



# Scalable deep learning to identify brick kilns and aid regulatory capacity

Jihyeon Lee<sup>a,1</sup> , Nina R. Brooks<sup>b,1,2</sup> , Fahim Tajwar<sup>a</sup> , Marshall Burke<sup>c,d</sup> , Stefano Ermon<sup>a,c</sup> , David B. Lobell<sup>c,d</sup> , Debashish Biswas<sup>e</sup> , and Stephen P. Luby<sup>c,f</sup>

<sup>a</sup>Computer Science Department, Stanford University, Stanford, CA 94305; <sup>b</sup>Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, CA 94305; <sup>c</sup>Woods Institute for the Environment, Stanford University, Stanford, CA 94305; <sup>d</sup>Department of Earth System Science, Stanford University, Stanford, CA 94305; <sup>e</sup>International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka 1212, Bangladesh; and <sup>f</sup>Center for Innovation in Global Health, Stanford University, Stanford, CA 94305

Edited by Arild Underdal, University of Oslo, Oslo, Norway, and approved February 3, 2021 (received for review September 8, 2020)

**Improving compliance with environmental regulations is critical for promoting clean environments and healthy populations. In South Asia, brick manufacturing is a major source of pollution but is dominated by small-scale, informal producers who are difficult to monitor and regulate—a common challenge in low-income settings. We demonstrate a low-cost, scalable approach for locating brick kilns in high-resolution satellite imagery from Bangladesh. Our approach identifies kilns with 94.2% accuracy and 88.7% precision and extracts the precise GPS coordinates of every brick kiln across Bangladesh. Using these estimates, we show that at least 12% of the population of Bangladesh (>18 million people) live within 1 km of a kiln and that 77% and 9% of kilns are (illegally) within 1 km of schools and health facilities, respectively. Finally, we show how kilns contribute up to 20.4  $\mu\text{g}/\text{m}^3$  of  $\text{PM}_{2.5}$  (particulate matter of a diameter less than 2.5  $\mu\text{m}$ ) in Dhaka when the wind blows from an unfavorable direction. We document inaccuracies and potential bias with respect to local regulations in the government data. Our approach demonstrates how machine learning and Earth observation can be combined to better understand the extent and implications of regulatory compliance in informal industry.**

deep learning | satellite imagery | environmental regulations | Bangladesh | air pollution

**E**nforcing environmental regulations is notoriously difficult. Regulatory agencies often lack information on compliance and the ability to effectively punish or sanction violators (1–4). A growing literature demonstrates how advances in Earth observation and machine learning can reduce the costs of monitoring, enforcement, and compliance of environmental regulations and support regulatory agencies (5–8). However, most of this literature focuses on applications in highly developed countries where the location of firms is readily observed. In developing countries where informal industries such as brick kilns, tanneries, and small-scale mines are responsible for substantial pollution, these problems are exacerbated both because the location and activity of these firms is hard to observe and because resources for regulatory enforcement are often limited (9).

Brick manufacturing is an excellent case to develop data-driven, low-cost approaches for environmental compliance. First and foremost, this industry is associated with substantial environmental and health harm. Traditional kilns emit a number of pollutants and greenhouse gases into the atmosphere, including particulate matter of a diameter less than 2.5  $\mu\text{m}$  ( $\text{PM}_{2.5}$ ), which can be inhaled deep inside the lungs and cause significant disease and mortality, and black carbon, which is a major driver of climate change (10–13). Models estimate that brick manufacturing is responsible for 30 to 50% of the  $\text{PM}_{2.5}$  in Dhaka during the winter months when kilns operate (11, 12) and is responsible for 17% of Bangladesh's total annual  $\text{CO}_2$  emissions (14). Recent estimates suggest that life expectancy in Bangladesh is reduced on average by 1.87 y per person due to air pollution, which is more than any other country in the world (15). In Bangladesh,

which lacks domestic sources of many other construction materials, brick production is central to construction (16, 17). As a result, traditional kilns have proliferated rapidly over the past few decades (17, 18).

Second, brick kilns lend themselves to an object detection approach. In Bangladesh, and across South Asia, bricks are produced in traditional, highly inefficient, coal-fired kilns (16–18). Brick kilns are very visually distinct, reasonably large, and surrounded by rows of drying bricks, making them visible in satellite images (see *SI Appendix*, Fig. S1 for example images of kilns).

Knowing precisely where all of the brick kilns in Bangladesh are located can help regulatory agencies or civil society target violations, identify populations who are at risk, and support better policy and regulatory decisions. To legally establish a kiln in Bangladesh, an owner must register with the government and obtain environmental clearance, but many eschew this formal process; enforcement against unregistered kilns is minimal (16, 19). The Government of Bangladesh (GoB) recently launched a study to manually map and verify the locations of brick kilns across the country (20)—an incredibly time- and labor-intensive effort that underscores the importance of obtaining this information for the government. As informal kilns are constantly being

## Significance

**Monitoring compliance with environmental regulations is a global challenge. It is particularly difficult for governments in low-income countries, where informal industry is responsible for a large amount of pollution, because the governments lack the ability to locate and monitor large numbers of dispersed polluters. This study demonstrates an accurate, scalable machine-learning approach for identifying brick kilns, a highly polluting informal industry in Bangladesh, in satellite imagery. Our data reveal widespread violations of the national regulations governing brick manufacturing, which has implications for the health and well-being of the country. Our approach offers a low-cost, replicable method for regulatory agencies to generate information on key pollution sources.**

Author contributions: J.L., N.R.B., M.B., S.E., D.B.L., and S.P.L. designed research; J.L., N.R.B., and F.T. performed research; N.R.B. and D.B. contributed primary data used in the analysis; J.L., N.R.B., and F.T. analyzed data; and J.L., N.R.B., F.T., M.B., S.E., D.B.L., and S.P.L. wrote the paper.

Competing interest statement: M.B., D.B.L., and S.E. are cofounders of AtlasAI, a company that uses machine learning to measure infrastructure quality and economic outcomes in the developing world.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nd/4.0/).

<sup>1</sup>J.L. and N.R.B. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [nbrooks2@stanford.edu](mailto:nbrooks2@stanford.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2018863118/-DCSupplemental>.

Published April 22, 2021.

constructed, this manual mapping approach is also a highly inefficient way to obtain the high-frequency information necessary for monitoring. Moreover, government data in low-income countries is often inaccurate and subject to bias (21).

Here we develop a scalable, deep-learning pipeline for accurately locating brick kilns across a broad geography. Brick kilns are visually distinct in satellite imagery, but it is expensive and time-consuming to annotate images with their exact locations. We make improvements on two recent papers that use machine-learning methods to identify brick kilns (22, 23) by presenting an approach that only requires classification-level annotations (indicating whether an image contains a brick kiln or not) to train a single model that performs the more difficult task of detection (indicating where within an image the brick kiln is located). Building on past applications of deep-learning to environmental monitoring (5, 6, 23), and on specific efforts to use deep learning to identify brick kilns (22, 23), we develop a weakly supervised learning approach that learns to localize kilns given lower cost classification-level data, rebuilds kilns that have been fragmented across multiple images, an inherent problem with satellite imagery, and distinguishes between two kiln technologies, based on their shape. To produce this final model, we conducted two main model development steps to improve the accuracy of our approach, first training an initial model on a small, limited dataset, using it to classify images from different parts of the country, and then hand-validating the pre-

dictions to incorporate them into an expanded and improved training dataset (*Materials and Methods*). We show how our approach is robust to the visual, geographic diversity of satellite imagery. We also demonstrate its scalability. Although our approach required human input during model development, this is a one-time fixed cost; our model is virtually free to run repeatedly and generate predictions across huge geographies without further input.

We then demonstrate how this map of brick kiln locations can be used to study compliance with environmental regulations, show how populations are exposed to brick manufacturing nationally, and examine the contribution of kilns to air pollution in Dhaka.

