Dear Author,

During the copyediting of your manuscript, the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

AUTHOR: Please note that missing content in references have been updated where we have been able to match the missing elements without ambiguity against a standard citation database, to meet the reference style requirements of the journal. It is your responsibility to check and ensure that all listed references are complete and accurate.

Query reference	Query	Remarks
1	AUTHOR: Please verify that the linked ORCID identifier is correct for the given author.	
2	AUTHOR: Please check that authors and their affiliations are correct.	
3	AUTHOR: Please check and confirm that all heading levels have been set correctly.	
4	AUTHOR: Please check and confirm whether "Tables 2 and 12" citations have been set correctly.	
5	AUTHOR: Please note that "Tables 8 and 7" have been renumbered to "Tables 7 and 8" to maintain sequential order. Kindly check and confirm.	
6	AUTHOR: Please note that "Figures 10, 11, 9" have been renumbered to "Figures 9, 10, 11" to maintain sequential order, kindly check and confirm.	
7	AUTHOR: Please check the placement of Appendix section and amend if necessary.	
8	AUTHOR: Please note that Tables 10 and 11 captions are identical. Kindly check and amend if necessary.	
9	AUTHOR: As per style, reference citations have been changed from "Numbered" to "Name-year" style. And also, the reference list has been reordered accordingly. Please confirm.	
10	AUTHOR: Please provide the "volume number" for reference Eckstein, 2001.	
11	AUTHOR: Please provide the "page range" for reference Shakur et al., 2010.	
12	AUTHOR: Figure 4 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf	

13	**AUTHOR: Figure 5 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
14	**AUTHOR: Figure 6 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
15	**AUTHOR: Figure 7 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
16	**AUTHOR: Figure 8 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
17	**AUTHOR: Figure 9 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
18	**AUTHOR: Figure 11 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	
19	**AUTHOR: Figure 13 is of poor quality. Please check required artwork specifications at https://authorservices.wiley.com/asset/photos/electronic_artwork_guidelines.pdf**	

Transactions in GIS

WILEY

Transactions in GIS

Straive | Journal Name: TGIS | Manuscript No.: 13236 | WILEY | Dispatch: 19-8-2024 | No. of pages: 19 | CE: | PE:

RESEARCH ARTICLE OPEN ACCESS

Predicting Protests and Riots in Urban Environments With Satellite Imagery and Deep Learning

Scott Warnke[1] | Daniel Runfola[2]

[1]Department of Applied Science, William & Mary, Williamsburg, Virginia, USA | [2]Department of Applied Science & Data Science Program, William & Mary, Williamsburg, Virginia, USA

Correspondence: Daniel Runfola (danr@wm.edu)

Received: 15 March 2024 | **Revised:** 17 June 2024 | **Accepted:** 31 July 2024

Funding: This work was supported by National Science Foundation (Grant 2317591).

Keywords: conflict | convolutional modeling | deep learning | non-permissive areas | satellite imagery

ABSTRACT

Conflict, manifesting as riots and protests, is a common occurrence in urban environments worldwide. Understanding their likely locations is crucial to policymakers, who may (for example) seek to provide overseas travelers with guidance on safe areas, or local policymakers with the ability to pre-position medical aid or police presences to mediate negative impacts associated with riot events. Past efforts to forecast these events have focused on the use of news and social media, restricting applicability to areas with available data. This study utilizes a ResNet convolutional neural network and high-resolution satellite imagery to estimate the spatial distribution of riots or protests within urban environments. At a global scale ($N = 18{,}631$ conflict events), by training our model to understand relationships between urban form and riot events, we are able to predict the likelihood that a given urban area will experience a riot or protest with accuracy as high as 97%. This research has the potential to improve our ability to forecast and understand the relationship between urban form and conflict events, even in data-sparse regions.

1 | Introduction

Instances of social unrest, often manifesting as riots or protests, wield significant influence on the communities, regions, and nations where they unfold (Bencsik 2018). The repercussions of such events are wide-ranging, ranging from geopolitical transformations (i.e., riots in Egypt in 2011 (Joya 2011), and Hong Kong in 2019 (Purbrick 2019)) to substantial economic losses (exemplified by the hundreds of millions of dollars incurred during the 2011 riots in the UK (Bencsik 2018)). These events may result in human casualties, as evidenced by food riots in Africa in 2007–2008 (Berazneva and Lee 2013) and riots caused by garbage collection issues in Beirut in 2015 (El Warea et al. 2019). These events impact cities across the entire globe, with recent examples in Latin America (Eckstein 2001), Asia (Purbrick 2019), Africa (Joya 2011; Berazneva and Lee 2013), and Europe (Andronikidou and Kovras 2012). Because of the importance of these events, scholars across multiple disciplines have sought to both predict and understand them, using a wide range of data sources and techniques (Pond and Lewis 2019; Snow, Vliegenthart, and Corrigall-Brown 2007; Davies et al. 2013). However, most of these approaches have relied on sources that may not be available or reliable in geographies of interest, such as news articles. Here, we explore the capability of satellite imagery to aid in the prediction of protest and riot events, explicitly seeking to understand the degree to which this globally available source of information may be able to augment existing predictive methodologies. This approach exploits correlations between the human-built environment—that is, urban form (Fox and Bell 2016)—and the likelihood of a protest or conflict event at a given geographic location.

One of the core innovations that enable us to estimate social events (such as conflict) from satellite imagery is convolutional modeling (Goodman, BenYishay, and Runfola 2021). Deep learning, including the use of Convolutional Neural Networks (CNNs), is being used in a wide range of applications from detecting changes in urban environments (Daudt et al. 2018) to tracking typhoons (Rüttgers et al. 2019). This includes innovations from the field of computer vision, which have shown the capability of CNNs to detect objects (Shin et al. 2016) and classify images (Krizhevsky, Sutskever, and Hinton 2017; Chauhan, Ghanshala, and Joshi 2018). Deep learning can be used in conjunction with satellite imagery to perform many different classification and detection tasks, such as detecting infrastructure destruction in conflict environments (Nabiee et al. 2022), identifying ships (Leclerc et al. 2018; Patel, Bhatt, and Mazzeo 2022), land cover and land use analysis (Helber et al. 2019; Kussul et al. 2017; Carranza-Garca, Garca-Gutiérrez, and Riquelme 2019; Lv et al. 2024), urban expansion (Zhang et al. 2018, 2019; He et al. 2019), and road quality analysis (Brewer et al. 2021). Building on this work, in this piece we combine global-scope high-resolution satellite imagery sourced from *Planet* with information on the spatial distribution of protest and riot events from *ACLED*, seeking to establish the degree to which satellite information can be used to directly predict the geospatial locations of protest events.

This paper is structured as follows. In Section 2, we introduce background literature pertaining to modeling civil unrest, deep learning, and satellite imagery. In Section 3, we discuss our data collection and methodology. In Section 4, we present our results. In Section 5, we provide some conclusions and discussion.

2 | Background

2.1 | Measurement and Modeling of Riots and Protests

Riots and protests constitute integral components of democratic societies (U.S. Constitution 1791; Anderson and Mendes 2006), yet it is imperative for government authorities to effectively mitigate the economic and human costs that may be associated with these events to maintain stable governance (Klein and Regan 2018). This is accentuated by the heightened prevalence of protests and riots on a global scale in recent years (Ciorciari and Weiss 2016). One viable strategy for authorities to temper the negative impacts of these events is through preemptive allocation of resources, such as medical units (Gong and Batta 2007) or increased international presence (i.e., UN peacekeepers) in anticipation of unrest (Greer and McLaughlin 2010). On the international scale, in an attempt to protect citizens who are traveling abroad, responsible governmental foreign offices (the US Department of State as an example) may also issue travel warnings for particular areas to avoid (Löwenheim 2007). However, proactive approaches necessitate the capacity to predict both the time and location of potential conflict events (Wu and Gerber 2017).

A number of approaches exist which aid in the measurement and prediction of protests or riots (Wu and Gerber 2017). Past literature, for instance, has demonstrated the utility of news reports in providing valuable insights into civil conflict, such as riots and protests in response to rising food prices (Heslin 2021). Using this approach, studying riots in France, researchers were able to replicate the spread of riots using an epidemic-like model with as few as six parameters that included population demographics, police reports, and spatial information (Bonnasse-Gahot et al. 2018). Social media platforms represent another venue for authorities to detect and analyze real-world events, including social unrest like riots and protests (Becker, Naaman, and Gravano 2011; Korolov et al. 2016; Petrović, Osborne, and Lavrenko 2010). X (formerly Twitter) is a common focus of these studies, and can be used as a near real-time reporting source, distinguishing between real-world events and random tweets with 83% accuracy (Becker, Naaman, and Gravano 2011). Analysis of Twitter data demonstrates the correlative relationship between daily hashtag use and protests, enabling predictions 24–48 h prior to protests in Baltimore and New York City during 2015 (Korolov et al. 2016). Prior work in this field has show the ability to predict the probability of fatalities associated with conflict events using satellite imagery, within conflict areas in Nigeria, with accuracy rates of 80% when combining Landsat imagery and CNNs (Goodman, BenYishay, and Runfola 2021).

Much of the current research in forecasting social unrest is focused on the likelihood of a future event (Renaud et al. 2019; Phillips et al. 2017; Cadena et al. 2015; Filchenkov, Azarov, and Abramov 2014; Compton et al. 2013). There are other efforts to better understand and model the characteristics of smaller subevents within broader riots, such as shooting or fires (Alsaedi, Burnap, and Rana 2017). Mathematical modeling of riots demonstrates an ability to accurately simulate many of the spatial characteristics of riots, including the distance participants will travel within contiguous riot areas (Davies et al. 2013). X (formerly Twitter) text analysis demonstrates the ability to detect and discriminate between disruptive events and normal information dissemination (Alsaedi, Burnap, and Rana 2015). Social media has been studied to demonstrate not only how information is distributed about future and concurrent protests, but also how individuals are recruited into protesting through the spread of information in their social network (González-Bailón et al. 2011).

The accuracy and spatial specificity of alternative riot and protest forecasting techniques vary. Previous research has shown that leveraging information from social media (i.e., Tweets) can result in the accurate prediction of riots in some cities (i.e., Baltimore and New York City), but these models require location-specific information or hashtags which inhibit their use in other settings (i.e., San Francisco) (Korolov et al. 2016). Related tweet-based analyses have shown that accurate temporal estimates across broad geographies are possible but without spatial specificity in where riots or protests are likely to occur (González-Bailón et al. 2011). Other researchers have used a broader range of sources to achieve higher spatiotemporal accuracy, such as police reports, but these techniques are inherently limited to a small number of areas in which such information is available (Bonnasse-Gahot et al. 2018; Korolov et al. 2016; Alsaedi, Burnap, and Rana 2017, 2015; González-Bailón et al. 2011).

2.2 | Convolutional Modeling and Satellite Imagery

In this study, we rely on convolutional neural networks, a type of deep learning designed for analyzing image data. These techniques are effective at detecting, labeling, and differentiating objects (Krizhevsky, Sutskever, and Hinton 2017; Simonyan and Zisserman 2014; Zhang, Zhang, and Du 2016; He et al. 2016; Voulodimos et al. 2018; Gorban, Mirkes, and Tyukin 2020). CNNs represent a family of deep learning techniques implementing convolutional layers to extract features from an image (Zhang, Zhang, and Du 2016). Many types of CNN architectures perform well across a wide range of computer vision tasks (Simonyan and Zisserman 2014; Voulodimos et al. 2018; Szegedy et al. 2015; Bressem et al. 2020).

