



## Article

# Spatiotemporal Prediction of Conflict Fatality Risk Using Convolutional Neural Networks and Satellite Imagery

Seth Goodman <sup>1,\*</sup> , Ariel BenYishay <sup>1,2</sup> and Daniel Runfola <sup>3,4</sup>

<sup>1</sup> AidData, Global Research Institute, William & Mary, Williamsburg, VA 23185, USA; abenyishay@aiddata.wm.edu

<sup>2</sup> Department of Economics, William & Mary, Williamsburg, VA 23185, USA

<sup>3</sup> Department of Applied Science, William & Mary, Williamsburg, VA 23185, USA; danr@wm.edu

<sup>4</sup> Geospatial Evaluation and Observation Lab, Data Science Program, William & Mary, Williamsburg, VA 23185, USA

\* Correspondence: smgoodman@wm.edu

**Abstract:** As both satellite imagery and image-based machine learning methods continue to improve and become more accessible, they are being utilized in an increasing number of sectors and applications. Recent applications using convolutional neural networks (CNNs) and satellite imagery include estimating socioeconomic and development indicators such as poverty, road quality, and conflict. This article builds on existing work leveraging satellite imagery and machine learning for estimation or prediction, to explore the potential to extend these methods temporally. Using Landsat 8 imagery and data from the Armed Conflict Location & Event Data Project (ACLED) we produce subnational predictions of the risk of conflict fatalities in Nigeria during 2015, 2017, and 2019 using distinct models trained on both yearly and six-month windows of data from the preceding year. We find that predictions at conflict sites leveraging imagery from the preceding year for training can predict conflict fatalities in the following year with an area under the receiver operating characteristic curve (AUC) of over 75% on average. While models consistently outperform a baseline comparison, and performance in individual periods can be strong (AUC > 80%), changes based on ground conditions such as the geographic scope of conflict can degrade performance in subsequent periods. In addition, we find that training models using an entire year of data slightly outperform models using only six months of data. Overall, the findings suggest CNN-based methods are moderately effective at detecting features in Landsat satellite imagery associated with the risk of fatalities from conflict events across time periods.

**Keywords:** machine learning; convolutional neural networks; GeoAI; Landsat; satellite imagery; conflict



**Citation:** Goodman, S.; BenYishay, A.; Runfola, D. Spatiotemporal Prediction of Conflict Fatality Risk Using Convolutional Neural Networks and Satellite Imagery. *Remote Sens.* **2024**, *16*, 3411. <https://doi.org/10.3390/rs16183411>

Academic Editor: Silvia Liberata Ullo

Received: 15 July 2024

Revised: 30 August 2024

Accepted: 10 September 2024

Published: 13 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Conflict in Sub-Saharan nations such as Nigeria can have many drivers such as ethnic and religious strife (including terrorist groups), climate change and drought, violence against civilians, and general corruption [1]. The resulting conflicts can range in form from violent clashes, explosions, or attacks on civilians to one-sided protests, riots, or strategic developments [2,3]. Methods to forecast conflict and the risk of violence (i.e., fatalities) can provide utility to humanitarian organizations whose efforts are needed to support people in conflict-impacted areas, yet are also impeded by ongoing conflict (As an example, see <https://www.usaid.gov/humanitarian-assistance/nigeria> for an overview of USAID's work in Nigeria. Accessed on 1 July 2024). Advances in the use of satellite imagery and deep learning provide opportunities to remotely develop forecasts of future deadly conflict that can support humanitarian efforts in conflict-affected areas.

Human and environmental systems can be observed with satellite imagery in many ways that may be directly relevant for conflict, such as detecting drought conditions or

agricultural encroachment into pastoral lands [4]. The severity of conflict events in particular, in terms of whether the event results in fatalities, can be linked to observable conditions. Data around conflict event sites in Nigeria reveals that the median population of non-fatal event locations is nearly triple that of fatal event locations, and non-fatal event areas are on average 40% urban while fatal event areas are on average only 20% urban (Based on population [5], land cover [6], and precipitation [7] data within a 3 km buffer around conflict event sites in Nigeria from 2015, 2017, and 2019 from the Armed Conflict Location & Event Data Project (ACLED) [3]. Population, land cover, and precipitation data prepared following methods in [8]). In addition, the median coverage of rainfed cropland in fatal event locations is nearly 40% yet only 7.5% in non-fatal locations. Irrigated cropland however is only present in around 1% of all conflict event areas. Given the increased occurrence of fatal events in areas with rainfed cropland, it is notable that average rainfall (See World Bank Climate Change Knowledge Portal <https://climateknowledgeportal.worldbank.org/country/nigeria/climate-data-historical>. Accessed on 1 July 2024) at non-fatal event locations is in line with national rainfall averages, while fatal event locations are approximately 15% lower (based on fatal and non-fatal event locations where at least 25% of the nearby area was rainfed cropland). These geospatial characteristics provide meaningful insight into potential information (e.g., populous urban areas, dry cropland) that can be detected from even moderate-resolution satellite imagery over time in order to forecast fatal conflict events.

Convolutional neural networks (ConvNets or CNNs) have been a focus of recent work that is connecting advancements in geospatial AI with research focused on human development [9]. CNNs are a class of neural networks well-suited for a variety of computer vision tasks that require detecting spatial patterns or features within images [10–13]. Numerous implementations of CNNs have been developed, dealing with image classification, object detection, and more [14].

Geospatial applications of CNNs include using spectral signatures and phenological trends captured by satellite imagery to perform pixel and image classification [15,16], land use classification [17], and the detection of objects such as vehicles or buildings in satellite imagery [18,19]. In recent years, researchers focused on development economics have leveraged these methods to improve upon existing methods to fill the gaps in sparse datasets such as household surveys [20]. By using abundant remotely sensed data sources (satellite imagery) as inputs, CNNs are trained using existing sparse data (e.g., poverty metrics from household surveys) as labels. The resulting trained network can be used to predict the sparse data in additional geographical areas based on satellite imagery alone.

Building on work within the geospatial community [21], Jean et al. showed that CNN-based estimates of poverty offered improvements on previous methods of using nighttime lights to estimate poverty [20]. Over the past several years there have been a number of additional contributions to the literature utilizing CNNs and satellite imagery to predict human development indicators [9]. Much of this work has focused on exploring effectiveness in new geographic contexts, along with advancements based on incorporating new data or improvements to machine learning algorithms (e.g., [22]). Poverty estimates have been produced using a variety of available satellite imagery across most of Africa [23], as well as Mexico [24], Bangladesh & India [25], and Sri Lanka [26]. In addition, literature addressing the usage of CNN and satellite image-based methods to predict other development indicators has explored infrastructure quality assessments [27–29], crop yield [30], education & health-related metrics [31,32], population mapping [33,34], and conflict [35]. The potential for application, improvement, and operationalization of these methods across varying sectors remains a topic of discussion [9,36].

While much of this work has shown promising results, the literature has highlighted some limitations [32] as well as the abundance of opportunities for future research to advance the practical application of these methods [9,35,36]. One important area yet to be fully addressed when predicting development indicators using CNNs and satellite imagery is the capacity for accurate predictions across varying and discrete time periods.

Producing time series estimates is necessary for many applications [9] such as geospatial impact evaluations that aim to leverage geospatial data over time to study the impact of development projects and other interventions [37–42]. To date, the literature has largely focused on using imagery that is a combination of a broad temporal range, or imagery tied to a single, specific period based on available training data, with limited evaluation of temporal transferability of models [9]. Recent work has indicated that existing approaches with strong cross-sectional performance may not perform as well across multiple points in time [43].

More broadly, a range of existing literature has highlighted the capability of deep learning methods such as CNNs to leverage time series data [44,45]. The Time Series Land Cover Challenge (TiSeLaC) [46] is one effort that engages with these problems and has led to the advancement of time series CNNs for classifying land cover types using satellite imagery [47]. Notably, [48], the winners of the 2017 TiSeLaC, were able to accurately classify multispectral time series Landsat 8 imagery over 99% of the time, using a combination of MLPs (multilayer perceptrons) and CNNs [48]. Related time series crop classification efforts have also led to the development of novel neural architectures [49]. Numerous applications in recent literature provide opportunities to explore whether existing time series approaches for deep learning can be extended to using CNNs to estimate development indicators [20,23,24,26,36]. In particular, recent work demonstrating that for a discrete temporal window, CNNs can detect features from moderate resolution imagery associated with the risk of conflict fatalities is well suited to be expanded as a time series application [35].

A large body of existing work in the conflict prediction space has explored a range of methods, including machine learning, to forecast conflict in varying geographical contexts [50]. Applications of existing methods have often been limited in either their temporal or spatial precision, scope, or reliability [51,52]. While some approaches incorporate explicitly spatial modeling and/or data [53], many use non-spatial inputs such as text analysis from media content and other sources [54]. Several recent efforts using geospatial data have been a part of the Violence Early Warning System's (ViEWS) prediction competition. The models created have leveraged an array of geospatial data with modeling techniques including logistic regressions, random forests, a variety of neural networks, and other ML algorithms [55,56]. A defining feature of the ViEWS-related efforts is the use of either the country level or coarse (approximately  $50 \times 50$  km) grid cells as the sampling/prediction unit along with predefined features (e.g., demographics, GDP, past conflict data) as input to train models [57–59]. Recent work in the conflict prediction space has sought to improve the resolution of predictions, using  $25 \times 25$  km grid cells, along with predefined features from a range of remote sensing data sources such as land cover, population, and crop data [60]. A critical distinction between these approaches and existing imagery-based deep learning efforts is the use of predefined input features and the resolution of the sampling/prediction units [35].

Building on existing applications of CNNs and satellite imagery in order to produce multi temporal predictions of future fatal conflict has the potential to benefit a range of development, security, and other applications that are currently limited by available data. Estimates of the spatiotemporal risk of conflict fatalities—as opposed to only the spatial location of any future conflict events—are particularly valuable for practicable applications including humanitarian and aid efforts, which require determining whether it is safe for personnel to operate in potentially dangerous regions. Models that estimate the risk of conflict fatalities should a conflict event occur in the future could be used by organizations to assess the risk of deploying personnel to those locations. Finer resolution predictions, as well as depending on only multispectral imagery for model inputs, can provide additional operational value over existing approaches which depend on many inputs and use coarse-resolution spatial grids for predictions.

In this piece, we explicitly focus on the degree to which the fatality of potential future conflict events can be forecast using satellite imagery and deep learning. The research builds

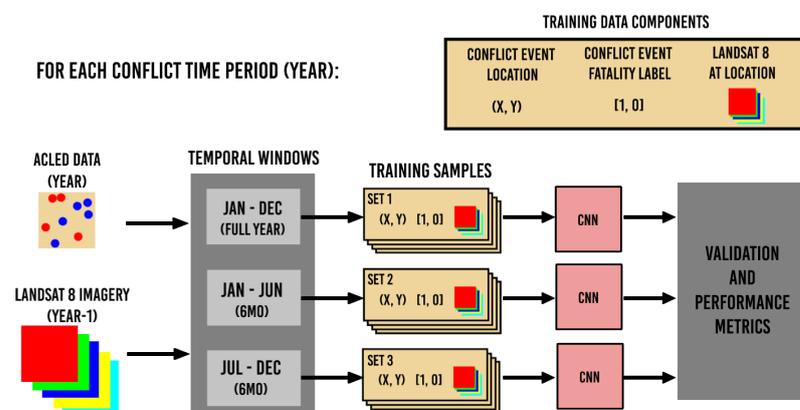
on previous work to estimate conflict fatality in order to further explore the effectiveness and challenges of detecting precursors to fatal conflict from satellite imagery at multiple discrete points in time. We train CNN models using satellite imagery and conflict event data from both yearly and six-month windows from 2014 to 2019 to address two research questions: (RQ1) *Can CNNs detect features from moderate resolution satellite imagery associated with conflict fatality at multiple discrete points in time?* And if so, (RQ2) *how does model performance vary based on the temporal window size used for training?* In doing so, we also aim to contribute to the broader literature by exploring the potential of CNN-based approaches to estimate development indicators using satellite imagery at multiple points in time in order to support practical time series applications.

The paper is structured as follows. Section 2 presents the methods used to prepare Landsat 8 imagery and ACLED data samples for training, as well as the approach for training and validating CNN models. The results of the training and validation are reviewed in Section 3. Finally, Section 4 discusses potential paths forward based on this work and a conclusion.

## 2. Materials and Methods

To assess whether CNNs can detect features from satellite imagery associated with conflict fatality at multiple discrete points in time, and how model performance varies based on the temporal window size used for training, we train a series of ResNet (residual network) [14] CNNs using imagery from Landsat 8 around ACLED conflict event locations which are labeled based on whether a conflict fatality occurred during the associated event. ResNets are a well-established type of CNN that have been used in similar applications leveraging satellite imagery to estimate socioeconomic factors [26]. Landsat 8 imagery provides sufficient resolution to detail a wide range of features including environmental conditions and land use with broad spatial and temporal coverage including Nigeria during the study window. Finally, ACLED offers extensive coverage of conflict events and related fatality counts that can be used to label imagery [35].

Throughout this paper, predicting conflict fatality refers specifically to determining the risk of a fatality *if conflict were to occur at a given location*. Individual CNNs are trained for three distinct time periods of conflict event data (2015, 2017, 2019) using: (1) twelve months of imagery from one full calendar year prior to the conflict event year (January–December), (2) six months of imagery starting twelve months prior to the event year (January–June), or (3) six months of imagery starting six months prior to the event year (July–December) (see Figure 1). Testing both yearly and six-month windows will provide insight into both the general effectiveness of temporal models, as well as whether smaller windows (A) provide sufficient data to train models, and (B) can improve models by enabling the detection of seasonal trends.



**Figure 1.** Overview of the methodology applied for training and validating conflict fatality models at each discrete time period.

The resulting models are assessed based on their ability to accurately classify validation data left out from the original training. Finally, the performance of models from each discrete time period using both yearly and six-month windows of data are compared with one another, as well as with a naive baseline conflict prediction approach based only on the location of fatal conflict from the previous year. Each of these steps is detailed further in this section.

### 2.1. Data Preparation

NASA and USGS have made Landsat 8 imagery free and publicly available since its launch in 2013. Like its predecessor Landsat 7, Landsat 8 imagery captures data for the entire planet every 16 days at 30 m resolution [61]. It consists of eleven spectral bands of imagery, captured using its operational land imager (OLI) and thermal infrared sensor (TIRS) [62] (Table 1).

**Table 1.** Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Bands [35].

Bands	Wavelength (Micrometers)	Resolution (Meters)
Band 1—Coastal aerosol	0.433–0.453	30
Band 2—Blue	0.450–0.515	30
Band 3—Green	0.525–0.600	30
Band 4—Red	0.630–0.680	30
Band 5—Near Infrared (NIR)	0.845–0.885	30
Band 6—Shortwave Infrared (SWIR) 1	1.560–1.660	30
Band 7—Shortwave Infrared (SWIR) 2	2.100–2.300	30
Band 8—Panchromatic	0.500–0.680	15
Band 9—Cirrus	1.360–1.390	30
Band 10—Thermal Infrared (TIRS) 1	10.60–11.20	30 *
Band 11—Thermal Infrared (TIRS) 2	11.50–12.50	30 *

\* Collected at 100 m but resampled to 30 m to match OLI bands.