## Results

**Identifying Brick Kilns.** The model we trained on the initial, limited dataset attained a low precision (the proportion of true positives among all images predicted positive) and very low recall (the proportion of true positives among all positives) on the final validation set (respectively 82% and 56%; *SI Appendix, Fig S3A*). After steps to improve the model's ability to generalize over the entire country (*Materials and Methods*), we launched the final kiln classification model over the 1.8 million images that cover Bangladesh at a spatial resolution of 1 m per pixel. Our model predicted 7,380 images contained kilns (positive) and 1,785,904 did not contain kilns (negative). To evaluate the

**A Confusion Matrix for Kiln Detection**

		PREDICTION	
		Kiln	No Kiln
ACTUAL	Kiln	6547	5
	No Kiln	833	6995

Precision: 0.887  
Recall: 0.999  
Accuracy: 0.942  
Prob. of missing a kiln: 0.07% (95% CI: [0.0088%, 0.13%])

**B Confusion Matrix for Shape Classification**

		PREDICTION	
		Zigzag	FCK
ACTUAL	Zigzag	944	96
	FCK	208	375

Zigzag Accuracy: 0.908 (LB: 0.809, UB: 0.918)  
FCK Accuracy: 0.643 (LB: 0.528, UB: 0.707)  
Overall Accuracy: 0.813 (LB: 0.754, UB: 0.826)

**Fig. 1.** Confusion matrices and performance summary for imagery from all of Bangladesh. (A) The confusion matrix for the performance of the final model from *SI Appendix, Fig. S2* launched over the imagery from all of Bangladesh, after being hand-validated by our team. Note that all 7,380 images the model detected as containing kilns (positive images) and a division-stratified sample of 7,000 nonkiln (negative images) from around 1.7 million were hand-validated. The rows indicate the actual label (kiln, no kiln), while the columns indicate the prediction from the model. The green cells indicate correct predictions, where the top left are true positives (images with kilns that were correctly predicted by the model to contain kilns) and bottom right are true negatives (images with no kilns that were correctly predicted by the model to contain no kiln). The red cells are incorrect predictions, where the top right indicates false negatives (where the model predicted an image did not contain a kiln but the image did contain a kiln) and the bottom left indicates false positives (images that contained no kiln but were predicted by the model to contain a kiln). Performance statistics (precision, recall, and accuracy are presented beneath the confusion matrix. "Accuracy" is defined as the number of images correctly identified divided by the total number of images, "precision" is defined as the number of true positives divided by the total number of images predicted as positive, and "recall" is defined as the number of true positives divided by all positive images. Because the validation of negative images was performed on a sample of images, we constructed an estimate and 95% CI for the probability of missing a kiln. (B) The comparable confusion matrix for the performance of the shape classification model on a hand-validated sample of 1,623 images (note that the original sample contained 1,750 images, but 127 were dropped due to the inability to distinguish the shape during hand validation). Lower and upper bounds on the predictions were calculated by assigning the 127 "unclassifiable" images to different cells. For example, to get the lower bound for "ZZK accuracy" we assume all of the 127 images contain ZZK but were incorrectly classified by our model as FCK.

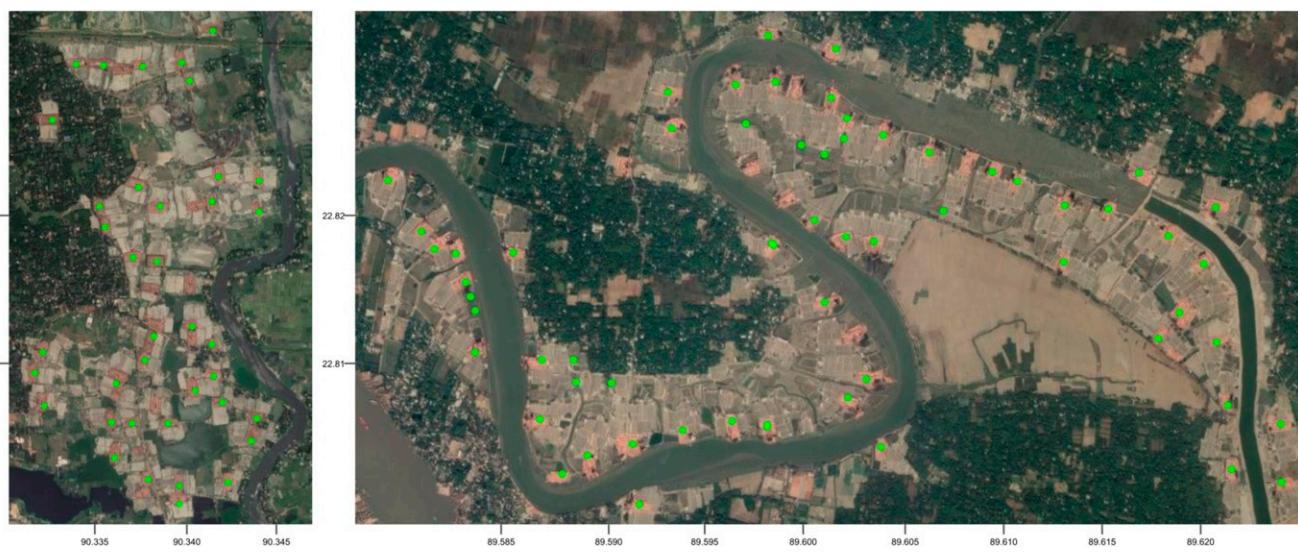
accuracy of this final model, we hand-validated all 7,380 positive images and an administrative division-stratified, random sample of 7,000 negative images. To ensure consistency across the hand-validated results each image was hand-validated by two people and any conflicts were resolved by a third reviewer. Our final model achieved 94.2% accuracy, 88.7% precision, and 99.9% recall (calculated from the hand-validated sample 7,000 negatives) on the imagery from the entire country, as demonstrated in Fig. 1A. Although we cannot report complete accuracy statistics on the final model run because we did not hand-validate all of the 1.79 million negative images, we can obtain an approximation by extrapolating from the 7,000 negative images we hand-validated. Our false omission rate (ratio of false negatives out of all negatives) of 0.07% (5 out of 7,000) implies that out of 1.79 million negative images our model likely missed around 1,276 images with kilns (95% CI: 158 to 2,393; see calculation details in *Materials and Methods*).

Because identifying the precise location of brick kilns is important for regulatory compliance (e.g., is the kiln within the allowable distance from a school), we developed a weakly supervised approach to identify precise geographic coordinates of each kiln (*Materials and Methods* and *SI Appendix*, Fig. S3). From the 6,547 true positive images and 5 false negative images (Fig. 1A), we identified 6,978 centroids in total, which were passed to the shape-classification step of the pipeline to distinguish between the two dominant kiln technologies. The shape-classification model made predictions for each kiln, classifying 1,967 as fixed-chimney kilns (FCK) and 5,011 as zigzag kilns (ZZK). We hand-validated a randomly selected 25% sample of the shape predictions (an initial total of 1,750 images, which was reduced to 1,623 as the shape could not be distinguished in 127 images during hand-validation due to poor image quality), which again involved two people examining each image, with any conflicts resolved by a third person. Based on the hand-validated sample, our shape classification model achieved 81.3% accuracy overall, 90.8% accuracy for ZZK, and 64.3% accuracy for FCK (Fig 1B). Lower and upper bounds on the accuracies were calculated by assigning the 127 “unclassifiable” kilns to the different cells of the confusion matrix. For example, to get the lower bound for “ZZK accuracy” we assume all of the 127 images contain ZZK but were incorrectly classified by our model as FCK. The proportion of ZZK among all hand-validated kilns is 64% (95% CI: 61.7

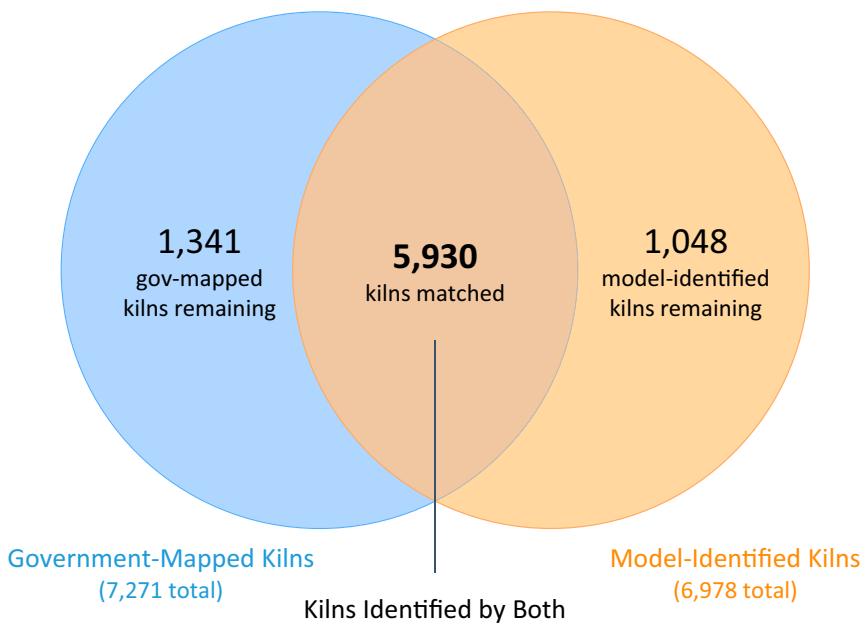
to 66.5%). This estimate, however, does not account for the 127 kilns that our team could not classify due to poor image quality, which adds additional uncertainty. To address this, we reestimated the proportion of ZZK and 95% CIs under two extreme cases: 1) assuming all 127 kilns are FCK and 2) assuming all 127 kilns are ZZK. Under the first scenario when the 127 “unclassifiable” kilns are assumed to be FCK, the proportion of ZZK in the sample becomes 59.4% (95% CI: 57.1%, 61.7%). In the alternative scenario where the 127 kilns are assumed to be ZZK, the proportion of ZZK in the sample becomes 66.7% (95% CI: 64.5%, 68.9%).