There is a long history of utilizing satellite imagery in research that is based on visually observable characteristics, such as habitat and land cover change (Alo and Pontius Jr 2008; Stow et al. 2008; Rogan and Chen 2004), soil evaluation (Foody and Mathur 2004), and urban land cover (Zhou and Troy 2008). When satellite imagery is used in conjunction with deep learning techniques, including CNNs, researchers are able to learn about topics not normally associated with traditional satellite imagery uses, such as predicting crime (Najjar, Kaneko, and Miyanaga 2018) or the prevalence of cancer (Bibault et al. 2020). Other examples include estimating human migratory flows (Runfola et al. 2022), estimating educational outcomes (Runfola, Stefanidis, and Baier 2022), tracking economic growth in China (Brewer, Lv, and Runfola 2023), predicting road quality (Brewer et al. 2021), and estimating socioeconomic census variables from satellite imagery (Runfola et al. 2024).

In scenarios where data is challenging or impossible (i.e., historic time periods) to collect, there is increasing evidence that satellite imagery can aid in filling data gaps (Goodman, BenYishay, and Runfola 2021; Jean et al. 2016; Bharti and Tatem 2018; Hu et al. 2019; Aung et al. 2021). The capability of satellite information becomes particularly important in the context of studying riots and protests, given that the majority of literature we identify focuses on news or social media sources (Purbrick 2019; Ciorciari and Weiss 2016; Greer and McLaughlin 2010; Wu and Gerber 2017; Becker, Naaman, and Gravano 2011; Korolov et al. 2016; Renaud et al. 2019; Phillips et al. 2017; Cadena et al. 2015; Filchenkov, Azarov, and Abramov 2014; Compton et al. 2013; Alsaedi, Burnap, and Rana 2017). Our approach aims to leverage the availability of satellite imagery as a data source, increasing the application to predicting events when other traditional data sources are restricted. There are many countries of research interest that do not allow free access to social media or control the news narrative, such as Russia (Gehlbach 2010), China (Tai 2014), Iran (Rahimi 2015), and Venezuela (Pain and Korin 2021). Satellite imagery provides a unique capability to access data in a country that might restrict access to social media or control news sources, motivating us to use satellite imagery to predict conflict.

3 | Data and Methods

The primary objective of this work is to predict if a riot or protest will occur in a specific urban area, based solely on data from satellite imagery. In order to accomplish this objective, we leverage convolutional neural networks in combination with two data sources, ACLED (Raleigh, Kishi, and Linke 2023) and Planet (Planet Team 2023a). We use these data to generate two different sets of information: the first set is satellite imagery of locations where riots occurred, and the second is a set of images of proximate areas (within the same city) that did not experience a riot event. Our deep learning model then seeks to disambiguate between these two cases, based on satellite imagery alone. This section provides details of our data processing and analytic approach.

3.1 | Data

3.1.1 | Selecting Riot Locations

Determining the locations where riots and protests have occurred is the first step in developing a data set for this work. To identify these locations, we leverage The Armed Conflict Location Event Data Project (ACLED), an open-source database, which contains information on a wide range of conflict types from across the globe (Raleigh, Kishi, and Linke 2023). ACLED contains more than 1.5 million events from 1997 to 2023, which we aggregate, categorize, and curate to create a data source that can specify time and location for conflict. We filter this database according to a number of criteria:

1. *Type of event*. We focus our analysis on protests and riots, which primarily represent urban unrest.
2. *Date*. We only leverage protest or riot events with a known date of occurrence.
3. *Geography*. Only events with a neighborhood-level geographic footprint are selected.[1]

After filtering events, we are left with a resultant database of 53,307 events. In order to prevent overrepresentation of any single unique location in the database, a maximum of 500 events are randomly selected from each neighborhood (i.e., "Seoul—Jongno"). After this stage, a total of 37,728 events across 1089 unique locations are leveraged to construct our dataset of the location of conflict events.

3.1.2 | Satellite Data

Once we identify the location of riot events, we retrieve relevant *PlanetScope* satellite imagery both (a) 24–48 h prior to each event, and (b) in similar, nearby geographic locations that did not experience unrest. PlanetScope—an integrated collection of images from the Dove, Dove-R, and SuperDove satellites—provides four-band (RGB and NIR), approximately 3–4 m spatial resolution satellite imagery with a daily temporal resolution (Planet Team 2023b; see Table 1). For both cases of imagery (with and without riot), we consider images that contain < 50% cloud cover. An example of the imagery available can be seen in Figure 1.

For each of the 37,728 instances of riots in our filtered ACLED dataset, we first retrieve a full scene of imagery from 24 to 48 h prior to the event (Table 2). These scenes are guaranteed to encompass the latitude and longitude representing the centroid of the neighborhood at which a conflict occurred; in cases where multiple images were available for a given event, we chose the one closest in time to the event (with a minimum of 24 h prior to the event). Ultimately, this process resulted in 19,902 satellite scenes being downloaded, with an average spatial dimension that can vary depending on the generation of satellite 2 and geographic latitude of collection. Because riots may occur at the same location, but at multiple points in time, some locations (i.e., a seat of government and culturally significant locations) may appear in the database multiple times; the most common of these occurrences are summarized in Table 3.

From the satellite scene retrieved for each conflict event, we extract two types of data. First, we extract a 1 km by 1 km box centered on the conflict event neighborhood. This box is saved and identified as the location of the unrest in our database.

Second, we extract a number of cases to serve as null events—that is, locations from the same urban area, but where no unrest occurred. To generate these null cases, we follow a multiple step process in which we:

1. *Identify urban areas*. We only consider areas in the scene that have a population density over 300 inhabitants per kilometer.
2. *Exclude areas that are within 10 km of the conflict event*. We isolate the conflict event by removing the urban areas that are within 10 km of the centroid of the neighborhood in which conflict occurred.
3. *Sample*. With the remaining urban areas in the satellite scene, we generate a list of random centroids which are constrained to be a minimum of 2 km apart, and select a maximum of 10 of these to generate 1 km box "null" locations at which no protest or conflict occurred. The 2 km separation ensures that none of our null boxes overlap.

TABLE 1 | Technical wavelength specifications for RGB bands of PlanetScope sensors (Planet Team 2023b).

Band	Dove classic Wavelength (nm)	Dove-R Wavelength (nm)	SuperDove Wavelength (nm)
Red	590–670	650–682	650–680
Green	500–590	547–585	547–583
Blue	455–515	464–517	465–515

In step 1, we overlay information about the degree of urbanization (Schiavina, Melchiorri, and Pesaresi 2023; European Commission and Statistical Office of the European Union 2021) onto each satellite scene to determine what portions are urban, and which parts are not. This is accomplished by using the DEGURB dataset (Schiavina, Melchiorri, and Pesaresi 2023), which was developed by the European Commission's Joint Research Centre. This data categorizes geographical areas into Urban Centre, Urban Clusters (including towns and suburbs), and Rural Grid Cells (including villages and dispersed rural) zones based on population density and contiguity of dense areas (European Commission and Statistical Office of the European Union 2021). The DEGURB dataset used in this work is representative of 2020 (see Figure 2; Schiavina, Melchiorri, and Pesaresi 2023). This binary representation of urban areas is then applied to each satellite scene as a mask, allowing us to select null cases from proximate urban areas.

In step two, in order to ensure the areas selected for null cases are distinct from the areas of unrest, we exclude all urban areas up to 10 km away from the centroid of the riot neighborhood from consideration, as illustrated in Figure 3.

Third, after excluding the 10 km region around each unrest event, from the remaining urban regions in the satellite scene we select random locations for null-riots. We accomplish this by generating a list of random latitudes and longitudes that are within the available regions. We ensure that each of these random locations is at least 2 km away from any other locations on our random list. We then take a maximum of ten of these locations and construct a 1 km box around each one. We construct up to 10 null cases (that do not overlap) from the eligible urban regions from each scene (noting that less dense urban areas are occasionally represented by < 10 null cases due to a lack of

FIGURE 1 | Satellite image of Athens Greece, taken 31 January 2018. Imagery © Planet Labs PBC 2023. All rights reserved. Background map from OpenStreetMap (OpenStreetMap Contributors 2024).

proximate urban areas). A visualization of the results from this process can be seen in Figure 3.

After this process is completed, for each conflict event we are left with a set of one (1 km^2) kilometer box representative of where unrest occurred, and up to 10 (1 km^2) km boxes representative of urban areas proximate to the unrest event, but with no known activity. Across our full dataset of 19,902 unrest locations, 18,634 (93.6%) had 10 null cases available; the distribution of null cases across images can be seen in Figure 13. Our final

TABLE 2 | PlanetScope constellation (Planet Team 2023b).

Instrument	Image area	Availability
Dove classic	25×11.5 sq km	July 2014 to April 2022
Dove-R	25×23 sq km	March 2019 to April 2022
SuperDove	32.5×19.6 sq km	March 2020 to present

TABLE 3 | Neighborhood locations that occur most frequently.

Country	Neighborhood	Count	Earliest date	Latest date
Pakistan	Karachi—Saddar	278	7 October 2017	30 September 2022
Iran	Tehran—District 6	270	9 October 2017	26 September 2022
Iran	Tehran—District 12	268	9 October 2017	28 September 2022
Lebanon	Beirut—Port	252	7 October 2017	26 September 2022
Greece	Athens—Central Athens	247	18 January 2018	28 September 2022
South Korea	Seoul—Jongno	240	18 January 2018	21 September 2022
South Korea	Seoul—Jung	226	18 January 2018	26 September 2022
Italy	Rome—City Center	222	7 January 2018	23 September 2022
India	Delhi—New Delhi	220	2 October 2017	4 September 2022
South Korea	Seoul—Seocho	220	8 January 2018	28 September 2022

Note: "Earliest" and "Latest" date refer to the earliest and latest date of a protest event for each neighborhood. For example, in the neighborhood of Seocho in Seoul, 220 independent protest or riot events occurred from 8 January 2018 to 28 September 2022. In our analysis, this would be represented by 220 individual satellite tiles, each taken between 24 and 48 h before the actual event.

