Targeting Direct Cash Transfers to the Extremely Poor

Brian Abelson
Enigma
New York, NY USA
brianabelson@gmail.com

Kush R. Varshney
IBM Research
Yorktown Heights, NY USA
krvarshn@us.ibm.com

Joy Sun*
GiveDirectly
New York, NY USA
joy@givedirectly.org

ABSTRACT

Unconditional cash transfers to the extreme poor via mobile telephony represent a radical, new approach to giving. GiveDirectly is a non-governmental organization (NGO) at the vanguard of delivering this proven and effective approach to reducing poverty. In this work, we streamline an important step in the operations of the NGO by developing and deploying a data-driven system for locating villages with extreme poverty in Kenya and Uganda. Using the type of roof of a home, thatched or metal, as a proxy for poverty, we develop a new remote sensing approach for selecting extremely poor villages to target for cash transfers. We develop an analytics algorithm that estimates housing quality and density in patches of publicly-available satellite imagery by learning a predictive model with sieves of template matching results combined with color histograms as features. We develop and deploy a crowdsourcing interface to obtain labeled training data. We deploy the predictive model to construct a fine-scale heat map of poverty and integrate this discovered knowledge into the processes of GiveDirectly's operations. Aggregating estimates at the village level, we produce a ranked list from which top villages are included in GiveDirectly's planned distribution of cash transfers. The automated approach increases village selection efficiency significantly.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding; I.5.4 [**Pattern Recognition**]: Applications; J.4 [**Social and Behavioral Sciences**]: Economics

Keywords

poverty economics; remote sensing; social good

*This project was conducted under the auspices of the DataKind DataCorps, Brooklyn, NY USA.

KDD'14, August 24–27, 2014, New York, NY, USA.

http://dx.doi.org/10.1145/2623330.2623335.

1. INTRODUCTION

GiveDirectly is a non-profit, non-governmental organization (NGO) that aims to help people living in extreme poverty by making unconditional cash transfers to them via mobile telephony. Unlike other charitable giving, unconditional cash transfers do not presuppose that food, livestock, or any other provisions are best for the recipients; recipients are free to spend the funds in any way they choose. Evidence from a randomized control trial shows that this method of charity has large positive effects on multiple measures of recipients' well-being [7]. Additionally, by forgoing large intermediary infrastructures, donations can be used extremely efficiently; to date, more than 90% of each donated dollar has reached a recipient. Currently operating in Kenya and Uganda, GiveDirectly takes an end-to-end operations model and does not outsource or subcontract work to other organizations.

Several steps comprise GiveDirectly's operations. First, funds are solicited from individual and institutional donors, and collected in GiveDirectly's bank account. Once a sufficient amount has been collected, e.g. US$ 1 million, the NGO initiates a campaign to disburse the donations to the extremely poor. The first part of a campaign is enrolling extremely poor households to be recipients; the second is transferring funds to recipients via mobile money systems such as M-Pesa [8]; and the third is conducting telephone and in-person follow-up with each recipient.

Targeting extremely poor households starts by identifying regions of the country with high poverty rates through data from the national census. Within those regions, villages and households are selected through a transparent criterion that is associated with extreme poverty. In Kenya and Uganda, where the NGO currently operates, the extremely poor tend to live in homes with thatched roofs whereas the less poor tend to live in homes with metal roofs. Villages with an abundance of low-quality housing and access to a mobile money agent are targeted.

After this selection stage, using a rigorous process of audits and security measures to prevent errors and fraud, money that has been transferred to GiveDirectly's local mobile money account is transferred directly to the recipient's account.[1] A text message is sent to the recipient who then goes to a mobile money agent—usually a shopkeeper in the village or nearby town. The recipient transfers the new funds in his or her account to the agent's account in return for cash. The

[1]SIM cards are given to those recipients that do not have mobile phones.

recipient is then free to use the cash for any purpose of his or her choosing.

In the current operations, the village selection phase involves multiple rounds of site visits with their attendant costs and logistical difficulties because governmental or other poverty data does not exist at village-level granularity. In this work, we formulate, implement, and deploy data-driven approaches to streamline this part of a campaign, thereby both increasing the efficiency of direct cash transfers and improving targeting of transfers to the extremely poor.