In the final step of the pipeline we converted the centroids to geographic coordinates. Fig. 2 shows the coordinates generated by our model in two locations, which demonstrates the pipeline’s ability to geolocate the kilns, often landing on the kiln itself. However, there were also cases when our model failed to obtain the coordinates for every single kiln in an area, usually in dense clusters of kilns. When comparing the model predictions to the nearest-neighbor kiln mapped by the government (Fig. 3), we found the average distance between a model identified kiln and a GoB kiln was 69 m (SD 236 m), while the median was 18 m, suggesting that generally the predicted coordinates fall quite close to the kilns mapped by the government when they could be matched.

**Scale and Scope of Brick Manufacturing.** After removing false positives and adding false negatives identified through hand-validation, our model located 6,978 kilns in Bangladesh. This figure is around 300 fewer than the total number of traditional kilns mapped by the Department of Environment ( $n = 7,271$ ) (20). The recall of our model implies we would miss around 1,276 kilns (95% CI: 158 to 2,393), which implies that the government’s data are also missing kilns. Moreover, because the recall is an approximation based on hand-validating a random sample of negative images, the number of kilns detected by our model should be considered a lower bound on the true number of kilns. We compared the model predictions to the coordinates produced by the GoB through manual mapping and verification (20), by identifying the nearest-neighbor matches within 2 km between the GoB-mapped kilns and the model-predicted kilns. We matched 5,390 kilns, leaving 1,341 unmatched kilns in the government data and 1,048 unmatched kilns in our model



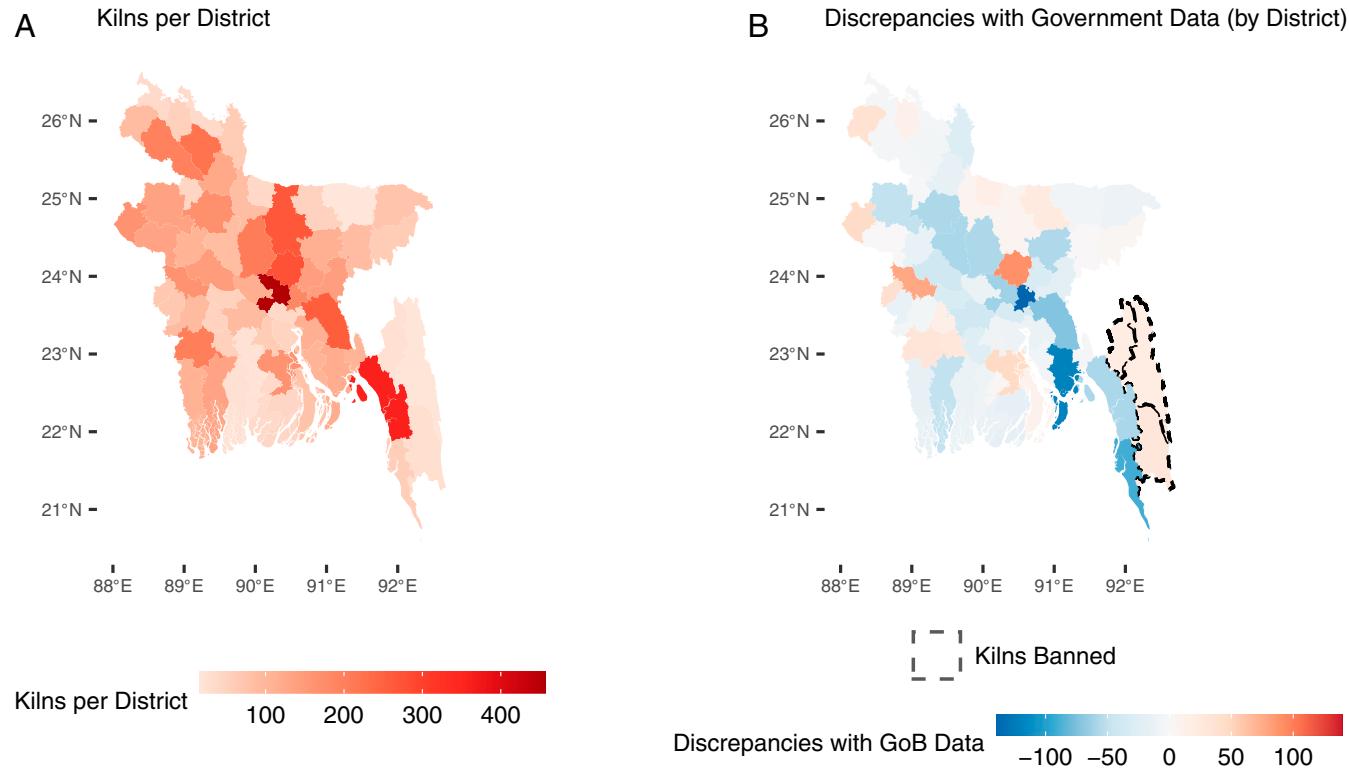
**Fig. 2.** Map of model-generated kiln coordinates. This map plots the coordinates identified by our model (shown in green) over satellite imagery for a cluster of kilns just north of Dhaka, near Gazipur, Bangladesh (Left) and Khulna District (Right). The kilns themselves are the red objects, which are surrounded by rows of drying bricks that appear as the brown color around each kiln.



**Fig. 3.** Comparison of kilns identified by the Bangladesh government (7,271 total) and kilns identified by our model (6,978 total). First, we matched model-identified kilns (6,978 total) to government-mapped kilns (7,271) by identifying the nearest neighbor within 2 km. A total of 5,930 kilns were matched between both sets, leaving 1,341 unmatched kilns among the government-mapped kilns and 1,048 unmatched kilns among the model-identified kilns.

predictions (Fig 3). The number of unmatched kilns in the government data are well within the interval we would expect, given our model's recall. However, since we hand-validated all of the positive images detected by our model, the 1,048 unmatched kilns in the model predictions are assumed to be illegal kilns con-

structed without government permits. The proportion of FCK among the detected illegal kilns (that is, the model-identified kilns that we could not match to a government-mapped kiln) is 28.7%. This is quite similar to the proportion of FCK in the overall predictions (28.2%) but lower than the proportion of FCK



**Fig. 4.** Distribution of kilns across Bangladesh and comparison to government data. (A) Plot of the number of kilns identified by our model per district. (B) Plot of the difference between kilns identified by our model and those registered with the government. Red districts indicate districts where our model identified more kilns than the government records contain, and blue districts indicate the opposite. Three districts where kilns are banned are indicated by the dashed line border (Bandarban, Khagrachhari, and Rangamati).

in the hand-validated sample (35.9%). However, it is important to note that we did not hand-validate the full sample of kiln type results but only a subsample, so this 28.7% is based on the nonvalidated predictions.