Individual scenes of Landsat 8 imagery covering Nigeria are acquired using the EarthExplorer platform and bulk download tools [63,64] for 2014, 2016, and 2018 to be paired with conflict data from 2015, 2017, and 2019. These scenes, acquired at sixteen-day intervals, are subsequently mosaiced and aggregated to six-month and yearly windows for each year by taking the average pixel values for each band. Cloud pixels within individual scenes are masked based on the Landsat quality assessment band, ignoring any pixels marked as medium or high confidence of clouds [65].

In this analysis, we seek to construct three different temporal aggregations of Landsat data: two six-month periods and the twelve-month period associated with the year prior to the detected conflict event. The six-month window aggregations are performed based on the first and second half of each calendar year (i.e., January–June and July–December). The twelve-month window aggregated uses all data from the year preceding the year in which the event took place (i.e., January–December). Of the eleven bands available from Landsat 8, all individual and non overlapping bands are utilized (i.e., the panchromatic band is dropped) to include any potentially relevant spatial features across the detectable spectrum. The result of this imagery preprocessing is an aggregate image of the entirety of Nigeria for each of the ten bands for each year at yearly and six-month windows.

Conflict locations and fatality counts are downloaded using ACLED's data export tool [3]. The dataset consists of a date, coordinates, and fatality count for each recorded conflict event. This data is aggregated into yearly and six-month windows for 2015, 2017, and 2019 to be paired with Landsat 8 imagery from the previous year. For each event, the conflict fatality count is converted into a binary value indicating whether or not a fatality occurred (1 or 0, respectively). This binary classification is used to label imagery associated with the conflict location during the training and validation of CNN models.

Imagery from 2014, 2016, and 2018 is paired with conflict fatality data from 2015, 2017, and 2019, respectively, at yearly and six-month windows as listed in Table 2. For each time period and temporal window, samples are prepared for training and validation of the CNNs. Individual samples consist of a location (longitude and latitude) based on a conflict event which can be used to load georeferenced imagery at that location, and a label of whether the associated conflict event had a fatality or not. Each conflict event location is treated as a unique sample as conflict event locations are not consistent across time periods, and therefore do not allow time-series sampling of locations.

**Table 2.** Temporal windows and associated conflict event counts used for training models.

Imagery Window	Conflict Window	Conflict Event Count (Training/Validation Splits)
2014 all	2015 all	1673 (1422/251)
2014 January–June	2015 January–June	997 (847/150)
2014 July–December	2015 July–December	676 (575/101)
2016 all	2017 all	1644 (1397/247)
2016 January–June	2017 January–June	817 (694/123)
2016 July–December	2017 July–December	827 (703/124)
2018 all	2019 all	2218 (1885/333)
2018 January–June	2019 January–June	1214 (1032/182)
2018 July–December	2019 July–December	1004 (853/151)

Imagery for each sample consists of a  $244 \times 244 \times 10$  image stack around the sample location (approximately a  $7 \times 7$  km area), representing each of the ten bands of Landsat 8 imagery detailed in Table 1. In the event that the classes within the sample sets are imbalanced (fatal and non-fatal events), the larger class is trimmed to the size of the smaller class (observations from the larger class are randomly dropped until the classes are balanced; on average, fatal events represent approximately 44% of the data). The balanced class samples are then split into separate datasets to be used during the training and validation stages (85% training, 15% validation), and the samples are duplicated nine times to increase model exposure during training (considerations regarding sampling schemes are included in Section 4). Finally, the process of producing the training and validation dataset is repeated five times to ensure the outcomes are reproducible regardless of randomization during sample generation.

## 2.2. Training

To train CNNs using the sampling data frames generated from Landsat imagery and ACLED conflict data, a ResNet CNN architecture is implemented using PyTorch (v1). This work employs ResNets pre-trained on ImageNet [66] to benefit from transfer learning, which has been shown to be effective in a range of applications within the broader machine learning community including satellite imagery-based application [20,35,67–73]. To utilize transfer learning, a CNN first learns generalizable features—such as basic lines, shapes, and other patterns from a much larger (although potentially unrelated) set of images such as ImageNet. The pre-trained CNN retains that knowledge and can then be fine-tuned using application-specific data (e.g., satellite imagery with conflict fatality labels). This process allows leveraging CNNs for applications that would not have sufficient data to properly train a newly initialized CNN.

The default ResNet architecture is modified to work with ten bands of imagery (compared to the three bands of red, green, and blue found in common images, including those in ImageNet). Following examples in the literature, the additional bands are initialized using an average of the weights for each of the red, green, and blue bands [20,35]. The final layer of the CNN responsible for classification is then replaced in order to utilize the binary classes identifying whether a conflict event was fatal. Five sample data frames for training and validation are created for each time period and temporal window listed in

Table 2 to mitigate the chance that results are influenced by the random generation of an individual sample data frame. For each sample data frame generated, 16 CNNs are trained during a grid search using combinations of the specific ResNet architecture, the learning rate, gamma, and step size as described in Table 3.

**Table 3.** Network parameters adjusted during testing, and associated values.

Parameter	Values
Network	Resnet18, Resnet50
Learning Rate	0.0001, 0.00001
Gamma	0.25, 0.5
Step Size	10, 15

The values tested for each of the hyperparameters, seen in Table 3, are based on results from earlier work which showed minimal variation in performance across hyperparameters Goodman et al. [35]. Despite the lack of hyperparameter impact in previous work, we include a range of hyperparameter values to assess potential variation in optimal hyperparameters across time periods and temporal windows tested in this paper. For all tests, stochastic gradient descent (SGD) optimization is used over 60 epochs with a batch size of 64. Across all time periods, temporal windows, and hyperparameter combinations a total of 720 CNNs are trained.

### 2.3. Validation

For each CNN, a set of validation data (15% of total data for each period, see Table 2) is withheld from the training data, and subsequently used to produce predictions. The result of each prediction is the probability of a fatality at the location of a conflict event.

A confusion matrix defining the relationship of all possible outcomes for a binary classifier is generated to assess the performance of the models. In the confusion matrix, actual and predicted values are either positive (a fatality occurred from a conflict event) or negative (no fatality occurred from a conflict event) as seen in Table 4.

**Table 4.** Confusion matrix definitions.

Actual	Predicted	Classification
Positive	Positive	true positive (tp)
Positive	Negative	false negative (fn)
Negative	Positive	false positive (fp)
Negative	Negative	true negative (tn)

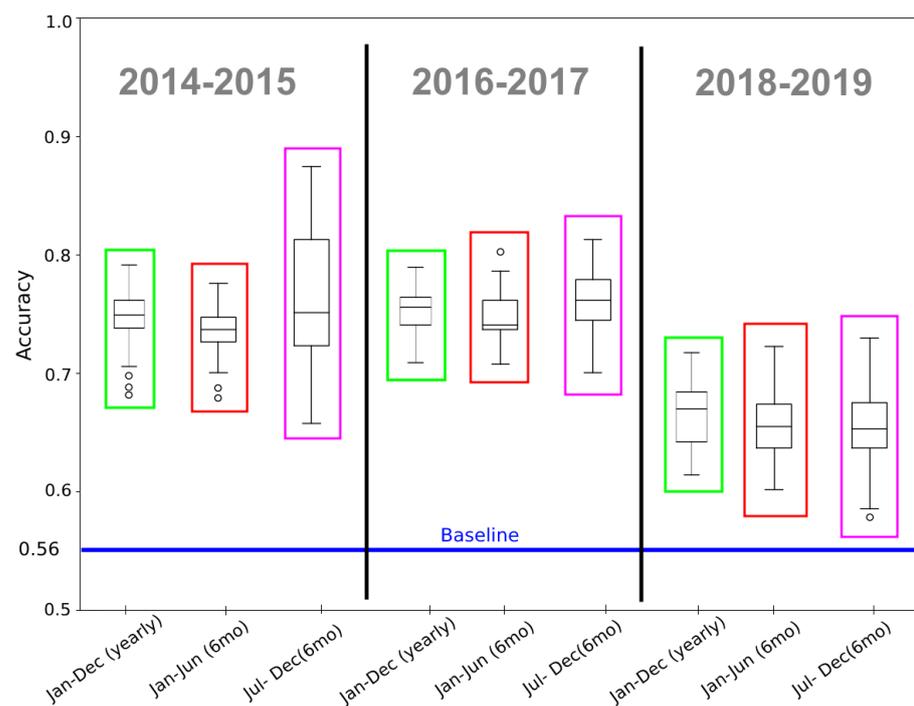
The results from the confusion matrix can be used to produce Receiver Operating Characteristic (ROC) curves, which are a useful visualization to understand model performance [74]. An ROC curve plots the true positive rate on the y-axis and the false positive rate on the x-axis as the threshold for detecting positive cases based on probability is varied from 0 to 1. Using the ROC curve, we then generate the area under the ROC curve (AUC). In an ideal classifier, which has a true positive rate of 1.0 regardless of the false positive rate, the AUC would be 1.0. For a random or “no skill” classifier, the AUC would be 0.5. The value of the AUC for a model can be used to assess overall performance, as well as to identify a threshold value suitable for a specific application, and has been used in other applications of satellite imagery and CNNs within the development community [27]. Identifying suitable threshold values is of particular importance in conflict and security applications given the potential implications of false negative predictions.

To establish a baseline for comparing model accuracy, we generate a naive predictor of conflict fatality risk based only on proximity to the locations of fatal conflict events from the prior year. For each conflict window listed in Table 2 we evaluate the percentage of

both fatal and non-fatal events correctly classified using this naive predictor. Four versions of the naive predictor are tested using proximity thresholds of 1 km, 5 km, 7 km, and 15 km. For example, when using the 5 km naive predictor for an event location in 2019, a fatal conflict event would be predicted if a fatal conflict event occurred within 5 km during 2018. Although the threshold distances used are not directly comparable to the imagery window size used in the CNN (approximately  $7.3 \times 7.3$  km), the threshold distances were selected to assess naive performance across a broad range. Smaller proximity distances are more likely to result in better true negative rates (and worse true positive rates) while larger distances will likely result in better true positive rates (and worse true negative rates).

### 3. Results

Using the methods described in Section 2, 720 CNNs were trained to predict the risk of conflict fatality across three distinct time periods using yearly and six-month windows of imagery and conflict data (Table 2). The models also explored the impact of 16 different CNN hyperparameter combinations (Table 3). To ensure model performance is not dependent on a particular random sample generation, testing included repeating the sample generation process five times. A baseline accuracy comparison was produced using the maximum accuracy of a naive predictor described in Section 2 employed across all time periods and temporal windows. The accuracy results from all CNN models, along with the baseline, can be seen in Figure 2.



**Figure 2.** Accuracy distribution across all tests based on time period (inset year range) and temporal window (x-axis). Boxplot outline colors reflect the same temporal window across the different time periods. The blue baseline is the average baseline accuracy across all time periods and temporal windows.

Average accuracy across time periods and temporal windows was approximately 70%, with models using yearly windows of data slightly outperforming models using six-month windows overall. Model performance in the 2018–2019 time period was lower as contrasted to the 2014–2015 and 2016–2017 periods. Despite the significant decline in performance of the 2019 models, the minimum accuracy of all CNN models exceeded the 56% baseline accuracy of the naive predictor.

The accuracy of the naive predictor when only classifying fatal events (true positive rate) (Table 5) is only slightly below that of the CNN models' average accuracy. However, the naive predictor's accuracy when classifying non-fatal events (true negative rate) falls significantly compared to the CNN (indicating a high false positive rate) and results in worse overall performance. The baseline comparison with the naive predictor provides some confirmation that the CNN is detecting meaningful features from imagery that extend beyond solely the geographic locations of past events or exact features from training data, which allows the models to identify conflict in areas outside of those on which it was trained. The impact of a significant geographic shift in the location of training data on performance is discussed further in Section 4.

**Table 5.** Accuracy of naive predictor based on proximity to previous fatal conflict locations when using varying proximity thresholds.

Category	Year	Proximity Threshold			
		1 km	5 km	7 km	15 km
Mean % Fatal Correct	2015	0.62	0.67	0.71	0.81
	2017	0.67	0.72	0.74	0.83
	2019	0.52	0.62	0.67	0.82
Mean % Non-Fatal Correct	2015	0.48	0.40	0.38	0.29
	2017	0.36	0.29	0.27	0.21
	2019	0.40	0.34	0.30	0.16

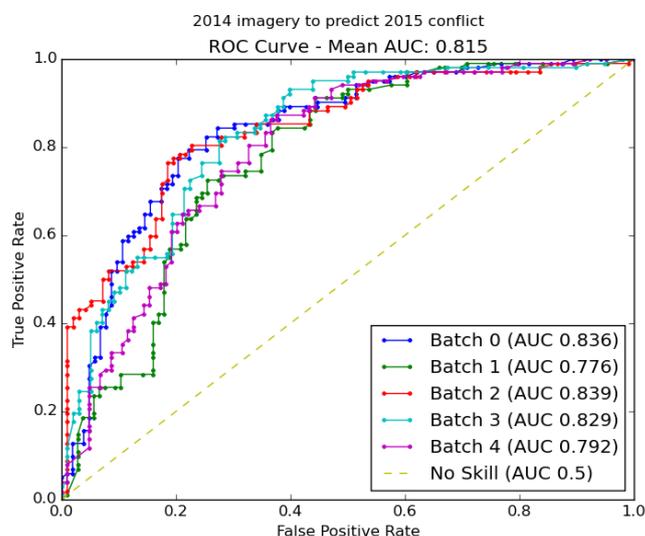
Across all tests, no noticeable changes based on hyperparameters were found and no notable changes in hyperparameter performance were associated with specific time periods or data windows. These results build upon findings from earlier work that found fine-tuning of hyperparameters within a generally acceptable range does not have a significant impact on model performance [35]. A notable additional finding is that a broader range between minimum and maximum values was seen across tests when using the six-month windows of data compared to yearly windows.

ROC curves were produced for the best-performing hyperparameter combination across all tests. The optimal hyperparameter combination consisted of ResNet50 CNN architecture, a learning rate of 0.0001, a gamma of 0.25, and a step size of 15. These hyperparameters produced a mean accuracy of 72% across all tests, with a maximum accuracy of 82% using six-month windows of data, and a maximum of 78% using yearly windows. ROC curves and an accompanying AUC metric were produced to gauge the ability of the models to effectively predict true positives without generating excessive false positives. An example of the ROC curves generated is seen in Figure 3 and shows the results based on using 2014 imagery from January–December to predict the 2015 (January–December) probability of a fatality from conflict. In this example, five curves are shown, showing the results from our five randomized permutations of test and training data. All ROC curves are included in Appendix A.3. The average AUC for each temporal pair is shown in Table 6.

AUC values when using the yearly data windows were consistently higher than when using the six-month windows. For both the 2015 and 2017 models, the AUC dropped by 2.5% when moving from yearly to six-month windows. For the 2019 model, the AUC when using a six-month data window dropped by nearly 5%. There was a significant drop in performance based on AUC in 2019 in line with earlier findings when compared to the 2015 and 2017 models. The 2019 model AUC values were approximately 10% less than the corresponding 2015 and 2017 models' AUC.