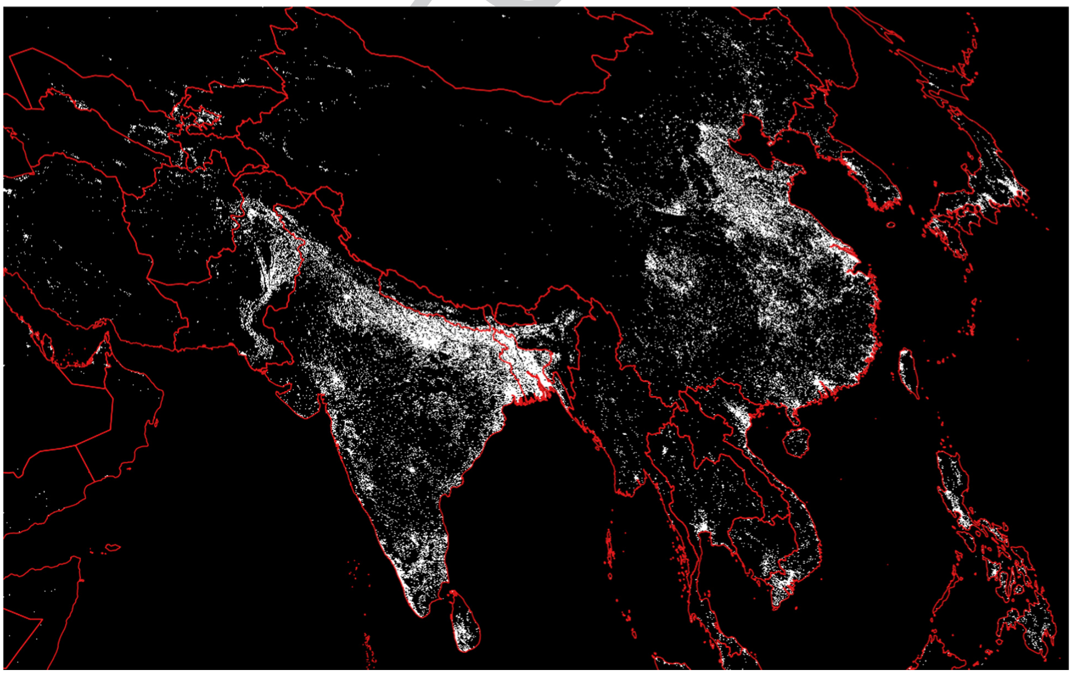

FIGURE 2 | A portion of the DEGURB data, highlighting areas of the world that are considered urban in our data set. DEGURB defines urban regions as those with a density more than 300 inhabitants per km (European Commission and Statistical Office of the European Union 2021) Red lines represent country-level boundaries (Runfola et al. 2020).

dataset includes only locations that have the full complement of null clips, for a total of 18,631 cases of unrest and 186,310 null cases. We then normalize all of these image clips based on a sample of the full satellite scenes (Goodman, BenYishay, and Runfola 2021; Lv et al. 2024; Runfola et al. 2022; Brewer, Lv, and Runfola 2023). Tests of different permutations of this dataset (i.e., models with a 1:1 ratio of null and riot cases) can be found in Section 6.1 of the appendix.

3.2 | Methods

Our overall modeling architecture is summarized in Figure 4. To estimate the likelihood of an unrest event occurring or not at each location, we leverage a ResNet18 (He et al. 2016) as our base model, but replace the fully connected layer with a series of dense layers that include 128, 64, and 32 hidden nodes. In order to improve the efficiency of our training, following other

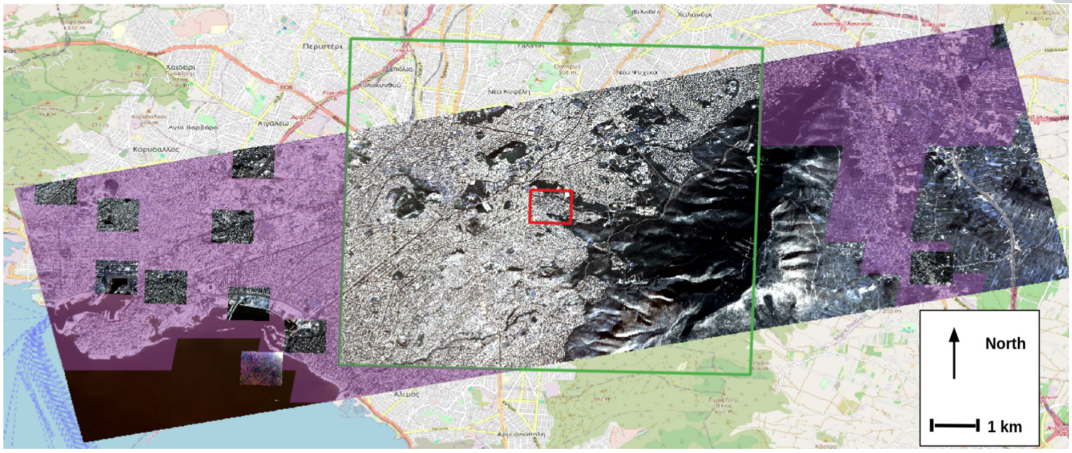

FIGURE 3 | Satellite Image of Athens Greece, taken 31 January 2018. The red box in the center of the image is a 1 km box around the riot location. The green box is a 10 km exclusionary area around the riot location, from which we do not draw "null" case contrasts. Areas which fall outside the green box, that are also urban, are eligible for selection (displayed in purple). From the potential null region, we sample random, non-overlapping 1 km boxes to generate null location clips. Imagery © Planet Labs PBC 2023. All rights reserved. Background map from OpenStreetMap (OpenStreetMap Contributors 2024).

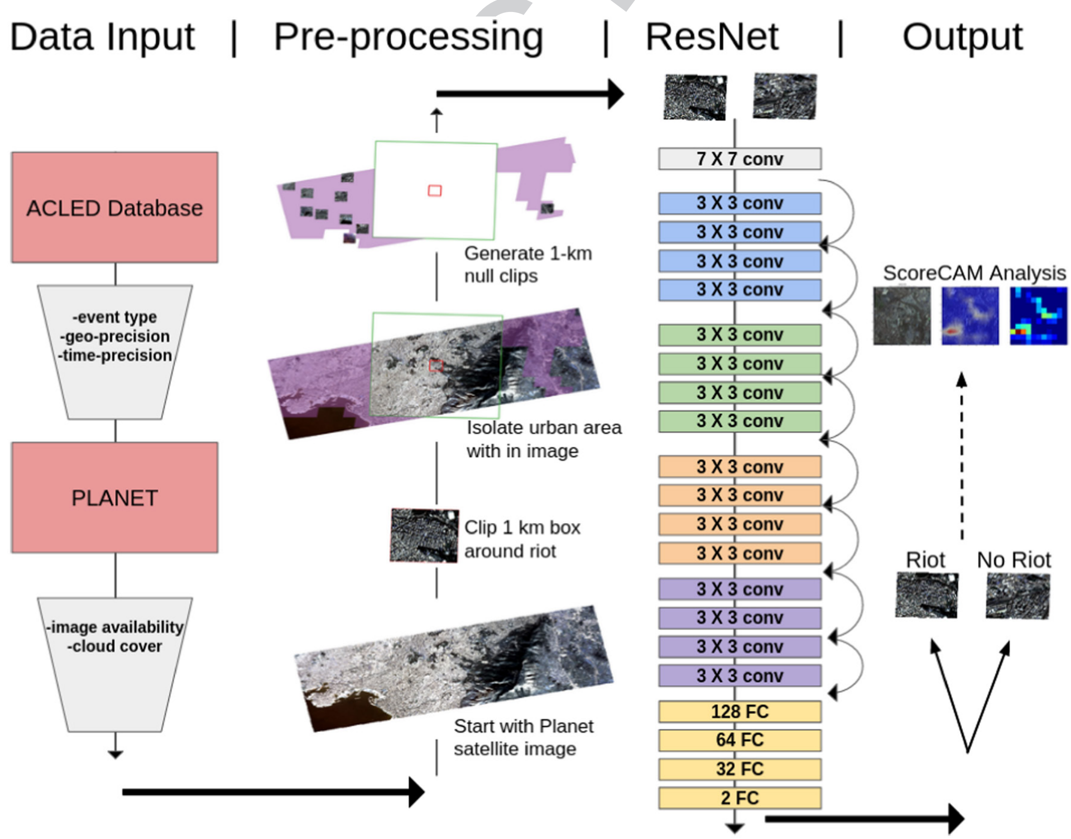

FIGURE 4 | A synopsis of our overall modeling architecture. Stages include the collection of data, pre-processing, network training, categorization, and explainability analysis. Imagery © Planet Labs PBC 2023. All rights reserved.

literature in the satellite imagery analysis space (Goodman, BenYishay, and Runfola 2021; Lv et al. 2024; Brewer et al. 2021; Runfola et al. 2022; Brewer, Lv, and Runfola 2023), we use pre-trained weights from ImageNet as our initial baseline.

3.2.1 | Hyperparameter Search

Prior to training on all 18,631 events, we first randomly select a subset of 1000 conflict events (1000 unrest cases and 10,000 null cases) to implement a grid search across hyper-parameters.[2] To account for class imbalance, we implement a weighted cross entropy loss (Ho and Wookey 2019) with an ADAM optimizer (Kingma and Ba 2014) for our training procedure.

Our hyperparameter search includes trials of different learning rates, L2 regularization, dropout, freezing layers (results and parameters from a sample of the trials can be seen in the appendix in Section 6.2). Results from a selection of three of the best performing cases in the hyperparameter testing are shown in Table 4. On the basis of these results, we select one model (denoted as Model C in Table 4) to test on the full dataset, which is described in Table 5.

We assess our model by interpreting the overall accuracy, precision, and recall. The precision is the ratio of true positives to the number of positive predictions our model made (Davis and Goadrich 2006), which will measure our model's ability to correctly predict riots when it does makes a prediction. The recall is the ratio of true positives to the number of riots in the data set, which measures our model's ability to identify how frequently riots are occurring (Davis and Goadrich 2006).

3.2.2 | Additional Analyses

In addition to identifying the best convolutional model performance, we implement two additional analysis to better understand the strengths and weaknesses of this approach. These include (a) generating information on the country-level performance of the model and (b) and explanatory model that seeks to identify the features within a given image that are correlated with conflict events (or the lack thereof).

To explore the spatial distribution of accuracy of the approach, we first filter our data to only consider countries that had 500 or more observations (a minimum of 250 riot clips and 250 null clips). This creates a validation set consisting of 32,548 clipped images, distributed across 24 countries (see Table 6). From this, we withhold 20% of each country's observations for validation after training. This ensures that each country has at least 100 observations (50 riot clips and 50 null clips) for validation. We then select the hyperparameters from our best performing model (model C, see Table 4), and train a ResNet18 using 80% of the validation data (26,058 images, half riot or protest and half null) for 50 epochs. We then use the withheld 20% of images (6490 images, half riot or protest and half null) to test for accuracy within each country.

To begin to explore the underlying drivers of model performance, we additionally take preliminary steps toward trying

TABLE 4 | Representative results from hyperparameter tuning efforts.

Model	Learning rate	L2 decay	Freeze layers	Drop out	Test accuracy (%)	TP	FP	FN	TN	Precision (%)	Recall (%)	F1 (%)
A	0.000001	None	None	No	92.5	56	34	122	1861	62.2	31.5	41.8
B	0.000015	0.01	First 5	No	90.5	95	103	93	1782	48.0	50.5	49.2
C	0.00001	0.001	First 5	No	92.2	95	55	107	1816	63.3	47.0	54.0

Note: All training iterations were based on the same ResNet18 architecture, training with the same 1000 satellite images from the full dataset, for 40 epochs.

to assess what features the model may be identifying and using in predictions. To implement this process, we leverage Score-CAM (Wang et al. 2020). Score-CAM is a Class Activation Mapping (CAM) method that attempts to explain, with a human interpretable visual display, the features within an image that determine classification. Score-CAM differs from traditional CAM methods that utilize gradients and instead use the forward pass scores of activation maps to determine the significance for target classes (Wang et al. 2020). For the purposes of this work, Wang et al. found that it outperforms other techniques when there are multiple objects of relevance in the scene (Wang et al. 2020), a nearly universal characteristic of satellite imagery.

4 | Results

4.1 | Full Data Set

In this section, we report our findings from our analysis of the full dataset ($N = 204{,}941$ clipped images), using the best-performing model from our hyper-parameter testing (model C, as described in Table 4). The results of this model are presented in Table 5.