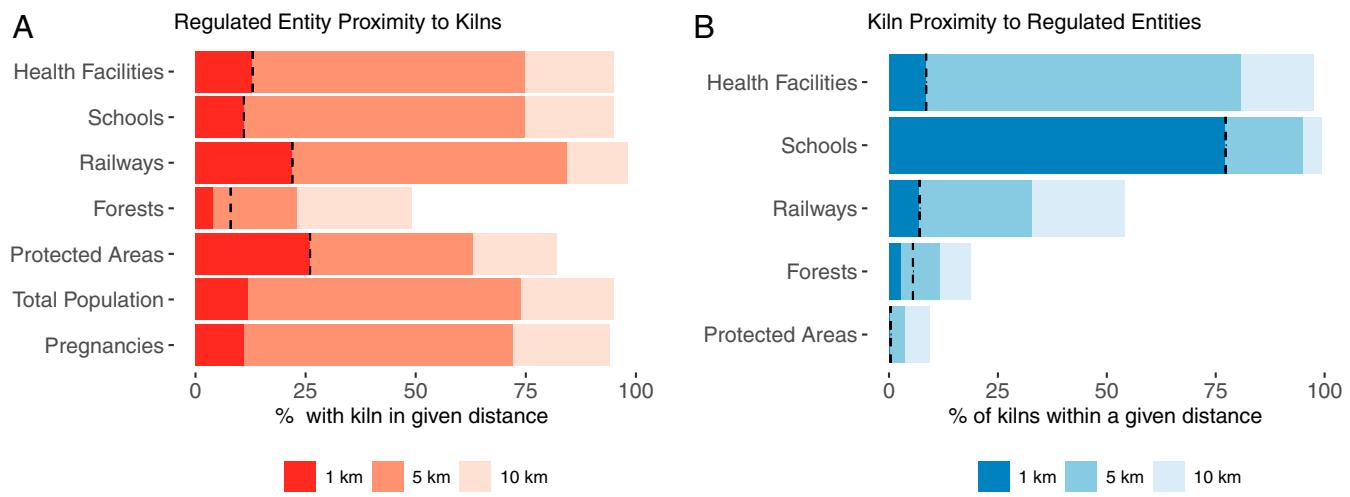
Although the overall number of kilns is similar to the government data, the distribution of kilns differs (Fig. 4A). For example, we found 77 more kilns in the Gazipur district of Dhaka Division than are registered with the Department of Environment (DoE) (the dark red district in the center of the map in Fig. 4B), and when we looked specifically in three hilly districts of Chittagong, where kilns are specifically banned, we found far more discrepancies (the three red districts in the southeast portion of Fig. 4B outlined in the dashed lines). In these three districts (Bandarban, Khagrachhari, and Rangamati), the government data report 1, 9, and 1 registered kilns, respectively, whereas our model detected 30, 27, and 19 kilns. Moreover, the districts that border the three banned districts are blue in Fig. 4B, which indicates fewer kilns predicted by our model than in the government data, suggesting some kilns are formally registered in the adjacent districts where they are legal but then constructed in the banned districts. When we compared the 1,623 kilns for which we hand-validated the shape to government registry of kiln type, we found that the government overreports the percentage of ZZK (newer, cleaner technology) relative to FCK (an older, banned technology) by 7%. In fact, the percentage of ZZK reported in the government data exceeds the upper bound of the 95% CI of our model predictions, which further implies an overreporting of kiln technologies that comply with the law. This systematic underreporting highlights the value of an external, transparent accounting of kilns.

With the precise locations of brick kilns mapped across the entire country, we identified the proportion of the population that is exposed to kilns, as well as the pregnancies that occur near kilns, which may be particularly vulnerable to the air pollution emitted. Using gridded population data from 2017, we found that over 18 million, or 12% of the entire population of Bangladesh, live within 1 km of a kiln, while over 150 million, or 95% of the population, lives within 10 km of a kiln (Fig. 5A). We see similar distributions for the number of pregnancies that occur in proximity to brick kilns.

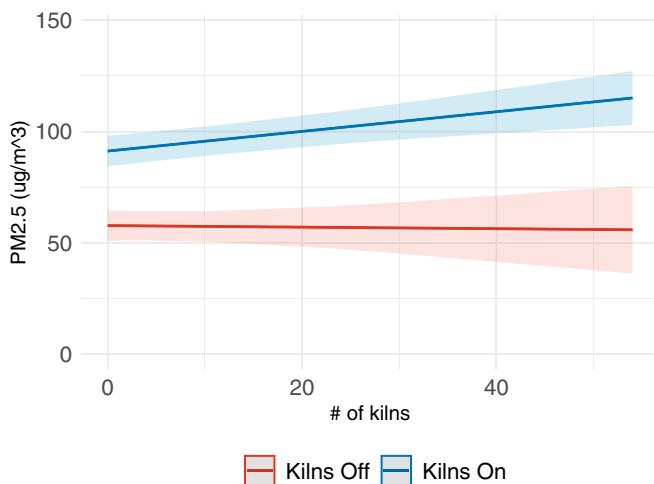
The spatial precision of our results also allows us to check whether kilns are in compliance with the Brick Manufacturing and Brick Kilns Establishment (Control) Act, which mandates

no kiln can built within 1 km from hospitals and clinics, educational facilities, protected areas, and railways and 2 km from public forests. We matched the kilns identified by our model to spatial data on each of these and found that 13% of health facilities, 11% of schools, 8% of forests, 22% of railway length, and 26% of protected areas are proximate to at least one kiln that violates this law (Fig. 5A). Because multiple kilns can be constructed too close to each entity, we also identified the percentage of kilns in violation of the law. We found that 77% of kilns are illegally constructed within 1 km of a school and 9% of kilns are illegally constructed too close to health facilities (Fig. 5B). The percent of kilns illegally constructed too close to railways, forests, and protected areas are 7%, 6%, and 0.5%, respectively. Taken together this represents very limited compliance with national regulations.

To understand whether our map of brick kiln locations could also be used to shed light on environmental outcomes, we combined the brick kiln locations with publicly available  $PM_{2.5}$  data from the air quality monitor at the US Embassy in Dhaka and a global wind reanalysis dataset to identify the prevailing wind direction each day. We then measured whether changes in pollution between the brick-producing season (approximately November to March) and the nonproducing season (the rest of the year) were larger on days when the embassy was downwind from more kilns as compared to days when the embassy was downwind from fewer kilns, controlling for meteorological factors and year fixed effects (*Materials and Methods*). We find that days where the wind blows smoke from one kiln within 15 km during the brick production season raises daily  $PM_{2.5}$  0.47  $\mu g/m^3$  (95% CI: 0.18 to 0.76; see *SI Appendix*, Table S1, column 4), while days where the wind carries pollution from an additional 12 kilns (equivalent to a one-SD increase) during the brick production season increased daily  $PM_{2.5}$  by 5.4  $\mu g/m^3$ . This effect is equivalent to 16% of the daily average increase in  $PM_{2.5}$  during the brick season (*SI Appendix*, Table S1). On the most unfavorable day in this time period, the wind blows pollution from 54 kilns toward the US Embassy; the result on such a day is an increase in  $PM_{2.5}$  by 24  $\mu g/m^3$ . Fig. 6 plots the effect of kilns on daily  $PM_{2.5}$  implied by the regression model and 95% CIs for change in  $PM_{2.5}$ , across the two seasons. The difference between the lines represents the additional effect of being downwind from kilns during the brick-production season (when they are operational) on  $PM_{2.5}$ . Effects of kilns in



**Fig. 5.** (A) The percent of a given regulated entity or population that has a brick kiln within a certain distance. For railways, A plots the average percentage of the rail length that is within 1, 5, and 10 km of a brick kiln. (B) The percent of kilns that are within a certain distance to each regulated entity. In both panels, the darkest color shows the percent within 1 km, while the lightest color shows the percent within 10 km. The dashed vertical lines mark the distance at which the government regulates brick kilns from each entity.



**Fig. 6.** Effect of brick kilns on daily PM<sub>2.5</sub> during the brick-production season. Predicted effects of daily upwind kilns within 15 km, holding all other variables at means, shown from the difference-in-difference regression, plotted in blue for the brick-production season and red for the nonproducing season. The shaded areas are 95% CIs. The difference between the lines represents the additional effect of upwind kilns during the brick production season (when they are operational).

other distances (12 and 20 km) were similar and are shown in *SI Appendix*.

## Discussion

We have demonstrated a successful and highly accurate deep-learning approach for identifying brick kilns in satellite imagery, as well as a series of applications using the model predictions that highlight the utility of obtaining spatially precise, external data on an informal industry. Compared to manual approaches that require either sending individuals out to collect GPS information or manually scanning satellite imagery, our approach is quick, inexpensive, reproducible, and easily scalable. Although the model development required human involvement, now that it has been trained and is highly accurate, it is virtually free to run again, provided there is access to computing power and imagery. Our model can now be applied “off-the-shelf” by government agencies or civil society to locate kilns over large geographies without further human input. Our approach also has the added benefit of objectivity and transparency for independent civil society organizations interested in environmental monitoring as it does not rely on government records, which we demonstrated contain inaccuracies, such as missing kilns (Fig. 3), and are sometimes biased with respect to regulations (Fig. 4B). Additionally, even when taking the upper bound of the shape classification predictions and assuming all 127 “hard” images were ZZK, the proportion of ZZK relative to (illegal) FCK in the government is overstated.