In particular, the main contribution of this work is a remote sensing system that estimates the quality and density of housing in large areas of central east Africa at a fine scale to facilitate village selection. Specifically, we apply a combination of image processing and machine learning methods on publicly-available electro-optical satellite imagery to construct features for a regression that estimates the number of houses of different roof types in an image patch. We use template matching to find roofs in images, and color histograms and supervised classification to differentiate roof types. We take advantage of a special characteristic of roofs in this part of the world: thresholding template matching results over a sequence of values reveals different roof types like a sieve. Therefore, all features are calculated over a sequence of thresholds.

We develop an interactive image labeling platform and deploy it to crowdsource the acquisition of training data on roof locations and types for the supervised classification and regression algorithms. We find that crowdsourcing results in rapid and high-quality completion of the labeling task, which we attribute to the intrinsic motivation of crowd members to contribute to social good. The final learned regression function allows us to construct heat maps of housing density and thatched-roof proportion from individual image patch-level estimates; areas with a preponderance of thatched-roof houses are to be targeted. From the heat maps, we aggregate village-level proportions of thatched roofs to produced a ranked list of villages to target.

Thus far, the data-driven prioritization of villages has been deployed in three districts in Kenya.[2] In total, the villages selected through the proposed approach have received or are in the process of receiving over US$ 4 million in direct cash transfers. Using the data mining approach saved approximately 100 person-hours of manual work that would have been incurred in manual ways of conducting village selection. The system offers a way to quickly scale GiveDirectly's operations and generate further cost savings in the future.

There is no specific prior work in the remote sensing literature on estimating the proportion of metal and thatched roofs in a village. However, two typical remote sensing problems are related: land use classification and building extraction. The most common feature for land use classification from single electro-optical images is color histograms as we also use [11]. Building extraction methods in the literature, including those based on the Hough transform [9], are primarily focused on urban and suburban environments having closely-spaced buildings with high variability in size and shape, and also sometimes attempt to reconstruct the three-dimensional geometry of the buildings. For our work, such approaches are excessve because we are concerned with the easier rural setting with roofs all of approximately the same size and of a similar approximately rotationally-invariant shape; therefore, we can use simpler template matching approaches with good performance.

[2]We do not report names of districts and villages in this paper to protect recipients and the process from harm.

(a) (b)

Figure 1: Homes in central east Africa with (a) metal and (b) thatched roofs.

The remainder of the paper is organized as follows. In Section 2, we further discuss the penury criterion of thatched and metal roofs. Section 3 discusses data available from satellites and previous campaigns, and the crowdsourcing we instituted to obtain further data. We present a feature construction method from satellite imagery based on template matching and use the constructed features for estimating the density of homes and the proportion that have thatched roofs in Section 4. Section 5 describes the deployment of the estimation system in GiveDirectly's campaigns for village targeting. Finally, Section 6 presents the results and impact of the deployment to GiveDirectly and Section 7 provides a summary and suggestions for future research.

2. THATCHED AND METAL ROOFS

In many villages of central east Africa, the roofs of homes are typically constructed in one of two ways: they are either made of grass, i.e., are thatched, or are made of metal like tin or iron. Photographs of the two roof types are shown in Fig. 1. Metal roofs provide many benefits over thatched roofs. For example, thatched roofs leak and collapse regularly, requiring replacement one or two times a year at a cost of US$ 100 or US$ 150 per year in materials and labor, whereas metal roofs last ten to fifteen years. Gutters to collect rain water are sometimes installed with metal roofs which lessens the incidence of waterborne diseases and reduces the labor involved with fetching water from a potentially large distance. Thatched roofs are also a comfortable habitat for mosquitoes, leading to higher incidence of mosquito-borne diseases among people living under thatched roofs [5].

Households that can afford to invest in a metal roof—which can cost as much as US$ 564 purchasing power parity in western Kenya [7]—typically do. Therefore the type of roof, thatched or metal, is a reliable proxy of the poverty level of a given area. In order to maximize operating efficiency and target the poorest, the NGO seeks to operate in administrative units with a high proportion of thatched roofs and low proportion of metal roofs.