Overall, these results indicate that the CNNs are able to leverage sufficient information from satellite imagery to consistently outperform a baseline naive prediction at multiple time periods and temporal windows. Using six-month windows of data can achieve similar performance to using yearly windows of data, but both may be susceptible to degraded performance due to changes in the underlying data based on ground conditions. The

shifts in performance seen across the time periods explored during tests—primarily poorer performance in 2019—along with other limitations and potential future research directions are discussed in the following section.



**Figure 3.** ROC curve produced from CNNs trained on 2014 imagery (January–December) to predict the probability of a fatality if there is a conflict event in 2015 (January–December).

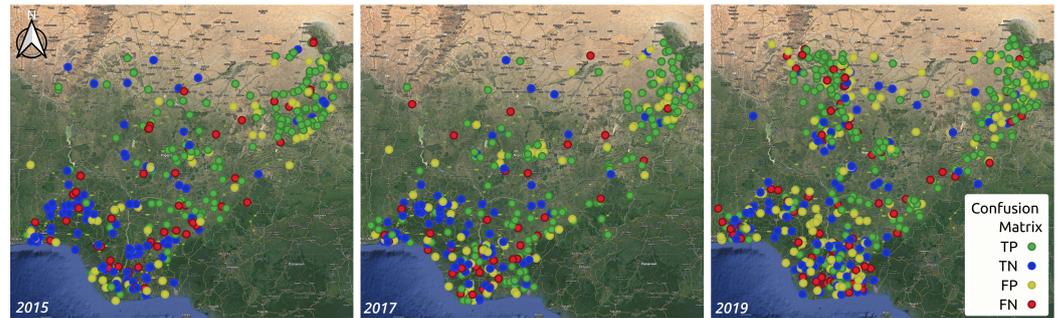
**Table 6.** Average ROC AUC from models for each time period and temporal window.

Imagery Temporal	Conflict Temporal	AUC ROC
2014 January–December	2015 January–December	0.815
2014 January–June	2015 January–June	0.790
2014 July–December	2015 July–December	0.796
2016 January–December	2017 January–December	0.807
2016 January–June	2017 January–June	0.782
2016 July–December	2017 July–December	0.797
2018 January–December	2019 January–December	0.722
2018 January–June	2019 January–June	0.681
2018 July–December	2019 July–December	0.678

#### 4. Discussion

This paper presented evidence that CNN-based methods are moderately effective at identifying features in satellite imagery associated with the risk of fatalities from conflict events across time periods. These findings contribute to a broader set of literature using CNNs and satellite imagery to predict development indicators at a single year or coarser time periods [20,24,26,32,35].

As a byproduct of testing performance over multiple time periods, this work was able to observe the model’s behavior when the geographic distribution of samples shifted over time. The decrease in performance seen in the models using 2018/19 data—compared to models using data from 2014/15 or 2016/17—was one of the most pronounced changes found throughout testing. As shown in Figure 4, this change in performance was associated with a notable shift in the geographic distribution of conflict locations used for training (an increase in conflict events in the northwest of the country). This reinforces the importance of testing out-of-sample predictions, across both space and time, presented in literature [26], and suggests a need for additional training and validation data to improve the robustness of the models.



**Figure 4.** Location of conflict events [3] and prediction results in 2015 (left), 2017 (center), 2019 (right). Imagery ©2024 TerraMetrics, Map data ©2024 Google.

Addressing the amount and spatial coverage of samples used for training and validation has been handled in multiple ways in literature exploring CNN-based approaches to predict development indicators. Previous work to produce estimates of poverty using CNNs and satellite imagery incorporated data across multiple geographic extents (countries), and shown varying performance based on which areas were used to train models, as well as which country was being evaluated [20]. Other work has advised caution when extending predictions across geographic areas and also suggests that temporal variation in imagery may impact performance [26]. Given the potential impact of shifting the geographic distribution of locations over time on model performance, applications of these methods should consider the geographic distribution of training and validation data as well as the distribution of locations where the model will be used to generate predictions. The lack of considerable variation in performance across hyperparameters seen during optimization may be one indicator of the generalizability of the model and could facilitate training models on broader geographic areas.

The amount of data available also impacts the approaches that are used to train models. Transfer learning and pre-trained CNNs can be leveraged to reduce the dependence on large training data sets, yet there is still a need for enough data to sufficiently fine-tune pre-trained networks and represent all geographic regions where the model may be used for predictions. For example, insufficient training data may be one of the reasons for a decrease in performance when moving from using a full-year window of data to a six-month window. Across tests, there was a slightly larger range between minimum and maximum accuracy when using six-month windows which may be indicative of the instability of model performance when relying on smaller training sets. The potential ability and value of narrower temporal windows to reflect seasonal trends (by aggregating less underlying data and therefore containing less noise from seasonal variation) may have either been negated by the associated drop in samples, not realized due to seasonal misalignment, or simply not been meaningful when predicting the fatality of conflict events.

One possible solution to help address limitations around sample size is to train a model using a combination of samples across time periods. In addition to increasing geographic coverage and sample size, combining data across time periods would enable exploring the model's ability to predict conflict fatality beyond the temporal range on which they were trained. Related to this, the specific seasonality of sub-annual time periods used for training could impact model performance and would require further research to evaluate. The extensibility of models to time periods beyond those on which they were trained would be necessary to support many real-world applications (i.e., using historical data to predict future conflict fatality risk). While the current approach could be used to gain general insight into extensibility, the lack of repeat observations across time periods does limit our ability to gauge the performance of a time series of estimates (i.e., using true panel data). Similarly, the current work provides little explanatory power as to the drivers of conflict/fatalities, yet may be able to contribute to broader literature exploring the nature of conflict through further research into the underlying features in satellite imagery associated with conflict fatalities.

Another aspect of real-world applications not explored in this paper is the spatial scale of predictions. This paper focuses on predictions and validation in terms of known conflict events at specific locations. Future estimates would likely be made in the form of either a continuous surface or a grid across an entire region. In these scenarios, the resolution of the surface or distance between grid points used would play a significant role in the accuracy of predictions. Given possible concerns with the accuracy of the underlying georeferencing of conflict events datasets [75], and that a single coordinate is often a simplification of the extent of a conflict event, it may also be practical to aggregate predictions to relevant local units such as administrative zones.

As methods for predicting conflict and fatalities progress, applications may aim to generate predictions for narrower windows of time (e.g., three-month windows) or further examine the existence of seasonal trends. Using the presented methodology may result in issues of (a) producing cloud-free aggregates of satellite imagery without missing data and (b) suffering reduced performance due to limitations regarding the temporal precision of recorded conflict events. Despite the sixteen-day revisit time of Landsat 8 imagery, not all images result in usable scenes due to cloud cover or other quality issues. As a result, it is necessary to have a sufficient temporal window over which to aggregate scenes and achieve full coverage. A possible solution to this problem would be to aggregate a longer period of imagery than the prediction interval. For example, a model could be trained to use a six-month window of imagery to produce predictions for a three-month window of conflict fatality. Limitations based on the temporal precision of the underlying conflict data used for training would be more difficult to overcome. Potential solutions include identifying new sources of data with improved temporal precision or working with data providers to improve data collection. In addition, while the arbitrary six-month windows did not benefit from seasonal trends in imagery, they also only showed minimal decreases in performance and it is possible more specific windows (e.g., harvest seasons, dry season) could produce better results.

In addition to the areas of future research discussed above, multiple opportunities exist for improving the existing methodology based on progress within the broader machine learning community. Examples of pathways towards improving model performance include the use of ensemble models [76–78], 3D convolutional layers [79,80], and other CNN architectures such as ResNeXt [81]. Applying transformations such as rotating or flipping the input satellite image are other viable means of increasing exposure to data variations when using limited training data [82,83]. These methods have been used widely and may offer an improvement of the basic sampling scheme presented in this paper [84].

Similarly, opportunities may exist to leverage alternative sources of data for imagery and conflict records. Finer resolution imagery products such as Sentinel 2 (European Space Agency) or PlanetScope (Planet Labs) may support detecting features associated with conflict that cannot be detected from Landsat 8 imagery. In addition, improvements to conflict datasets to produce more precise location records (or utilizing only subsets of datasets such as ACLED that were recorded at precise levels) may enable more effective training of deep learning models.

## 5. Conclusions

The work presented in this paper sought to address two research questions: (RQ1) *Can CNNs detect features from moderate resolution satellite imagery associated with conflict fatality at multiple discrete points in time?* And if so, (RQ2) *how does model performance vary based on the temporal window size used for training?* Using Landsat 8 imagery and ACLED conflict data, this paper showed that models trained using six-month and yearly windows of data were able to achieve comparable performance at multiple points in time. While predictive ability varied, overall metrics of both accuracy as well as ROC area under the curve both reached approximately 75% on average and outperformed a naive baseline predictor. These results advance existing work in the field by providing evidence that features in moderate-resolution satellite imagery that are associated with conflict fatalities can be

detected using CNNs at multiple points in time, as well as when using varying windows of data. Combined, these findings provide a pathway forward for producing future estimates of the risk of conflict fatalities that may be useful in a variety of development and security applications.

**Author Contributions:** S.G. conceptualized the project, with input from D.R.; A.B. and S.G. wrote code to prepare data and implement machine learning methods, analysis, and other components. D.R. assisted in algorithm and methodology development. This piece was written by S.G. and D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the USAID Global Development Lab under Grant AID-OAA-A-12-00096.

**Data Availability Statement:** All data necessary to reproduce the work in this paper—specifically Landsat 8 imagery and ACLED conflict data—are from publicly available sources and can be acquired as detailed in the paper, but are either too large to redistribute or not permitted to be redistributed. Code used for model development can be accessed on GitHub ([https://github.com/aiddata/geo-cnn/tree/nigeria\\_npe\\_prediction](https://github.com/aiddata/geo-cnn/tree/nigeria_npe_prediction), accessed on 1 July 2024).

**Acknowledgments:** 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.

## Appendix A. Additional Metrics

This appendix section consists of a full confusion matrix (Table A1) and associated performance metrics (Table A2) using varying thresholds for the models referenced in Section 3. For Tables A1 and A2, the *version* column references each of the five versions of training/validation data generated randomly generated to ensure model performance is not dependent on a particular random sample generation.

### Appendix A.1. Confusion Matrix

**Table A1.** Full confusion matrices for all models generated.

Data	Version	Thresh	tp	tn	fp	fn
2016 imagery, 2017 conflict [July–December]	v1001	0.3	0.879	0.484	0.516	0.121
2016 imagery, 2017 conflict [July–December]	v1001	0.35	0.879	0.516	0.484	0.121
2016 imagery, 2017 conflict [July–December]	v1001	0.4	0.879	0.548	0.452	0.121
2016 imagery, 2017 conflict [July–December]	v1001	0.45	0.862	0.629	0.371	0.138
2016 imagery, 2017 conflict [July–December]	v1001	0.5	0.845	0.645	0.355	0.155
2016 imagery, 2017 conflict [July–December]	v1002	0.3	0.897	0.559	0.441	0.103
2016 imagery, 2017 conflict [July–December]	v1002	0.35	0.845	0.627	0.373	0.155
2016 imagery, 2017 conflict [July–December]	v1002	0.4	0.845	0.644	0.356	0.155
2016 imagery, 2017 conflict [July–December]	v1002	0.45	0.845	0.661	0.339	0.155
2016 imagery, 2017 conflict [July–December]	v1002	0.5	0.81	0.695	0.305	0.19
2016 imagery, 2017 conflict [July–December]	v1003	0.3	0.947	0.346	0.654	0.053
2016 imagery, 2017 conflict [July–December]	v1003	0.35	0.895	0.462	0.538	0.105
2016 imagery, 2017 conflict [July–December]	v1003	0.4	0.895	0.519	0.481	0.105
2016 imagery, 2017 conflict [July–December]	v1003	0.45	0.877	0.596	0.404	0.123
2016 imagery, 2017 conflict [July–December]	v1003	0.5	0.842	0.654	0.346	0.158
2016 imagery, 2017 conflict [July–December]	v1004	0.3	0.914	0.582	0.418	0.086
2016 imagery, 2017 conflict [July–December]	v1004	0.35	0.897	0.709	0.291	0.103
2016 imagery, 2017 conflict [July–December]	v1004	0.4	0.879	0.745	0.255	0.121
2016 imagery, 2017 conflict [July–December]	v1004	0.45	0.793	0.764	0.236	0.207
2016 imagery, 2017 conflict [July–December]	v1004	0.5	0.707	0.855	0.145	0.293
2016 imagery, 2017 conflict [July–December]	v100	0.3	0.877	0.517	0.483	0.123
2016 imagery, 2017 conflict [July–December]	v100	0.35	0.825	0.6	0.4	0.175
2016 imagery, 2017 conflict [July–December]	v100	0.4	0.825	0.617	0.383	0.175
2016 imagery, 2017 conflict [July–December]	v100	0.45	0.807	0.617	0.383	0.193

Table A1. Cont.

