

Mapping Urban Land Cover: A Novel Machine Learning Approach Using Landsat and Nighttime Lights

Abstract

Reliable large-scale representations of contemporary urban extent remain limited, hindering scientific progress across a range of disciplines aimed at helping create functional and sustainable cities. We present a novel, efficient, and low-cost machine-learning approach to map urban areas at large scales. Our methodology combines nighttime-lights data and Landsat 8 imagery using a transfer-learning approach that overcomes the lack of extensive ground-truth data. We demonstrate the effectiveness of our methodology through the development of high-quality 30m resolution maps that characterize urban areas in three diverse countries: India, Mexico, and the US. We implement our methodology in Google Earth Engine and show that it produces accurate maps of built-up land cover at high resolution over large spatial extents. Our approach highlights the usefulness of machine-learning techniques for studying the built environment, with broad implications for identification of urbanization drivers and effects on earth-system processes.

Keywords: Urbanization, built-up land cover, nighttime light, image classification

1. Introduction

Urbanization has been a fundamental trend of the past two centuries and a key force shaping almost every dimension of the modern world. In the period between 1950 and 2014, the share of the global population living in urban areas increased from 30% to 54%, and in the next few decades is projected to expand by an additional 2.5 billion urban dwellers, primarily in Asia and Africa (Seto et al., 2012; UN, 2014). Urban population growth is accompanied by a dramatic

23 increase in the land area incorporated in cities (Georgescu et al., 2015). While urbanization in
24 rapidly growing nations is helping lift hundreds of millions of people out of poverty, it is also
25 creating immense societal challenges by increasing greenhouse-gas emissions, destabilizing
26 fragile ecosystems and creating new demands on public services and infrastructure. Despite the
27 importance of understanding the drivers of urban growth, we are still unable to quantify the
28 magnitude and pace of urbanization in a consistent manner at high resolution and global scale.
29 Standard empirical approaches use data from household surveys that are expensive to collect,
30 produced infrequently, and subject to measurement problems.

31 The revolution in geospatial data transforms how we study cities. Since the 1970s, terrestrial
32 Earth-observation data has been continuously collected in various spectral, spatial and temporal
33 resolutions. As improved satellite imagery becomes available, new remote-sensing methods and
34 machine-learning approaches have been developed to convert terrestrial Earth-observation data
35 into meaningful information about the nature and pace of change of urban landscapes and human
36 settlements (CIESIN, 2005; Gaughan et al., 2013; Pesaresi et al., 2016; Potere et al., 2009; Seto et al.,
37 2011; Taubenböck et al., 2012). Existing maps of urban land show considerable disagreement on
38 the location and extent of urbanization (Potere et al., 2009; Seto et al., 2011) and are further subject
39 to limitations across space and time. These inconsistencies may arise in part because the
40 delineation of urban land depends on the nature of the input data (Schneider et al., 2010), which
41 may capture different dimensions of urbanization, such as built-up land cover or land use and
42 population density (Bagan and Yamagata, 2014; Stevens et al., 2015; Tatem et al., 2007).

43 Since the early 1990's, data on nighttime lights have been amassed, primarily from sensors on
44 board the Operational Line-scan System of the Defense Meteorological Satellite Program (DMSP-
45 OLS). DMSP-OLS sensors capture artificial lighting, which is associated with developed land
46 (Elvidge et al., 2014; Levin and Duke, 2012; Sutton, 2003) and can be used to infer the extent of
47 urban areas (Bagan and Yamagata, 2015; Small and Elvidge, 2013; Zhang and Seto, 2013), as well
48 as economic activity at the local, regional and national levels (Elvidge et al., 2014; Henderson et
49 al., 2012; Keola et al., 2015). According to this approach, a pixel is considered urbanized if its
50 magnitude exceeds a threshold, where the appropriate threshold may vary across countries
51 (Small and Elvidge, 2013) and even across regions within a country (Henderson et al., 2003; Liu
52 et al., 2016; Su et al., 2015; Wei et al., 2014; Zhou et al., 2015, 2014). Thus, inference using nighttime-
53 light data is often inaccurate, especially in low-density urban areas (Zhang and Seto, 2013).
54 DMSP-OLS can also exaggerate the extent of urban areas (Henderson et al., 2003; Small et al.,
55 2005) while overlooking small or developing settlements. In addition, the extent and intensity of
56 lit areas cannot directly delimit urban regions due to the "blooming" effect (Imhoff et al., 1997)
57 and the "saturation" of the pixels (Hsu et al., 2015). "Blooming" refers to the identification of lit
58 areas as consistently larger than the settlements they are associated with (Small et al., 2005);
59 "saturation" occurs when pixels in bright areas, such as in city centers, reach the highest possible
60 digital number (DN) value (i.e., 63) and no further details can be recognized (Hsu et al., 2015).

61 *1.1. Detecting Urbanization Processes by Means of Machine Learning Approaches*

62 Urban areas can be detected in satellite imagery using various machine-learning approaches (e.g.,
63 supervised, unsupervised and semi-supervised). These approaches typically rely on ground-

64 truth data that mark urban features, either for training or for validation. Several datasets have
65 been previously proposed to serve as ground truth for urban research. These include Landsat-
66 based urban maps (Potere et al., 2009), census-based population databases (Stevens et al., 2015)
67 and hand-labeled examples (Goldblatt et al., 2016), as well as data collected via crowd-source
68 platforms, such as OpenStreetMap (OSM) (Belgiu and Dr ăgu

69 Despite significant progress in machine learning and the increasing availability of satellite data
70 that can be used as input for classification, there remains a paucity of ground-truth datasets that
71 have been developed to detect urban areas (Miyazaki et al., 2011). Previous studies have used
72 ground-truth datasets that are of limited size (Goldblatt et al., 2016; Trianni et al., 2015). Transfer
73 learning (or knowledge transfer) has emerged as a useful framework for machine learning and
74 image classification, including in remote sensing, in order to address the scarcity of ground-truth
75 data and to minimize the need for expensive labeling efforts. Transfer learning aims to transfer
76 knowledge between task domains, even where the training and test data are drawn from a
77 different feature space (Pan and Yang, 2010). Poverty prediction is an example of previous
78 application of transfer learning in the remote-sensing field (Jean et al., 2016).

79 Until recently, the majority of studies that analyze urbanization have been limited in scale
80 because of the lack of extensive high-resolution satellite data, scarcity of ground-truth data, and
81 computational constraints. However, emerging cloud-based computational platforms allow for
82 scaling analysis across space and time. Google Earth Engine (GEE) is an example of a platform
83 that leverages cloud-computational services to achieve planetary-scale utility. GEE has been
84 previously used for various research applications, including mapping population (Patel et al.,
85 2015; Trianni et al., 2015), urban areas (Goldblatt et al., 2016) and forest cover (Hansen et al., 2013).

86 **1.2. Research Objective**

87 This paper develops a novel machine-learning methodology for supervised image classification
88 of built-up areas that leverages high-resolution satellite data for analysis of large-scale regions
89 using GEE's cloud-based computational platform. Our methodology utilizes nighttime-light data
90 as the source of training data for classification of built-up areas, yielding high precision without
91 relying on expensive hand-labeled examples. It can be applied for any region on Earth.

92 Our methodology integrates two datasets: DSMP-OLS nighttime-light data and Landsat high-
93 resolution daytime satellite imagery. We infer the spatial distribution of human activity and built-
94 up land cover from nighttime-light data to collect examples for supervised image classification in
95 Landsat's imagery. We assess the accuracy of the methodology using a dataset of 60,000 hand-
96 labeled polygons characterizing built up (BU) and not built-up (NBU) pixels for each of the three
97 study areas.

98 **1.3. Study Area**

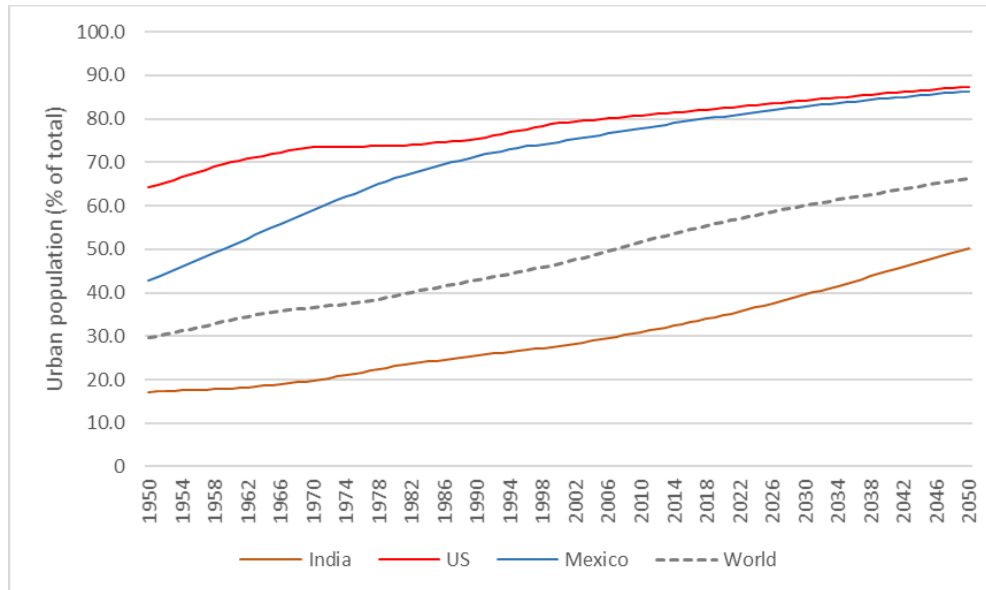
99 To illustrate our methodology and its applicability in heterogeneous and diverse geographical
100 conditions, we map the built-up land cover in three countries that are characterized by distinct
101 geographical conditions (e.g., land cover, topography, climate, soil, landform, and fauna): India,
102 Mexico, and the US.