As Table 5 shows, the approach outlined in this paper achieves an overall accuracy of 97.39%—that is, of the 40,989 images in the test dataset, 39,921 were correctly identified as the site of a riot or not.[3] There are 3646 riot or protest images in the testing set and the model correctly identifies 2741 of these, resulting in a recall score of 75.18%. This demonstrates the model's ability to distinguish riot/protest events from non-riot events. The model predicts there will be a riot in 2904 of the images and is only incorrect 163 times producing a precision score of 94.30%. In the context of our scenario, when the model predicts there will be a riot or protest in an image, it is correct over 94% of the time.

TABLE 5 | Results from ResNet18 using the full data set.

Test accuracy	**97.39%**
True positives (predict riot)	2741
False positives	163
False negatives (missed riot)	905
True negatives	37,180
Precision	94.39%
Recall	75.18%
*F*1 score	83.69%

4.2 | Balanced Validation Data Set

We validate the performance of our model with a data set that withholds data from training and testing, and has a one-to-one riot-to-null ratio. The results of this validation training are displayed in Table 7. The accuracy of the validation testing was 89.41%; 5803 of the 6490 images were correctly identified. This validation testing has very similar false positive and false negative rates, resulting in precision, recall, and *F*1 scores that are similar to the test accuracy. Of note, the re-trained model which withheld data for each individual country had a slightly lower global accuracy than our full results, of 89%. While this 89% accuracy is lower than the accuracy from the full data set shown in Table 5, the testing circumstances of the validation are more challenging due to the even split between riot and null in the validation data set.

In addition to the global accuracy, we also subset our data by country and report accuracy within each based on a validation

TABLE 7 | Results from validation testing; 6490 images are withheld during training, half of which are from a riot and half from a non-riot.

Test accuracy	**89.41%**
True positives (predict riot)	2903
False positives	345
False negatives (missed riot)	342
True negatives	2900
Precision	89.38%
Recall	89.46%
*F*1 score	89.42%

Note: These results do not have the class imbalance present in the full data set, instead there is a single riot clipped image and a single non-riot clipped image for every full satellite image. The accuracy of the network approaches 90%, with similar capabilities to distinguish among false positives and false negatives.

TABLE 6 | There are 32,548 clipped images in the validation data set.

Country	Images	Country	Images	Country	Images	Country	Images
South Korea	7494	South Africa	1480	Ukraine	924	Greece	634
Pakistan	2622	Chile	1302	Thailand	890	Yemen	604
Iran	2334	Japan	1256	Italy	728	United Kingdom	566
Lebanon	1656	India	1148	Indonesia	678	Taiwan	562
Palestine	1572	Brazil	1112	Russia	668	Peru	522
China	1550	Bangladesh	1092	Venezuela	648	Iraq	506

Note: Half of these are from riots/protests, and half are null clips. Only countries that have at least 500 images are included. Twenty percent of each county's images will be withheld from training and testing, and used in validation.

TABLE 8 | Results from country level accuracy after validation testing.

Country	Count	Accuracy (%)	TP	FP	TN	FN
Lebanon	330	94.54	159	12	153	6
Iran	466	94.42	225	18	215	8
Pakistan	524	92.56	247	24	238	15
South Korea	1498	92.12	700	69	680	49
Ukraine	184	91.85	84	7	85	8
Chile	260	91.15	120	13	117	10
Iraq	100	91.00	45	4	46	5
China	310	90.00	140	16	139	15
Palestine	314	89.49	141	17	140	16
Venezuela	128	89.06	58	8	56	6
Bangladesh	218	88.99	90	5	104	19
India	228	88.60	97	9	105	17
Italy	144	88.19	62	7	65	10
Greece	126	87.30	62	15	48	1
Thailand	178	87.08	75	9	80	14
Indonesia	134	86.57	62	13	54	5
Japan	250	85.60	105	16	109	20
Brazil	222	85.14	94	16	95	17
United Kingdom	112	83.93	50	12	44	6
South Africa	296	82.43	114	18	130	34
Taiwan	112	82.14	45	9	47	11
Yemen	120	78.33	41	7	53	19
Russia	132	78.03	50	13	53	16
Peru	104	77.88	37	8	44	15

Note: These results are listed from highest accuracy to lowest accuracy. We have also included the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) for each country.

set (see Section 3.2.2 of our methods). The results of this country-specific validation testing are shown in Table 8. Lebanon (94.5%), Iran (94.4%), and Pakistan (92.6%) were the most accurate in this analysis, while Yemen (78.3%), Russia (78.0%), and Peru (77.9%) were the least accurate countries. No clear regional patterns existed, though some evidence suggests that accuracy and total number of observations may be correlated (i.e., less accurate news media reporting in Russia may be attributable to the lower accuracy in that context).

Of note, we observe a strong correlation between our softmax classification scores and accuracy within each country around the world, suggesting that softmax scores can be used as a proxy for prediction confidence (see Figure 5). While softmax may bias toward higher degrees of confidence (Pearce, Brintrup, and Zhu 2021; Subramanya, Srinivas, and Babu 2017), as a relative metric it may provide helpful guidance to policymakers seeking to use these types of methods.

4.3 | Explainability of Results

For our best performing model (model C in the Table 5), we implement Score-CAM on a subset of randomly selected, paired locations, ultimately consisting of 1089 riot locations, and 1089 null locations.[4] The Score-CAM results are then visually reviewed in an attempt to discern patterns in what the trained ResNet prioritized in classification. Understanding the results of utilizing Score-CAM on our data is inherently difficult to interpret or understand, making this a rich area for future work; we discuss this limitation further in Section 4.3 of the discussion.

While this analysis is inherently qualitative, visual interpretation indicates a few clear patterns. An example of the first of these is displayed in Figure 6. We can observe a large sports stadium in the image in the southeast region of Figure 6. This large stadium is the location which Score-CAM identifies as the portion of the image which leads toward the classification (indicated through brighter values in the displayed heatmap). In this case, the sports stadium leads the ResNet to classify the scene as a non-riot. We can see another example in Figure 7, in which again, the ResNet identifies the sports stadium as the reason to classify the scene as a non-riot. We do not offer any explanation for why the sports stadiums are indicative of a non-riot scene, but these stadiums provide an example of the specific features which ResNet is learning to make classification decisions.

Another example highlighted in the Score-CAM analysis is shown in Figure 8. We can see a densely populated area, with a large open park or green space in the center of the image. The trained network correctly predicts this image is from a riot or protest. When we reference the ACLED data, this image is from a protest in the Lalbagh neighborhood of Dhaka, Bangladesh. Lalbah is a fort built during the Mughal period in 1678, which was used subsequently by the British and Bangladesh governments as a location of governance and influence (Shakur, Islam, and Masood 2010). Today, it is a location containing monument's and statues symbolizing rulers and regimes of the past, that is known as a common location for protests in the city of Dhaka (Begum 2018). While the deep learning model was not aware of these historic contexts, the unusual land use and associated image features were sufficient to classify this as a likely location of riots.

5 | Discussion and Conclusions

The results presented in Table 5 provide evidence that satellite information alone can provide useful information for the purposes of predicting where, within an urban environment, protests and riots are most likely to occur. While this finding is likely to be of interest to those operating in data-sparse environments, it is well supported by past social science literature highlighting the interconnected nature of urban form and social processes (Fox and Bell 2016; Begum 2018). By engaging in a global-scope study, here we are able to exploit this correlation by learning what these patterns are, and then leveraging them in estimation. This finding held true across multiple model and data permutations (see Tables 5 and 9), indicating that—even in some of the most

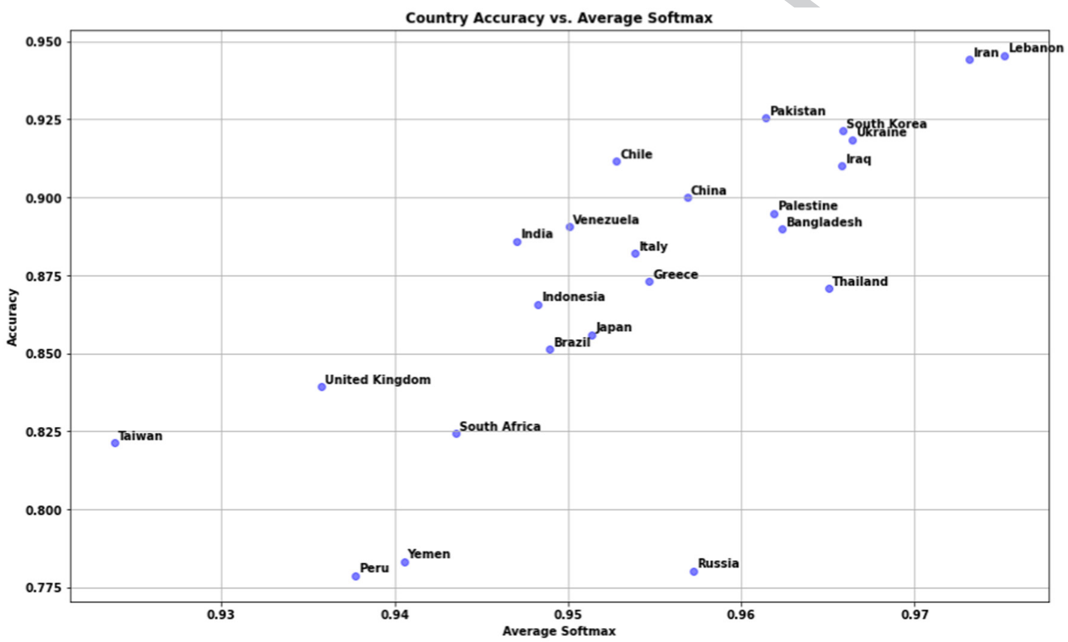

FIGURE 5 | The average softmax for each country when compared to the average accuracy of prediction of each country. Of note, the axis's do not begin at 0, but instead focus in on the domain and range of the values in the data. 13

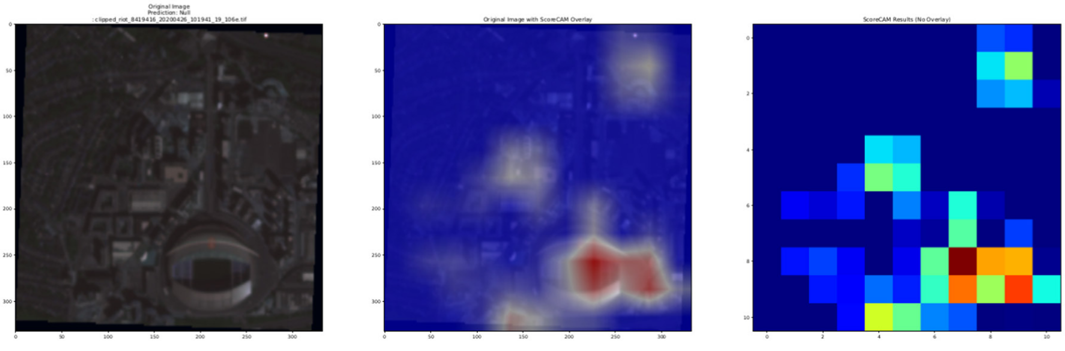

FIGURE 6 | Example clipped image on the left. The clipped image, a 1 km box around a riot location. The Score-CAM overlayed on top of the image is shown in the middle. The Score-CAM visual is displayed on the right. Imagery © Planet Labs PBC 2023. All rights reserved. 14