In this work, we would like to quantify the proportion of thatched and metal roofs in administrative units as efficiently as possible. Toward this end, we turn to remote sensing because it is possible to differentiate roof type in satellite imagery. Leaving aside all other possible proxies for poverty, the remainder of the paper is concerned with estimating the

Figure 2: Example of metal roof in center of satellite image.

Figure 3: Example of thatched roof in center of satellite image.

proportion of thatched roofs in administrative units and using these estimates for village selection operations.

3. AVAILABLE DATA AND CROWDSOURCING DEPLOYMENT

In this section, we describe various data sources available to us including their limitations and also describe a crowdsourcing platform we deployed to obtain additional data.

3.1 Satellite Imagery

Various satellites take images of the earth with differing modalities (optical, hyperspectral, radar, etc.), spatial resolutions, temporal resolutions, and costs of imagery. In our application, we require the cost of obtaining imagery to be minimal, preferably zero, and the spatial resolution to be sufficient so that homes occupy more than a handful of pixels. Temporal resolution is a secondary concern for us and the electro-optical modality is sufficient. Under these desiderata, we choose to pull 400×400 pixel images at the highest zoom level available in this part of the world from the Google Maps API, which is made available for free. Metal and thatched roofs are distinguishable in these images, as seen in Fig. 2 and Fig. 3.

Satellite images from Google Maps are not without issue. In the part of Africa in which we are working, some small regions are obscured by clouds. An example is shown in Fig. 4(a). Since only one temporal snapshot is available per spatial location, we have no way of removing the clouds. Additionally, we have no indication on the provenance of the images, specifically the time stamp of acquisition. In examining the images for east central Africa, it is clear that images have been taken in different seasons, which results in different visual characteristics of roofs and the other components of the scene. As shown in Fig. 4(b)–(d), some images are from a wet season, some are from a dry season, and some are from a season in which there is an abundance of haze. These differences have a direct consequence on image processing and machine learning, as we detail in Section 4. Moreover, without a time stamp, we cannot say how out-of-date our estimates may be in various small regions.

3.2 Data from Previous Censuses and Campaigns

During the process of selecting individual households to receive cash transfers, GiveDirectly sends staff to villages to conduct censuses of the households there. Collected data includes geographic coordinates of individual homes obtained using GPS receivers and an observation on roof type. Censuses include all households in a village, irrespective of roofing material, which may have provided excellent training data for a supervised learning algorithm had it not been for the following reason. Homes are often located at close proximity to each other; the GPS coordinates did not have enough precision and accuracy to allow us to draw a correspondence between homes seen in satellite images and homes recorded in censuses. Therefore, this data was not usable as a labeled training set. This same reason prevented us from using similar data from previous campaigns along with the fact that only data on actual recipients (living under thatched roofs) was collected in campaigns. Although not amenable to supervised learning, the census coordinates proved valuable in illuminating the geographic location of typical populated villages as we discuss in Section 3.3, where we describe a crowdsourcing application we developed to obtain labeled training data.

There are several levels of administrative units in Kenya, going from coarse to fine as province, district, division, location, sublocation, and village. Prioritizing at the level of the finest administrative unit, the village level in Kenya, requires us to have the geographic coordinates of the boundaries of

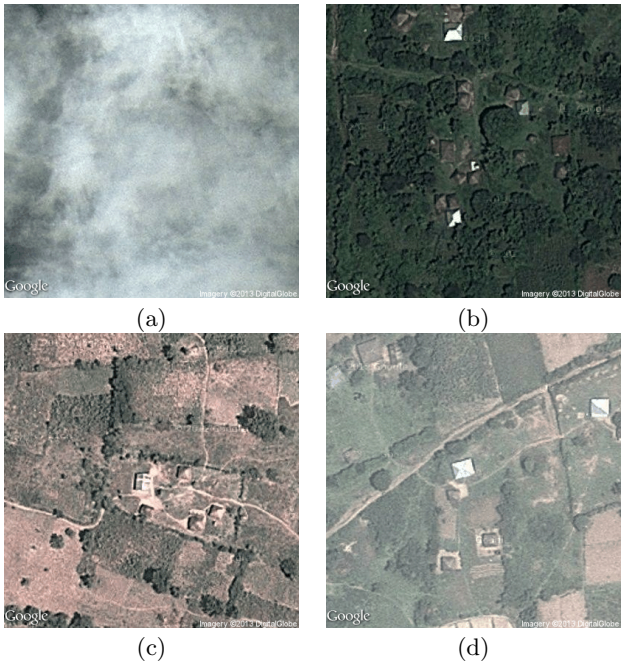

Figure 4: Satellite images taken under different conditions: (a) cloudy, (b) wet, (c), dry, and (d) hazy.