103 **India.** The share of India's population living in urban areas in 2015 was 33%, which is much lower
104 than the corresponding values of 79% and 82% for Mexico and the US, respectively. However,
105 India is urbanizing at a relatively rapid rate (Figure 1). For example, between 2010 and 2015,
106 India's average annual rate of change of the urban population was 1.14%, compared to 0.36% and

107 0.21% in Mexico and in the US, respectively. By 2050, half of India's population is likely to be
108 urban. In the last decade, the growth of India's urban population outpaced the growth of its rural
109 population by 31.8% to 12.2% (H. S. Sudhira and K. V. Gururaja, n.d.)—primarily the result of
110 natural urban population growth and secondarily because of rural-to-urban migration (Buhaug
111 and Urdal, 2013)—a trend which is expected to continue (H. S. Sudhira and K. V. Gururaja, n.d.).

112 *Mexico.* Mexico has gone through three major phases of urbanization. In the first phase, 1900-
113 1940, urban growth slowly incorporated 10% to 20% of the population. In the second phase, 1940-
114 1980, rapid urban expansion, particularly in Mexico City, increased the share of urban population
115 to 55%. In the current phase, since the 1980s, more dispersed moderate urban growth increased
116 the urban share of the population to over 70% (Consejo Nacional de Población, 2012). The current
117 phase of urbanization is characterized by informal urbanization on the city periphery,
118 representing 65% of all new housing construction in Mexico City, and is even higher for small to
119 mid-sized cities (Connolly, 2014). Informal settlements tend to be marginalized in terms of lower
120 socio-economic development, access to services like water and electricity, and are more
121 vulnerable to risks like water scarcity and flooding (Aguilar, 2008; Aguilar and Guerrero, 2013;
122 Consejo Nacional de Población, 2012; Eakin et al., 2016).

123 *US.* In the context of the US, the initial urban growth occurred from 1790 to 1890 and the country
124 has become increasingly urban since (Census Bureau, 2012). In 1910, the Census Bureau defined
125 an urban area as one with a population above 2,500, and the 1920 census marked the first time
126 that 50% of the US population lived in an urban area (US Census Bureau, 2016).

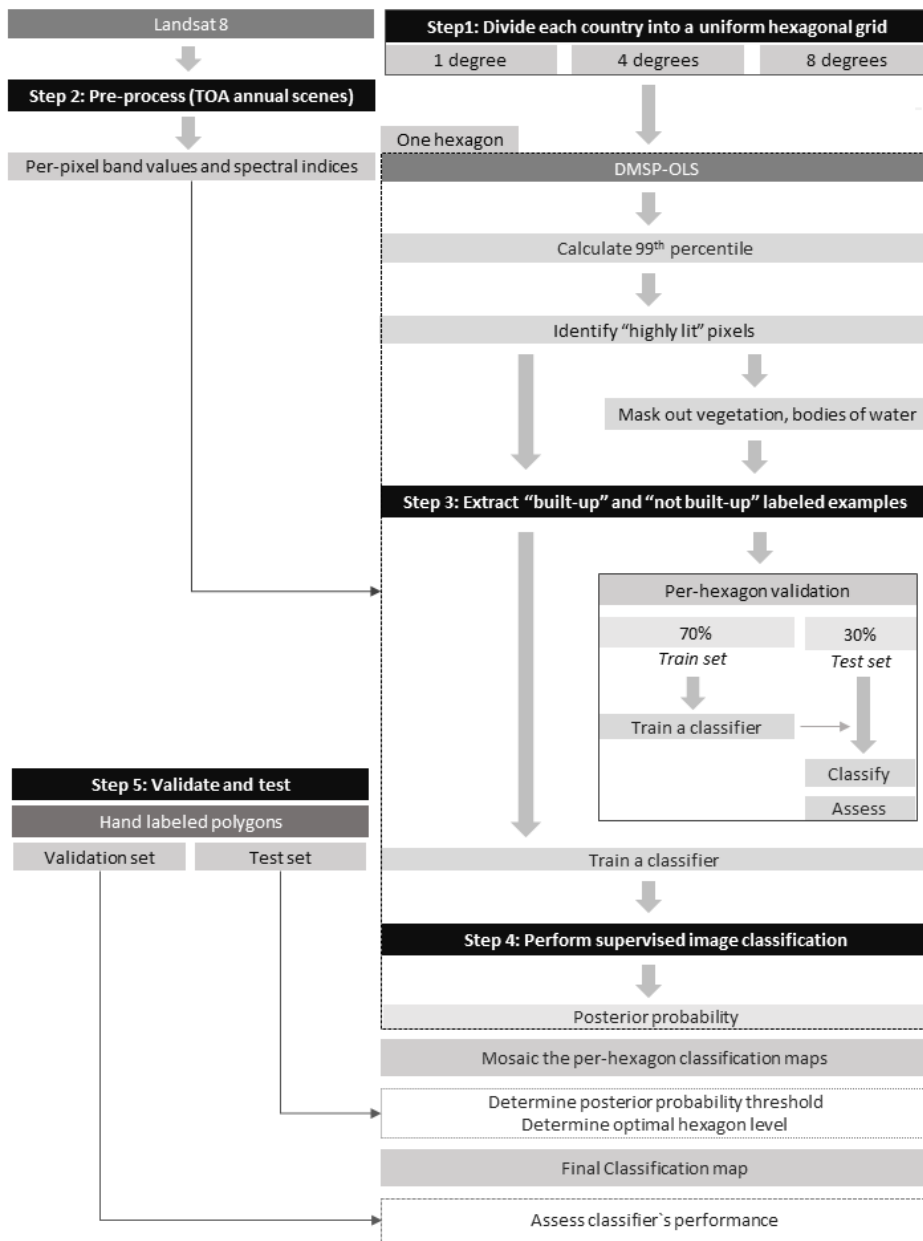


127

128 *Figure 1:* Annual changes in share of urban population in India, US and Mexico compared to world
 129 average (UN, 2014).

130 **2. Conceptual Framework: Infused DMSP-OLS / Landsat Methodology**

131 Our methodology relies on infusion of Earth-observation datasets from two domains: DMSP-OLS
 132 (which is used to extract examples of areas associated with human activity and built-up land-
 133 cover) and Landsat (which is used as the input for supervised image classification). We proceed
 134 in five steps (Figure 2): (1) Divide each country into a uniform hexagonal grid; (2) Pre-process
 135 Landsat 8 images; (3) Extract labeled examples from DMSP-OLS; (4) Perform supervised image
 136 classification; (5) Validate and test. We next describe these steps in detail.



137

138

Figure 2: Schematic illustration of our infused DMSP-OLS / Landsat methodology.

139

2.1. Divide each country into an equal-area hexagonal grid ("mapping zones")

140

Mapping large-scale heterogeneous land cover requires partitioning the region of interest into a

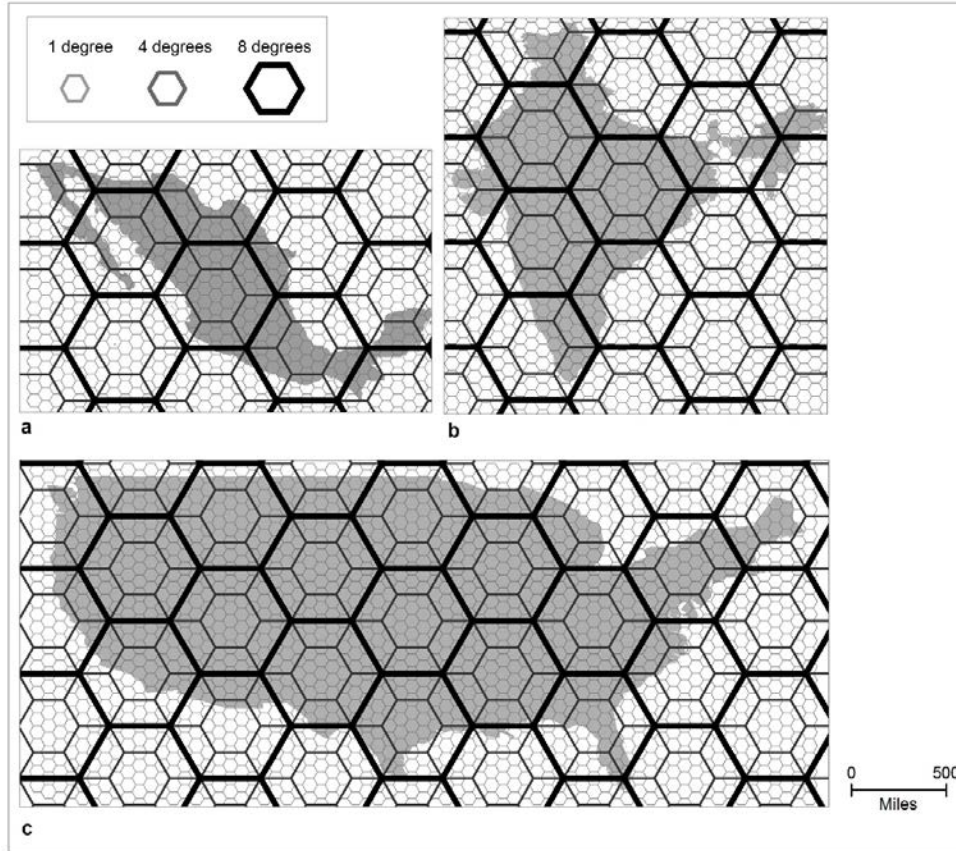
141

finite number of relatively homogenous sub-regions, or zones, that are characterized by similar