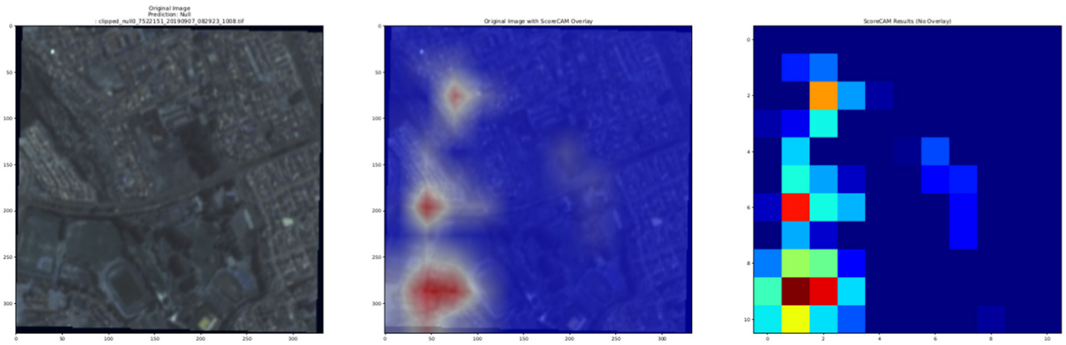

FIGURE 7 | Example clipped image on the left. The clipped image, a 1 km box around a non-riot location. The Score-CAM overlayed on top of the image is shown in the middle. The Score-CAM visual is displayed on the right. Imagery © Planet Labs PBC 2023. All rights reserved. 15

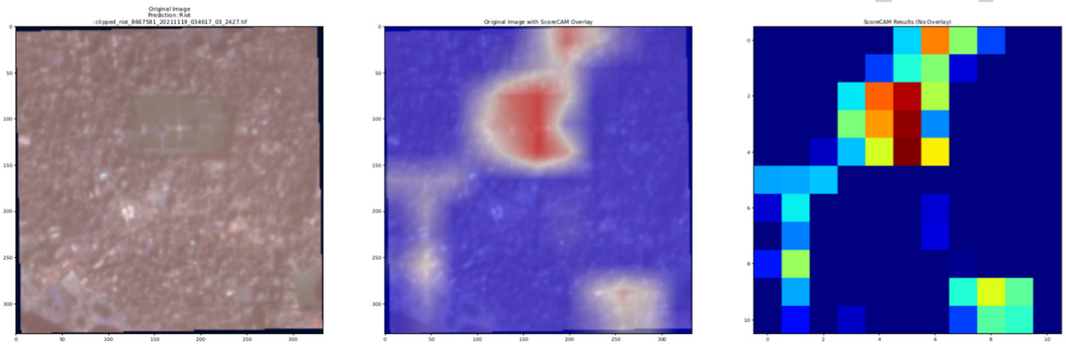

FIGURE 8 | The image on the left is centered on Lalbagh Fort in Dhaka Bangladesh, taken on 19 November 2021, < 48 h before a protest at that location. The Score-Cam visual is displayed on the right. Imagery © Planet Labs PBC 2023. All rights reserved. 16

challenging situations (i.e., relatively small training and validation sets), model accuracy can approach or exceed 90%.

Furthermore, this technique performs well across the globe. As highlighted in Table 7, there do not seem to be any regions that under perform. Many countries with a relative low accuracy score (i.e., Russia) are in close proximity to a country with a higher accuracy score (i.e., China). This pattern holds across the globe in South America, Asia, the Middle East, and Europe.

Of note, in our softmax analysis seeking to correlate scores to accuracy, a single outlier, Russia, is observed in Figure 5 and Table 7. Russia has a lower comparative accuracy to other countries with similar softmax results. This might be indicative of Russia's control of news sources (Gehlbach 2010), or inherit in ACLED's collection of data which relies on news sources and non-governmental observation organizations that might not be focused on Russia.

5.1 | Limitations

5.1.1 | Satellite Information

The satellite imagery we incorporate into this study has a number of notable limitations. First, while a satellite scene might contain 50% or less cloud cover (see Figure 9), the clipped images might be completely covered in clouds (see, e.g., Figure 10). Further, in some cases the conflict event selected may be at the edge of a scene, with no valid scene available to fill in null information, resulting in a partially clipped image (see Figure 11). 6

TABLE 9 | Results from ResNet18 using only a single riot clip and single null riot clip per location.

Test accuracy	**65.37%**
True positives (predict riot)	154
False positives	105
False negatives (missed riot)	46
True negatives	131
Precision	59.46%
Recall	77.0%
*F*1 score	67.1%

Additionally, some of the clips contain interference or distortion, such as the clip at the bottom of Figure 11.

Inter-related with these challenges, in many scenes, we were unable to identify enough geographic locations to support the creation of 10 null cases. For example, in Figure 12 we can see that the riot location in consideration does not have any null location possibilities due to the riot's proximity to the coast, and the concomitant lack of proximate urban areas eligible for building null (no-protest) cases. There are similar limitations that cause the distribution of clipped images in Figure 13.

Another limitation is in our definition of where conflict events occurred, as the definition of a "neighborhood" is inherently imprecise. We used OpenStreetMap (OpenStreetMap

Contributors 2024) to visually compare the size of our ten most repeated locations 3. We were able to confirm that the sizes of neighborhoods were inconsistent, but rarely of a size greater than our 10 km^2 exclusionary zone (see Figure 3).

5.1.2 | Explainability

Currently, the majority of explainability techniques in the literature focus on datasets consisting of object-centric images. For example, two common data sets CIFAR-10 and CIFAR-100 (Krizhevsky and Hinton 2009) are used in many computer vision tasks and competitions, but those data sets only have objects centered in the middle of the picture, taking up most of the image space. This differs significantly from our satellite imagery. Our images contain all of the spatial information within a square kilometer of a city. As opposed to an image of a cat or dog, our images have multiple buildings, cars, streets, parks, etc. So while current explainability techniques can highlight portions of our image that lead to classification which are easily human interpretable, it is challenging for us to determine what in the image is being highlighted. The example we discuss in Section 4.3, highlights sports stadiums in Score-CAM outputs as easily identified visually in the satellite image. There are other patterns that emerge in our Score-CAM analysis; however it is very difficult to describe many of the features Score-CAM identifies with easily identifiable semantic definitions. While we were able to identify a few other patterns, such as transitions from one zone to another zone (residential to commercial as an example), we are not confident in interpreting what these different types of zones are at this time. The field of explainability, as it relates to satellite images, has very little published in literature and remains a strong avenue for future inquiry.

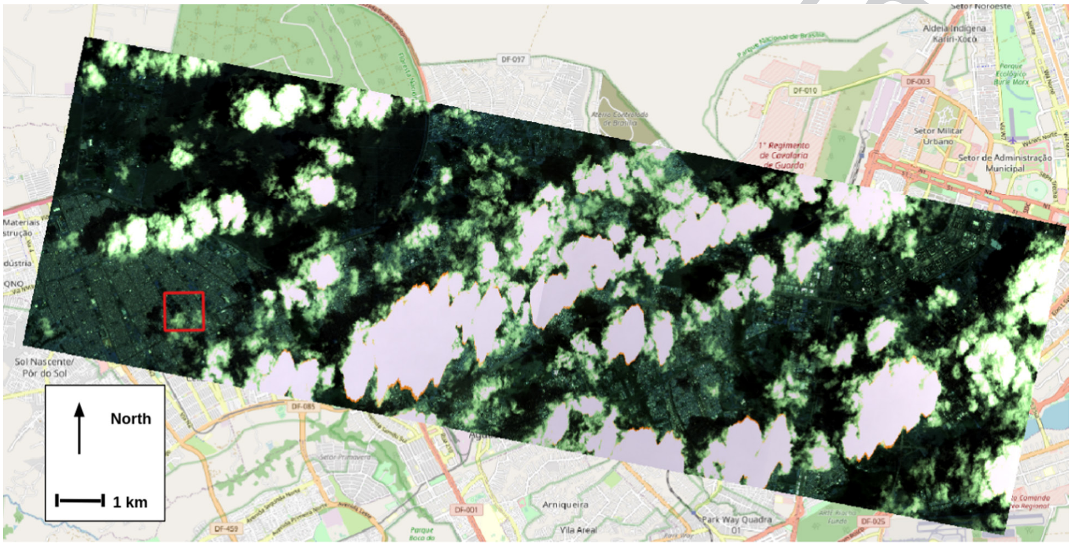

FIGURE 9 | Satellite image of Brazil collected on 1 November 2018. This image contains < 50% cloud cover for the full satellite scene. The riot location indicated in the red square has minimal cloud cover, but other locations in the scene will be impacted by the cloud cover as seen in Figure 10. 17
Imagery © Planet Labs PBC 2023. All rights reserved. Background map from OpenStreetMap (OpenStreetMap Contributors 2024).

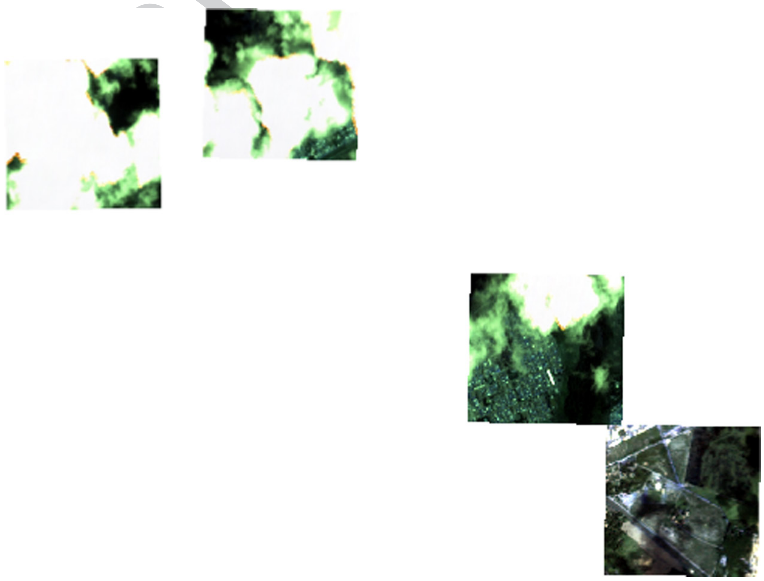

FIGURE 10 | Clips from a satellite image of Brazil collected on 1 November 2018. While the full image contains < 50% cloud cover, many of the clips are partially or completely obscured. Imagery © Planet Labs PBC 2023. All rights reserved.