Data	Version	Thresh	tp	tn	fp	fn
2016 imagery, 2017 conflict [July–December]	v100	0.5	0.807	0.667	0.333	0.193
2014 imagery, 2015 conflict [July–December]	v1001	0.3	0.95	0.575	0.425	0.05
2014 imagery, 2015 conflict [July–December]	v1001	0.35	0.925	0.675	0.325	0.075
2014 imagery, 2015 conflict [July–December]	v1001	0.4	0.925	0.7	0.3	0.075
2014 imagery, 2015 conflict [July–December]	v1001	0.45	0.875	0.7	0.3	0.125
2014 imagery, 2015 conflict [July–December]	v1001	0.5	0.875	0.7	0.3	0.125
2014 imagery, 2015 conflict [July–December]	v1002	0.3	0.821	0.632	0.368	0.179
2014 imagery, 2015 conflict [July–December]	v1002	0.35	0.744	0.658	0.342	0.256
2014 imagery, 2015 conflict [July–December]	v1002	0.4	0.692	0.658	0.342	0.308
2014 imagery, 2015 conflict [July–December]	v1002	0.45	0.692	0.711	0.289	0.308
2014 imagery, 2015 conflict [July–December]	v1002	0.5	0.667	0.789	0.211	0.333
2014 imagery, 2015 conflict [July–December]	v1003	0.3	0.775	0.615	0.385	0.225
2014 imagery, 2015 conflict [July–December]	v1003	0.35	0.725	0.744	0.256	0.275
2014 imagery, 2015 conflict [July–December]	v1003	0.4	0.7	0.795	0.205	0.3
2014 imagery, 2015 conflict [July–December]	v1003	0.45	0.675	0.795	0.205	0.325
2014 imagery, 2015 conflict [July–December]	v1003	0.5	0.65	0.821	0.179	0.35
2014 imagery, 2015 conflict [July–December]	v1004	0.3	0.825	0.61	0.39	0.175
2014 imagery, 2015 conflict [July–December]	v1004	0.35	0.825	0.634	0.366	0.175
2014 imagery, 2015 conflict [July–December]	v1004	0.4	0.775	0.659	0.341	0.225
2014 imagery, 2015 conflict [July–December]	v1004	0.45	0.7	0.683	0.317	0.3
2014 imagery, 2015 conflict [July–December]	v1004	0.5	0.7	0.683	0.317	0.3
2014 imagery, 2015 conflict [July–December]	v100	0.3	0.9	0.576	0.424	0.1
2014 imagery, 2015 conflict [July–December]	v100	0.35	0.9	0.697	0.303	0.1
2014 imagery, 2015 conflict [July–December]	v100	0.4	0.875	0.818	0.182	0.125
2014 imagery, 2015 conflict [July–December]	v100	0.45	0.85	0.848	0.152	0.15
2014 imagery, 2015 conflict [July–December]	v100	0.5	0.85	0.848	0.152	0.15
2014 imagery, 2015 conflict [January–December]	v1011	0.3	0.951	0.456	0.544	0.049
2014 imagery, 2015 conflict [January–December]	v1011	0.35	0.902	0.534	0.466	0.098
2014 imagery, 2015 conflict [January–December]	v1011	0.4	0.853	0.66	0.34	0.147
2014 imagery, 2015 conflict [January–December]	v1011	0.45	0.824	0.728	0.272	0.176
2014 imagery, 2015 conflict [January–December]	v1011	0.5	0.775	0.777	0.223	0.225
2014 imagery, 2015 conflict [January–December]	v1012	0.3	0.912	0.509	0.491	0.088
2014 imagery, 2015 conflict [January–December]	v1012	0.35	0.892	0.557	0.443	0.108
2014 imagery, 2015 conflict [January–December]	v1012	0.4	0.843	0.623	0.377	0.157
2014 imagery, 2015 conflict [January–December]	v1012	0.45	0.765	0.651	0.349	0.235
2014 imagery, 2015 conflict [January–December]	v1012	0.5	0.725	0.736	0.264	0.275
2014 imagery, 2015 conflict [January–December]	v1013	0.3	0.931	0.485	0.515	0.069
2014 imagery, 2015 conflict [January–December]	v1013	0.35	0.892	0.526	0.474	0.108
2014 imagery, 2015 conflict [January–December]	v1013	0.4	0.853	0.619	0.381	0.147
2014 imagery, 2015 conflict [January–December]	v1013	0.45	0.804	0.732	0.268	0.196
2014 imagery, 2015 conflict [January–December]	v1013	0.5	0.765	0.814	0.186	0.235
2014 imagery, 2015 conflict [January–December]	v1014	0.3	0.971	0.469	0.531	0.029
2014 imagery, 2015 conflict [January–December]	v1014	0.35	0.951	0.531	0.469	0.049
2014 imagery, 2015 conflict [January–December]	v1014	0.4	0.931	0.582	0.418	0.069
2014 imagery, 2015 conflict [January–December]	v1014	0.45	0.863	0.643	0.357	0.137
2014 imagery, 2015 conflict [January–December]	v1014	0.5	0.814	0.724	0.276	0.186
2014 imagery, 2015 conflict [January–December]	v101	0.3	0.922	0.529	0.471	0.078
2014 imagery, 2015 conflict [January–December]	v101	0.35	0.912	0.558	0.442	0.088
2014 imagery, 2015 conflict [January–December]	v101	0.4	0.873	0.587	0.413	0.127
2014 imagery, 2015 conflict [January–December]	v101	0.45	0.863	0.625	0.375	0.137
2014 imagery, 2015 conflict [January–December]	v101	0.5	0.755	0.692	0.308	0.245
2014 imagery, 2015 conflict [January–June]	v1001	0.3	0.742	0.725	0.275	0.258
2014 imagery, 2015 conflict [January–June]	v1001	0.35	0.726	0.739	0.261	0.274
2014 imagery, 2015 conflict [January–June]	v1001	0.4	0.726	0.754	0.246	0.274
2014 imagery, 2015 conflict [January–June]	v1001	0.45	0.726	0.768	0.232	0.274
2014 imagery, 2015 conflict [January–June]	v1001	0.5	0.726	0.783	0.217	0.274
2014 imagery, 2015 conflict [January–June]	v1002	0.3	0.79	0.597	0.403	0.21
2014 imagery, 2015 conflict [January–June]	v1002	0.35	0.79	0.645	0.355	0.21

Table A1. Cont.

Data	Version	Thresh	tp	tn	fp	fn
2014 imagery, 2015 conflict [January–June]	v1002	0.4	0.79	0.661	0.339	0.21
2014 imagery, 2015 conflict [January–June]	v1002	0.45	0.774	0.726	0.274	0.226
2014 imagery, 2015 conflict [January–June]	v1002	0.5	0.758	0.742	0.258	0.242
2014 imagery, 2015 conflict [January–June]	v1003	0.3	0.81	0.623	0.377	0.19
2014 imagery, 2015 conflict [January–June]	v1003	0.35	0.778	0.623	0.377	0.222
2014 imagery, 2015 conflict [January–June]	v1003	0.4	0.762	0.672	0.328	0.238
2014 imagery, 2015 conflict [January–June]	v1003	0.45	0.746	0.689	0.311	0.254
2014 imagery, 2015 conflict [January–June]	v1003	0.5	0.73	0.689	0.311	0.27
2014 imagery, 2015 conflict [January–June]	v1004	0.3	0.645	0.716	0.284	0.355
2014 imagery, 2015 conflict [January–June]	v1004	0.35	0.629	0.731	0.269	0.371
2014 imagery, 2015 conflict [January–June]	v1004	0.4	0.597	0.791	0.209	0.403
2014 imagery, 2015 conflict [January–June]	v1004	0.45	0.581	0.806	0.194	0.419
2014 imagery, 2015 conflict [January–June]	v1004	0.5	0.565	0.866	0.134	0.435
2014 imagery, 2015 conflict [January–June]	v100	0.3	0.887	0.464	0.536	0.113
2014 imagery, 2015 conflict [January–June]	v100	0.35	0.871	0.5	0.5	0.129
2014 imagery, 2015 conflict [January–June]	v100	0.4	0.855	0.571	0.429	0.145
2014 imagery, 2015 conflict [January–June]	v100	0.45	0.806	0.625	0.375	0.194
2014 imagery, 2015 conflict [January–June]	v100	0.5	0.774	0.75	0.25	0.226
2018 imagery, 2019 conflict [January–June]	v1001	0.3	0.708	0.5	0.5	0.292
2018 imagery, 2019 conflict [January–June]	v1001	0.35	0.652	0.556	0.444	0.348
2018 imagery, 2019 conflict [January–June]	v1001	0.4	0.618	0.611	0.389	0.382
2018 imagery, 2019 conflict [January–June]	v1001	0.45	0.562	0.656	0.344	0.438
2018 imagery, 2019 conflict [January–June]	v1001	0.5	0.551	0.722	0.278	0.449
2018 imagery, 2019 conflict [January–June]	v1002	0.3	0.798	0.551	0.449	0.202
2018 imagery, 2019 conflict [January–June]	v1002	0.35	0.764	0.551	0.449	0.236
2018 imagery, 2019 conflict [January–June]	v1002	0.4	0.719	0.584	0.416	0.281
2018 imagery, 2019 conflict [January–June]	v1002	0.45	0.685	0.618	0.382	0.315
2018 imagery, 2019 conflict [January–June]	v1002	0.5	0.663	0.697	0.303	0.337
2018 imagery, 2019 conflict [January–June]	v1003	0.3	0.843	0.371	0.629	0.157
2018 imagery, 2019 conflict [January–June]	v1003	0.35	0.831	0.416	0.584	0.169
2018 imagery, 2019 conflict [January–June]	v1003	0.4	0.787	0.517	0.483	0.213
2018 imagery, 2019 conflict [January–June]	v1003	0.45	0.764	0.562	0.438	0.236
2018 imagery, 2019 conflict [January–June]	v1003	0.5	0.742	0.573	0.427	0.258
2018 imagery, 2019 conflict [January–June]	v1004	0.3	0.867	0.333	0.667	0.133
2018 imagery, 2019 conflict [January–June]	v1004	0.35	0.833	0.398	0.602	0.167
2018 imagery, 2019 conflict [January–June]	v1004	0.4	0.811	0.43	0.57	0.189
2018 imagery, 2019 conflict [January–June]	v1004	0.45	0.744	0.484	0.516	0.256
2018 imagery, 2019 conflict [January–June]	v1004	0.5	0.711	0.548	0.452	0.289
2018 imagery, 2019 conflict [January–June]	v100	0.3	0.876	0.443	0.557	0.124
2018 imagery, 2019 conflict [January–June]	v100	0.35	0.831	0.545	0.455	0.169
2018 imagery, 2019 conflict [January–June]	v100	0.4	0.809	0.58	0.42	0.191
2018 imagery, 2019 conflict [January–June]	v100	0.45	0.775	0.602	0.398	0.225
2018 imagery, 2019 conflict [January–June]	v100	0.5	0.764	0.67	0.33	0.236
2018 imagery, 2019 conflict [January–December]	v1011	0.3	0.743	0.503	0.497	0.257
2018 imagery, 2019 conflict [January–December]	v1011	0.35	0.694	0.531	0.469	0.306
2018 imagery, 2019 conflict [January–December]	v1011	0.4	0.667	0.559	0.441	0.333
2018 imagery, 2019 conflict [January–December]	v1011	0.45	0.625	0.614	0.386	0.375
2018 imagery, 2019 conflict [January–December]	v1011	0.5	0.611	0.676	0.324	0.389
2018 imagery, 2019 conflict [January–December]	v1012	0.3	0.972	0.239	0.761	0.028
2018 imagery, 2019 conflict [January–December]	v1012	0.35	0.951	0.331	0.669	0.049
2018 imagery, 2019 conflict [January–December]	v1012	0.4	0.917	0.486	0.514	0.083
2018 imagery, 2019 conflict [January–December]	v1012	0.45	0.882	0.606	0.394	0.118
2018 imagery, 2019 conflict [January–December]	v1012	0.5	0.729	0.711	0.289	0.271
2018 imagery, 2019 conflict [January–December]	v1013	0.3	0.917	0.329	0.671	0.083
2018 imagery, 2019 conflict [January–December]	v1013	0.35	0.896	0.4	0.6	0.104
2018 imagery, 2019 conflict [January–December]	v1013	0.4	0.833	0.465	0.535	0.167
2018 imagery, 2019 conflict [January–December]	v1013	0.45	0.771	0.548	0.452	0.229
2018 imagery, 2019 conflict [January–December]	v1013	0.5	0.722	0.632	0.368	0.278

Table A1. Cont.

Data	Version	Thresh	tp	tn	fp	fn
2018 imagery, 2019 conflict [January–December]	v1014	0.3	0.855	0.385	0.615	0.145
2018 imagery, 2019 conflict [January–December]	v1014	0.35	0.814	0.413	0.587	0.186
2018 imagery, 2019 conflict [January–December]	v1014	0.4	0.779	0.455	0.545	0.221
2018 imagery, 2019 conflict [January–December]	v1014	0.45	0.703	0.566	0.434	0.297
2018 imagery, 2019 conflict [January–December]	v1014	0.5	0.634	0.636	0.364	0.366
2018 imagery, 2019 conflict [January–December]	v101	0.3	0.951	0.338	0.662	0.049
2018 imagery, 2019 conflict [January–December]	v101	0.35	0.924	0.441	0.559	0.076
2018 imagery, 2019 conflict [January–December]	v101	0.4	0.882	0.531	0.469	0.118
2018 imagery, 2019 conflict [January–December]	v101	0.45	0.84	0.593	0.407	0.16
2018 imagery, 2019 conflict [January–December]	v101	0.5	0.757	0.641	0.359	0.243
2016 imagery, 2017 conflict [January–December]	v1011	0.3	0.957	0.291	0.709	0.043
2016 imagery, 2017 conflict [January–December]	v1011	0.35	0.94	0.325	0.675	0.06
2016 imagery, 2017 conflict [January–December]	v1011	0.4	0.915	0.41	0.59	0.085
2016 imagery, 2017 conflict [January–December]	v1011	0.45	0.872	0.521	0.479	0.128
2016 imagery, 2017 conflict [January–December]	v1011	0.5	0.821	0.641	0.359	0.179
2016 imagery, 2017 conflict [January–December]	v1012	0.3	0.847	0.648	0.352	0.153
2016 imagery, 2017 conflict [January–December]	v1012	0.35	0.788	0.697	0.303	0.212
2016 imagery, 2017 conflict [January–December]	v1012	0.4	0.78	0.738	0.262	0.22
2016 imagery, 2017 conflict [January–December]	v1012	0.45	0.763	0.779	0.221	0.237
2016 imagery, 2017 conflict [January–December]	v1012	0.5	0.737	0.803	0.197	0.263
2016 imagery, 2017 conflict [January–December]	v1013	0.3	0.829	0.645	0.355	0.171
2016 imagery, 2017 conflict [January–December]	v1013	0.35	0.803	0.685	0.315	0.197
2016 imagery, 2017 conflict [January–December]	v1013	0.4	0.795	0.734	0.266	0.205
2016 imagery, 2017 conflict [January–December]	v1013	0.45	0.769	0.774	0.226	0.231
2016 imagery, 2017 conflict [January–December]	v1013	0.5	0.769	0.79	0.21	0.231
2016 imagery, 2017 conflict [January–December]	v1014	0.3	0.803	0.653	0.347	0.197
2016 imagery, 2017 conflict [January–December]	v1014	0.35	0.803	0.669	0.331	0.197
2016 imagery, 2017 conflict [January–December]	v1014	0.4	0.778	0.702	0.298	0.222
2016 imagery, 2017 conflict [January–December]	v1014	0.45	0.744	0.727	0.273	0.256
2016 imagery, 2017 conflict [January–December]	v1014	0.5	0.718	0.785	0.215	0.282
2016 imagery, 2017 conflict [January–December]	v101	0.3	0.846	0.622	0.378	0.154
2016 imagery, 2017 conflict [January–December]	v101	0.35	0.821	0.63	0.37	0.179
2016 imagery, 2017 conflict [January–December]	v101	0.4	0.821	0.647	0.353	0.179
2016 imagery, 2017 conflict [January–December]	v101	0.45	0.795	0.681	0.319	0.205
2016 imagery, 2017 conflict [January–December]	v101	0.5	0.786	0.739	0.261	0.214
2018 imagery, 2019 conflict [July–December]	v1001	0.3	0.655	0.567	0.433	0.345
2018 imagery, 2019 conflict [July–December]	v1001	0.35	0.655	0.633	0.367	0.345
2018 imagery, 2019 conflict [July–December]	v1001	0.4	0.636	0.65	0.35	0.364
2018 imagery, 2019 conflict [July–December]	v1001	0.45	0.6	0.7	0.3	0.4
2018 imagery, 2019 conflict [July–December]	v1001	0.5	0.564	0.75	0.25	0.436
2018 imagery, 2019 conflict [July–December]	v1002	0.3	0.704	0.49	0.51	0.296
2018 imagery, 2019 conflict [July–December]	v1002	0.35	0.685	0.51	0.49	0.315
2018 imagery, 2019 conflict [July–December]	v1002	0.4	0.63	0.569	0.431	0.37
2018 imagery, 2019 conflict [July–December]	v1002	0.45	0.63	0.627	0.373	0.37
2018 imagery, 2019 conflict [July–December]	v1002	0.5	0.593	0.686	0.314	0.407
2018 imagery, 2019 conflict [July–December]	v1003	0.3	0.782	0.583	0.417	0.218
2018 imagery, 2019 conflict [July–December]	v1003	0.35	0.764	0.625	0.375	0.236
2018 imagery, 2019 conflict [July–December]	v1003	0.4	0.727	0.625	0.375	0.273
2018 imagery, 2019 conflict [July–December]	v1003	0.45	0.636	0.667	0.333	0.364
2018 imagery, 2019 conflict [July–December]	v1003	0.5	0.6	0.771	0.229	0.4
2018 imagery, 2019 conflict [July–December]	v1004	0.3	0.8	0.328	0.672	0.2
2018 imagery, 2019 conflict [July–December]	v1004	0.35	0.782	0.426	0.574	0.218
2018 imagery, 2019 conflict [July–December]	v1004	0.4	0.745	0.492	0.508	0.255
2018 imagery, 2019 conflict [July–December]	v1004	0.45	0.691	0.623	0.377	0.309
2018 imagery, 2019 conflict [July–December]	v1004	0.5	0.564	0.721	0.279	0.436
2018 imagery, 2019 conflict [July–December]	v100	0.3	0.709	0.537	0.463	0.291
2018 imagery, 2019 conflict [July–December]	v100	0.35	0.673	0.593	0.407	0.327
2018 imagery, 2019 conflict [July–December]	v100	0.4	0.655	0.611	0.389	0.345

Table A1. Cont.