142

landform, soil, vegetation, spectral reflectance, and image footprints (Homer et al., 2004). This

143 practice is often referred to as 'zone mapping' (Homer and Gallant, 2001). The partition can be
144 according to different criteria, such as land cover and land use, socio-political definition, size
145 (Hunsaker et al., 1994; O'Neill et al., n.d.; Turner, 1989), or by means of an artificial grid system
146 where each element in the grid is treated as an independent region of interest. In this study, we
147 partition each country into an equal-area hexagonal grid (or a hexagonal tessellation). Hexagonal
148 grids are advantageous because they are characterized by elements that do not have gaps or
149 overlaps, the center-to-center distances between adjacent grid cells are approximately equal, the
150 topology of the cells is symmetrical and invariant, the cells are equal area, and the cells can be
151 recursively partitioned (Richards et al., 2000). Because the classification is subject to the size of
152 each hexagon in the grid, we examine grids of different sizes of hexagons (different center-to-
153 center distances): a distance of 1 decimal degree, 4 decimal degrees and 8 decimal degrees from
154 center to center (see Figure 3 for illustration).



155

156 *Figure 3: The three examined hexagon levels: 1°, 4°, and 8° from center to center, for (a) Mexico, (b) India,*
 157 *and (c) the US.*

158 **2.2. *Pre-process Landsat 8 images (classifier`s inputs)***

159 We use Landsat-8 imagery as classifier inputs (predictors). We apply a standard Top-of-
 160 Atmosphere (TOA) calibration on all USGS Landsat 8 Raw Scenes in one year (since DMSP-OLS
 161 is only available until 2013, we begin with mapping built-up areas in 2013). We assign a cloud
 162 score to each pixel and select the lowest possible range of cloud scores. Then we compute per-
 163 band percentile values from the accepted pixels and scale them to 8 bits. For each pixel we
 164 calculate additional spectral indices, which we use as additional predictors for the classifier:
 165 Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI),

166 Urban Index (UI), Enhanced Vegetation Index (EVI), Normalized Difference Built-up Index
167 (NDBI) (see description of these indices in Appendix 1).

168 2.3. *Extract “built-up” and “not built-up” labeled examples from DMSP-OLS*

169 Because “highly lit” pixels are associated with man-made structures that emit light, we assume
170 that pixels with DN values that exceed a given threshold represent locations with built-up land
171 cover and man-made structures. In this study we use DMSP-OLS (the “stable light” band of the
172 ‘F182013’ satellite) to identify “highly lit” pixels. To account for regional variations, we determine
173 this threshold for each hexagon independently by calculating the value of the 99th percentile of
174 all pixels in the hexagon. A pixel is “built-up” if its DN value exceeds the threshold. Note that we
175 only use hexagons that include at least one DMSP-OLS pixel with a value higher than 0. This
176 definition allows us to capture, on the one hand, small settlements in isolated low-density regions
177 (i.e., where the threshold is low), and on the other hand, only the core of cities in high-population-
178 density regions (i.e., where the threshold is relatively high).

179 Due to the spatial resolution of DMSP-OLS and the blooming effect, areas identified as “highly
180 lit” may potentially include non-built land cover. Thus, we also examine the effect of excluding
181 these types of land cover from the lit pixels (according to Landsat’s per-pixel NDVI and NDWI
182 values). In each hexagon we randomly sample 100,000 pixels. We create a point at the center of
183 each Landsat pixel and associate each point with the spectral values of the Landsat composite
184 and derived spectral indices. Each example includes a label and the spectral values from Landsat.
185 These per-hexagon training sets are used to build the local classification model.

186 **2.4. *Perform supervised image classification***

187 As noted above, we train and classify each hexagon, with Random forest (20 trees) as the classifier.
188 Random forests are tree-based classifiers that include k decision trees (k predictors). When
189 classifying an example, its variables are run through each of the k tree predictors, and the k
190 predictions are averaged to get a less noisy prediction (by voting on the most popular class). The
191 learning process of the forest involves some level of randomness. Each tree is trained over an
192 independently random sample of examples from the training set and each node's binary question
193 in a tree is selected from a randomly sampled subset of the input variables. We use Random
194 forest because previous studies find that the performance of Random forest is superior to other
195 classifiers (Goldblatt et al., 2016), especially when dealing with large-scale and noisy datasets
196 (Jean et al., 2016). We identify hexagons that remained unclassified as a result of no examples of
197 built-up areas being sampled (e.g., hexagons that include only a few lit pixels, and that have a
198 99th percentile value of DMSP-OLS pixels equal to 0). Because some of these hexagons do include
199 isolated small settlements, which we want to capture, we additionally perform a classification
200 using each country as one region of analysis (i.e., defining a single hexagon for a country) and
201 use this classification to map the built-up land cover within these hexagons. Finally, we post
202 process the classification maps by clipping the maps to the extent of the countries' borders. In
203 addition, we remove misclassified built-up pixels in remote, unlit regions (i.e., where the DN
204 value of the DMSP-OLS is 0).

205

206

207 2.5. *Validate and test*

208 We next assess the performance of the classifiers in each hexagon by dividing the sampled
209 examples into a training set and a test set (30% and 70% of the examples, respectively). First, in
210 each hexagon the classifiers are trained with the examples in the training set. We classify the
211 examples in the test set and assess the classifier’s performance. This validation procedure is
212 designed to evaluate the quality of the sampled examples and to indicate how well our classifiers
213 can predict their class. Then, we use the sampled examples for per-hexagon training and
214 classification of built-up area. Finally, the classifications in all hexagons are mosaicked.

215 In each country we assess the accuracy of the classification using a large dataset of hand-labeled
216 examples. We manually label these examples (polygons, 30m by 30m in size) as “built-up” or as
217 “not built-up”. We define polygons as built-up if the majority of their area (more than 50%) is
218 paved or covered by human-made surfaces and used for residential, industrial, commercial,
219 institutional, transportation, or other non-agricultural purposes. Similar definitions for urban
220 areas are proposed by Goldblatt et al. (2016), Potere et al. (2009) and Schneider et al. (2010) who
221 characterize a pixel as “urban” when the built environment spans the majority (50% or greater)
222 of the sub-pixel space. For India, we use Goldblatt et al.’s (2016) ground-truth dataset. This dataset
223 includes 20,030 examples (30m by 30m in size) labeled as “built-up” or as “not built-up” spanning
224 the entire country (4682 polygons labeled as built-up and 16,348 labeled as not-built-up). For the
225 US and Mexico, we construct a manually labeled ground-truth dataset of 20,000 examples
226 (polygons, 30m by 30m in size) per country. Between 22%-27% of the datasets’ polygons are
227 labeled as built-up. We describe the procedure to create the stratified sample and the distribution
228 of the examples in Appendix 2. We use half of the hand-labeled dataset to assess alternative

229 parameters for the classifiers (the test set) and the other half to evaluate and report its
230 performance (the validation set).

231 3. Results

232 3.1. *Optimal Hexagon Scale and Parameters to the Classifiers*

233 Our methodology produces high-quality, high-resolution maps of built-up areas for India,
234 Mexico, and the US. First, we partition each country into a uniform hexagonal grid and consider
235 each hexagon as an independent unit of analysis. To determine the optimal scale of the hexagons
236 in the hexagonal grid division (1, 4, and 8 decimal degrees from center to center) and to assess
237 various parameters to the classifiers, we evaluate the accuracy of the maps using half of the hand-
238 labeled examples in each country (the test set). We use several performance estimators (we refer
239 to the class “built up” as positive and to the class “not built-up” as negative): (1) True-Positive
240 Rate (TPR) (the percentage of actual BU examples classified correctly as BU); (2) True-Negative
241 Rate (TNR) (the percentage of actual NBU examples classified correctly as NBU); (3) Balanced
242 Accuracy (the average of TPR and TNR); Precision (the percentage of actual BU examples out of
243 all examples that were classified as BU) and F-Measure (the harmonic mean of the TPR and the
244 precision):

- 245 1. $TPR = TP / (TP + FN)$
- 246 2. $TNR = TN / (TN + FP)$
- 247 3. $Balanced\ Accuracy = (TPR + TNR) / 2$
- 248 4. $Precision = TP / (TP + FP)$
- 249 5. $F-Measure = 2 * ((Precision * TPR) / (Precision + TPR))$
- 250

251 Where TP is the number of the actual BU examples predicted to be BU; TN is the NBU examples
252 predicted as NBU; FN is the actual BU examples predicted as NBU and FP is the actual NBU
253 examples predicted as BU.