5.1.3 | Additional Limitations

There are a number of additional limitations of the presented work. First, our data is focuses on spatial information, not temporal, and thus we do not generate predictions of *when* a riot will occur, only the likely urban locations. Leveraging changes in images over time could help us overcome this challenge, but will necessitate new modeling strategies beyond those presented in this piece. Second, we have selected a ResNet18 as our base model, which could limit our model performance if alternative architectures are better performing. Another limitation of the presented work is in the limited scope of network architectures

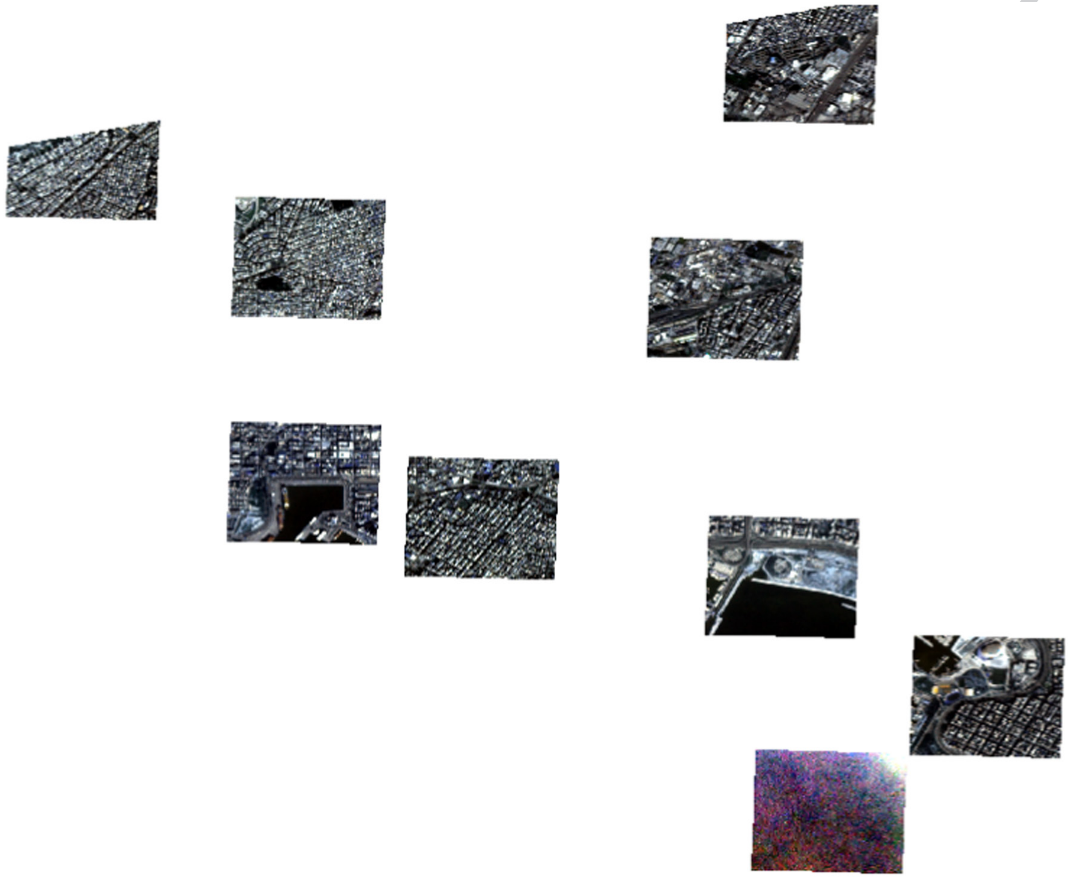

FIGURE 11 | Nine of the null riot clipped images from Athens, Greece. Imagery © Planet Labs PBC 2023. All rights reserved.

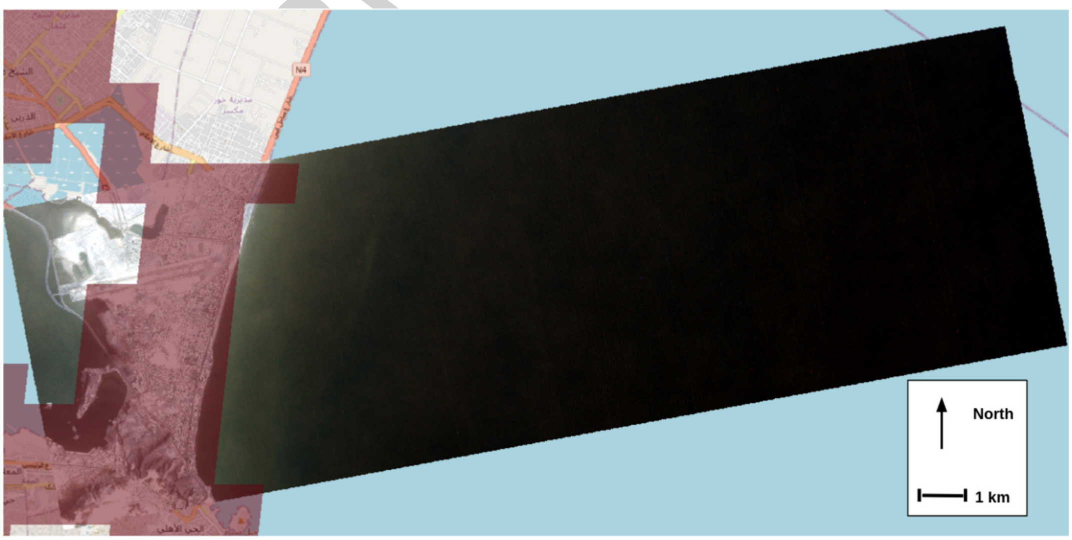

FIGURE 12 | Satellite image from Yemen collected on 14 September 2020. The urban areas are shown in red. Most of this image is not usable because of the lack of urban areas. Imagery © Planet Labs PBC 2023. All rights reserved. Background map from OpenStreetMap (OpenStreetMap Contributors 2024).

tested. Here, we focus explicitly on ResNet18, with anecdotal testing of ResNet50 as a part of initial model exploration. Future research in this area could benefit from testing a range of different CNN architectures (i.e., VGG, Xception, or ResNeXt) to determine their relative efficacy in feature detection.

Third, the ACLED database used to construct our imagery data set is drawn primarily from news sources (ACLED 2023). These come with some inherent challenges and limitations. If riots and protests are occurring in regions that traditional news sources are not reporting about, the events are not likely to populate the ACLED database. Further, the nature of civil unrest is sometimes difficult to delineate with clear definitions, and different news organizations may cover a protest in conflicting ways—for example, a protest that is met with armed government resistance (ACLED 2023). These challenges are not likely to be overcome in the near term, but are notable as they may impact the results presented in this study.

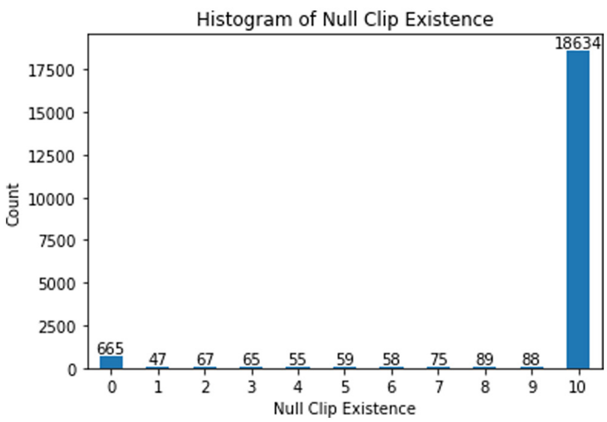

FIGURE 13 | Distribution of null clips from the full 19,902 images downloaded. Instances where <10 clips were taken are primarily due to the amount of urban area available in the satellite image. There were three additional locations that were eventually able to provide 10 null clips, but not included before the dataset was finalized with 18,631 locations at training time. 19

5.2 | Conclusions

In this work, we construct a data set consisting of 204,941 satellite images of riots and protests across the world. After subseting the images into two classes of riots and non-riots, we train a ResNet18 to identify which images are from locations associated with a riot. When fine-tuned, our model achieves an accuracy of over 97%, suggesting that satellite imagery has information of relevance and value to estimating the location of riot events. This was true across a wide range of different tests and permutations of the data. We further provide some initial exploration into the explainability of this model, leveraging ScoreCAM to identify features the model is leveraging in the classification task. This research has suggested a number of future directions, which may be valuable to the research community. First, given the promise of spatially predicting where conflict is likely to occur, research into the temporal domain using satellite data may be of value. Second, we note the relative lack of explainablity techniques applied to satellite imagery, and the importance of additional future research into that domain. Third, we anticipate future efforts can explore implementing this technique on full satellite images to localize protest and riot predictions. Finally, we note that future work that explores new model architectures, or integrating multiple data sources for conflict information, could provide high value.

6 | Appendix

6.1 | Deduplication Tests

In this section, we present a test that controls for both class imbalance and geographic bias in our data. Our methodology leverages a large set of training data, specifically relying on an arbitrary 10:1 7

TABLE 10 | All of the models in this table were tested with 100 random locations (100 riot clips and 1000 null clips). 8

Metric	A1	A2	B1	B2	C1	C2
Test accuracy (%)	91.59	91.12	90.65	93.93	93.46	86.45
True positives	0	0	0	0	0	0
False positives	0	0	4	0	0	0
False negatives	18	19	16	13	14	29
True negatives	196	195	194	201	200	185
Precision (%)	0.00	0.00	0.00	0.00	0.00	0.00
Recall (%)	0.00	0.00	0.00	0.00	0.00	0.00
*F*1 score (%)	0.00	0.00	0.00	0.00	0.00	0.00
Learning rate	1e−06	1e−06	1e−06	1e−06	1e−06	1e−06
Freeze layers	0	0	5	5	10	10
Drop out pair	(0, 0)	(0.1, 0.05)	(0, 0)	(0.1, 0.05)	(0, 0)	(0.1, 0.05)
L2 weight decay	0	0	0	0	0	0

Note: In this table, all models used a learning rate of 1e−06. Models froze either none of the ResNet layers (A1, A2), the first 5 layers (B1, B2), or the first 10 layers (C1, C2). Between the first two and the second two layers, none of the connections were dropped (A1, B1, C1), or 10% and 5% were dropped (A2, B2, C2).

TABLE 11 | All of the models in this table were tested with 100 random locations (100 riot clips and 1000 null clips).

Metric	D1	D2	E1	E2	F1	F2
Test accuracy (%)	88.78	86.92	91.12	91.59	88.32	88.78
True positives	5	1	7	5	0	0
False positives	7	13	4	6	0	0
False negatives	17	15	15	12	25	24
True negatives	185	185	188	191	189	190
Precision (%)	41.67	7.14	63.64	45.45	0.00	0.00
Recall (%)	22.73	6.25	31.82	29.41	0.00	0.00
*F*1 score (%)	29.41	6.67	42.42	35.71	0.00	0.00
Learning rate	1e−05	1e−05	1e−05	1e−05	1e−05	1e−05
Freeze layers	0	0	5	5	10	10
Drop out pair	(0, 0)	(0.1, 0.05)	(0, 0)	(0.1, 0.05)	(0, 0)	(0.1, 0.05)
L2 weight decay	0	0	0	0	0	0

Note: In this table, all models used a learning rate of 1e−05. Models froze either none of the ResNet layers (D1, D2), the first 5 layers (E1, E2), or the first 10 layers (F1, F2). Between the first two and the second two layers, none of the connections were dropped (D1, E1, F1), or 10% and 5% were dropped (D2, E2, F2).

TABLE 12 | Model performance metrics for configurations 1 to 6 with learning rate of 1e−05, with variations in L2 weight decay, freeze layer, and dropout pair settings.