Data	Version	Thresh	tp	tn	fp	fn
2018 imagery, 2019 conflict [July–December]	v100	0.45	0.636	0.63	0.37	0.364
2018 imagery, 2019 conflict [July–December]	v100	0.5	0.564	0.722	0.278	0.436
2016 imagery, 2017 conflict [January–June]	v1001	0.3	0.75	0.508	0.492	0.25
2016 imagery, 2017 conflict [January–June]	v1001	0.35	0.733	0.576	0.424	0.267
2016 imagery, 2017 conflict [January–June]	v1001	0.4	0.7	0.61	0.39	0.3
2016 imagery, 2017 conflict [January–June]	v1001	0.45	0.7	0.661	0.339	0.3
2016 imagery, 2017 conflict [January–June]	v1001	0.5	0.683	0.763	0.237	0.317
2016 imagery, 2017 conflict [January–June]	v1002	0.3	0.933	0.475	0.525	0.067
2016 imagery, 2017 conflict [January–June]	v1002	0.35	0.933	0.492	0.508	0.067
2016 imagery, 2017 conflict [January–June]	v1002	0.4	0.883	0.59	0.41	0.117
2016 imagery, 2017 conflict [January–June]	v1002	0.45	0.883	0.639	0.361	0.117
2016 imagery, 2017 conflict [January–June]	v1002	0.5	0.883	0.639	0.361	0.117
2016 imagery, 2017 conflict [January–June]	v1003	0.3	0.833	0.542	0.458	0.167
2016 imagery, 2017 conflict [January–June]	v1003	0.35	0.75	0.61	0.39	0.25
2016 imagery, 2017 conflict [January–June]	v1003	0.4	0.717	0.678	0.322	0.283
2016 imagery, 2017 conflict [January–June]	v1003	0.45	0.717	0.763	0.237	0.283
2016 imagery, 2017 conflict [January–June]	v1003	0.5	0.717	0.814	0.186	0.283
2016 imagery, 2017 conflict [January–June]	v1004	0.3	0.85	0.492	0.508	0.15
2016 imagery, 2017 conflict [January–June]	v1004	0.35	0.783	0.525	0.475	0.217
2016 imagery, 2017 conflict [January–June]	v1004	0.4	0.767	0.593	0.407	0.233
2016 imagery, 2017 conflict [January–June]	v1004	0.45	0.767	0.61	0.39	0.233
2016 imagery, 2017 conflict [January–June]	v1004	0.5	0.767	0.627	0.373	0.233
2016 imagery, 2017 conflict [January–June]	v100	0.3	0.833	0.557	0.443	0.167
2016 imagery, 2017 conflict [January–June]	v100	0.35	0.817	0.59	0.41	0.183
2016 imagery, 2017 conflict [January–June]	v100	0.4	0.8	0.59	0.41	0.2
2016 imagery, 2017 conflict [January–June]	v100	0.45	0.783	0.689	0.311	0.217
2016 imagery, 2017 conflict [January–June]	v100	0.5	0.783	0.705	0.295	0.217

## Appendix A.2. Performance Metrics

Table A2. Full performance metrics for all models generated.

Data	Version	Thresh	Accuracy	Precision	Recall	f1
2016 imagery, 2017 conflict [July–December]	v1001	0.3	0.675	0.614	0.879	0.723
2016 imagery, 2017 conflict [July–December]	v1001	0.35	0.692	0.63	0.879	0.734
2016 imagery, 2017 conflict [July–December]	v1001	0.4	0.708	0.646	0.879	0.745
2016 imagery, 2017 conflict [July–December]	v1001	0.45	0.742	0.685	0.862	0.763
2016 imagery, 2017 conflict [July–December]	v1001	0.5	0.742	0.69	0.845	0.76
2016 imagery, 2017 conflict [July–December]	v1002	0.3	0.726	0.667	0.897	0.765
2016 imagery, 2017 conflict [July–December]	v1002	0.35	0.735	0.69	0.845	0.76
2016 imagery, 2017 conflict [July–December]	v1002	0.4	0.744	0.7	0.845	0.766
2016 imagery, 2017 conflict [July–December]	v1002	0.45	0.752	0.71	0.845	0.772
2016 imagery, 2017 conflict [July–December]	v1002	0.5	0.752	0.723	0.81	0.764
2016 imagery, 2017 conflict [July–December]	v1003	0.3	0.661	0.614	0.947	0.745
2016 imagery, 2017 conflict [July–December]	v1003	0.35	0.688	0.646	0.895	0.75
2016 imagery, 2017 conflict [July–December]	v1003	0.4	0.716	0.671	0.895	0.767
2016 imagery, 2017 conflict [July–December]	v1003	0.45	0.743	0.704	0.877	0.781
2016 imagery, 2017 conflict [July–December]	v1003	0.5	0.752	0.727	0.842	0.78
2016 imagery, 2017 conflict [July–December]	v1004	0.3	0.752	0.697	0.914	0.791
2016 imagery, 2017 conflict [July–December]	v1004	0.35	0.805	0.765	0.897	0.825
2016 imagery, 2017 conflict [July–December]	v1004	0.4	0.814	0.785	0.879	0.829
2016 imagery, 2017 conflict [July–December]	v1004	0.45	0.779	0.78	0.793	0.786
2016 imagery, 2017 conflict [July–December]	v1004	0.5	0.779	0.837	0.707	0.766
2016 imagery, 2017 conflict [July–December]	v100	0.3	0.692	0.633	0.877	0.735
2016 imagery, 2017 conflict [July–December]	v100	0.35	0.709	0.662	0.825	0.734
2016 imagery, 2017 conflict [July–December]	v100	0.4	0.718	0.671	0.825	0.74

Table A2. Cont.

Data	Version	Thresh	Accuracy	Precision	Recall	f1
2016 imagery, 2017 conflict [July–December]	v100	0.45	0.709	0.667	0.807	0.73
2016 imagery, 2017 conflict [July–December]	v100	0.5	0.735	0.697	0.807	0.748
2014 imagery, 2015 conflict [July–December]	v1001	0.3	0.762	0.691	0.95	0.8
2014 imagery, 2015 conflict [July–December]	v1001	0.35	0.8	0.74	0.925	0.822
2014 imagery, 2015 conflict [July–December]	v1001	0.4	0.813	0.755	0.925	0.831
2014 imagery, 2015 conflict [July–December]	v1001	0.45	0.787	0.745	0.875	0.805
2014 imagery, 2015 conflict [July–December]	v1001	0.5	0.787	0.745	0.875	0.805
2014 imagery, 2015 conflict [July–December]	v1002	0.3	0.727	0.696	0.821	0.753
2014 imagery, 2015 conflict [July–December]	v1002	0.35	0.701	0.69	0.744	0.716
2014 imagery, 2015 conflict [July–December]	v1002	0.4	0.675	0.675	0.692	0.684
2014 imagery, 2015 conflict [July–December]	v1002	0.45	0.701	0.711	0.692	0.701
2014 imagery, 2015 conflict [July–December]	v1002	0.5	0.727	0.765	0.667	0.712
2014 imagery, 2015 conflict [July–December]	v1003	0.3	0.696	0.674	0.775	0.721
2014 imagery, 2015 conflict [July–December]	v1003	0.35	0.734	0.744	0.725	0.734
2014 imagery, 2015 conflict [July–December]	v1003	0.4	0.747	0.778	0.7	0.737
2014 imagery, 2015 conflict [July–December]	v1003	0.45	0.734	0.771	0.675	0.72
2014 imagery, 2015 conflict [July–December]	v1003	0.5	0.734	0.788	0.65	0.712
2014 imagery, 2015 conflict [July–December]	v1004	0.3	0.716	0.673	0.825	0.742
2014 imagery, 2015 conflict [July–December]	v1004	0.35	0.728	0.688	0.825	0.75
2014 imagery, 2015 conflict [July–December]	v1004	0.4	0.716	0.689	0.775	0.729
2014 imagery, 2015 conflict [July–December]	v1004	0.45	0.691	0.683	0.7	0.691
2014 imagery, 2015 conflict [July–December]	v1004	0.5	0.691	0.683	0.7	0.691
2014 imagery, 2015 conflict [July–December]	v100	0.3	0.753	0.72	0.9	0.8
2014 imagery, 2015 conflict [July–December]	v100	0.35	0.808	0.783	0.9	0.837
2014 imagery, 2015 conflict [July–December]	v100	0.4	0.849	0.854	0.875	0.864
2014 imagery, 2015 conflict [July–December]	v100	0.45	0.849	0.872	0.85	0.861
2014 imagery, 2015 conflict [July–December]	v100	0.5	0.849	0.872	0.85	0.861
2014 imagery, 2015 conflict [January–December]	v1011	0.3	0.702	0.634	0.951	0.761
2014 imagery, 2015 conflict [January–December]	v1011	0.35	0.717	0.657	0.902	0.76
2014 imagery, 2015 conflict [January–December]	v1011	0.4	0.756	0.713	0.853	0.777
2014 imagery, 2015 conflict [January–December]	v1011	0.45	0.776	0.75	0.824	0.785
2014 imagery, 2015 conflict [January–December]	v1011	0.5	0.776	0.775	0.775	0.775
2014 imagery, 2015 conflict [January–December]	v1012	0.3	0.707	0.641	0.912	0.753
2014 imagery, 2015 conflict [January–December]	v1012	0.35	0.721	0.659	0.892	0.758
2014 imagery, 2015 conflict [January–December]	v1012	0.4	0.731	0.683	0.843	0.754
2014 imagery, 2015 conflict [January–December]	v1012	0.45	0.707	0.678	0.765	0.719
2014 imagery, 2015 conflict [January–December]	v1012	0.5	0.731	0.725	0.725	0.725
2014 imagery, 2015 conflict [January–December]	v1013	0.3	0.714	0.655	0.931	0.769
2014 imagery, 2015 conflict [January–December]	v1013	0.35	0.714	0.664	0.892	0.762
2014 imagery, 2015 conflict [January–December]	v1013	0.4	0.739	0.702	0.853	0.77
2014 imagery, 2015 conflict [January–December]	v1013	0.45	0.769	0.759	0.804	0.781
2014 imagery, 2015 conflict [January–December]	v1013	0.5	0.789	0.813	0.765	0.788
2014 imagery, 2015 conflict [January–December]	v1014	0.3	0.725	0.656	0.971	0.783
2014 imagery, 2015 conflict [January–December]	v1014	0.35	0.745	0.678	0.951	0.792
2014 imagery, 2015 conflict [January–December]	v1014	0.4	0.76	0.699	0.931	0.798
2014 imagery, 2015 conflict [January–December]	v1014	0.45	0.755	0.715	0.863	0.782
2014 imagery, 2015 conflict [January–December]	v1014	0.5	0.77	0.755	0.814	0.783
2014 imagery, 2015 conflict [January–December]	v101	0.3	0.723	0.657	0.922	0.767
2014 imagery, 2015 conflict [January–December]	v101	0.35	0.733	0.669	0.912	0.772
2014 imagery, 2015 conflict [January–December]	v101	0.4	0.728	0.674	0.873	0.761
2014 imagery, 2015 conflict [January–December]	v101	0.45	0.743	0.693	0.863	0.769
2014 imagery, 2015 conflict [January–December]	v101	0.5	0.723	0.706	0.755	0.73
2014 imagery, 2015 conflict [January–June]	v1001	0.3	0.733	0.708	0.742	0.724
2014 imagery, 2015 conflict [January–June]	v1001	0.35	0.733	0.714	0.726	0.72
2014 imagery, 2015 conflict [January–June]	v1001	0.4	0.74	0.726	0.726	0.726
2014 imagery, 2015 conflict [January–June]	v1001	0.45	0.748	0.738	0.726	0.732
2014 imagery, 2015 conflict [January–June]	v1001	0.5	0.756	0.75	0.726	0.738
2014 imagery, 2015 conflict [January–June]	v1002	0.3	0.694	0.662	0.79	0.721

Table A2. Cont.