254 The results show differences between countries in the optimal hexagonal scale. In India and
255 Mexico, classification with the smallest level of hexagon (1°) results in the best performance,
256 indicated by high balanced accuracy rates of 79% and 84%, respectively (Table 1), as well as by
257 highest F-Measure scores (63% and 73%, respectively). Classification with larger hexagons results
258 in a lower balanced accuracy: for example, with 8° hexagons it drops to about 75% and 77% in
259 India and in Mexico, respectively. In the US, classification with the largest hexagons (8°) results
260 in a marginally greater accuracy than with the smallest hexagon (1°), indicated by a balanced
261 accuracy rate of 81.7% (compared to 81.5%). Classification with 8° hexagons also shows the
262 highest F-measure value (67%). When no hexagons are used for classification and the
263 classification is done with the entire country as one region of interest, both balanced accuracy and
264 F-Measure drop. The classifiers predict for each new example (pixel) the probability it is a positive
265 example (“built-up”) (a posterior probability, ranging between 0 and 1). We find that the best
266 performance in all three countries is achieved with a lower threshold on posterior probability
267 (around 0.2). Table 2 presents, as an example, the performance of the classifiers as a factor of the
268 posterior probability threshold, for classification with 1° hexagons. Both balanced accuracy and
269 F-measure decrease as the posterior probability threshold increases. Although the TNR increases
270 with higher posterior probability thresholds, the TPR decreases (i.e. the classification “misses”
271 urban pixels). Thus, an optimal balanced accuracy is achieved with lower thresholds.

272
273

Table 1: Performance measures as a factor of the hexagon level for India, the US, and Mexico (classification with a posterior probability threshold of 0.1, 0.1 and 0.2, respectively).

		Overall accuracy	TPR	TNR	Balanced
India	no hexagon	80.5%	53.9%	88.2%	71.0%
	8°	78.1%	68.4%	80.9%	74.6%
	4°	80.8%	69.1%	84.2%	76.7%
	1°	80.3%	76.5%	81.4%	79.0%
US	no hexagon	82.4%	77.4%	83.9%	80.7%
	8°	81.5%	81.9%	81.4%	81.7%
	4°	74.2%	86.2%	70.6%	78.4%
	1°	78.0%	88.0%	75.0%	81.5%
Mexico	no hexagon	84.9%	46.1%	97.4%	71.8%
	8°	82.4%	66.2%	87.6%	76.9%
	4°	84.0%	71.0%	88.2%	79.6%
	1°	85.9%	79.9%	87.8%	83.8%

274

275
276

Table 2: The effect of the posterior probability threshold on the classifier’s performance (classification using 1° hexagons).

		Threshold	0.1	0.2	0.4	0.6
India	Balanced		79.0%	75.6%	68.8%	63.6%
	F-Measure		63.2%	63.0%	53.3%	42.8%
US	Balanced		81.5%	80.2%	73.3%	67.1%
	F-Measure		65.0%	68.3%	61.0%	50.5%
Mexico	Balanced		81.4%	83.8%	78.6%	72.5%
	F-Measure		65.0%	73.2%	69.6%	61.1%

277

278 In the experiment described above, we sampled “built-up” examples from “highly lit” pixels
 279 (defined as all pixels with a DN value above the 99th percentile of the value across all pixels within
 280 a given hexagon). However, because these areas may also include other types of land cover, such
 281 as vegetation and bodies of water, in an additional experiment we mask out these types of land
 282 cover from the “highly-lit” pixels. We do this according to Landsat’s per-pixel NDVI value (above
 283 0.3 or 0.7) and NDWI values (a negative value). We find that excluding vegetation and bodies of

284 water from the lit pixels does not affect much the performance of the classifiers (with 1° hexagons,
285 performance only marginally improves by 0.2% and 0.6% in the US and Mexico, respectively).
286 Thus, in the sequential experiments we do not remove these types of land cover from the “highly
287 lit” pixels. Because the optimal hexagon level and posterior probability threshold varies between
288 countries, we choose for each the following optimal parameters: for India 1° hexagons, posterior
289 probability threshold 0.1; for Mexico 1° hexagons, posterior probability threshold 0.2; for the US
290 8° hexagons, posterior probability threshold 0.1.

291 3.2. *Internal Per-Hexagon Accuracy Assessment*

292 To evaluate the performance of the classifiers and the quality of our sampled training examples,
293 we perform per-hexagon accuracy assessment (using only the sampled examples). This analysis
294 is intended to assess the quality of the sampled examples and to provide an additional estimate
295 on the performance of the classifiers in relation to spatial scale (it only relies on the sampled
296 examples rather than on hand-labeled examples). In each hexagon, we randomly designate 70%
297 of the examples for training and 30% for testing. We find a high balanced accuracy rate of 71.5%
298 in India, 76.5% in the US. The balanced accuracy rate is lower in the case of Mexico (61%). These
299 results indicate that the sampling procedure generates accurate examples that are beyond
300 random. Although, by their nature, these examples are relatively “noisy”, the classifiers can
301 predict their class with reasonably high precision.

302 3.3. *Accuracy Assessment with Validation Set*

303 Based on the optimal hexagon level found for each country, we produce classification maps of
304 built-up areas spanning the three countries. We use the second subset of our hand-labeled

305 examples (the validation set) to assess the accuracy of the classification and find a high balanced
306 accuracy rate of 79%, 80% and 84% in India, the US and Mexico, respectively. Similar to the
307 accuracy measures found with the test set, the best performance is achieved with lower posterior
308 probability thresholds, again, indicating a lower TPR as the posterior probability thresholds
309 increase. Interestingly, while the TPR and TNR measures are relatively similar in India and the
310 US (77% and 81%, and 79% and 81%, with a posterior probability of 0.1), in the case of Mexico the
311 TPR is significantly higher than the TNR (92% and 71%, respectively, with a posterior probability
312 of 0.1) (Table 3 presents the accuracy measures of the validation set and a confusion matrix of the
313 classification performance).

314 Finally, we post-process the classification maps and clip them to the extent of each country. In
315 addition, “low-lit” hexagons that include only a few lit pixels (i.e. where the 99th percentile of the
316 lit pixels is 0) are mapped according to the classification we create for each country as one unit of
317 analysis. We find that although the post-processing procedure only marginally improves the
318 classification accuracy (see Appendix 3), a visual examination suggests that this procedure
319 removes misclassified pixels, primarily over bodies of water and distant bare land. Thus, the final
320 classification maps are post-processed.

321

322

323

324

325 **Table 3:** Description of accuracy measures and the confusion matrix using the validation set a
 326 (classification in India and Mexico is done with 1° hexagons, and in the US with 8° hexagons)

Posterior probability	Overall accuracy	TPR	TNR	Balanced	Precision	F-Measure					
India											
0.1*	80.0%	77.0%	80.9%	79.0%	53.9%	63.4%	Confusion matrix				
0.2	83.9%	60.8%	90.6%	75.7%	65.2%	62.9%	Predicted				
0.3	84.0%	56.2%	92.0%	74.1%	67.0%	61.1%	BU NBU Sum				
0.4	83.8%	44.7%	95.2%	69.9%	72.8%	55.4%	Actual	BU	1955	585	2540
0.5	83.5%	40.6%	95.9%	68.2%	74.3%	52.5%		NBU	1672	7101	8773
0.6	82.6%	32.2%	97.2%	64.7%	76.6%	45.3%		Sum	3627	7686	11313
US											
0.1*	80.4%	78.7%	81.0%	79.8%	54.5%	64.4%	Confusion matrix				
0.2	84.9%	58.9%	92.4%	75.7%	69.2%	63.6%	Predicted				
0.3	84.8%	52.8%	94.0%	73.4%	71.9%	60.9%	BU NBU Sum				
0.4	83.8%	38.2%	96.9%	67.6%	78.3%	51.4%	Actual	BU	1933	523	2456
0.5	83.3%	34.4%	97.4%	65.9%	79.4%	48.0%		NBU	1617	6873	8490
0.6	82.1%	26.3%	98.3%	62.3%	81.6%	39.8%		Sum	3550	7396	10946
Mexico											
0.1	76.1%	91.7%	71.1%	81.4%	50.3%	65.0%	Confusion matrix				
0.2*	85.9%	79.9%	87.8%	83.8%	67.6%	73.2%	Predicted				
0.3	86.5%	75.4%	90.1%	82.7%	70.8%	73.0%	BU NBU Sum				
0.4	86.8%	62.7%	94.4%	78.6%	78.2%	69.6%	Actual	BU	2123	534	2657
0.5	86.3%	57.4%	95.6%	76.5%	80.5%	67.0%		NBU	1018	7321	8339
0.6	85.2%	48.0%	97.1%	72.5%	83.9%	61.1%		Sum	3141	7855	10996

327 * denotes the highest balanced accuracy rate for which the confusion matrix is presented.