From a machine-learning perspective, we demonstrated the viability of a model that learns to perform the task of geolocating objects from only classification-level labels, a successful weakly supervised learning approach. However, there were limitations. Image quality was not always consistent, as clouds obscured certain areas or the satellite encountered issues in capturing imagery, causing parts of or entire images to be blurry or missing. Although our shape classification algorithm was fairly accurate, it performed worse than the overall prediction model, which was likely influenced by image quality. Future work could improve the accuracy of this step in the pipeline and addressing imagery quality would be an important step here. Nonetheless, even in its current state the shape-classification algorithm could provide results that are extremely useful for regulators

since they provide a predicted probability of whether a kiln is an FCK (banned since 2013) and could be used to target likely violators.

Additionally, there is a trade-off between spatial resolution and processing time. Satellite images taken at higher zoom levels cover less spatial area for a fixed image size than lower zoom levels (because they are more “zoomed in”), which means images are clearer but more images are necessary to cover an area, increasing processing time. One potential solution could employ methods that screen low-resolution images first to identify likely kiln areas then use high-resolution imagery only for those areas to make predictions, saving computational cost (24).

In our results, the final coordinates we identified often landed directly on the kiln, but there were instances when the model was not able to mark every single kiln, particularly in dense clusters. Although we were able to address the problem of double counting by rebuilding kilns split across multiple tiles, the problem of splitting clusters remains another area for future work. A limitation of our image classification approach is that it cannot distinguish between operational and nonoperational kilns if the structure of the kiln is still present. An interesting direction for future research is to incorporate satellite-based information on heat or emissions into the neural network to identify operational kilns.

We selected a few applications to demonstrate the benefits of this approach but also ones that shed light on the regulatory landscape for brick kilns in Bangladesh and provide information on who is affected by the harmful by-products of brick production. These were meant as examples of what is possible with this type of data and demonstrated the applicability to other industrial operations or public health hazards.

Our results suggest that a substantial proportion of the Bangladesh population lives in close proximity to brick kilns and almost the entire population lives within 10 km of a kiln. This is because brick kilns tend to be clustered just outside large urban areas where the majority of construction occurs in response to rapid urbanization and development. We also find that adherence to the Brick Manufacturing and Brick Kilns Establishment (Control) Act is low. Thirteen percent of health facilities and 11% of schools are within 1 km of a brick kiln, in violation of the Brick Manufacturing and Brick Kiln Establishment (Control) Act, while 75% of kilns are illegally constructed too close to schools (Fig. 5B). The brick sector is no anomaly, as corruption and limited capacity hinder enforcement and compliance with regulations across traffic and transportation, forests and conservation, food safety, building codes, and income taxes (25–30).

By examining the contribution of brick kilns to PM<sub>2.5</sub> concentrations, we shed light on the consequences for the exposed populations, school, and health facilities. On days where the wind blew pollution from 54 kilns toward the embassy, PM<sub>2.5</sub> concentrations were 24  $\mu\text{g}/\text{m}^3$  higher relative to days where the wind did not carry any kiln pollution. This is almost equivalent to the World Health Organization guideline for 24-h PM<sub>2.5</sub> exposure (25  $\mu\text{g}/\text{m}^3$ ) coming only from kilns on a bad wind day.

With the evidence presented on illegal proximity of kilns to schools, health facilities, and populations (including pregnancies), our findings suggest that large proportions of the population in Bangladesh are exposed to the air pollution generated by kilns. This exposure is particularly troubling when considering pregnant women and schools that are close to kilns. Pregnant women and school children are highly vulnerable to pollution exposure (31), and both in utero and early-life exposure to air pollution has been linked to poor pregnancy outcomes, poor child health status, lower academic performance and cognitive development, and other adverse adult outcomes (32, 33). Given that our model missed around 1,276 kilns, the extent of exposure and adverse impacts are likely underestimates. Ultimately,

our results demonstrate the regulations mandating kilns be built at least 1 km from critical areas are insufficient from a health perspective.

Easily identifying violations of existing law could help a constrained regulatory apparatus determine how to best allocate resources or set priorities regarding violations. Moreover, the ability to generate external and transparent data on violations introduces external transparency to a system plagued by corruption (16). For example, knowing the precise locations of brick kilns would be useful for civil society organizations concerned about child and bonded labor and other potential human rights violations. We documented fairly extensive violations of regulations governing where kilns can be built with respect to schools, health facilities, forests, railways, and three banned districts. Similarly, FCK have been banned in Bangladesh since 2013, yet many still exist (both according to our model predictions and the government's own data). Running an algorithm such as ours to identify likely violations, such as presence of FCK or kilns built too close to forests, schools, or health facilities, before manually verifying and sending government representatives to address violations could reduce the costs of oversight.

The current mode of brick manufacturing causes substantial harm to human health and the environment, yet as we have demonstrated in this analysis compliance with existing regulations is low. In the context of limited capacity to monitor and enforce regulations, low-cost tools that can enhance the government's ability to detect violations have great potential. Alternatively, in the hands of civil society organizations, this information could be used to publicly disclose violators and introduce a disincentive for hiding illegal activities, as well as pressure government officials to enforce the law. Ultimately, advances in machine learning and Earth observation hold great promise for promoting transparency and improving regulatory compliance, thus reducing human health and environmental implications of noncompliance across a range of industries and sectors.

## Materials and Methods

### Machine-Learning Pipeline.

**Initial data and limitations.** We downloaded satellite imagery from DigitalGlobe to cover the entire region of Bangladesh from the period of October 2018 to May 2019, filtering out images that were more than 5% obscured by clouds. Each tile was 224 by 224 pixels with three channels (representing the red, green, and blue color values of an image) at a zoom level of 17, which translates to approximately a 1-m-per-pixel resolution. The entire country is covered by 1.8 million images at this spatial resolution. The latitude and longitude coordinates of the top left corner of each image were included as part of the metadata. To form the initial training data, field workers collected GPS coordinates of kilns from Brahmanbaria, Jessore, Manikganj, Tangail, and Mymensingh Districts in Bangladesh and we hand-picked coordinates in the nearby vicinity without kilns. We identified the satellite images that contained these coordinates and removed duplicates, resulting in 52 "positive" images (images that contain a kiln) and 364 "negative" images (images that do not contain a kiln). Since kilns can be split across multiple images due to the nature of satellite imagery, we defined an image as positive if at least 25% of a kiln appeared in the image. We refer to this data as the initial dataset (*SI Appendix*, Fig. S2). The first task of our pipeline was to classify images as either positive or negative.

We used deep learning to train a convolutional neural network (CNN) and experimented with different architectures, including VGG-16 (34), VGG-19 (34), Inception (35), Xception (36), and ResNet50 (37). Each architecture comes with different strengths and weaknesses, e.g., speed of training and testing, accuracy, and size. We found that the VGG-16 architecture performed best in terms of precision, which was important for our specific task. Moving forward with a VGG-16 architecture for the CNN, we took a transfer learning approach, which means we initialized the network with weights that were produced by training on the ImageNet (38) dataset. Transfer learning is a widely used technique that allows a model to make use of features learned from training on a much larger dataset, which is then fine-tuned for specific task, i.e., detecting the presence of kilns. We split the

initial dataset into a training set (80%) and a validation set (20%), maintaining the ratio of positive to negative images from the overall set of images within the training and validation sets. As in standard machine-learning approaches, the model learned from the examples in the training set and made predictions on the validation set, the results of which allowed us to assess the performance of the model.

The model quickly reached 98% accuracy on the training data and 97% accuracy on the validation set. However, the initial dataset was limited in both size and geographic diversity. When we randomly sampled 10% of the entire country, a total of 176,000 images, the model predicted around 10,726 images as positive, but only 621 of them were true positives. The initial dataset provided useful examples, but they were concentrated in one region of the country. In order to make accurate predictions on millions of diverse images from across the country, we needed to expand our dataset to capture greater variation in the visual appearance of kilns and their surroundings. Thus, we utilized a data expansion technique for the purpose of improving the dataset. To expand the data, we conducted two rounds of hand validation to verify the model's predictions, once to acquire examples from other divisions and a second to label geographically contiguous regions to obtain examples of kilns split across adjacent satellite images.