Data	Version	Thresh	Accuracy	Precision	Recall	f1
2014 imagery, 2015 conflict [January–June]	v1002	0.35	0.718	0.69	0.79	0.737
2014 imagery, 2015 conflict [January–June]	v1002	0.4	0.726	0.7	0.79	0.742
2014 imagery, 2015 conflict [January–June]	v1002	0.45	0.75	0.738	0.774	0.756
2014 imagery, 2015 conflict [January–June]	v1002	0.5	0.75	0.746	0.758	0.752
2014 imagery, 2015 conflict [January–June]	v1003	0.3	0.718	0.689	0.81	0.745
2014 imagery, 2015 conflict [January–June]	v1003	0.35	0.702	0.681	0.778	0.726
2014 imagery, 2015 conflict [January–June]	v1003	0.4	0.718	0.706	0.762	0.733
2014 imagery, 2015 conflict [January–June]	v1003	0.45	0.718	0.712	0.746	0.729
2014 imagery, 2015 conflict [January–June]	v1003	0.5	0.71	0.708	0.73	0.719
2014 imagery, 2015 conflict [January–June]	v1004	0.3	0.682	0.678	0.645	0.661
2014 imagery, 2015 conflict [January–June]	v1004	0.35	0.682	0.684	0.629	0.655
2014 imagery, 2015 conflict [January–June]	v1004	0.4	0.698	0.725	0.597	0.655
2014 imagery, 2015 conflict [January–June]	v1004	0.45	0.698	0.735	0.581	0.649
2014 imagery, 2015 conflict [January–June]	v1004	0.5	0.721	0.795	0.565	0.66
2014 imagery, 2015 conflict [January–June]	v100	0.3	0.686	0.647	0.887	0.748
2014 imagery, 2015 conflict [January–June]	v100	0.35	0.695	0.659	0.871	0.75
2014 imagery, 2015 conflict [January–June]	v100	0.4	0.72	0.688	0.855	0.763
2014 imagery, 2015 conflict [January–June]	v100	0.45	0.72	0.704	0.806	0.752
2014 imagery, 2015 conflict [January–June]	v100	0.5	0.763	0.774	0.774	0.774
2018 imagery, 2019 conflict [January–June]	v1001	0.3	0.603	0.583	0.708	0.64
2018 imagery, 2019 conflict [January–June]	v1001	0.35	0.603	0.592	0.652	0.62
2018 imagery, 2019 conflict [January–June]	v1001	0.4	0.615	0.611	0.618	0.615
2018 imagery, 2019 conflict [January–June]	v1001	0.45	0.609	0.617	0.562	0.588
2018 imagery, 2019 conflict [January–June]	v1001	0.5	0.637	0.662	0.551	0.601
2018 imagery, 2019 conflict [January–June]	v1002	0.3	0.674	0.64	0.798	0.71
2018 imagery, 2019 conflict [January–June]	v1002	0.35	0.657	0.63	0.764	0.69
2018 imagery, 2019 conflict [January–June]	v1002	0.4	0.652	0.634	0.719	0.674
2018 imagery, 2019 conflict [January–June]	v1002	0.45	0.652	0.642	0.685	0.663
2018 imagery, 2019 conflict [January–June]	v1002	0.5	0.68	0.686	0.663	0.674
2018 imagery, 2019 conflict [January–June]	v1003	0.3	0.607	0.573	0.843	0.682
2018 imagery, 2019 conflict [January–June]	v1003	0.35	0.624	0.587	0.831	0.688
2018 imagery, 2019 conflict [January–June]	v1003	0.4	0.652	0.619	0.787	0.693
2018 imagery, 2019 conflict [January–June]	v1003	0.45	0.663	0.636	0.764	0.694
2018 imagery, 2019 conflict [January–June]	v1003	0.5	0.657	0.635	0.742	0.684
2018 imagery, 2019 conflict [January–June]	v1004	0.3	0.596	0.557	0.867	0.678
2018 imagery, 2019 conflict [January–June]	v1004	0.35	0.612	0.573	0.833	0.679
2018 imagery, 2019 conflict [January–June]	v1004	0.4	0.617	0.579	0.811	0.676
2018 imagery, 2019 conflict [January–June]	v1004	0.45	0.612	0.583	0.744	0.654
2018 imagery, 2019 conflict [January–June]	v1004	0.5	0.628	0.604	0.711	0.653
2018 imagery, 2019 conflict [January–June]	v100	0.3	0.661	0.614	0.876	0.722
2018 imagery, 2019 conflict [January–June]	v100	0.35	0.689	0.649	0.831	0.729
2018 imagery, 2019 conflict [January–June]	v100	0.4	0.695	0.661	0.809	0.727
2018 imagery, 2019 conflict [January–June]	v100	0.45	0.689	0.663	0.775	0.715
2018 imagery, 2019 conflict [January–June]	v100	0.5	0.718	0.701	0.764	0.731
2018 imagery, 2019 conflict [January–December]	v1011	0.3	0.623	0.598	0.743	0.663
2018 imagery, 2019 conflict [January–December]	v1011	0.35	0.612	0.595	0.694	0.641
2018 imagery, 2019 conflict [January–December]	v1011	0.4	0.612	0.6	0.667	0.632
2018 imagery, 2019 conflict [January–December]	v1011	0.45	0.619	0.616	0.625	0.621
2018 imagery, 2019 conflict [January–December]	v1011	0.5	0.644	0.652	0.611	0.631
2018 imagery, 2019 conflict [January–December]	v1012	0.3	0.608	0.565	0.972	0.714
2018 imagery, 2019 conflict [January–December]	v1012	0.35	0.643	0.591	0.951	0.729
2018 imagery, 2019 conflict [January–December]	v1012	0.4	0.703	0.644	0.917	0.756
2018 imagery, 2019 conflict [January–December]	v1012	0.45	0.745	0.694	0.882	0.777
2018 imagery, 2019 conflict [January–December]	v1012	0.5	0.72	0.719	0.729	0.724
2018 imagery, 2019 conflict [January–December]	v1013	0.3	0.612	0.559	0.917	0.695
2018 imagery, 2019 conflict [January–December]	v1013	0.35	0.639	0.581	0.896	0.705
2018 imagery, 2019 conflict [January–December]	v1013	0.4	0.642	0.591	0.833	0.692
2018 imagery, 2019 conflict [January–December]	v1013	0.45	0.656	0.613	0.771	0.683

Table A2. Cont.

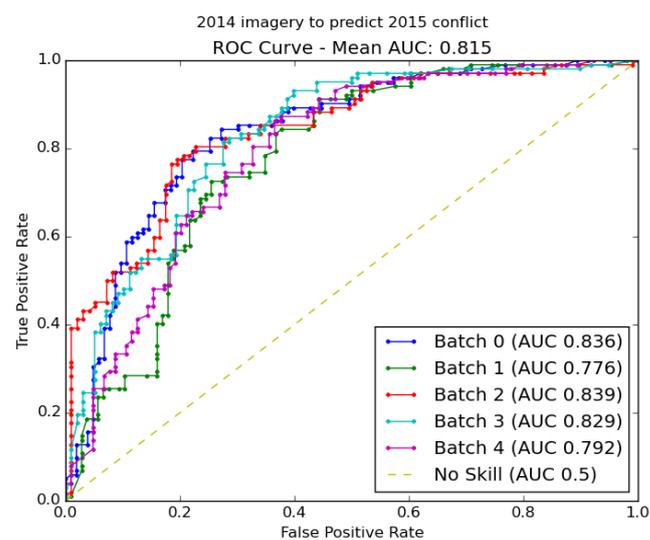
Data	Version	Thresh	Accuracy	Precision	Recall	f1
2018 imagery, 2019 conflict [January–December]	v1013	0.5	0.676	0.646	0.722	0.682
2018 imagery, 2019 conflict [January–December]	v1014	0.3	0.622	0.585	0.855	0.695
2018 imagery, 2019 conflict [January–December]	v1014	0.35	0.615	0.584	0.814	0.68
2018 imagery, 2019 conflict [January–December]	v1014	0.4	0.618	0.592	0.779	0.673
2018 imagery, 2019 conflict [January–December]	v1014	0.45	0.635	0.622	0.703	0.66
2018 imagery, 2019 conflict [January–December]	v1014	0.5	0.635	0.639	0.634	0.637
2018 imagery, 2019 conflict [January–December]	v101	0.3	0.644	0.588	0.951	0.727
2018 imagery, 2019 conflict [January–December]	v101	0.35	0.682	0.621	0.924	0.743
2018 imagery, 2019 conflict [January–December]	v101	0.4	0.706	0.651	0.882	0.749
2018 imagery, 2019 conflict [January–December]	v101	0.45	0.716	0.672	0.84	0.747
2018 imagery, 2019 conflict [January–December]	v101	0.5	0.699	0.677	0.757	0.715
2016 imagery, 2017 conflict [January–December]	v1011	0.3	0.624	0.574	0.957	0.718
2016 imagery, 2017 conflict [January–December]	v1011	0.35	0.632	0.582	0.94	0.719
2016 imagery, 2017 conflict [January–December]	v1011	0.4	0.662	0.608	0.915	0.73
2016 imagery, 2017 conflict [January–December]	v1011	0.45	0.697	0.646	0.872	0.742
2016 imagery, 2017 conflict [January–December]	v1011	0.5	0.731	0.696	0.821	0.753
2016 imagery, 2017 conflict [January–December]	v1012	0.3	0.746	0.699	0.847	0.766
2016 imagery, 2017 conflict [January–December]	v1012	0.35	0.742	0.715	0.788	0.75
2016 imagery, 2017 conflict [January–December]	v1012	0.4	0.758	0.742	0.78	0.76
2016 imagery, 2017 conflict [January–December]	v1012	0.45	0.771	0.769	0.763	0.766
2016 imagery, 2017 conflict [January–December]	v1012	0.5	0.771	0.784	0.737	0.76
2016 imagery, 2017 conflict [January–December]	v1013	0.3	0.734	0.688	0.829	0.752
2016 imagery, 2017 conflict [January–December]	v1013	0.35	0.743	0.707	0.803	0.752
2016 imagery, 2017 conflict [January–December]	v1013	0.4	0.763	0.738	0.795	0.765
2016 imagery, 2017 conflict [January–December]	v1013	0.45	0.772	0.763	0.769	0.766
2016 imagery, 2017 conflict [January–December]	v1013	0.5	0.78	0.776	0.769	0.773
2016 imagery, 2017 conflict [January–December]	v1014	0.3	0.727	0.691	0.803	0.743
2016 imagery, 2017 conflict [January–December]	v1014	0.35	0.735	0.701	0.803	0.749
2016 imagery, 2017 conflict [January–December]	v1014	0.4	0.739	0.717	0.778	0.746
2016 imagery, 2017 conflict [January–December]	v1014	0.45	0.735	0.725	0.744	0.734
2016 imagery, 2017 conflict [January–December]	v1014	0.5	0.752	0.764	0.718	0.74
2016 imagery, 2017 conflict [January–December]	v101	0.3	0.733	0.688	0.846	0.759
2016 imagery, 2017 conflict [January–December]	v101	0.35	0.725	0.686	0.821	0.747
2016 imagery, 2017 conflict [January–December]	v101	0.4	0.733	0.696	0.821	0.753
2016 imagery, 2017 conflict [January–December]	v101	0.45	0.737	0.71	0.795	0.75
2016 imagery, 2017 conflict [January–December]	v101	0.5	0.763	0.748	0.786	0.767
2018 imagery, 2019 conflict [July–December]	v1001	0.3	0.609	0.581	0.655	0.615
2018 imagery, 2019 conflict [July–December]	v1001	0.35	0.643	0.621	0.655	0.637
2018 imagery, 2019 conflict [July–December]	v1001	0.4	0.643	0.625	0.636	0.631
2018 imagery, 2019 conflict [July–December]	v1001	0.45	0.652	0.647	0.6	0.623
2018 imagery, 2019 conflict [July–December]	v1001	0.5	0.661	0.674	0.564	0.614
2018 imagery, 2019 conflict [July–December]	v1002	0.3	0.6	0.594	0.704	0.644
2018 imagery, 2019 conflict [July–December]	v1002	0.35	0.6	0.597	0.685	0.638
2018 imagery, 2019 conflict [July–December]	v1002	0.4	0.6	0.607	0.63	0.618
2018 imagery, 2019 conflict [July–December]	v1002	0.45	0.629	0.642	0.63	0.636
2018 imagery, 2019 conflict [July–December]	v1002	0.5	0.638	0.667	0.593	0.627
2018 imagery, 2019 conflict [July–December]	v1003	0.3	0.689	0.683	0.782	0.729
2018 imagery, 2019 conflict [July–December]	v1003	0.35	0.699	0.7	0.764	0.73
2018 imagery, 2019 conflict [July–December]	v1003	0.4	0.68	0.69	0.727	0.708
2018 imagery, 2019 conflict [July–December]	v1003	0.45	0.65	0.686	0.636	0.66
2018 imagery, 2019 conflict [July–December]	v1003	0.5	0.68	0.75	0.6	0.667
2018 imagery, 2019 conflict [July–December]	v1004	0.3	0.552	0.518	0.8	0.629
2018 imagery, 2019 conflict [July–December]	v1004	0.35	0.595	0.551	0.782	0.647
2018 imagery, 2019 conflict [July–December]	v1004	0.4	0.612	0.569	0.745	0.646
2018 imagery, 2019 conflict [July–December]	v1004	0.45	0.655	0.623	0.691	0.655
2018 imagery, 2019 conflict [July–December]	v1004	0.5	0.647	0.646	0.564	0.602
2018 imagery, 2019 conflict [July–December]	v100	0.3	0.624	0.609	0.709	0.655
2018 imagery, 2019 conflict [July–December]	v100	0.35	0.633	0.627	0.673	0.649

Table A2. Cont.

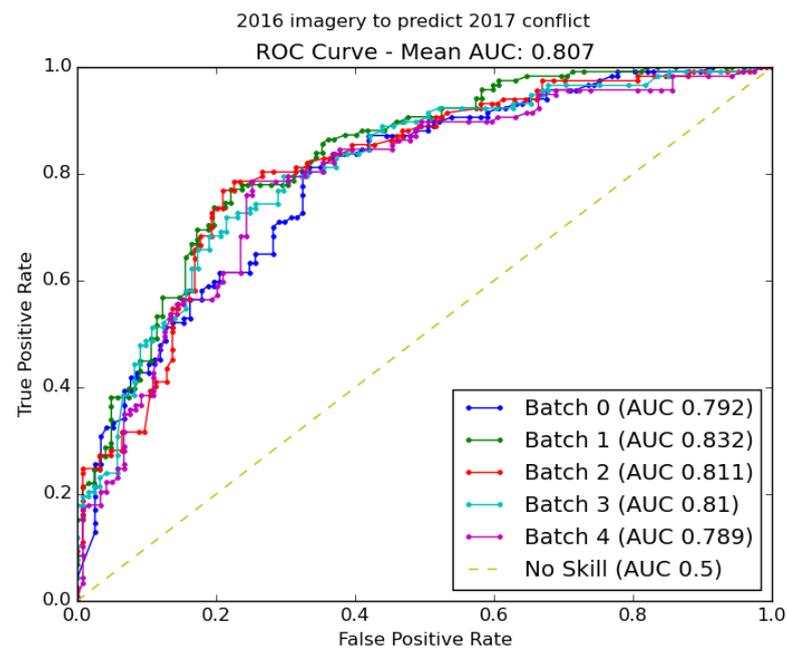
Data	Version	Thresh	Accuracy	Precision	Recall	f1
2018 imagery, 2019 conflict [July–December]	v100	0.4	0.633	0.632	0.655	0.643
2018 imagery, 2019 conflict [July–December]	v100	0.45	0.633	0.636	0.636	0.636
2018 imagery, 2019 conflict [July–December]	v100	0.5	0.642	0.674	0.564	0.614
2016 imagery, 2017 conflict [January–June]	v1001	0.3	0.63	0.608	0.75	0.672
2016 imagery, 2017 conflict [January–June]	v1001	0.35	0.655	0.638	0.733	0.682
2016 imagery, 2017 conflict [January–June]	v1001	0.4	0.655	0.646	0.7	0.672
2016 imagery, 2017 conflict [January–June]	v1001	0.45	0.681	0.677	0.7	0.689
2016 imagery, 2017 conflict [January–June]	v1001	0.5	0.723	0.745	0.683	0.713
2016 imagery, 2017 conflict [January–June]	v1002	0.3	0.702	0.636	0.933	0.757
2016 imagery, 2017 conflict [January–June]	v1002	0.35	0.711	0.644	0.933	0.762
2016 imagery, 2017 conflict [January–June]	v1002	0.4	0.736	0.679	0.883	0.768
2016 imagery, 2017 conflict [January–June]	v1002	0.45	0.76	0.707	0.883	0.785
2016 imagery, 2017 conflict [January–June]	v1002	0.5	0.76	0.707	0.883	0.785
2016 imagery, 2017 conflict [January–June]	v1003	0.3	0.689	0.649	0.833	0.73
2016 imagery, 2017 conflict [January–June]	v1003	0.35	0.681	0.662	0.75	0.703
2016 imagery, 2017 conflict [January–June]	v1003	0.4	0.697	0.694	0.717	0.705
2016 imagery, 2017 conflict [January–June]	v1003	0.45	0.739	0.754	0.717	0.735
2016 imagery, 2017 conflict [January–June]	v1003	0.5	0.765	0.796	0.717	0.754
2016 imagery, 2017 conflict [January–June]	v1004	0.3	0.672	0.63	0.85	0.723
2016 imagery, 2017 conflict [January–June]	v1004	0.35	0.655	0.627	0.783	0.696
2016 imagery, 2017 conflict [January–June]	v1004	0.4	0.681	0.657	0.767	0.708
2016 imagery, 2017 conflict [January–June]	v1004	0.45	0.689	0.667	0.767	0.713
2016 imagery, 2017 conflict [January–June]	v1004	0.5	0.697	0.676	0.767	0.719
2016 imagery, 2017 conflict [January–June]	v100	0.3	0.694	0.649	0.833	0.73
2016 imagery, 2017 conflict [January–June]	v100	0.35	0.702	0.662	0.817	0.731
2016 imagery, 2017 conflict [January–June]	v100	0.4	0.694	0.658	0.8	0.722
2016 imagery, 2017 conflict [January–June]	v100	0.45	0.736	0.712	0.783	0.746
2016 imagery, 2017 conflict [January–June]	v100	0.5	0.744	0.723	0.783	0.752