328 **4. Discussion**

329 We present a novel machine learning approach to map built-up areas at scale. Our methodology
 330 utilizes nighttime-light data (derived from DMSP-OLS) as a source for training examples of built-
 331 up and not built-up areas, which are then used for supervised image classification in Landsat 8
 332 imagery. This is the first study, to our knowledge, to present a practical and simple form of
 333 transfer learning that can be applied to map built-up areas across space. Although many
 334 classification products map urban land, they are typically limited in their temporal and/or spatial
 335 resolution. This limits their use to track urbanization processes over time.

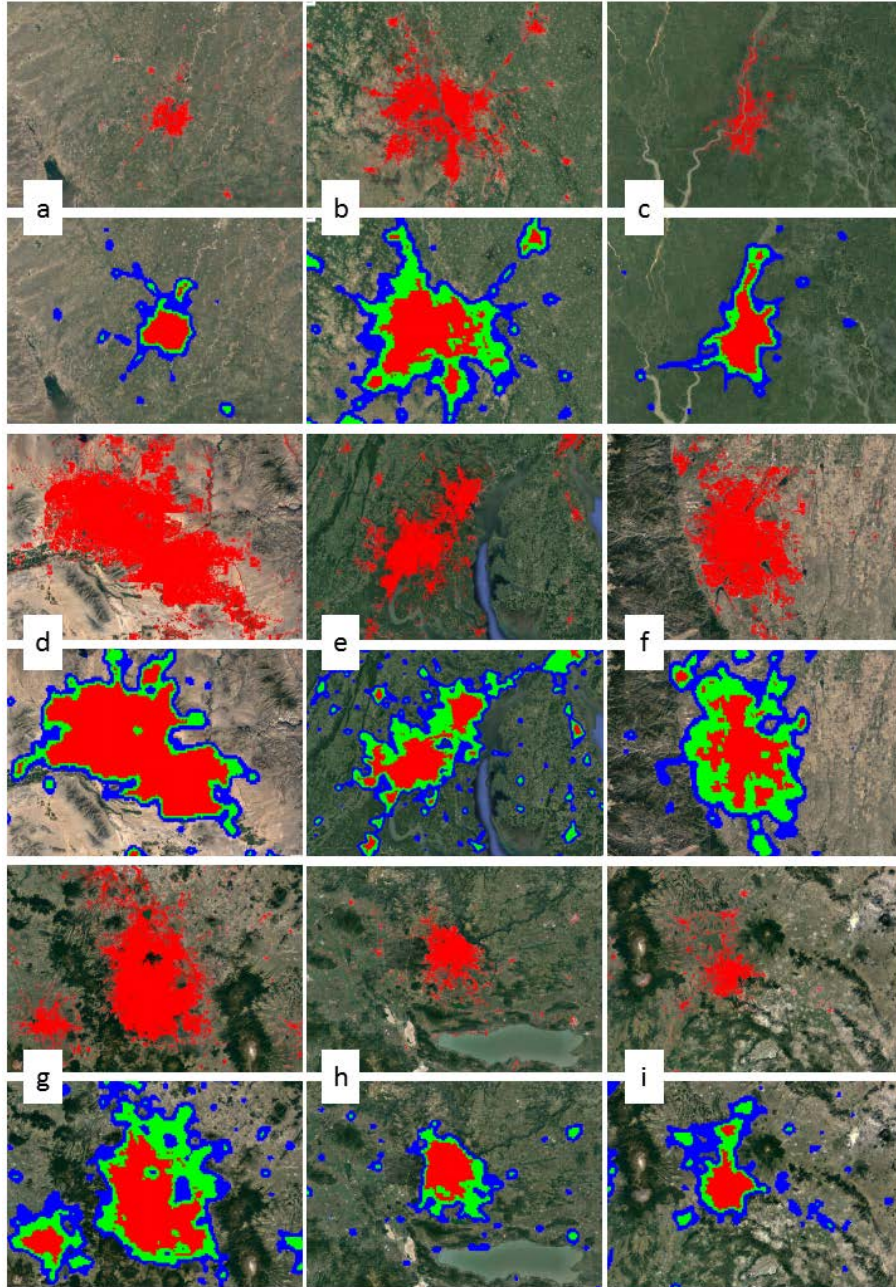
336 Mapping built-up areas at scale is challenging because of the scarcity of extensive ground-truth
337 data for supervised classification and validation. Crowd-sourced datasets, such as
338 OpenStreetMap (OSM) can also be used to map urban areas (Belgiu and Dr ăgu
339 Painho, 2015). OSM is a valuable source for ground-truth data, primarily because of its vast extent
340 and free availability. However, the completeness of OSM and its suitability for urban research is
341 subject to the number and reliability of OSM contributors (Schlesinger, 2015). The use of OSM for
342 supervised image classification remains challenging due to the risk of imbalanced distribution of
343 class labels (including their spatial coverage), the presence of errors or missing class assignments
344 (“class-noise”), and inaccurate polygon boundary delineations (Johnson and Iizuka, 2016). Our
345 methodology overcomes the lack of such data by utilizing low-resolution DMSP-OLS data for
346 classification of built-up areas in Landsat imagery. We collect examples of built-up and not-built-
347 up areas by identifying “highly lit” areas within small homogenous regions (or mapping zones).
348 These examples are used for image classification of built-up areas from Landsat imagery. By
349 partitioning countries into smaller regions, we allow the parameters of the classification model to
350 vary in what is determined as “built-up” pixels. We demonstrate that this flexibility is important
351 and show that countries differ in this optimal hexagon scale. Although many studies address the
352 effect of the classifiers’ hyper-parameters on their performance, in this study we show that
353 classifiers also have an optimal spatial scale, which can and should be discovered through
354 techniques similar to those we propose here.

355 We assess the validity of our approach using two procedures for accuracy assessment: internal
356 per-hexagon validation (assessing the classification of our sampled examples) and an external
357 validation that uses 60,000 hand-labeled examples. The results demonstrate the robustness of our

358 approach and its applicability in heterogeneous regions. We find that our classification performs
359 well with a high balanced accuracy rate of around 80%. Yet, the degree to which the “localization”
360 of our classifiers affects their performance varies between regions and depend on the
361 heterogeneous nature of the mapped landscape. Applying this method at a global scale will require
362 automatic methods for selecting, in any pixel, the scale of classification that maximizes accuracy.
363 However, due to the lack of on-board calibration and unstable radiometric performance of the
364 DMSP-OLS sensors, the absolute radiance of light cannot directly represent temporal changes in
365 the intensity of the light, and thus, inter-sensor calibration is required to make our approach
366 operational in time.

367 Our methodology overcomes the need for expensive hand-labeled data for supervised
368 classification as well as many of the limitations associated with DMSP-OLS data. As illustrated
369 in Figure 4, due to the blooming effect, lit areas are consistently larger than the built-up land cover
370 they are associated with. Our classification captures the fine boundaries of built-up areas with
371 high precision (Figure 5).

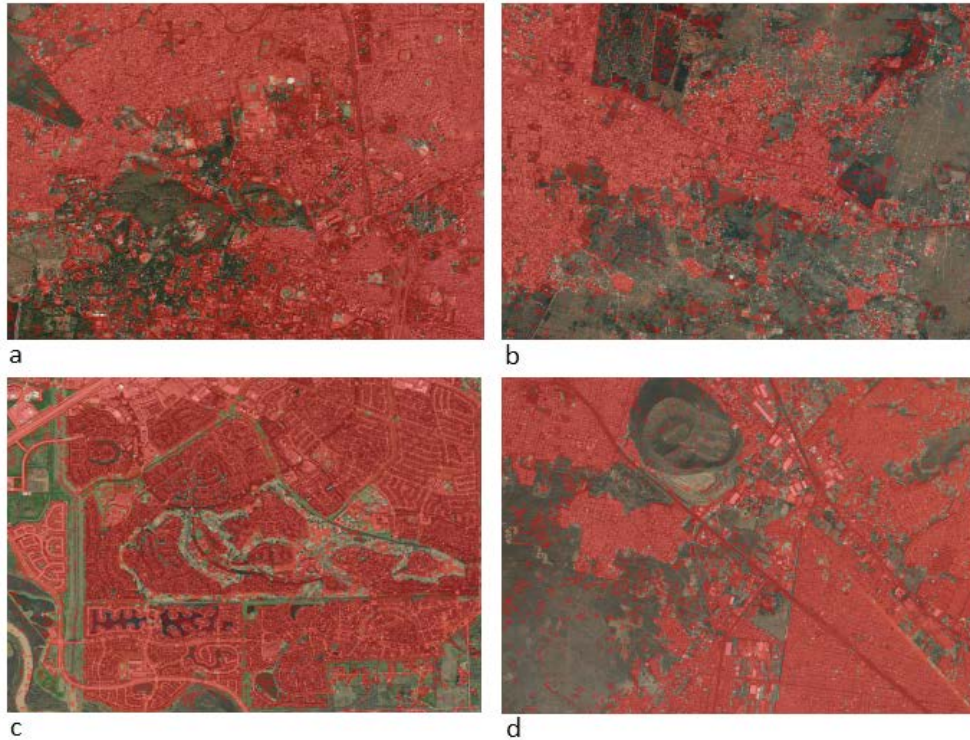
372



373

374 *Figure 4:* A comparison between our classification of built-up areas and lit pixels according to DMSP-OLS
 375 in (a) Ahmedabad, Gujarat, India; (b) New Delhi, Delhi, India; (c) Kolkata, West Bengal, India; (d)
 376 Phoenix, Arizona, US; (e) Washington DC, US; (f) Denver, Colorado, US; (g) Mexico City, Mexico; (h)
 377 Guadalajara, Mexico; (i) Puebla, Mexico. (the top figure in each city presents our classification; the bottom
 378 figure presents the DN values of DMSP-OLS stable lights band; 30-55 (red), 56-61 (green), 63 (blue))