**Data expansion.** To create a more nationally representative dataset, we randomly sampled 10,000 images from each administrative division (except Dhaka, from which we sampled 15,000 to account for its larger size and the fact that it was later split into two divisions) for a total of 75,000 images to label. We considered the model trained from the initial dataset as a "teacher" model that can educate a "student's" model. We used the teacher model to make predictions on the sample of 75,000 images and hand-validated the predictions. We uncovered 301 additional positive images and randomly sampled images from the false positives and true negatives for an additional 1,614 negative images. We incorporated these additional training examples and refer to this dataset as Augmented Dataset 1 or *AugData*<sup>1</sup> (*SI Appendix*, Fig. S2). We split *AugData*<sup>1</sup> into a training set with 80% of the data (*AugData*<sub>train</sub><sup>1</sup>) and a validation set with the remaining 20% (*AugData*<sub>val</sub><sup>1</sup>). The output predictions of the teacher model provided initial labels that were verified or corrected by our research team. We then trained a new model on this expanded dataset, which is considered the student model as it learned from the results of the teacher model. This process can be automated and repeated to expand the dataset further [known as data distillation (39)]. We conducted one round of data distillation with "humans in the loop" (40), which means we hand-verified predictions to reduce noise and ensure the data were high-quality.

**Robustness experiments.** Before launching the model across the entire country, we verified the robustness of our model in terms of two factors: 1) its ability to generalize and perform well in geographically distinct parts of the country and 2) its ability to detect kilns that are split across multiple images and kilns that are clustered together. For the robustness experiments, we first hand-validated images covering two contiguous regions from different parts of the country (Sylhet Division and Tangail District). The use of contiguous regions helped to address the issue of kiln splitting since all adjacent images were contained in the sample, allowing us to rebuild kilns across multiple images.

We ran the model trained on *AugData*<sup>1</sup> over images from the Sylhet and Tangail contiguous regions. We observed that model performance dropped substantially in Sylhet, which is a hilly region in northeast Bangladesh that is different in terrain and brick kiln density, obtaining only 47% precision compared to the 98% it achieved on *AugData*<sub>val</sub><sup>1</sup>. This raised concerns that our model was overfitting. We used a hard example mining (41) approach, in which we identified images that the model classified incorrectly ("mistakes") and thus could be considered hard or difficult examples. These were included to expand the diversity of the training data and thus the model's knowledge.

To construct the most effective training dataset, we tested incorporating different combinations of the model's mistakes in Sylhet and Tangail:

- 1) *AugData*<sup>1</sup>: only examples from Augmented Dataset 1, which includes examples from the initial data and 75,000 sample
- 2) *AugData*<sup>1</sup> + *S*: examples from Augmented Dataset 1 and misclassifications from Sylhet (*S*)
- 3) *AugData*<sup>1</sup> + *T*: examples from Augmented Dataset 1 and misclassifications from Tangail (*T*)
- 4) *AugData*<sup>1</sup> + *S* + *T*: examples from Augmented Dataset 1 and misclassifications from both Sylhet (*S*) and Tangail (*T*)

We split each of the above datasets into a training (80%) and validation (20%) set and trained four models, one for each dataset. We tested the

models on all four validation sets, as well as the entire contiguous regions of Sylhet and Tangail to check their performance at scale. The model trained on *AugData*<sup>1</sup> + *S* performed best. We moved forward with this final training dataset, which we call *AugData*<sup>2</sup> (*SI Appendix*, Fig. S2). We conducted more iterations of training to tune hyperparameters (parameters that are set before training and affect aspects such as how the model learns, how quickly, and for how long) to optimize performance gains, resulting in our final model. To compare the different models we developed throughout these steps to improve the robustness and generalizability, we evaluated each one on the same validation set (*SI Appendix*, Figs. S4 and S5). These results demonstrate the gains in performance achieved by the training data expansion and robustness experiments and the high level of accuracy our final model obtained.

**Kiln localization.** We developed a weakly supervised approach in which the model learns from the image-level classification labels of positives and negatives to perform the harder task of localization (i.e., identifying where within an image the kiln is located). This approach is weakly supervised precisely because it does not require the training data to be annotated with exact kiln locations, which can be very time-intensive to produce. First, we modified the architecture of the CNN. We added a global average pooling layer as in ref. 42, which takes the feature maps the model has learned in past layers and weighs them in terms of how much they contribute to the final prediction. By summing the contributions, it is possible to produce a class activation map (CAM), which visualizes the model's signals, like a heat map of which pixels contribute most strongly to the model's prediction (*SI Appendix*, Fig. S3). These clusters of activated kilns provided an approximation of where the kilns were located in a positive image.

Kilns can be split across multiple satellite image tiles, resulting in the same kiln's being detected in multiple images and thus counted more than once. Therefore, we developed an approach to rebuilding split kilns by generating "connected components," in which spatially adjacent tiles are stitched together to form a contiguous area. Starting from a given positive image, our approach used breadth first search to branch out to find all its neighboring positive images. By stitching the images together, we also connected activated areas from the CAMs.

However, the CAM output was very noisy because features in an image other than the kilns contributed to the model's prediction. To focus the model signals on just the kilns, we conducted postprocessing on the connected components. Each pixel had a value between 0 and 255, where a higher value meant a stronger "kiln" signal. The first step filtered out weak signals. We subtracted the mean pixel value from all pixels, which zeroed out the areas with low values. Then, we calculated the mean pixel value from only nonzero pixels and subtracted that value as well, leaving only the strongest signals in the image. We then normalized the values, such that the lowest value was 0 and highest was 255. Next, we concentrated the signals and filtered out any remaining noise. We increased contrast to amplify the signals and passed the image through a median filter and Gaussian filter to smooth areas and remove any small patches of signals that were too small to be a kiln. We iterated this second step of amplifying and smoothing signals. Finally, since the smoothing could unintentionally expand the size of kilns, we used a morphological operator that erodes pixels around the edges of the remaining clusters of signals. The centroids of each of these clusters provided the locations of the kilns.

In cases with multiple kilns in a single image, the activated pixels could be conjoined as one large mass in the CAM output. We kept track of a moving average of the size of these activated areas to estimate the average size of a kiln. When an activated area was larger than the average, we marked more centroids. For example, when an area was twice as large as the average kiln size, we collected two centroids spread evenly across the pixel mass.

**Shape classification and generating coordinates.** Equipped with the centroids of the kiln clusters, we implemented the final step of the pipeline: classifying the shape of each kiln. First, we prepared a dataset by labeling the hand-validated positive images detected across the country with kiln shapes. There were two classes: FCK, which are ovoidal, and ZZK, which are rectangular. Of the 3,345 images we labeled, 1,193 were FCK and 2,152 were ZZK, which were split into a training and validation set with the same ratio of FCK to ZZK. Using deep learning and transfer learning again, we trained a second CNN to classify the kiln shape, which achieved 98% accuracy on the training set and 96% accuracy on the validation set. With an accurate shape classification model trained, we then launched our end-to-end pipeline on unseen imagery covering the entire country. For each centroid identified from the localization step of the pipeline, we cropped an image

centered around the centroid of size 224 × 224. We input each of these recentered and cropped images to the shape classification CNN to produce the final predictions. Finally, we translated the pixel location (x,y) of the centroid within the image to its geographic coordinates, modeling the distance in x as a difference in longitude and the distance in y as a difference in latitude. The entire pipeline is portrayed visually in *SI Appendix*, Fig. S3.

**Evaluating final model performance.** In a standard prediction problem, to evaluate the performance of an algorithm one would compare the model predictions to the true values in a test dataset. During model development, we used data with underlying labels (kiln or no kiln) split into training, validation, and test sets to evaluate the model's performance (*SI Appendix*, Figs. S4 and S5). However, our final model was run on imagery covering the country, which does not come with labels. To assess the model's performance, we hand-validated all images predicted as "kiln" (7,380 images) and a division-stratified, random sample of 7,000 negative predictions. We generated a confusion matrix with the results of this hand validation (Fig. 1). However, because we did not hand-validate all 1.79 million negative images, the accuracy statistics are an estimate. To account for this uncertainty, we used the false-negative rate we found in the hand-validated sample to approximate the number of false negatives across the full set of negative predictions and construct a CI around the estimation.

To do this, we used the maximum likelihood estimator to estimate the false-negative rate as the mean and variance of a Bernoulli variable and then constructed a 95% CI using a critical value of 1.96. Specifically, let  $p$  be the probability of any negative image ("no kiln") of being a false negative (image that actually contains a kiln). If in a sample of negative images of size  $N$ , there are  $n$  false negatives, then the maximum likelihood estimate (MLE) of  $p$  is

$$\hat{p} = \frac{n}{N}. \quad [1]$$

Also, for a large sample size,  $\hat{p}$  converges to  $p$  in distribution (asymptotic distribution of the MLE), i.e.,

$$\sqrt{NI(p)}(\hat{p} - p) \sim \mathcal{N}(0, 1), \quad [2]$$

where  $I(p)$  is the Fisher information. For a Bernoulli variable, we have

$$I(p) = \frac{1}{p(1-p)}, \quad [3]$$

which is also the inverse of the variance. With this, we construct a 95% CI around  $\hat{p}$  using a critical value of 1.96 ( $z_{\alpha/2}$ , with  $\alpha = 0.05$ ):

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}. \quad [4]$$

Similarly, to construct a CI for the performance of the shape detection algorithm, which distinguished between FCK and ZZK, we used the same approach. If we let  $n$  be the number of ZZK in our hand-validated sample of  $N$  brick kilns, then the estimated ratio of ZZK to all kilns,  $\hat{p}$ , can be derived using the same formula (Eq. 1), and the 95% CI can be obtained using Eq. 4.