## Appendix A.3. ROC Curves

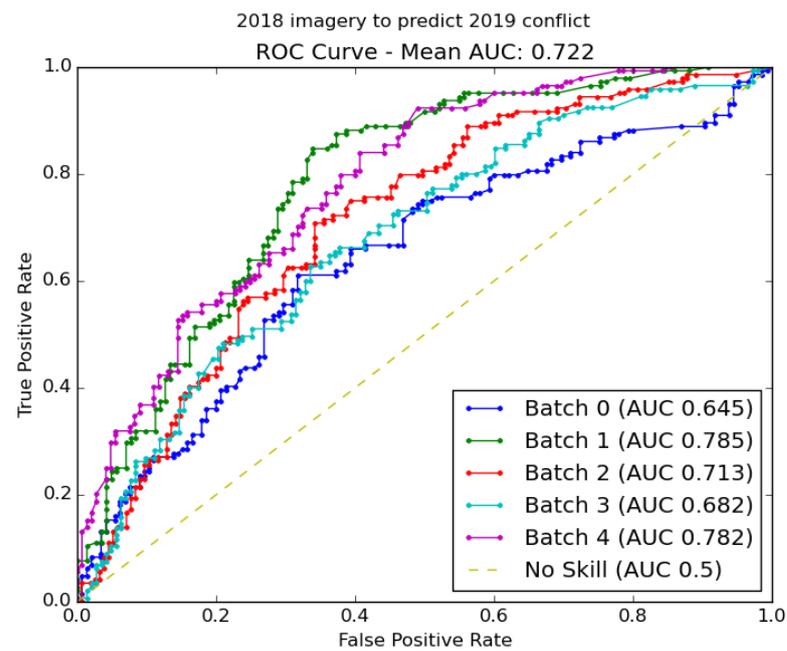
Each Figure includes the curve and AUC value for each of the five sample data frames tested for a combination of time periods and temporal windows of data. *h1* for a given year refers to January–June and *h2* refers to July–December.



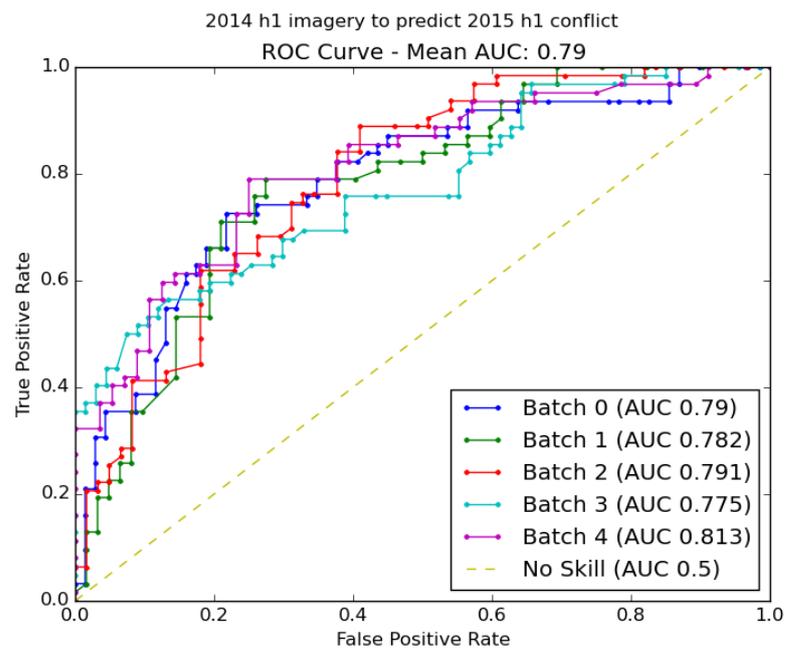
**Figure A1.** ROC curve produced from model train on 2014 (January–December) imagery to predict the probability of a fatality if there is a conflict event in 2015 (January–December).



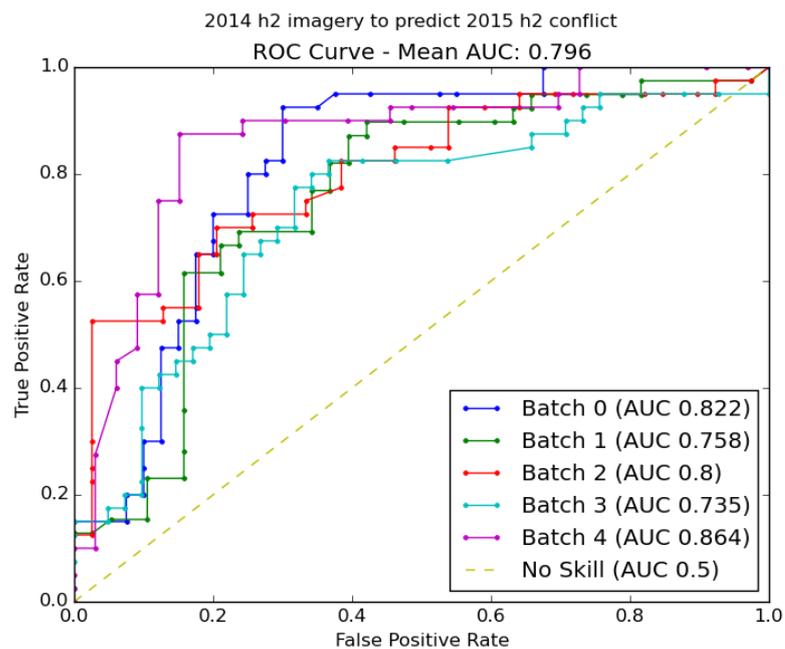
**Figure A2.** ROC curve produced from model train on 2015 (January–December) imagery to predict the probability of a fatality if there is a conflict event in 2016 (January–December).



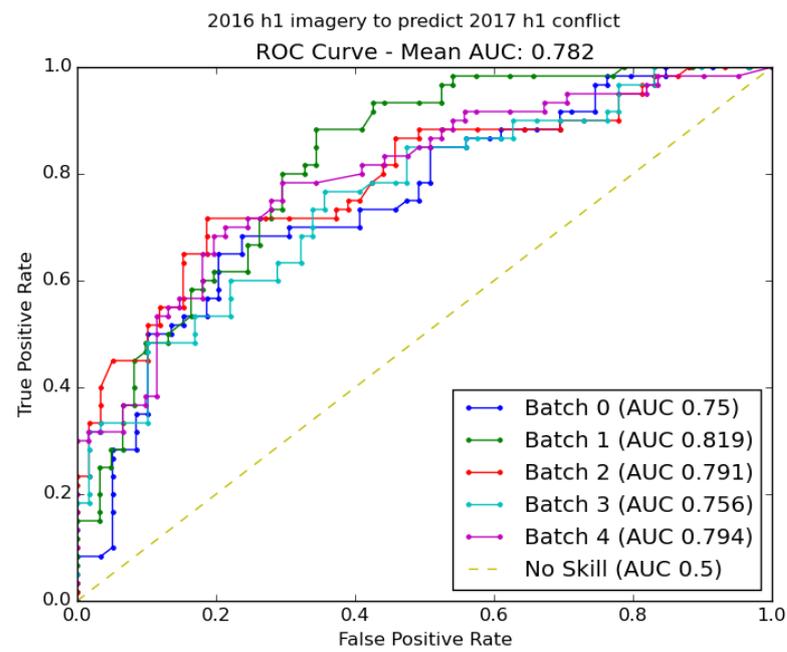
**Figure A3.** ROC curve produced from model train on 2018 (January–December) imagery to predict the probability of a fatality if there is a conflict event in 2019 (January–December).



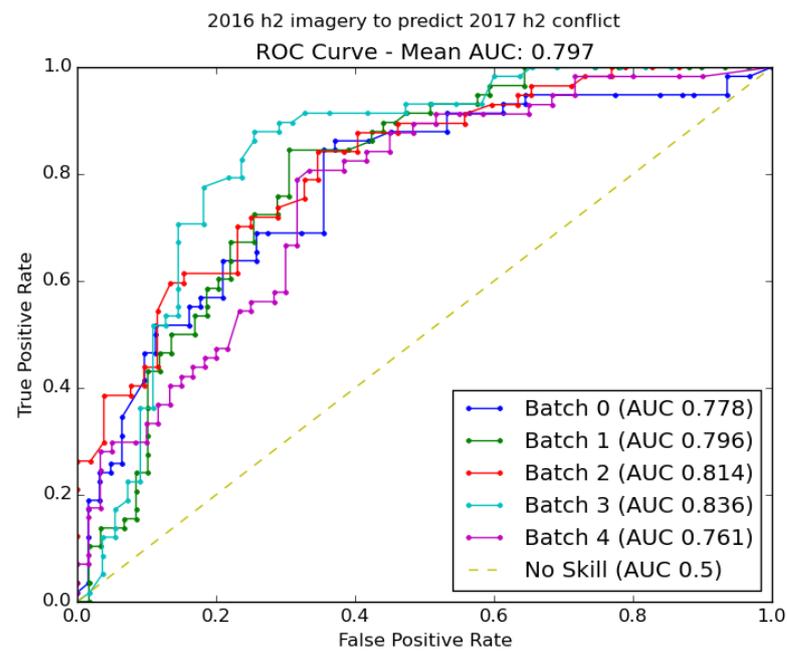
**Figure A4.** ROC curve produced from model train on 2014 h1 (January–June) imagery to predict the probability of a fatality if there is a conflict event in 2015 h1 (January–June).



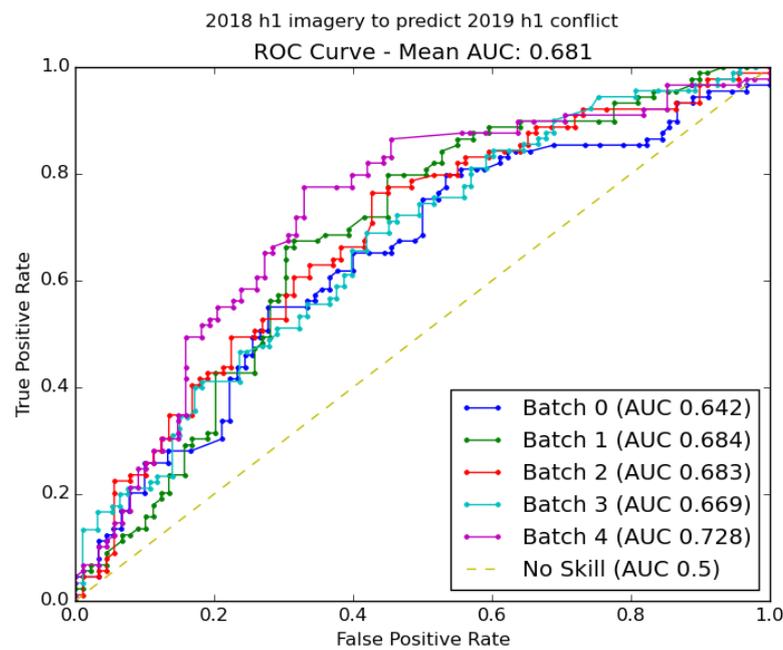
**Figure A5.** ROC curve produced from model train on 2014 h2 (July–December) imagery to predict the probability of a fatality if there is a conflict event in 2015 h2 (July–December).



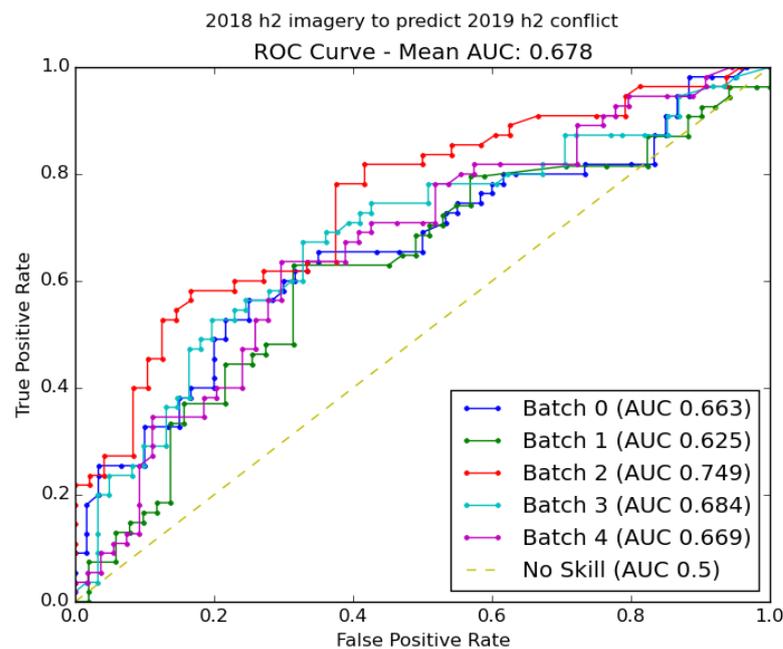
**Figure A6.** ROC curve produced from model train on 2016 h1 (January–June) imagery to predict the probability of a fatality if there is a conflict event in 2017 h1 (January–June).



**Figure A7.** ROC curve produced from model train on 2016 h2 (July–December) imagery to predict the probability of a fatality if there is a conflict event in 2017 h2 (July–December).



**Figure A8.** ROC curve produced from model train on 2018 h1 (January–June) imagery to predict the probability of a fatality if there is a conflict event in 2019 h1 (January–June).



**Figure A9.** ROC curve produced from model train on 2018 h2 (July–December) imagery to predict the probability of a fatality if there is a conflict event in 2019 h2 (July–December).