these localities. As part of previous campaigns, GiveDirectly obtained ESRI shapefiles of province, division, district, and sublocation boundaries in Kenya, but not village boundaries. The NGO also has PDF files of maps showing the village boundaries, but not in any georeferenced format, as seen in Fig. 5.

We had two options: work at the sublocation level only or try to georeference the PDF files. Since villages are the level at which the NGO desired to work, we set about converting the PDFs into shapefiles. The first step was to convert the PDF maps into a format we could open in a GIS application. Using Inkscape, an open source vector graphics editing program [1], we opened the PDF files and deleted all unnecessary features (rivers, roads, labels, etc.). We then exported the stripped-down maps as DXF files. We then opened these files in QGIS—an open source mapping application—and converted the DXF files to GEOJSON. Luckily, the PDF maps had been rendered in WGS 84, a commonly-used projection system; therefore to georeference the village-level map for each division, we wrote a simple Python script that applied a linear transformation on the coordinate system of the village-level maps to fit the bounds of the division-level shapefile.

At this point, we had a fairly good geographic representation of villages. However, since villages are fairly small, any discrepancies between their actual location and their location in our shapefiles would yield innacurate results. To address this problem, we opened our transformed shapefiles in CartoDB, a web application for creating and editing maps [12], and manually aligned the shapefiles to prominent features like roads, rivers, and district boundaries. Finally, so as to ensure the interpretability of our maps, we hand-labeled each village polygon by referencing our shapefiles against the original PDF images. While this process was somewhat inexact, NGO staff members in Kenya were satisfied with the results and confident that our village polygons faithfully represented reality.

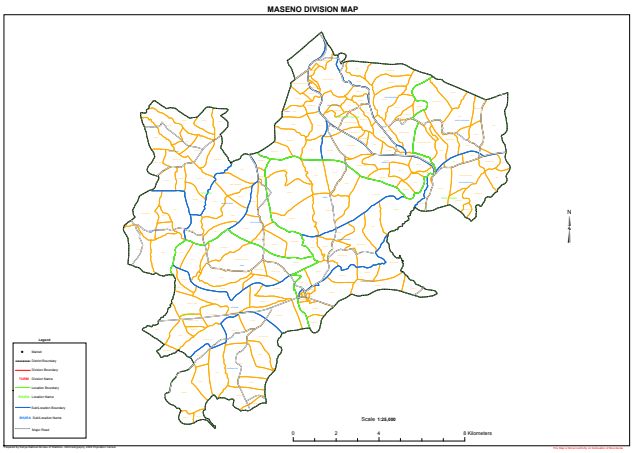

Figure 5: Example village map.

3.3 Crowdsourcing Application and Deployment

To create a collection of training data for our machine learning algorithms, we develop an interactive image labeling platform in Python using Flask—a web development framework [4]. As shown in Fig. 6, the application serves a random satellite image to a user and allows him or her to click on roofs in the image and indicate whether they are thatched or metal. When the user is done with an image, a new, unlabeled image is served and he or she replicates the process. On the backend, we record the total number of roofs of each type in each image, as well as the pixel coordinates and type of each roof in each image. The platform is deployed to the web using Heroku so that people can label images in a distributed fashion.

Working with our partners at DataKind, we recruited volunteers to help with the labeling process. DataKind initially announced the need for volunteers via meetup.com.[3] Within an hour, more than 20 people signed up. Ultimately, 10 people were selected for the crowdsourcing process. Volunteers were trained to use the platform by telephone and via a screencast we made and posted to YouTube.[4] Once volunteers were comfortable with the platform, we opened it up. Within one weekend, 10 users labeled all 1468 images in our training set.