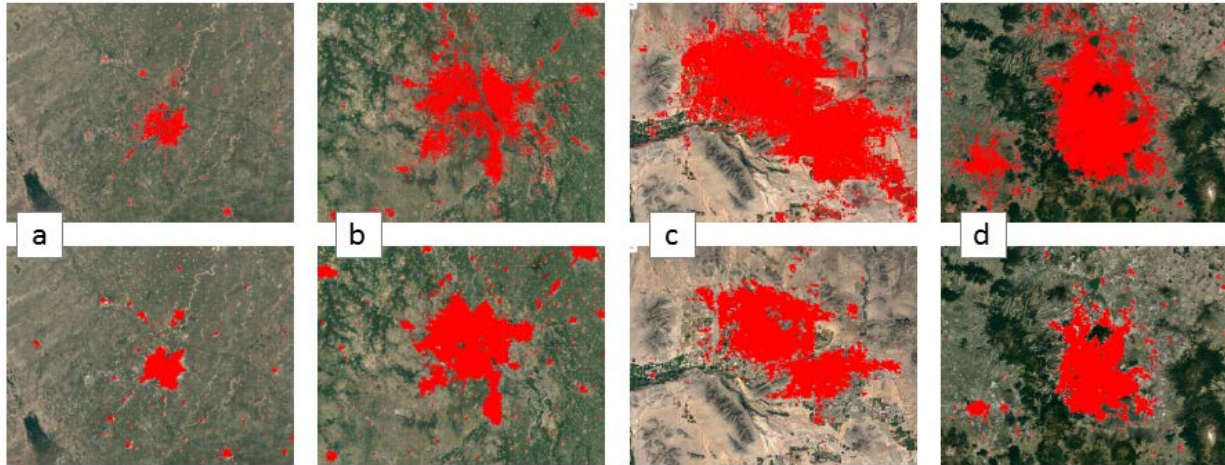
379



380

381 *Figure 5:* Classification of built-up areas (in red) in (a) Nagpur, Maharashtra, India; (b) Hyderabad,
382 Telangana, India; (c) Houston, Texas, US; (d) Mexico city, Mexico

383 To evaluate our classification and to compare it with other products, we used the 60,000 labeled
384 polygons to assess the accuracy of MCD12Q1 UMD MODIS classification scheme and DMSP-OLS
385 (we define here a highly lit pixel similar to the definition we propose in this study, i.e., a value
386 that is greater than the 99th percentile value of all pixels within a given hexagon). The accuracy of
387 our classification exceeds both datasets; Table 4 shows that we achieve a higher balanced accuracy
388 rate (between 6%-10%). As illustrated in Figure 6, the extent of the built-up land cover that we
389 detect with our methodology is smaller than the extent of the land cover that is classified as urban
390 and built-up by the MCD12Q1 classification.



391

392 *Figure 6:* A comparison between areas classified as built-up using our methodology (top) and areas
 393 classified as built up and urban by MCD12Q1 UMD MODIS classification scheme (bottom), in (a)
 394 Ahmedabad, Gujarat, India; (b) New Delhi, Delhi, India; (c) Phoenix, Arizona, US; (d) Mexico City,
 395 Mexico

396 Our classification also exceeds other national high-resolution land-cover and land-use maps. To
 397 illustrate, a comparison between our examples and the US National Land Cover Database
 398 (NLCD) classification map showed a lower balanced accuracy rate in the NLCD product (72.3%)
 399 (we define a polygon as “built-up” if more than 50% of its area is built. Thus, we relate to the
 400 NLCD classes “Developed, Medium Intensity” and “Developed High Intensity” as “built-up”).
 401 Similar findings were found in the case of Mexico, where a comparison between our examples
 402 and the urban classification of the Instituto Nacional de Estadística y Geografía (INEGI) resulted
 403 in a lower balanced accuracy rate of around 79%.

404

405

406

407 **Table 4:** Accuracy assessment (balanced accuracy) of our infused methodology for classification, MODIS-
 408 MCD12Q1 (MODIS) and DMSP-OLS “highly lit” areas (defined as pixels with a DN value above the 99th
 409 percentile). Accuracy assessment with 60,000 labeled examples.

		Our BU classification	DMSP- OLS	MODIS UMD
India	TPR	74.9%	62.8%	61.7%
	TNR	81.8%	82.6%	84.3%
	Balanced	78.3%	72.7%	73.0%
US	TPR	87.5%	67.5%	64.4%
	TNR	74.8%	80.7%	87.3%
	Balanced	81.2%	74.1%	75.8%
Mexico	TPR	81.5%	77.6%	53.4%
	TNR	87.0%	71.2%	94.1%
	Balanced	84.3%	74.4%	73.7%

410

411 To summarize, we have developed a conceptual framework whereby utility of our transfer-
 412 learning methodology results in high-resolution, high quality depictions of built-up areas across
 413 three highly diverse countries. In today’s era of big data, a globally consistent and data-driven
 414 method of defining and classifying urban areas has extensive applications. Economics, urban
 415 planning, climate modeling, water-resource management, hazard-response efforts, and urban-
 416 eco-system assessments all use geographic data on urban areas. With earth’s rapidly urbanizing
 417 population, having information on urban extent that is spatially and temporally consistent and
 418 defined at high resolution is both relevant to a wide range of disciplines and essential for helping
 419 society better understand the drivers of urbanization.

420 **5. References**

421 Aguilar, A.G., 2008. Peri-urbanization, illegal settlements and environmental impact in Mexico
 422 City. *Cities* 25, 133–145. doi:10.1016/j.cities.2008.02.003

423 Aguilar, A.G., Guerrero, F.L., 2013. Poverty in peripheral informal settlements in Mexico City:
424 The case of Magdalena Contreras, Federal District. *Tijdschr. Voor Econ. En Soc. Geogr.*
425 104, 359–378. doi:10.1111/tesg.12012

426 Bagan, H., Yamagata, Y., 2015. Analysis of urban growth and estimating population density using
427 satellite images of nighttime lights and land-use and population data. *GIScience Remote*
428 *Sens.* 52, 765–780. doi:10.1080/15481603.2015.1072400

429 Bagan, H., Yamagata, Y., 2014. Land-cover change analysis in 50 global cities by using a
430 combination of Landsat data and analysis of grid cells. *Environ. Res. Lett.* 9, 064015.
431 doi:10.1088/1748-9326/9/6/064015

432 Belgiu, M., Drăguț, L., 2014. Comparing supervised and unsupervised multiresolution
433 segmentation approaches for extracting buildings from very high resolution imagery.
434 *ISPRS J. Photogramm. Remote Sens.* 96, 67–75. doi:10.1016/j.isprsjprs.2014.07.002

435 Buhaug, H., Urdal, H., 2013. An urbanization bomb? Population growth and social disorder in
436 cities. *Glob. Environ. Change* 23, 1–10. doi:10.1016/j.gloenvcha.2012.10.016

437 Census Bureau, U.S., 2012. *Increasing Urbanization: Population Distribution by City Size, 1790 to*
438 *1890.*

439 CIESIN, 2005. *Gridded Population of the World, Version 3 (GPWv3) Data Collection.*

440 Connolly, P., 2014. *Latin American Informal Urbanism: Contexts, Concepts, and Contributions*
441 *with Specific Reference to Mexico*, in: Becerra, A., Hernandez, F. (Eds.), (Re)viewing Latin
442 *American Cities.* Cambridge Scholars Publishing, Cambridge.

443 Consejo Nacional de Población, 2012. *Sistema Urbano Nacional 2012.*

444 Eakin, H., Lerner, A.M., Manuel-Navarrete, D., Hernández Aguilar, B., Martínez-Canedo, A.,
445 Tellman, B., Charli-Joseph, L., Fernández Álvarez, R., Bojórquez-Tapia, L., 2016. Adapting
446 to risk and perpetuating poverty: Household's strategies for managing flood risk and
447 water scarcity in Mexico City. *Environ. Sci. Policy*. doi:10.1016/j.envsci.2016.06.006

448 Elvidge, C., Hsu, F.-C., Baugh, K., Ghosh, T., 2014. National Trends in Satellite-Observed
449 Lighting: 1992-2012, in: *Global Urban Monitoring and Assessment through Earth
450 Observation, Remote Sensing Applications Series*. CRC Press, pp. 97–120.

451 Estima, J., Painho, M., 2015. Investigating the Potential of OpenStreetMap for Land Use/Land
452 Cover Production: A Case Study for Continental Portugal, in: Arsanjani, J.J., Zipf, A.,
453 Mooney, P., Helbich, M. (Eds.), *OpenStreetMap in GIScience, Lecture Notes in
454 Geoinformation and Cartography*. Springer International Publishing, pp. 273–293.