## References

- Herbert, S.; Husaini, S. *Conflict, Instability and Resilience in Nigeria. Rapid Literature Review*; Technical Report; University of Birmingham: Birmingham, UK, 2018.
- The Armed Conflict Location & Event Data Project. ACLED Data Dashboard. 2019. Available online: <https://acleddata.com/data/> (accessed on 1 July 2024).
- Raleigh, C.; Linke, A.; Hegre, H.; Karlsen, J. Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. *J. Peace Res.* **2010**, *47*, 651–660. [[CrossRef](#)]
- Usman, M.; Nichol, J.E. Changes in agricultural and grazing land, and insights for mitigating farmer-herder conflict in West Africa. *Landsc. Urban Plan.* **2022**, *222*, 104383. [[CrossRef](#)]

5. WorldPop; CIESIN. Global High Resolution Population Denominators Project. 2018. Available online: <https://hub.worldpop.org/doi/10.5258/SOTON/WP00647> (accessed on 1 July 2024).
6. European Space Agency. ESA Land Cover. 2017. Available online: <https://www.esa-landcover-cci.org/> (accessed on 1 July 2024).
7. Matsuura, K.; Willmott, C.J. Terrestrial Precipitation: 1900–2014 Gridded Monthly Time Series. 2015. Available online: <https://climate.geog.udel.edu/> (accessed on 1 July 2024).
8. Goodman, S.; BenYishay, A.; Lv, Z.; Runfola, D. GeoQuery: Integrating HPC systems and public web-based geospatial data tools. *Comput. Geosci.* **2019**, *122*, 103–112. [[CrossRef](#)]
9. Burke, M.; Driscoll, A.; Lobell, D.B.; Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **2021**, *371*, 6535. [[CrossRef](#)]
10. Farabet, C.; Couprie, C.; Najman, L.; LeCun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [[CrossRef](#)]
11. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
13. Albelwi, S.; Mahmood, A. A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy* **2017**, *19*, 242. [[CrossRef](#)]
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [[CrossRef](#)]
16. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1349–1362. [[CrossRef](#)]
17. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.Q. Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [[CrossRef](#)]
18. Chen, X.; Xiang, S.; Liu, C.L.; Pan, C.H. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1797–1801. [[CrossRef](#)]
19. Vakalopoulou, M.; Karantzas, K.; Komodakis, N.; Paragios, N. Building detection in very high resolution multispectral data with deep learning features. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2015; pp. 1873–1876. [[CrossRef](#)]
20. Jean, N.; Burke, M.; Xie, M.; Davis, W.M.; Lobell, D.B.; Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* **2016**, *353*, 790–794. [[CrossRef](#)] [[PubMed](#)]
21. Marmanis, D.; Datcu, M.; Esch, T.; Stilla, U. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 105–109. [[CrossRef](#)]
22. Chi, G.; Fang, H.; Chatterjee, S.; Blumenstock, J.E. Microestimates of wealth for all low- and middle-income countries. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2113658119. [[CrossRef](#)]
23. Yeh, C.; Perez, A.; Driscoll, A.; Azzari, G.; Tang, Z.; Lobell, D.; Ermon, S.; Burke, M. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nat. Commun.* **2020**, *11*, 2583. [[CrossRef](#)]
24. Babenko, B.; Hersh, J.; Newhouse, D.; Ramakrishnan, A.; Swartz, T. Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico. *arXiv* **2017**, arXiv:1711.06323.
25. Subash, S.P.; Kumar, R.R.; Aditya, K.S. Satellite data and machine learning tools for predicting poverty in rural India. *Agric. Econ. Res. Rev.* **2018**, *31*, 231. [[CrossRef](#)]
26. Engstrom, R.; Hersh, J.S.; Newhouse, D.L. *Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being*; Technical Report; The World Bank: Washington, DC, USA, 2017.
27. Oshri, B.; Hu, A.; Adelson, P.; Chen, X.; Dupas, P.; Weinstein, J.; Burke, M.; Lobell, D.; Ermon, S. Infrastructure Quality Assessment in Africa Using Satellite Imagery and Deep Learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18, London, UK, 19–23 August 2018; ACM: New York, NY, USA, 2018; pp. 616–625. [[CrossRef](#)]
28. Brewer, E.; Lin, J.; Kemper, P.; Hennin, J.; Runfola, D. Predicting road quality using high resolution satellite imagery: A transfer learning approach. *PLoS ONE* **2021**, *16*, e0253370. [[CrossRef](#)]
29. Lv, Z.; Nunez, K.; Brewer, E.; Runfola, D. pyShore: A deep learning toolkit for shoreline structure mapping with high-resolution orthographic imagery and convolutional neural networks. *Comput. Geosci.* **2023**, *171*, 105296. [[CrossRef](#)]
30. Lobell, D.B.; Di Tommaso, S.; You, C.; Djima, I.Y.; Burke, M.; Kilic, T. Sight for sorghums: Comparisons of satellite-and ground-based sorghum yield estimates in Mali. *Remote Sens.* **2020**, *12*, 100. [[CrossRef](#)]
31. Runfola, D.; Stefanidis, A.; Baier, H. Using satellite data and deep learning to estimate educational outcomes in data-sparse environments. *Remote Sens. Lett.* **2021**, *13*, 87–97. [[CrossRef](#)]
32. Head, A.; Manguin, M.; Tran, N.; Blumenstock, J.E. Can Human Development Be Measured with Satellite Imagery? In Proceedings of the Ninth International Conference on Information and Communication Technologies and Development, ICTD '17, Lahore, Pakistan, 16–19 November 2017; ACM: New York, NY, USA, 2017; pp. 8:1–8:11. [[CrossRef](#)]

33. Hu, W.; Novosad, P.; Burke, M.; Patel, J.H.; Asher, S.; Lobell, D.; Robert, Z.A.; Tang, Z.; Ermon, S. Mapping missing population in rural India: A deep learning approach with satellite imagery. In Proceedings of the AIES 2019—2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–28 January 2019; Association for Computing Machinery, Inc.: New York, NY, USA, 2019; pp. 353–359. [\[CrossRef\]](#)
34. Runfola, D.; Baier, H.; Mills, L.; Naughton-Rockwell, M.; Stefanidis, A. Deep learning fusion of satellite and social information to estimate human migratory flows. *Trans. GIS* **2022**, *26*, 2495–2518. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Goodman, S.; BenYishay, A.; Runfola, D. A convolutional neural network approach to predict non-permissive environments from moderate-resolution imagery. *Trans. GIS* **2020**, *25*, 674–691. [\[CrossRef\]](#)
36. Runfola, D.; Stefanidis, A.; Lv, Z.; O'Brien, J.; Baier, H. A multi-glimpse deep learning architecture to estimate socioeconomic census metrics in the context of extreme scope variance. *Int. J. Geogr. Inf. Sci.* **2024**, *38*, 726–750. [\[CrossRef\]](#)
37. Buntaine, M.T.; Hamilton, S.E.; Millones, M. Titling community land to prevent deforestation: An evaluation of a best-case program in Morona-Santiago, Ecuador. *Glob. Environ. Chang.* **2015**, *33*, 32–43. [\[CrossRef\]](#)
38. BenYishay, A.; Runfola, D.; Trichler, R.; Dolan, C.; Goodman, S.; Parks, B.; Tanner, J.; Heuser, S.; Batra, G.; Anand, A. *A Primer on Geospatial Impact Evaluation Methods, Tools, and Applications*; AidData Working Paper #44; AidData at William & Mary: Williamsburg, VA, USA, 2017.
39. Runfola, D.; BenYishay, A.; Tanner, J.; Buchanan, G.; Nagol, J.; Leu, M.; Goodman, S.; Trichler, R.; Marty, R. A Top-Down Approach to Estimating Spatially Heterogeneous Impacts of Development Aid on Vegetative Carbon Sequestration. *Sustainability* **2017**, *9*, 409. [\[CrossRef\]](#)
40. Marty, R.; Goodman, S.; LeFevre, M.; Dolan, C.; BenYishay, A.; Runfola, D. Assessing the causal impact of Chinese aid on vegetative land cover in Burundi and Rwanda under conditions of spatial imprecision. *Dev. Eng.* **2019**, *4*, 100038. [\[CrossRef\]](#)
41. Runfola, D.; Batra, G.; Anand, A.; Way, A.; Goodman, S. Exploring the Socioeconomic Co-benefits of Global Environment Facility Projects in Uganda Using a Quasi-Experimental Geospatial Interpolation (QGI) Approach. *Sustainability* **2020**, *12*, 3225. [\[CrossRef\]](#)
42. BenYishay, A.; Sayers, R.; Singh, K.; Goodman, S.; Walker, M.; Traore, S.; Rauschenbach, M.; Noltze, M. Irrigation strengthens climate resilience: Long-term evidence from Mali using satellites and surveys. *PNAS Nexus* **2024**, *3*, pga022. [\[CrossRef\]](#)
43. Bansal, C.; Jain, A.; Barwaria, P.; Choudhary, A.; Singh, A.; Gupta, A.; Seth, A. Temporal prediction of socio-economic indicators using satellite imagery. In Proceedings of the CoDS COMAD 2020: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, Hyderabad, India, 5–7 January 2020; ACM International Conference Proceeding Series; pp. 73–81. [\[CrossRef\]](#)
44. Signoroni, A.; Savardi, M.; Baronio, A.; Benini, S. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. *J. Imaging* **2019**, *5*, 52. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [\[CrossRef\]](#)
46. Gaetano, R.; Ienco, D. *TiSeLaC: Time Series Land Cover Classification Challenge Dataset*; UMR TETIS: Montpellier, France, 2017.
47. Martino, T.D. Time Series Land Cover Challenge: A Deep Learning Perspective. 2020. Available online: <https://towardsdatascience.com/time-series-land-cover-challenge-a-deep-learning-perspective-6a953368a2bd> (accessed on 1 July 2024)
48. Mauro, N.D.; Vergari, A.; Basile, T.; Ventola, F.; Esposito, F. End-to-end Learning of Deep Spatio-temporal Representations for Satellite Image Time Series Classification. In Proceedings of the DC@PKDD/ECML 2017, Skopje, Macedonia, 18–22 September 2017.
49. Garnot, V.S.F.; Landrieu, L.; Giordano, S.; Chehata, N. Satellite Image Time Series Classification with Pixel-Set Encoders and Temporal Self-Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 12325–12334.
50. Hegre, H.; Metternich, N.W.; Nygård, H.M.; Wucherpfennig, J. Introduction: Forecasting in peace research. *J. Peace Res.* **2017**, *54*, 113–124. [\[CrossRef\]](#)
51. Cederman, L.E.; Weidmann, N.B. Predicting armed conflict: Time to adjust our expectations? *Science* **2017**, *355*, 474–476. [\[CrossRef\]](#)
52. Bazzi, S.; Blair, R.A.; Blattman, C.; Dube, O.; Gudgeon, M.; Merton Peck, R. *The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia*; Technical Report; The National Bureau of Economic Research: Cambridge, MA, USA, 2019.
53. Weidmann, N.B.; Ward, M.D. Predicting Conflict in Space and Time. *J. Confl. Resolut.* **2010**, *54*, 883–901. [\[CrossRef\]](#)
54. Mueller, H.; Rauh, C. The Hard Problem of Prediction for Conflict Prevention. *J. Eur. Econ. Assoc.* **2022**, *20*, 2440–2467. [\[CrossRef\]](#)
55. Hegre, H.; Allansson, M.; Basedau, M.; Colaresi, M.; Croicu, M.; Fjelde, H.; Hoyles, F.; Hultman, L.; Höglbladh, S.; Jansen, R.; et al. ViEWS: A political violence early-warning system. *J. Peace Res.* **2019**, *56*, 155–174. [\[CrossRef\]](#)
56. Vesco, P.; Hegre, H.; Colaresi, M.; Jansen, R.B.; Lo, A.; Reisch, G.; Weidmann, N.B. United they stand: Findings from an escalation prediction competition. *Int. Interact.* **2022**, *48*, 860–896. [\[CrossRef\]](#)
57. Radford, B.J. High resolution conflict forecasting with spatial convolutions and long short-term memory. *Int. Interact.* **2022**, *48*, 739–758. [\[CrossRef\]](#)
58. Brandt, P.T.; D’Orazio, V.; Khan, L.; Li, Y.F.; Osorio, J.; Sianan, M. Conflict forecasting with event data and spatio-temporal graph convolutional networks. *Int. Interact.* **2022**, *48*, 800–822. [\[CrossRef\]](#)
59. D’Orazio, V.; Lin, Y. Forecasting conflict in Africa with automated machine learning systems. *Int. Interact.* **2022**, *48*, 714–738. [\[CrossRef\]](#)

60. Racek, D.; Thurner, P.W.; Davidson, B.I.; Zhu, X.X.; Kauermann, G. Conflict forecasting using remote sensing data: An application to the Syrian civil war. *Int. J. Forecast.* **2024**, *40*, 373–391. [[CrossRef](#)]
61. United States Geological Survey. *Landsat 7*; United States Geological Survey: Reston, VA, USA, 2018.
62. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [[CrossRef](#)]
63. United States Geological Survey. *USGS EarthExplorer*; United States Geological Survey: Reston, VA, USA, 2017.
64. United States Geological Survey. *USGS Landsat Bulk Download*; United States Geological Survey: Reston, VA, USA, 2017.
65. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [[CrossRef](#)]
66. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the Computer Vision and Pattern Recognition, CVPR 2009, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
67. Weiss, K.R.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40.
68. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
69. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 3320–3328.
70. Razavian, A.S.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 512–519.
71. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732.
72. Hoo-Chang, S.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285.
73. Lv, Z.; Nunez, K.; Brewer, E.; Runfola, D. Mapping the tidal marshes of coastal Virginia: A hierarchical transfer learning approach. *GISci. Remote Sens.* **2024**, *61*, 2287291. [[CrossRef](#)]
74. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
75. Eck, K. In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Coop. Confl.* **2012**, *47*, 124–141. [[CrossRef](#)]
76. Li, H.; Fu, K.; Xu, G.; Zheng, X.; Ren, W.; Sun, X. Scene classification in remote sensing images using a two-stage neural network ensemble model. *Remote Sens. Lett.* **2017**, *8*, 557–566. [[CrossRef](#)]
77. Zhang, C.; Pan, X.; Li, H.; Gardiner, A.; Sargent, I.; Hare, J.; Atkinson, P.M. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 133–144. [[CrossRef](#)]
78. Li, W.; Liu, H.; Wang, Y.; Li, Z.; Jia, Y.; Gui, G. Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas. *IEEE Access* **2019**, *7*, 36274–36284. [[CrossRef](#)]
79. Li, Y.; Zhang, H.; Shen, Q. Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
80. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* **2018**, *10*, 75. [[CrossRef](#)]
81. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 5987–5995. [[CrossRef](#)]
82. Flusser, J.; Suk, T. Pattern recognition by affine moment invariants. *Pattern Recognit.* **1993**, *26*, 167–174. [[CrossRef](#)]
83. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.
84. Han, D.; Liu, Q.; Fan, W. A new image classification method using CNN transfer learning and web data augmentation. *Expert Syst. Appl.* **2018**, *95*, 43–56. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.