#### Scale and Scope of Brick Manufacturing.

**Data.** We compiled numerous sources of data to assess the scope of brick manufacturing in Bangladesh and the extent of regulatory compliance. First, we used two sources of data from the DoE of the GoB on the official number of kilns. One is the aggregate number of formally registered kilns in each district from 2017 to 2019, split out by the type of kiln and whether the kiln owner secured environmental clearance (18). The other is GPS coordinates of kilns that were manually mapped by the Clean Air and Sustainable Environment (CASE) project (20). Second, we used globally gridded datasets on total population and pregnancies produced by WorldPop (43) at a 100-m spatial resolution to identify exposed populations, particularly populations of interest to policymakers.

Third, we utilized publicly available data on the Humanitarian Data Exchange (44) to assess compliance with the GoB regulations (details on the regulations are given below). Specifically, we used shapefiles containing nationally designated forests and national parks, railways, health facilities, and schools, and administrative boundaries, as well as shapefiles from the World Database on Protected Areas (45).

Finally, to assess the contribution of brick kilns to air quality, we utilized publicly available PM<sub>2.5</sub> data spanning 2017 to 2019 from the air-quality monitor at the US Embassy in Dhaka. The data contained observations every

3 h, which were aggregated to a daily average. We also used gridded data on hourly precipitation, temperature, and wind speed and direction (represented by the  $u$  and  $v$  vector components) from the Climate Date Source ERA5-Land reanalysis data at a spatial resolution of  $0.1^\circ$  latitude  $\times$   $0.1^\circ$  longitude. We extracted all meteorological data for a 15-km buffer circle around the US Embassy in Dhaka and aggregated all data to a daily average. Our analytic approach (see below) utilizes daily changes in wind direction to determine whether pollution from kilns traveled downwind and affected the readings at the embassy. To do this, we converted the  $u$  and  $v$  wind components to wind speed and direction (see Eq. 5, where  $\theta$  is wind direction in degrees and  $s$  is wind speed) and identified the daily wind direction blowing toward the embassy using  $45^\circ$  and  $30^\circ$  sectors. Then, we calculated the pairwise distance and bearing between the embassy and each kiln identified by the model. Given the bearing between each kiln and embassy, we determined whether it was upwind of the embassy (in both the  $45^\circ$  and  $30^\circ$  sectors). We aggregated these by distance to determine the number of upwind kilns in increasing distances from the embassy (12 to 50 km). *SI Appendix, Fig. S6* presents a diagram that shows two different wind directions and how kilns would be classified. The resulting data are a daily-level dataset with  $PM_{2.5}$ , rain, temperature, wind speed, wind direction, and the number of upwind kilns in different distances—all varying daily:

$$\theta = 180 + \frac{180}{\pi} \arctan 2(v, u) \quad [5]$$

$$s = \sqrt{u^2 + v^2}$$

**Analysis.** We first conducted an external check on the GoB reported kiln numbers by district and type of kiln, by aggregating the kilns identified by our machine-learning model to the district level. We calculate the number of kilns per district and the difference between the GoB data and our model predictions. We also compared the percentage of kilns with environmental clearance (ZZK) reported in GoB data to the shape predictions from our model. To compare our model predictions to the kiln coordinates manually mapped by the GoB, we used a nearest-neighbor matching algorithm that matches to kilns that were within 2 km of each other. Then, to understand who is exposed to brick kilns nationally, we matched the kiln coordinates to the gridded population data from WorldPop on total number of people and total number of pregnancies. We summed the grid cells that were within varying distances to kilns to identify the exposed populations.

The Brick Manufacturing and Brick Kiln Establishment (Control) Act of 2013 restricts kilns from being built within 1 km of hospitals and clinics, educational facilities, protected areas, railways, and wetlands and 2 km of public forests. Additionally, per the Brick Kiln Policy of 2008 no brick kilns can be set up in three hilly districts (Rangamati, Khagrachari, and Bandarbans) (19). Using the publicly available data on each of these regulated entities, we calculated the distance between each forest, rail line, protected area, school, and health facility to every kiln identified in the country. We checked whether the minimum distances violate the GoB regulations. This allowed us to say what percentage of each regulated entity is exposed to kilns and also what percentage of kilns are in violation of the law. For forests, protected areas, and the three hilly districts we also checked whether any brick kilns fell within their designated geographic boundaries.

Finally, we assumed the primary rationale of regulating the proximity of brick kilns in areas of importance, such as city centers, schools, and health facilities, is to reduce the population's exposure to air pollution generated by the kilns. To test whether  $PM_{2.5}$  is higher around areas that are downwind from more brick kilns, we leveraged daily changes in wind direction that resulted in differing numbers of kilns being upwind from the embassy each day and seasonal variation in the timing of brick production (kilns operate only in the dry, winter months between November and March),

1. A. Heyes, Making things stick: Enforcement and compliance. *Oxf. Rev. Econ. Pol.* **14**, 50–63 (1998).
2. A. Heyes, Implementing environmental regulation: Enforcement and compliance. *J. Regul. Econ.* **17**, 107–129 (2000).
3. N. Gunningham, Enforcing environmental regulation. *J. Environ. Law* **23**, 169–201 (2011).
4. J. P. Shimshack, The economics of environmental monitoring and enforcement. *Annu. Rev. Res. Econ.* **6**, 339–360 (2014).
5. C. Handan-Nader, D. E. Ho, Deep learning to map concentrated animal feeding operations. *Nat. Sustain.* **2**, 298–306 (2019).
6. E. Benami, D. E. Ho, A. McDonough, “Innovations for environmental compliance: Emerging evidence and opportunities” (SIEPR policy brief, Stanford Institute for Economic Policy Research, Stanford, CA, 2020).
7. M. Hino, E. Benami, N. Brooks, Machine learning for environmental monitoring. *Nat. Sustain.* **1**, 583–588 (2018).
8. R. Purdy, Using earth observation technologies for better regulatory compliance and enforcement of environmental laws. *J. Environ. Law* **22**, 59–87 (2010).
9. A. Blackman *et al.*, The benefits and costs of informal sector pollution control: Mexican brick kilns. *Environ. Dev. Econ.* **11**, 603–627 (2006).
10. B. A. Begum, Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh. *Atmos. Environ.* **38**, 3025–3038 (2004).
11. B. A. Begum, P. K. Hopke, A. Markwitz, Air pollution by fine particulate matter in Bangladesh. *Atmos. Pollut. Res.* **4**, 75–86 (2013).
12. S. K. Guttikunda, B. A. Begum, Z. Wadud, Particulate pollution from brick kiln clusters in the Greater Dhaka region, Bangladesh. *Air Quality Atmos. Health* **6**, 357–365 (2013).

to estimate the causal effect of kilns on downwind  $PM_{2.5}$ . This approach is similar to other recent studies that have used wind direction to isolate the effect of air pollution from other confounding sources of pollution (46, 47), with the additional advantage of seasonality in brick production. See *SI Appendix, Fig. S6* for a diagram of how changes in wind direction were used to identify upwind kilns on different days. We estimated the following regression:

$$Y_d = \beta_0 + \beta_1 \text{Upwind Kilns}_d + \beta_2 \text{Brick Season}_d + \beta_3 \text{Upwind Kilns}_d \times \text{BrickSeason}_d + \theta X_d + \delta_y + \epsilon_d, \quad [6]$$

where  $Y_d$  is the daily  $PM_{2.5}$  measured at the US Embassy in Dhaka;  $\text{UpwindKilns}_d$  measures the number of upwind kilns relative to the embassy on day  $d$ ;  $\text{BrickSeason}_d$  is an indicator whether the day falls during the brick production season (November to March);  $\text{UpwindKilns}_d \times \text{BrickSeason}_d$  is the interaction between the number of upwind kilns and the brick season indicator;  $X_d$  is a vector of controls calculated for the US Embassy including quadratic polynomials of precipitation, temperature, and wind speed;  $\delta_y$  represents a year fixed effect (to control for secular trends in air pollution unrelated to kilns); and  $\epsilon_d$  represents the error term adjusted for serial autocorrelation between daily observations. The difference-in-difference approach isolates the causal effect of more upwind brick kilns on a given day during the brick season (represented by  $\beta_3$  in Eq. 6). We tested the effect of upwind kilns in various distances (12, 15, and 20 km). Results for upwind kilns within 15 km (identified using a  $30^\circ$  upwind area are shown in *Results*, and other distances and  $45^\circ$  upwind sectors are presented in the *SI Appendix*.