455 Gaughan, A.E., Stevens, F.R., Linard, C., Jia, P., Tatem, A.J., 2013. High resolution population
456 distribution maps for Southeast Asia in 2010 and 2015. *PloS One* 8, e55882.
457 doi:10.1371/journal.pone.0055882

458 Georgescu, M., Chow, W.T.L., Wang, Z.H., Brazel, A., Trapido-Lurie, B., Roth, M., Benson-Lira,
459 V., 2015. Prioritizing urban sustainability solutions: coordinated approaches must
460 incorporate scale-dependent built environment induced effects. *Environ. Res. Lett.* 10,
461 061001. doi:10.1088/1748-9326/10/6/061001

462 Goldblatt, R., You, W., Hanson, G., Khandelwal, K.A., 2016. Detecting the Boundaries of Urban
463 Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine.
464 *Remote Sens.* 8. doi:10.3390/rs8080634

465 H. S. Sudhira, K. V. Gururaja, n.d. Population crunch in India: is it urban or still rural. *Curr. Sci.*
466 103, 37–40.

467 Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau,
468 D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L.,
469 Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century
470 Forest Cover Change. *Science* 342, 850–853. doi:10.1126/science.1244693

471 Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring Economic Growth from Outer
472 Space. *Am. Econ. Rev.* 102, 994–1028. doi:10.1257/aer.102.2.994

473 Henderson, M., Yeh, E.T., Gong, P., Elvidge, C., Baugh, K., 2003. Validation of urban boundaries
474 derived from global night-time satellite imagery. *Int. J. Remote Sens.* 24, 595–609.
475 doi:10.1080/01431160304982

476 Homer, C., Gallant, A., 2001. Partitioning the Conterminous United States into Mapping Zones
477 for Landsat TM Land Cover Mapping. USGS.

478 Homer, C., Huang, C., Yang, L., Wylie, B.K., Coan, M., 2004. Development of a 2001 National
479 Land-Cover Database for the United States. *Photogramm. Eng. Remote Sens.* 7, 829–840.

480 Hsu, F.-C., Baugh, K.E., Ghosh, T., Zhizhin, M., Elvidge, C.D., 2015. DMSP-OLS Radiance
481 Calibrated Nighttime Lights Time Series with Intercalibration. *Remote Sens.* 7, 1855–1876.
482 doi:10.3390/rs70201855

483 Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the
484 radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens.*
485 *Environ.* 83, 195–213. doi:10.1016/S0034-4257(02)00096-2

486 Hunsaker, C.T., O'Neill, R.V., Jackson, B.L., Timmins, S.P., Levine, D.A., Norton, D.J., 1994.
487 Sampling to characterize landscape pattern. *Landsc. Ecol.* 9, 207–226.
488 doi:10.1007/BF00134748

489 Imhoff, M., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. A technique for using composite
490 DMSP/OLS “City Lights” satellite data to map urban area. *Remote Sens. Environ.* 61, 361–
491 370. doi:10.1016/S0034-4257(97)00046-1

492 Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite
493 imagery and machine learning to predict poverty. *Science* 353, 790–794.
494 doi:10.1126/science.aaf7894

495 Johnson, B.A., Iizuka, K., 2016. Integrating OpenStreetMap crowdsourced data and Landsat time-
496 series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna
497 de Bay area of the Philippines. *Appl. Geogr.* 67, 140–149. doi:10.1016/j.apgeog.2015.12.006

498 Kawamura, M., Jayamanna, S., Tsujiko, Y., 1996. Relation between social and environmental
499 conditions in Colombo Sri Lanka and the urban index estimated by satellite remote
500 sensing data. *Int. Arch. Photogramm. Remote Sens.* 31, 321–326.

501 Keola, S., Andersson, M., Hall, O., 2015. Monitoring Economic Development from Space: Using
502 Nighttime Light and Land Cover Data to Measure Economic Growth. *World Dev.* 66, 322–
503 334. doi:10.1016/j.worlddev.2014.08.017

504 Levin, N., Duke, Y., 2012. High spatial resolution night-time light images for demographic and
505 socio-economic studies. *Remote Sens. Environ.* 119, 1–10. doi:10.1016/j.rse.2011.12.005

506 Liu, Y., Delahunty, T., Zhao, N., Cao, G., 2016. These lit areas are undeveloped: Delimiting
507 China's urban extents from thresholded nighttime light imagery. *Int. J. Appl. Earth Obs.*
508 *Geoinformation* 50, 39–50. doi:10.1016/j.jag.2016.02.011

509 McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the
510 delineation of open water features. *Int. J. Remote Sens.* 17, 1425–1432.

511 Miyazaki, H., Iwao, K., Shibasaki, R., 2011. Development of a New Ground Truth Database for
512 Global Urban Area Mapping from a Gazetteer. *Remote Sens.* 3, 1177–1187.
513 doi:10.3390/rs3061177

514 O'Neill, R. V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Riitters, K.H., Wickham,
515 J.D., n.d. Scale problems in reporting landscape pattern at the regional scale. *Landsc. Ecol.*
516 11, 169–180. doi:10.1007/BF02447515

517 Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–
518 1359. doi:10.1109/TKDE.2009.191

519 Patel, N.N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F.R., Tatem, A.J., Trianni, G.,
520 2015. Multitemporal settlement and population mapping from Landsat using Google
521 Earth Engine. *Int. J. Appl. Earth Obs. Geoinformation* 35, Part B, 199–208.
522 doi:10.1016/j.jag.2014.09.005

523 Pesaresi, M., EHRlich Daniele, FERRI Stefano, FLORCZYK Aneta, CARNEIRO FREIRE Sergio
524 Manuel, HALKIA Stamatia, JULEA Andreea Maria, KEMPER Thomas, SOILLE Pierre,
525 SYRRIS Vasileios, 2016. Operating procedure for the production of the Global Human
526 Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Publications
527 Office of the European Union, Ispra (VA), Italy.

528 Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., Stenseth, N.C., 2005. Using the
529 satellite-derived NDVI to assess ecological responses to environmental change. *Trends*
530 *Ecol. Evol.* 20, 503–510. doi:10.1016/j.tree.2005.05.011

531 Potere, D., Schneider, A., Angel, S., Civco, D.L., 2009. Mapping urban areas on a global scale:
532 which of the eight maps now available is more accurate? *Int. J. Remote Sens.* 30, 6531–
533 6558. doi:10.1080/01431160903121134

534 Richards, T., Gallego, J., Achard, F., 2000. Sampling for forest cover change assessment at the pan-
535 tropical scale. *Int. J. Remote Sens.* 21, 1473–1490. doi:10.1080/014311600210272

536 schlesinger, J., 2015. Using Crowd-Sourced Data to Quantify the Complex Urban Fabric -
537 OpenStreetMap and the Urban–Rural Index, in: Arsanjani, J., Zipf, A., Mooney, P.,
538 Helbich, M. (Eds.), *OpenStreetMap in GIScience*. Springer International Publishing,
539 Switzerland, pp. 295–315.

540 Schneider, A., Friedl, M.A., Potere, D., 2010. Mapping global urban areas using MODIS 500-m
541 data: New methods and datasets based on “urban ecoregions.” *Remote Sens. Environ.*
542 114, 1733–1746. doi:10.1016/j.rse.2010.03.003

543 Seto, K.C., Fragkias, M., Güneralp, B., Reilly, M.K., 2011. A Meta-Analysis of Global Urban Land
544 Expansion. *PLOS ONE* 6, e23777. doi:10.1371/journal.pone.0023777

545 Seto, K.C., Güneralp, B., Hutyra, L.R., 2012. Global forecasts of urban expansion to 2030 and direct
546 impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci.* 109, 16083–16088.
547 doi:10.1073/pnas.1211658109

548 Small, C., Elvidge, C.D., 2013. Night on Earth: Mapping decadal changes of anthropogenic night
549 light in Asia. *Int. J. Appl. Earth Obs. Geoinformation* 22, 40–52.
550 doi:10.1016/j.jag.2012.02.009

551 Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS
552 night lights. *Remote Sens. Environ.* 96, 277–291. doi:10.1016/j.rse.2005.02.002

553 Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating Census Data for
554 Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data.
555 *Plos One* 10, e0107042. doi:10.1371/journal.pone.0107042

556 Su, Y., Chen, X., Wang, Chongyang, Zhang, H., Liao, J., Ye, Y., Wang, Changjian, 2015. A new
557 method for extracting built-up urban areas using DMSP-OLS nighttime stable lights: a
558 case study in the Pearl River Delta, southern China. *GIScience Remote Sens.* 52, 218–238.
559 doi:10.1080/15481603.2015.1007778

560 Sutton, P.C., 2003. A scale-adjusted measure of “Urban sprawl” using nighttime satellite imagery.
561 *Remote Sens. Environ., Urban Remote Sensing* 86, 353–369. doi:10.1016/S0034-
562 4257(03)00078-6

563 Tatem, A.J., Noor, A.M., Hagen, C. von, Gregorio, A.D., Hay, S.I., 2007. High Resolution
564 Population Maps for Low Income Nations: Combining Land Cover and Census in East
565 Africa. *PLOS ONE* 2, e1298. doi:10.1371/journal.pone.0001298

566 Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring
567 urbanization in mega cities from space. *Remote Sens. Environ.* 117, 162–176.
568 doi:10.1016/j.rse.2011.09.015