While developing the platform, one of the authors of this paper labeled a random subset of 70 images. As a rough indication of the crowd's labeling quality, we find the correlation coefficient between the number of thatched roofs identified by the author and the crowd to be 0.783. Similarly, the correlation coefficient for iron roofs is 0.698. Both quantities are quite good; manual verification of other images also suggested that we could be highly confident in the results.

We conjecture that the excellent response time and quality are consistent with findings by Mason and Watts [10]: "When it is possible to use non-financial rewards, such as

[3] `http://www.meetup.com/DataKind-NYC/events/159235222/`

[4] `https://www.youtube.com/watch?v=0mg-37JhAL4`

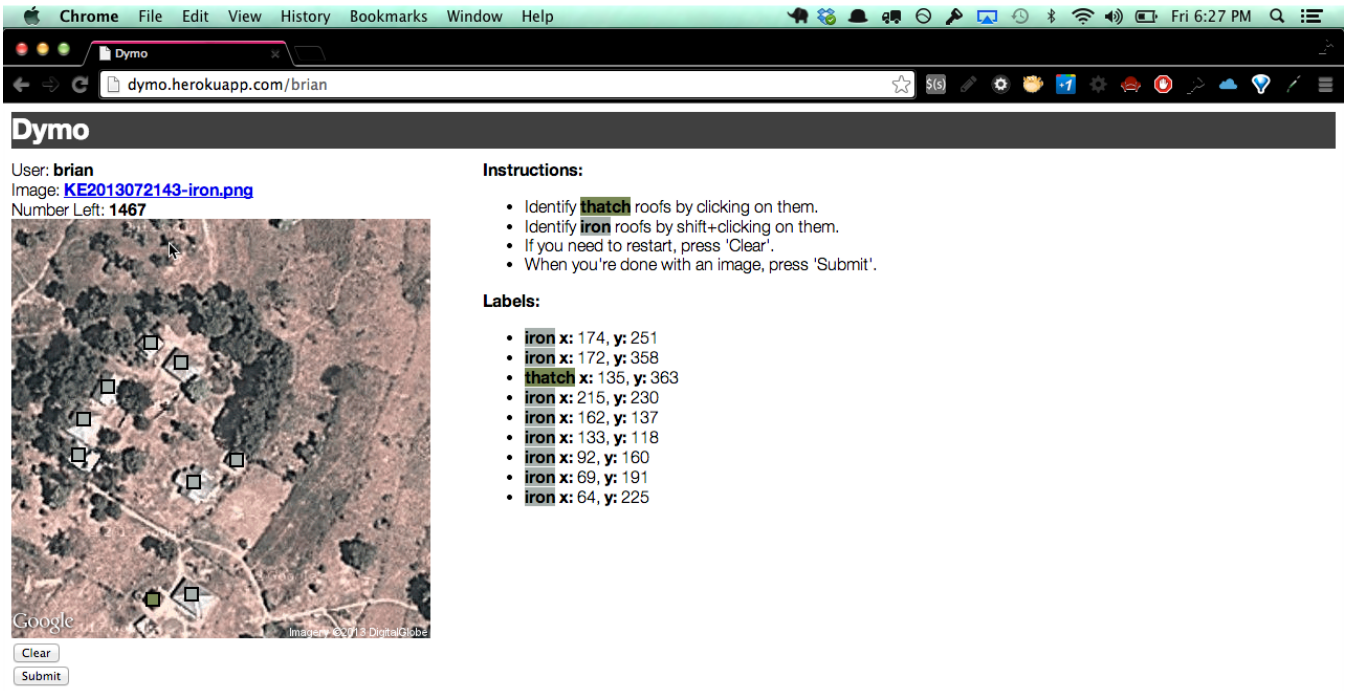

Figure 6: Screen shot of application deployed for crowdsourced labeling of roofs in satellite images.

harnessing intrinsic motivation, the quality of the work will be as good or better than using financial rewards, and therefore work can be accomplished as effectively for little to no cost." Helping to reduce poverty in the world, like other projects for social good, provides much intrinsic motivation.