The strength of this approach lies in the comparison of days that vary in the number of upwind kilns across the two time periods. Because brick kilns operate only between November and March, they should not affect air quality during the off-season, and in particular only kilns that are upwind should be responsible for air pollution downwind, whereas kilns that are downwind from the embassy should not. Additionally, the inclusion of year fixed effects controls for secular trends in air pollution that are unrelated to brick kilns, further strengthening the causal interpretation of this result.

**Data Availability.** We used DigitalGlobe (now Maxar Technologies) imagery, which is proprietary. However, we are able to share a “de-geo-identified” version of the imagery used in our machine learning pipeline. This data is stripped of its geospatial metadata, which means certain steps of our pipeline, such as translating the pixel locations of kilns detected within an image to a latitude and longitude will not be reproducible with this data. We also provide the final output of the machine learning pipeline (the GPS coordinates of all model predicted kilns), so the remainder of the analysis is replicable. The Government of Bangladesh data on brick kilns is proprietary and can be requested directly from the Department of Environment. All other geospatial data used to assess the scale and scope of brick manufacturing (population distributions, forests, schools, health facilities, protected areas,  $PM_{2.5}$ , wind, precipitation, temperature, and administrative boundaries) are publicly available and we provide details on how to obtain them. All data along with Python and R scripts for the entire machine learning pipeline and spatial analysis are available on the Harvard Dataverse at <https://dataVERSE.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HVGW8L> (48).

**ACKNOWLEDGMENTS.** We gratefully acknowledge DigitalGlobe for making the satellite imagery available; Christina Sintek and Zhongyi Tang for support with accessing the imagery; and Chris Tan, Bora Erden, and Sergio Camelo for research assistance with earlier iterations of this work. We acknowledge the funding from the Stanford King Center for Global Development, the Stanford Woods Institute for the Environment (EVP Grant), and NSF grants 1651565 and 1522054.

13. C. Weyant *et al.*, Emissions from South Asian brick production. *Environ. Sci. Technol.* **48**, 6477–6483 (2014).
14. A. Eil, J. Li, P. Baral, E. Saikawa, “Dirty stacks, high stakes: An overview of brick sector in South Asia” (World Bank, 2020).
15. J. S. Apte, M. Brauer, A. J. Cohen, M. Ezzati, C. A. Pope, Ambient PM<sub>2.5</sub> reduces global and regional life expectancy. *Environ. Sci. Technol. Lett.* **5**, 546–551 (2018).
16. S. P. Luby, D. Biswas, E. S. Gurley, I. Hossain, Why highly polluting methods are used to manufacture bricks in Bangladesh. *Energy Sustain. Dev.* **28**, 68–74 (2015).
17. World Bank, “Introducing energy-efficient clean technologies in the brick Sector of Bangladesh” (Technical report, IBRD/World Bank, Washington, DC, 2011).
18. Department of Environment, “Registered brick kilns” (DOE report, Government of Bangladesh, Dhaka, Bangladesh, 2019).
19. Department of Environment, “National strategy for sustainable brick production in Bangladesh” (Technical report, Dhaka, Bangladesh, 2017).
20. M. M. Rana, Sources of air pollution in Bangladesh: Brick kiln & vehicle emission scenario (Technical report, Department of Environment, Bangladesh, 2019).
21. A. Kinyondo, R. Pelizzo, Poor quality of data in Africa: What are the issues?: Politics and policy. *Polit. Pol.* **46**, 851–877 (2018).
22. U. Nazir, U. K. Mian, M. U. Sohail, M. Taj, M. Uppal, Kiln-net: A gated neural network for detection of brick kilns in South Asia. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **13**, 3251–3262 (2020).
23. G. Foody, F. Ling, D. Boyd, X. Li, J. Wardlaw, Earth observation and machine learning to meet sustainable development goal 8.7: Mapping sites associated with slavery from space. *Rem. Sens.* **11**, 266 (2019).
24. K. Ayush, B. Uzkent, M. Burke, D. Lobell, S. Ermon, Generating interpretable poverty maps using object detection in satellite images. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2002.01612> (Accessed 9 October 2020).
25. T. Sarker, Who bears the burden of taxes in developing countries? A case of income taxation in Bangladesh. *Pakistan Econ. Soc. Rev.* **44**, 181–207 (2006).
26. R. Rezaur, Road traffic accident situation in Khulna city, Bangladesh. *Proc. Eastern Asia Soc. Transport. Studies* **5**, 65–74 (2005).
27. M. Chowdhury, M. Koike, N. Muhammed, Embracing collaborative protected area management for conservation: An analysis of the development of the forest policy of Bangladesh. *Int. For. Rev.* **11**, 359–374 (2009).
28. M. Murshed, Cutting of hills and human tragedy: An example of law enforcement loophole in the southeastern region of Bangladesh. *Int. J. Geomatics Geosci.* **4**, 410 (2013).
29. I. Ahmed *et al.*, Compliance to building codes for disaster resilience: Bangladesh and Nepal. *Procedia Eng.* **212**, 986–93 (2018).
30. A. Ali, Food safety and public health issues in Bangladesh: A regulatory concern. *Euro. Food Feed Law Rev.* **8**, 31–40 (2013).
31. World Health Organization, “Effects of air pollution on children’s health and development: A review of the evidence” (Technical Report EUR/05/5046027, WHO Regional Office for Europe, Copenhagen, 2005).
32. J. Currie, J. G. Zivin, J. Mullins, M. Neidell, What do we know about short- and long-term effects of early-life exposure to pollution? *Annu. Rev. Res. Econ.* **6**, 217–247 (2014).
33. J. G. Zivin, M. Neidell, Environment, health, and human capital. *J. Econ. Lit.* **51**, 689–730 (2013).
34. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint]* (2015). <https://arxiv.org/abs/1409.1556> (Accessed 9 October 2020).
35. C. Szegedy *et al.*, Going deeper with convolutions. *arXiv [Preprint]* (2014). <https://arxiv.org/abs/1409.4842> (Accessed 9 October 2020).
36. F. Chollet, Xception: Deep learning with depthwise separable convolutions. *arXiv [Preprint]* (2016). <https://arxiv.org/abs/1610.02357> (Accessed 9 October 2020).
37. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *arXiv [Preprint]* (2015). <https://arxiv.org/abs/1512.03385> (Accessed 9 October 2020).
38. J. Deng *et al.*, “ImageNet: A large-scale hierarchical image database” in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 248–255.
39. I. Radovanic, P. Dollár, R. Girshick, G. Gkioxari, K. He, “Data distillation: Towards omni-supervised learning” in *2018 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 4119–4128.
40. G. Wang, Humans in the loop: The design of interactive AI systems. *J. Artificial Intelligence Res.* **64**, 243–252 (2019).
41. A. Shrivastava, A. Gupta, R. Girshick, “Training region-based object detectors with online hard example mining” in *2016 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 761–769.
42. B. Zhou, A. Khosla, Á. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization. *arXiv [Preprint]* (2015). <https://arxiv.org/abs/1512.04150> (Accessed 9 October 2020).
43. WorldPop, Global High Resolution Population Denominators Project. <https://dx.doi.org/10.5258/SOTON/WP00674>. Accessed 2 April 2021.
44. Humanitarian Data Exchange, Bangladesh. <https://data.humdata.org/group/bgd>. Accessed 2 April 2021.
45. UNEP-WCMC, Protected area profile for Bangladesh from the World Database of Protected Areas, April 2021. <https://www.protectedplanet.net/>. Accessed 2 April 2021.
46. M. A. Rangel, T. S. Vogl, Agricultural fires and health at birth. *Rev. Econ. Stat.* **101**, 616–630 (2019).
47. T. Deryugina, G. Heutel, NH. Miller, D. Molitor, J. Reif, The mortality and medical costs of air pollution: Evidence from changes in wind direction. *Am. Econ. Rev.* **109**, 4178–4219 (2019).
48. J. Lee, *et al.*, Replication Data for: Scalable deep learning to identify brick kilns and aid regulatory capacity. Harvard Dataverse. <https://doi.org/10.7910/DVN/HVGW8L>. Deposited 9 April 2021.