Metric	Config 1	Config 2	Config 3	Config 4	Config 5	Config 6
Test accuracy (%)	91.80	91.80	91.27	90.69	92.52	91.85
True positives	62	39	43	77	25	78
False positives	54	11	41	65	7	62
False negatives	116	159	140	128	148	107
True negatives	1841	1864	1849	1803	1893	1826
Precision	0.5345	0.7800	0.5119	0.5423	0.7812	0.5571
Recall	0.3483	0.1970	0.2350	0.3756	0.1445	0.4216
*F*1 score	0.4218	0.3145	0.3221	0.4438	0.2439	0.4800
Learning rate			1e−05			
L2 weight decay		0.1			0.01	
Freeze layer	0	0	0	5	5	5
Dropout pair	(0, 0)	(0.1, 0.05)	(0.5, 0.1)	(0, 0)	(0.1, 0.05)	(0.5, 0.1)

ratio of 10 null cases (no conflict event) to 1 positive case (a location where a conflict occurred). Furthermore, some geographic locations are in the database multiple times—that is, there may have been multiple protests at the same geographic location, even if they are on different dates (see Table 3). This results in both class imbalance (10 null cases for every 1 positive case), and geographic biases from where we draw our events. The class imbalance will potentially inflate accuracy scores, given a 10 to 1 ratio of null clips to riot clips—that is, an untrained model could simply predict null for all images, and achieve an accuracy of 90.9%. Additionally, with repeated locations, the model will see the riot clip locations multiple times (i.e., even when each satellite scene has unique spatial information as it is drawn from a different date, the 1-km box centered on the latitude and longitude of the neighborhood will be the same). This might allow our network to learn the specifics of a location, and over-fit to particular locations, instead of learning what features in urban areas predict riots and protests. Therefore we construct a limited data set to control for these issues.

To test if these attributes of our data result in bias, we construct a new dataset that limits the data to a single riot image (1089 1-km boxes) and a single non-riot image (1089 1-km boxes) per location. This means that our model is only able to analyze a riot location a single time during training, regardless of how frequently riots might happen at that location. This should be a much harder training task for the model, with far less data available (2178 images in total; these 2178 images represent roughly 1% of the data available for training in the full data set of 204,941 images). Under these

TABLE 13 | Model performance metrics for configurations 7 to 12 with learning rate of 1e−05, transitioning from L2 weight decay settings of 0.01 to 0.001, including variations in freeze layer and dropout pair settings.

Metric	Config 7	Config 8	Config 9	Config 10	Config 11	Config 12
Test accuracy (%)	91.51	89.77	92.91	92.33	90.16	92.76
True positives	85	86	54	86	92	80
False positives	64	106	34	32	101	49
False negatives	112	106	113	127	103	101
True negatives	1812	1775	1872	1828	1777	1843
Precision	0.5705	0.4479	0.6136	0.7288	0.4767	0.6202
Recall	0.4315	0.4479	0.3234	0.4038	0.4718	0.4420
*F*1 score	0.4913	0.4479	0.4235	0.5196	0.4742	0.5161
Learning rate			1e−05			
L2 weight decay		0.01			0.001	
Freeze layer	0	0	0	5	5	5
Dropout pair	(0, 0)	(0.1, 0.05)	(0.5, 0.1)	(0, 0)	(0.1, 0.05)	(0.5, 0.1)

constraints, the maximum classification accuracy we observed was 67.37% 9. Of note, the recall scores for our full data set and limited data set were very similar (75.18% and 77.0%, respectively), despite the different size and scope of the training data.

These results suggest that—even under extremely challenging, small-N circumstances—deep learning models can still identify meaningful features that are correlated with protest and riot events from satellite imagery.

6.2 | All Results

While we focus on our best performing models throughout this piece, there were a number of additional permutations and tests we perform while identifying the best modeling strategies, which we present here. We begin a grid search across select hyperparameters, using a small test set of 100 random samples from our full data set. Initially we are concerned with narrowing down the selection of the best performing learning rates, freezing layers of the ResNet, and dropping out connections between our fully connected layers. The results of a sample of these are shown in Tables 10 and 11.

After the initial grid search, we increase the size of data set to 1000 locations (1000 riot clips, and 10,000 null clips) (Table 12). We also refine the hyperparameter grid search space. Our best performing model referred to as Model C in Table 4, is Config 10 in Table 13. Config 10 has the highest F1 score across these grid search results, reflecting the best balance between Precision and Recall. Due to this strong performance, these parameters are used to train with the full dataset.

Acknowledgments

This work utilized data made available through the NASA Commercial Smallsat Data Acquisition (CSDA) program. Funding provided by National Science Foundation, Award Number: 2317591. The authors acknowledge William & Mary Research Computing for providing computational resources and/or technical support that have contributed to the results reported within this paper.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from Planet Labs PBC. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of Planet Labs PBC.

Endnotes

[1] For example, some riots are known to have occurred in Beriut, while others occurred within neighborhoods in Beriut. There are 12 neighborhoods listed within some of the ALCED entries for Beriut (Ras Beirut, Port, Mazraa, Achrafieh, Mousseitbeh, Saifi, Minet El Hosn, Rmeil, Bachoura, Medawar, Ain Mreisseh, and Zokak El Blat). These neighborhood specific entries have neighborhood specific latitudes and longitudes, and we use these neighborhood specific events to construct our data set.

[2] Training was performed using pyTorch on 8 RTX 6000 NVIDIA GPUs. On average, models trained using the hyperparameter dataset took approximately 6.5 h to complete 40 epochs; our full model across all images took 321 h for 100 epochs.

[3] It is important to note that our data set is constructed in a manner that would result in relatively high test accuracy. We have one riot and ten null riot clips per satellite scene. This means that if our model predicted no riot for every clipped image, the model would be correct 90.9% of the time. Even given imbalance in the data set, our trained model achieves better results, accurately predicting riots and null riots over 97% of the time. Further explorations of the value of the model in the context of imbalance are described in Section 6.1 of the appendix.

[4] Data were randomly selected from data used to train the model in appendix Section 6.1.

References

ACLED. 2023. Codebook, 2023. www.acleddata.com.

Alo, C. A., and R. G. Pontius Jr. 2008. "Identifying Systematic Land-Cover Transitions Using Remote Sensing and GIS: The Fate of Forests Inside and Outside Protected Areas of Southwestern Ghana." *Environment and Planning B: Planning and Design* 35, no. 2: 280–295.

Alsaedi, N., P. Burnap, and O. Rana. 2015. "Identifying Disruptive Events From Social Media to Enhance Situational Awareness." In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 934–941.

Alsaedi, N., P. Burnap, and O. Rana. 2017. "Can We Predict a Riot? Disruptive Event Detection Using Twitter." *ACM Transactions on Internet Technology (TOIT)* 17, no. 2: 1–26.

Anderson, C. J., and S. M. Mendes. 2006. "Learning to Lose: Election Outcomes, Democratic Experience and Political Protest Potential." *British Journal of Political Science* 36, no. 1: 91–111.

Andronikidou, A., and I. Kovras. 2012. "Cultures of Rioting and Anti-Systemic Politics in Southern Europe." *West European Politics* 35, no. 4: 707–725.

Aung, T. S., I. Overland, R. Vakulchuk, and Y. Xie. 2021. "Using Satellite Data and Machine Learning to Study Conflict-Induced Environmental and Socioeconomic Destruction in Data-Poor Conflict Areas: The Case of the Rakhine Conflict." *Environmental Research Communications* 3, no. 2: 025005.

Becker, H., M. Naaman, and L. Gravano. 2011. "Beyond Trending Topics: Real-World Event Identification on Twitter." In *Proceedings of the International AAAI Conference on web and Social Media*, vol. 5, 438–441.

Begum, S. 2018. "Changing Scenarios of Public Open Space in a British Colonial City: The Case of the Ramna Area, Dhaka." *Nakhara: Journal of Environmental Design and Planning* 14: 39–56.

Bencsik, P. 2018. "The Non-Financial Costs of Violent Public Disturbances: Emotional Responses to the 2011 Riots in England."
9 *Journal of Housing Economics* 40: 73–82.

Berazneva, J., and D. R. Lee. 2013. "Explaining the African Food Riots of 2007–2008: An Empirical Analysis." *Food Policy* 39: 28–39.

Bharti, N., and A. J. Tatem. 2018. "Fluctuations in Anthropogenic Nighttime Lights From Satellite Imagery for Five Cities in Niger and Nigeria." *Scientific Data* 5, no. 1: 1–9.

Bibault, J. E., M. Bassenne, H. Ren, and L. Xing. 2020. "Deep Learning Prediction of Cancer Prevalence From Satellite Imagery." *Cancers* 12, no. 12: 3844.

Bonnasse-Gahot, L., H. Berestycki, M. A. Depuiset, et al. 2018. "Epidemiological Modelling of the 2005 French Riots: A Spreading Wave and the Role of Contagion." *Scientific Reports* 8, no. 1: 107.

Bressem, K. K., L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek. 2020. "Comparing Different Deep Learning Architectures for Classification of Chest Radiographs." *Scientific Reports* 10, no. 1: 13590.

Brewer, E., J. Lin, P. Kemper, J. Hennin, and D. Runfola. 2021. "Predicting Road Quality Using High Resolution Satellite Imagery: A Transfer Learning Approach." *PLoS ONE* 16, no. 7: e0253370.

Brewer, E., Z. Lv, and D. Runfola. 2023. "Tracking the Industrial Growth of Modern China With High-Resolution Panchromatic Imagery: A Sequential Convolutional Approach." *arXiv preprint arXiv:230109620*.

Cadena, J., G. Korkmaz, C. J. Kuhlman, A. Marathe, N. Ramakrishnan, and A. Vullikanti. 2015. "Forecasting Social Unrest Using Activity Cascades." *PLoS ONE* 10, no. 6: e0128879.

Carranza-Garca, M., J. Garca-Gutiérrez, and J. C. Riquelme. 2019. "A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks." *Remote Sensing* 11, no. 3: 274.

Chauhan, R., K. K. Ghanshala, and R. Joshi. 2018. "Convolutional Neural Network (CNN) for Image Detection and Recognition." In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 278–282. IEEE.

Ciorciari, J. D., and J. C. Weiss. 2016. "Nationalist Protests, Government Responses, and the Risk of Escalation in Interstate Disputes." *Security Studies* 25, no. 3: 546–583.

Compton, R., C. Lee, T. C. Lu, L. De Silva, and M. Macy. 2013. "Detecting Future Social Unrest in Unprocessed Twitter Data: "Emerging Phenomena and Big Data"." In *2013 IEEE International Conference on Intelligence and Security Informatics IEEE*, 56–60.

Daudt, R. C., B. Le Saux, A. Boulch, and Y. Gousseau. 2018. "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks." In *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium IEEE*, 2115–2118.

Davies, T. P., H. M. Fry, A. G. Wilson, and S. R. Bishop. 2013. "A Mathematical Model of the London Riots and Their Policing." *Scientific Reports* 3, no. 1: 1303.

Davis, J., and M. Goadrich. 2006. "The Relationship Between Precision-Recall and ROC Curves." In *Proceedings of the 23rd International Conference on Machine Learning*, 233–240.

Eckstein, S. 2001. "Power and Popular Protest in Latin." *Power and* 10
Popular Protest: Latin American Social Movements: 1.

El Warea, M., R. Sasso, R. Bachir, and M. El Sayed. 2019. "Riots in Beirut: Description of the Impact of a New Type of Mass Casualty Event on the Emergency System in Lebanon." *Disaster Medicine and Public Health Preparedness* 13, no. 5–6: 849–852.

European Commission and Statistical Office of the European Union. 2021. *Applying the Degree of Urbanisation—A Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons—2021 Edition*. Publications Office of the European Union. https://doi.org/10.2785/706535.