569 Trianni, G., Lisini, G., Angiuli, E., Moreno, E.A., Dondi, P., Gaggia, A., Gamba, P., 2015. Scaling
570 up to National/Regional Urban Extent Mapping Using Landsat Data. *IEEE J. Sel. Top.*
571 *Appl. Earth Obs. Remote Sens.* 8, 3710–3719. doi:10.1109/JSTARS.2015.2398032

572 Turner, M.G., 1989. Landscape Ecology: The Effect of Pattern on Process. *Annu. Rev. Ecol. Syst.*
573 20, 171–197.

574 UN, 2014. World Urbanization Prospects: The 2014 Revision (No. ST/ESA/SER.A/352).

575 US Census Bureau, 2016. Urban and Rural Areas [WWW Document]. URL
576 [https://www.census.gov/history/www/programs/geography/urban_and_rural_areas.ht](https://www.census.gov/history/www/programs/geography/urban_and_rural_areas.html)
577 [ml](https://www.census.gov/history/www/programs/geography/urban_and_rural_areas.html)

578 Wei, Y., Liu, H., Song, W., Yu, B., Xiu, C., 2014. Normalization of time series DMSP-OLS nighttime
579 light images for urban growth analysis with Pseudo Invariant Features. *Landsc. Urban*
580 *Plan.* 128, 1–13. doi:10.1016/j.landurbplan.2014.04.015

581 Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically
582 mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24, 583–594.
583 doi:10.1080/01431160304987

584 Zhang, Q., Seto, K.C., 2013. Can Night-Time Light Data Identify Typologies of Urbanization? A
585 Global Assessment of Successes and Failures. *Remote Sens.* 5, 3476–3494.
586 doi:10.3390/rs5073476

587 Zhou, Y., Smith, S.J., Elvidge, C.D., Zhao, K., Thomson, A., Imhoff, M., 2014. A cluster-based
588 method to map urban area from DMSP/OLS nightlights. *Remote Sens. Environ.* 147, 173–
589 185. doi:10.1016/j.rse.2014.03.004

590 Zhou, Y., Smith, S.J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., Asrar, G.R., Zhang,
591 X., He, C., Elvidge, C.D., 2015. A global map of urban extent from nightlights. Environ.
592 Res. Lett. 10, 054011. doi:10.1088/1748-9326/10/5/054011

593

594

595 **Appendix 1: Indices used in Machine Learning Classifier**

596 *I. NDVI (Normalized Difference Vegetation Index)*

597 NDVI expresses the relation between red visible light (which is typically absorbed by a plant's
598 chlorophyll) and near-infrared wavelength (which is scattered by the leaf's mesophyll structure)
599 (Pettorelli et al., 2005). It is computed as:

$$600 \quad (\text{NIR-RED}) / (\text{NIR+RED}) \quad (1)$$

601 where NIR is the near infra-red wavelength and RED is the red wavelength. The values of NDVI
602 range between (-1) and (+1). An average NDVI value in 2014 was calculated for each pixel (with
603 Landsat 7 32-Day NDVI Composite).

604 *II. NDBI (Normalized Difference Built-up Index)*

605 NDBI expresses the relation between the medium infra-red and the near infra-red wavelengths
606 (Zha et al., 2003). It is computed as:

$$607 \quad (\text{MIR-NIR}) / (\text{MIR+NIR}) \quad (2)$$

608 where MIR is the medium infra-red and NIR is the near infra-red wavelength. The index assumes
609 a higher reflectance of built-up areas in the medium infra-red wavelength range than in the near
610 infra-red.

611 *III. NDWI (Normalized Difference Water Index)*

612 NDWI expressed the relation between the green (G) and the NIR (near infra-red), with a scaling
613 of -1 to +1(McFeeters, 1996). It is computed as:

614 $(G-NIR) / (G+NIR) \quad (3)$

615 The positive values are typically open water areas with reflect green light but not NIR
616 wavelengths, while negative values are non-water features, like soil and vegetation, which reflect
617 higher NIR values than green wavelengths.

618 *IV. UI (Urban Index)*

619 UI is the difference between the short infra-red (SWIR) and the near-infrared wavelengths
620 (Kawamura et al., 1996). It is computed as:

621 $(SWIR-NIR) / (SWIR+NIR) \quad (4)$

622 Like NDBI, UI assumed high brightness in SWIR in urban areas as opposed to the NIR.

623 *VI. EVI (Enhanced Vegetation Index)*

624 EVI is an improved vegetation index with higher sensitivity in high biomass regions where NDVI
625 tends to saturate, reduces atmospheric influences, and removes the canopy background
626 signal(Huete et al., 2002). It is computed as:

627 $(2.5 * ((NIR-R) / (NIR + 6R - 7.5B + 1))) \quad (5)$

628 This is a similar formula to the NDVI, which takes advantage of high reflectance of vegetation in
629 the NIR band as opposed to the R band. The blue band (B) is used (with a coefficient of 7.5) to
630 correct for aerosol influences in the red band (R). There is a integer of 1 added to the denominator
631 to adjustment for nonlinear NIR and R radiant transfer through canopies. 2.5 is applied as a gain
632 to the index.

Table 6. Bands in landsat and indices used

	Spectral band	Wavelength (micrometers)	Resolution (meters)
Landsat 8			
B1	Band 1 – Ultra blue	0.43 - 0.45	30
B2	Band 2 - Blue	0.45 - 0.51	30
B3	Band 3 - Green	0.53 - 0.59	30
B4	Band 4 – Red	0.64 - 0.67	30
B5	Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
B6	Band 6 - SWIR 1	1.57 - 1.65	30
B7	Band 7 - SWIR 2	2.11 - 2.29	30
B8	Band 8 - Panchromatic	0.50 - 0.68	15
B10	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 (resampled to 30)
B11	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 (resampled to 30)
NDVI	$(B5-B4)/(B5+B4)$		30
NDWI	$(B3-B5)/(B3+B5)$		30
NDBI	$(B6-B5)/(B6+B5)$		30
EVI	$2.5*((B5/B4)/(B5+6*B4-7.5*B2+1))$		30
UI	$(B7-B5)/(B7+B5)$		30

634

635

636

637 **Appendix 2: Sampling Scheme for Testing and Validation**

638 We use a random stratified sampling procedure according to the nighttime light intensity. In each

639 country, we identify the NTL (night time lights) pixels whose value is 63 (the highest possible

640 value). Then we calculate a 5 pixels radius circle-shaped boolean kernel (a buffer of

641 approximately 5 Km). We found this method to result in an approximate distribution of 75%

642 urban points, our class of interest, and 25% non-urban points. We randomly sample ~10,000

643 points in this buffer zone of high NTL and its periphery. This process was repeated twice, to

644 generate two 10,000 point datasets for each country- only used in testing to determine optimal

645 algorithm parameters, and the second used to validation the final urban class map to assess its

646 accuracy.

647 A 30 by 30 meter square buffer (the size of a Landsat pixel) is drawn around each random sample
 648 point. These polygons are overlaid with high resolution imagery in Google Earth. The interior of
 649 each polygon is compared to the imagery it overlays by an analyst. Based on the imagery, the
 650 polygons are manually labeled as built-up or as not-built-up. A polygon is built-up when 50% or
 651 more of the contents of polygon are man-made. Polygons labeled not built-up may still contain
 652 man-made structures, such as roads or buildings, but they make up less than 50%.

653 Table 7: the distribution of the built-up and not-built-up points for each country.

	BU points	NBU points	Total
India	4682 (22.26%)	16,348 (77.74%)	21,030
US	4386 (21.6%)	15,898 (78.4%)	20,284
Mexico	5477 (27.4%)	14,523 (72.6%)	20,000

654

655 **Appendix 3: Accuracy Assessment of Post Processed Country Maps**

656

657 All samples (~20,000 per country) described in appendix 2 were used to assess accuracy when
 658 adding in pre-processing and after post-processing. Pre-processing involved replacing hexagons
 659 with low NTL in each country with the urban classification map produced without hexagons (e.g.
 660 using the entire country boundary to extract training data and build a classifier). This is because
 661 hexagons with low NTL do not have enough training data to make a good classification. Post-
 662 processing includes the pre-processing steps in addition to removing pixels with 0 NTL values.
 663 Results are shown in table 8. India has marginal accuracy improvements, Mexico shows no effect,
 664 and the US improves accuracy around 1%. Table 3 shows the results for various accuracy metrics.

665

666

667

Table 8: Accuracy assessment pre- and post-processing (removing pixels with 0 NTL values)

	India		Mexico		US	
	Pre-Process	Post- Process	Pre-Process	Post- Process	Pre-Process	Post- Process
Accuracy =	0.807372748	0.810091743	0.862046781	0.862046781	0.819810714	0.833104772
Precision =	0.54927557	0.554166061	0.735567091	0.735567091	0.575346505	0.600739372
Recall =	0.751963865	0.748428908	0.779723195	0.779723195	0.789239482	0.786448881
TPR =	0.751963865	0.748428908	0.779723195	0.779723195	0.789239482	0.786448881
TNR =	0.823244824	0.827733453	0.893364628	0.893364628	0.828795625	0.846781344
Balanced =	0.787604345	0.788081181	0.836543912	0.836543912	0.809017553	0.816615113
F-Measure =	0.634833789	0.636811764	0.757001781	0.757001781	0.665529592	0.681163217

668

669

670

671