4. REGRESSION OF ROOF COUNT AND ROOF TYPE PROPORTION

As we have discussed, the ultimate use of the data analytics we are performing is in selecting administrative units with the smallest proportion of metal roofs. To achieve this objective, we pose two regression problems: estimating the total number of roofs in a 400×400 pixel satellite image and estimating the proportion of metal roofs in the same image. With both estimates, we can then obtain an aggregate proportion of metal roofs for any region by taking a weighted average of estimates from a tiling of image patches from that region. Let us note that accurately classifying individual roofs is not the ultimate objective because we are tackling the village selection problem.

From the crowdsourcing, we have a training set of 1468 satellite images labeled by the number of thatched roofs and the number of metal roofs, from which we can derive the total number of roofs and the fraction that are metal. These are the response variables to predict. We use random forests with 50 trees per ensemble to perform the supervised regression because of their typical effectiveness, speed in prediction, and ease of use [2]. We note that the fraction of metal roofs response variable takes values in the range $[0, 1]$, but since all training samples fall in this range and we use random forests, the constraint does not need to be explicit in the learning. The features for the regressions are constructed from satellite image patches as we discuss next. We provide tenfold cross-validation accuracies for different feature choices at the end of the section.

4.1 Feature Construction

In constructing features that relate to the number of roofs and proportion of metal roofs in an image, we rely on two concepts: color histograms and template matching. Since metal roofs, thatched roofs, and the background all have different color distributions, we can use counts of pixels at different color values as features. In particular, we quantize each of the three color channels (red, green and blue) into eight equal-width bins to obtain twenty-four color histogram features per image patch. Color histogram features have long been used in image retrieval applications [6].

We also have a fairly good idea of the size and appearance of central east African roofs, which is fairly stationary across the region, so we can apply a basic template matching approach to find individual roofs in image patches [3]. We use the sum of absolute differences matching score on grayscale versions of image patches with two different templates: a 10×10 pixel white square and a 12×12 pixel template which is a central portion of an actual thatched roof from one of our satellite images. Specifically, denoting the image patch as I and the template as T, the matching score at pixel j of the image patch is:

$$d(I_j, T) = \sum_{i=1}^{n} |I_{i,j} - T_i| , \quad (1)$$

where n is the total number of pixels in the template (100 for the white square template and 144 for the thatched roof template). An example of applying the white square template to a satellite image is shown in Fig. 7. The input image is shown in Fig. 7(a) and the resulting matching score is shown in Fig. 7(b).

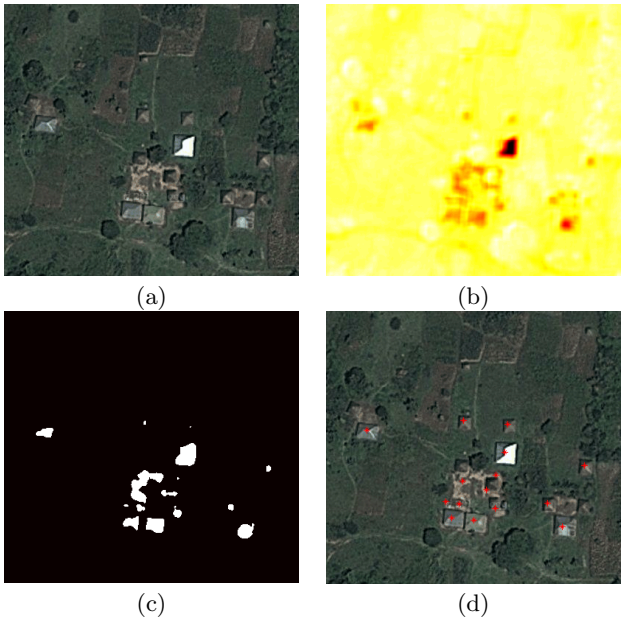

Figure 7: Template matching steps, from (a) the input image, to the (b) matching score, to the (c) thresholded matching score, to the (d) centroids of connected components in the thresholded matching score.