Filchenkov, A. A., A. A. Azarov, and M. V. Abramov. 2014. "What Is More Predictable in Social Media: Election Outcome or Protest Action?" In *Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia*, 157–161.

Foody, G. M., and A. Mathur. 2004. "Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for SVM Classification." *Remote Sensing of Environment* 93, no. 1–2: 107–117.

Fox, S., and A. Bell. 2016. "Urban Geography and Protest Mobilization in Africa." *Political Geography* 53: 54–64.

Gehlbach, S. 2010. "Reflections on Putin and the Media." *Post-Soviet Affairs* 26, no. 1: 77–87.

Gong, Q., and R. Batta. 2007. "Allocation and Reallocation of Ambulances to Casualty Clusters in a Disaster Relief Operation." *IIE Transactions* 39, no. 1: 27–39.

González-Bailón, S., J. Borge-Holthoefer, A. Rivero, and Y. Moreno. 2011. "The Dynamics of Protest Recruitment Through an Online Network." *Scientific Reports* 1, no. 1: 1–7.

Goodman, S., A. BenYishay, and D. Runfola. 2021. "A Convolutional Neural Network Approach to Predict Non-permissive Environments From Moderate-Resolution Imagery." *Transactions in GIS* 25, no. 2: 674–691.

Gorban, A. N., E. M. Mirkes, and I. Y. Tyukin. 2020. "How Deep Should Be the Depth of Convolutional Neural Networks: A Backyard dog Case Study." *Cognitive Computation* 12: 388–397.

Greer, C., and E. McLaughlin. 2010. "We Predict a Riot? Public Order Policing, New Media Environments and the Rise of the Citizen Journalist." *The British Journal of Criminology* 50, no. 6: 1041–1059.

He, C., Z. Liu, S. Gou, Q. Zhang, J. Zhang, and L. Xu. 2019. "Detecting Global Urban Expansion Over the Last Three Decades Using a Fully Convolutional Network." *Environmental Research Letters* 14, no. 3: 034008.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Helber, P., B. Bischke, A. Dengel, and D. Borth. 2019. "Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12, no. 7: 2217–2226.

Heslin, A. 2021. "Riots and Resources: How Food Access Affects Collective Violence." *Journal of Peace Research* 58, no. 2: 199–214.

Ho, Y., and S. Wookey. 2019. "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling." *IEEE Access* 8: 4806–4813.

Hu, W., J. H. Patel, Z. A. Robert, et al. 2019. "Mapping Missing Population in Rural India: A Deep Learning Approach With Satellite Imagery." In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 353–359.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353, no. 6301: 790–794.

Joya, A. 2011. "The Egyptian Revolution: Crisis of Neoliberalism and the Potential for Democratic Politics." *Review of African Political Economy* 38, no. 129: 367–386.

Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv preprint arXiv:14126980*.

Klein, G. R., and P. M. Regan. 2018. "Dynamics of Political Protests." *International Organization* 72, no. 2: 485–521.

Korolov, R., D. Lu, J. Wang, et al. 2016. "On Predicting Social Unrest Using Social Media." In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 89–95. IEEE.

Krizhevsky, A., and G. Hinton. 2009. *Learning Multiple Layers of Features from Tiny Images*. http://www.cs.toronto.edu/kriz/cifar.html.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. "ImageNet Classification With Deep Convolutional Neural Networks." *Communications of the ACM* 60, no. 6: 84–90.

Kussul, N., M. Lavreniuk, S. Skakun, and A. Shelestov. 2017. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14, no. 5: 778–782.

Leclerc, M., R. Tharmarasa, M. C. Florea, A. C. Boury-Brisset, T. Kirubarajan, and N. Duclos-Hindié. 2018. "Ship Classification Using Deep Learning Techniques for Maritime Target Tracking." In *21st International Conference on Information Fusion (FUSION)*, 737–744. IEEE.

Löwenheim, O. 2007. "The Responsibility to Responsibilize: Foreign Offices and the Issuing of Travel Warnings." *International Political Sociology* 1, no. 3: 203–221.

Lv, Z., K. Nunez, E. Brewer, and D. Runfola. 2024. "Mapping the Tidal Marshes of Coastal Virginia: A Hierarchical Transfer Learning Approach." *GIScience & Remote Sensing* 61, no. 1: 2287291.

Nabiee, S., M. Harding, J. Hersh, and N. Bagherzadeh. 2022. "Hybrid U-Net: Semantic Segmentation of High-Resolution Satellite Images to Detect War Destruction." *Machine Learning with Applications* 9: 100381.

Najjar, A., S. Kaneko, Y. Miyanaga. 2018. "Crime Mapping From Satellite Imagery via Deep Learning." *arXiv preprint arXiv:181206764*.

OpenStreetMap Contributors. 2024. *OpenStreetMap*. https://www.openstreetmap.org.

Pain, P., and E. Korin. 2021. "'Everything Is Dimming Out, Little by Little': Examining Self-Censorship Among Venezuelan Journalists." *Communication Research and Practice* 7, no. 1: 71–88.

Patel, K., C. Bhatt, and P. L. Mazzeo. 2022. "Deep Learning-Based Automatic Detection of Ships: An Experimental Study Using Satellite Images." *Journal of Imaging* 8, no. 7: 182.

Pearce, T., A. Brintrup, and J. Zhu. 2021. "Understanding Softmax Confidence and Uncertainty." *arXiv preprint arXiv:210604972*.

Petrović, S., M. Osborne, and V. Lavrenko. 2010. "Streaming First Story Detection With Application to Twitter." In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 181–189.

Phillips, L., C. Dowling, K. Shaffer, N. Hodas, and S. Volkova. 2017. "Using Social Media to Predict the Future: A Systematic Literature Review." *arXiv preprint arXiv:170606134*.

Planet Team. 2023a. *Planet Application Program Interface: In Space for Life on Earth*. https://api.planet.com.

Planet Team. 2023b. *PlanetScope: Constellation and Sensor Overview*. https://developers.planet.com/docs/data/planetscope/.

Pond, P., and J. Lewis. 2019. "Riots and Twitter: Connective Politics, Social Media and Framing Discourses in the Digital Public Sphere." *Information, Communication & Society* 22, no. 2: 213–231.

Purbrick, M. 2019. "A Report of the 2019 Hong Kong Protests." *Asian Affairs* 50, no. 4: 465–487.

Rahimi, B. 2015. "Censorship and the Islamic Republic: Two Modes of Regulatory Measures for Media in Iran." *The Middle East Journal* 69, no. 3: 358–378.

Raleigh, C., R. Kishi, and A. Linke. 2023. "Political Instability Patterns Are Obscured by Conflict Dataset Scope Conditions, Sources, and Coding Choices." *Humanities and Social Sciences Communications* 10: 74.

Renaud, M., R. Korolov, D. Mendonça, and W. Wallace. 2019. "Social Network Structure as a Predictor of Social Behavior: The Case of Protest in the 2016 Us Presidential Election." In *Recent Developments in Data Science and Intelligent Analysis of Information: Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information, June 4–7, 2018*, 267–278. Kyiv: Ukraine Springer.

Rogan, J., and D. Chen. 2004. "Remote Sensing Technology for Mapping and Monitoring Land-Cover and Land-Use Change." *Progress in Planning* 61, no. 4: 301–325.

Runfola, D., A. Anderson, H. Baier, et al. 2020. "geoBoundaries: A Global Database of Political Administrative Boundaries." *PLoS ONE* 15, no. 4: e0231866.

Runfola, D., H. Baier, L. Mills, M. Naughton-Rockwell, and A. Stefanidis. 2022. "Deep Learning Fusion of Satellite and Social Information to Estimate Human Migratory Flows." *Transactions in GIS* 26, no. 6: 2495–2518.

Runfola, D., A. Stefanidis, and H. Baier. 2022. "Using Satellite Data and Deep Learning to Estimate Educational Outcomes in Data-Sparse Environments." *Remote Sensing Letters* 13, no. 1: 87–97.

Runfola, D., A. Stefanidis, Z. Lv, J. O'Brien, and H. Baier. 2024. "A Multi-Glimpse Deep Learning Architecture to Estimate Socioeconomic Census Metrics in the Context of Extreme Scope Variance." *International Journal of Geographical Information Science* 38: 1–25.

Rüttgers, M., S. Lee, S. Jeon, and D. You. 2019. "Prediction of a Typhoon Track Using a Generative Adversarial Network and Satellite Images." *Scientific Reports* 9, no. 1: 6057.

Schiavina, M., M. Melchiorri, and M. Pesaresi. 2023. *GHS-SMOD R2023A—GHS Settlement Layers, Application of the Degree of*

Urbanisation Methodology (Stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, Multitemporal (1975–2030). European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/a0df7a6f-49de-46ea-9bde-563437a6e2ba, https://doi.org/10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA.

11 Shakur, T., I. Islam, and J. Masood. 2010. "What Culture, Whose Space and Which Technology? The Contested Transformation and the Changing Historic Built Environments of South Asia." *ArchNet-IJAR* 4, no. 1.

Shin, H. C., H. R. Roth, M. Gao, et al. 2016. "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning." *IEEE Transactions on Medical Imaging* 35, no. 5: 1285–1298.

Simonyan, K., and A. Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv preprint arXiv:14091556*.

Snow, D. A., R. Vliegenthart, and C. Corrigall-Brown. 2007. "Framing the French Riots: A Comparative Study of Frame Variation." *Social Forces* 86, no. 2: 385–415.

Stow, D., Y. Hamada, L. Coulter, and Z. Anguelova. 2008. "Monitoring Shrubland Habitat Changes Through Object-Based Change Identification With Airborne Multispectral Imagery." *Remote Sensing of Environment* 112, no. 3: 1051–1061.

Subramanya, A., S. Srinivas, and R. V. Babu. 2017. "Confidence Estimation in Deep Neural Networks via Density Modelling." *arXiv preprint arXiv:170707013*.

Szegedy, C., W. Liu, Y. Jia, et al. 2015. "Going Deeper With Convolutions." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.

Tai, Q. 2014. "China's Media Censorship: A Dynamic and Diversified Regime." *Journal of East Asian Studies* 14, no. 2: 185–210.

U.S. Constitution. 1791. Amend. I.

Voulodimos, A., N. Doulamis, A. Doulamis, and E. Protopapadakis. 2018. "Deep Learning for Computer Vision: A Brief Review." *Computational Intelligence and Neuroscience* 2018: 7068349.

Wang, H., Z. Wang, M. Du, et al. 2020. "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 24–25.

Wu, C., and M. S. Gerber. 2017. "Forecasting Civil Unrest Using Social Media and Protest Participation Theory." *IEEE Transactions on Computational Social Systems* 5, no. 1: 82–94.

Zhang, C., S. Wei, S. Ji, and M. Lu. 2019. "Detecting Large-Scale Urban Land Cover Changes From Very High Resolution Remote Sensing Images Using CNN-Based Classification." *ISPRS International Journal of Geo-Information* 8, no. 4: 189.

Zhang, L., L. Zhang, and B. Du. 2016. "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the art." *IEEE Geoscience and Remote Sensing Magazine* 4, no. 2: 22–40.

Zhang, P., Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang. 2018. "Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery." *Sensors* 18, no. 11: 3717.

Zhou, W., and A. Troy. 2008. "An Object-Oriented Approach for Analysing and Characterizing Urban Landscape at the Parcel Level." *International Journal of Remote Sensing* 29, no. 11: 3119–3135.