Once we have the template matching score for the image patch, we apply a threshold to obtain a binary image, as shown in Fig. 7(c) in which we can identify roofs. Based on morphological processing, we find all of the connected components in the binary image. The number of connected compounds is equivalent to the number of roofs identified. The centroids of the connected components that are found are superimposed on the input image in Fig. 7(d).

Of course, the number of roofs identified is a function of the threshold value applied to the template matching score, as illustrated in Fig. 8. A unique characteristic of central east African roofs that is not usually apparent in general image processing or remote sensing applications is that changing the threshold sequentially results in a sieve or filtration of roof type. When the white square template is used, as the threshold is increased, the metal roofs are revealed first and the thatched roofs are revealed later. Finally, non-roof items are revealed. The opposite is true with the thatched roof template. Therefore, the sequence of roof counts as a function of threshold is a sieve-like feature that not only provides information on the total number of roofs, but also on the proportion of metal and thatched roofs. A compound of dwellings may be identified as a single roof or as several roofs depending on the threshold, but according to our conception of the problem and our instructions to the crowd, compounds should be counted as multiple roofs.

The final set of features we construct combines template matching with color histograms within a roof classification scheme. Based on the crowdsourced labeling, we not only have the counts of thatched and metal roofs in the 1468 images, but we also have the pixel locations of all of the roofs with their labels. Among the crowdsourced image set, there are 8536 roofs: 5999 thatched and 2537 metal. Taking image patches surrounding the crowdsourced clicks, we can construct color histogram features and use those to classify the roofs. Fig. 9 shows the average color histogram by roof type in wet, dry, and hazy images. From the figure, it is apparent that on average, these color histogram features are able to distinguish thatched from metal. The baseline classification accuracy is $5999/8536 = 70.28\%$. With a random forests classifier with 50 trees (again chosen due to typical effectiveness, speed, and ease of use), we are able to achieve a tenfold cross-validation test accuracy of 89.67%.

To actually construct the final set of features, we use the output of the template matching at each threshold, take a 15×15 patch centered at each roof that is found, use the random forest classifier on color histogram features calculated on those patches, and compute the classification score output by the random forests (average vote among all constituent decision trees) [2]. The average classification score among all roofs in the image at a given template matching threshold is the feature for that image at that threshold, which is a good indicator of the proportion of metal and thatched roofs in the image.

Thus overall, we have three sets of features derived from the color pixel values in a 400×400 image patch. First is the color histogram features for the entire image patch. Second is a vector of roof counts at different template matching thresholds for two different templates. Third is the average individual roof classification score at the same thresholds for the two templates. The feature sets are summarized in Table 1.

4.2 Cross-Validation Accuracy

As mentioned at the beginning of the section, we learn regression functions for two different response variables using random forests: total roofs in an image and proportion of roofs that are metal in an image using the three sets of features described above. We examine two options for dealing with the different conditions during satellite image acquisition: learning a single estimator for all conditions, and learning separate estimators for the different conditions.

The figure of merit we use is mean absolute error (MAE) because of its easier interpretability than root mean squared error. However, we should note that MAE at the satellite image level is not the ultimate metric of interest. Since we ultimately average together estimates from all images in a village (which should reduce the error assuming that the image-level estimates are close to unbiased), and then rank the villages, the ultimate metrics of interest are MAE at the village level and especially rank correlation at the village level. However, we unfortunately have no ground truth for these quantities to report accuracy against.

Table 2 presents tenfold cross-validation testing error for different sets of features used and the two approaches for dealing with differing satellite conditions. As a point of comparison, if we use the mean value of the training response variables as the estimates for all test samples, the baseline MAEs are 2.8392 for total roofs and 0.2121 for the proportion of roofs that are metal. In examining the table, we see that separate regressions for the different conditions are better than a single regression across the board. Among the different options for feature sets, the combination of all features yields the best performance in predicting the proportion of metal roofs, with mean absolute error 0.1621. The average roof classification score feature set does not help in predicting the total number of roofs, which is expected

Figure 8: Template matching results at different thresholds.

Table 1: Summary of Feature Sets

Feature Set	Number of Features	Number of Templates	Number of Thresholds	Total
full image color histogram	24	—	—	24
template matching roof count	1	2	14	28
average classifications	1	2	14	28

because by averaging together the classification scores, we lose all information about the total number of roofs. The best performance is achieved by the combination of color histograms features and template matching roof count features with mean absolute error 1.9511. As we see in the next section, these error values are sufficiently small for our application and deployment; we use these feature sets going forward.

5. DEPLOYMENT

Having learned regression functions as described in Section 4, we now describe how we deploy the image processing and machine learning algorithm as part of the NGO's operations workflow. Based on previous experience and national-level data, GiveDirectly already identified which districts and divisions it is considering operating in. The question to answer is which villages within those divisions to target.

For any new 400×400 pixel image that was not seen by the predictive model during training, we can calculate its features and estimate the total number of roofs and the proportion metal using the learned model. The different satellite conditions in images persist over fairly large contiguous regions, so it is a quick and straightforward exercise to define bounding boxes corresponding to wet, dry, and hazy conditions, and use the appropriate regression function.

One may ask why we do not crowdsource the entire labeling effort instead of using the crowd only to provide training data for a machine learning algorithm. Our initial deployment scored approximately thirty times as many images as in the training set and future deployments may have hundreds of times as many images. Therefore relying on the crowd for the entire labeling task is not scalable, especially because we require consistency in the labels to allow comparisons between villages.

We download satellite images from the Google Maps API that tile the entire region of interest (and could do so for all of Kenya and Uganda if desired). We use equally-spaced latitudes and longitudes to do so since the API is called using latitude and longitude inputs. Since our region of interest straddles the equator, equally-spaced longitudes are an extremely close approximation to equally-spaced ground locations. We estimate the two quantities of interest for each downloaded image. We also perform a post-processing equalization step to match estimated values across satellite condition boundaries. The estimates for all image patches are visualized as heat maps, as shown in Fig. 10 and Fig. 11.

The northwest part of the region of interest contains clouds in the satellite images, which is reflected in both estimates. The estimated total number of roofs in these parts is very close to zero and the metal proportion is very high. A central large town in the middle of the region of interest shows up with a high roof count and high metal proportion. Lakes and reservoirs in the southwest appear as areas of low roof count and low metal proportion. A few thin horizontal and vertical lines in the heatmaps are artifacts of images containing a blend of satellite conditions.

The next step is to aggregate these fine-scale estimations within village boundaries. We take the average metal proportion weighted by total number of roofs of all image patches whose centers fall within the village boundary polygon. The resulting village-level metal roof proportions are mapped in Fig. 12. The cloud image patches do not contribute to the village-level metal proportion estimate because their total roof estimate is almost zero; the clouds do have the effect of occluding anything beneath them, but if the occluded portion of a village is similar to the unoccluded portion, then the effect is not too severe. The thin horizontal and vertical line artifacts have negligible effect on village-level estimates.

The village-level estimates are also presented as a ranked list, as in Table 3, which shows the top twelve (obfuscated) villages for targeting. Village C14C has the highest proportion of thatched roofs in the region of interest and is a prime village to target. Although proportion of thatched roofs is the main criterion in village targeting, there are some logistical considerations as well, such as access to M-Pesa agents and reachability from roads. We may incorporate these other criteria into an overall village score in future iterations of the algorithm for future campaigns, but in the initial campaigns for which the data-driven approach was applied, the rankings as in Table 3 were used as a priority list for manual selection that takes the other considerations into account.

In fact, many of the villages at the top of the list had already been enrolled in previous campaigns, providing a validation to the analytics. Additionally, several villages that constitute the main township area of this region were estimated to have high metal proportion, which is consistent

Table 2: Tenfold Cross-Validation Test Mean Absolute Error

Feature Set	Total Roofs (1)	Prop. Metal (1)	Total Roofs (Sep.)	Prop. Metal (Sep.)
full image color histogram (A)	2.1020	0.1755	2.0423	0.1744
template matching roof count (B)	2.0084	0.1755	1.9915	0.1755
average classifications (C)	2.4696	0.1704	2.4090	0.1670
(A) + (B)	1.9743	0.1712	1.9511	0.1720
(B) + (C)	2.0494	0.1658	2.0203	0.1630
(A) + (B) + (C)	2.0028	0.1656	1.9731	0.1621

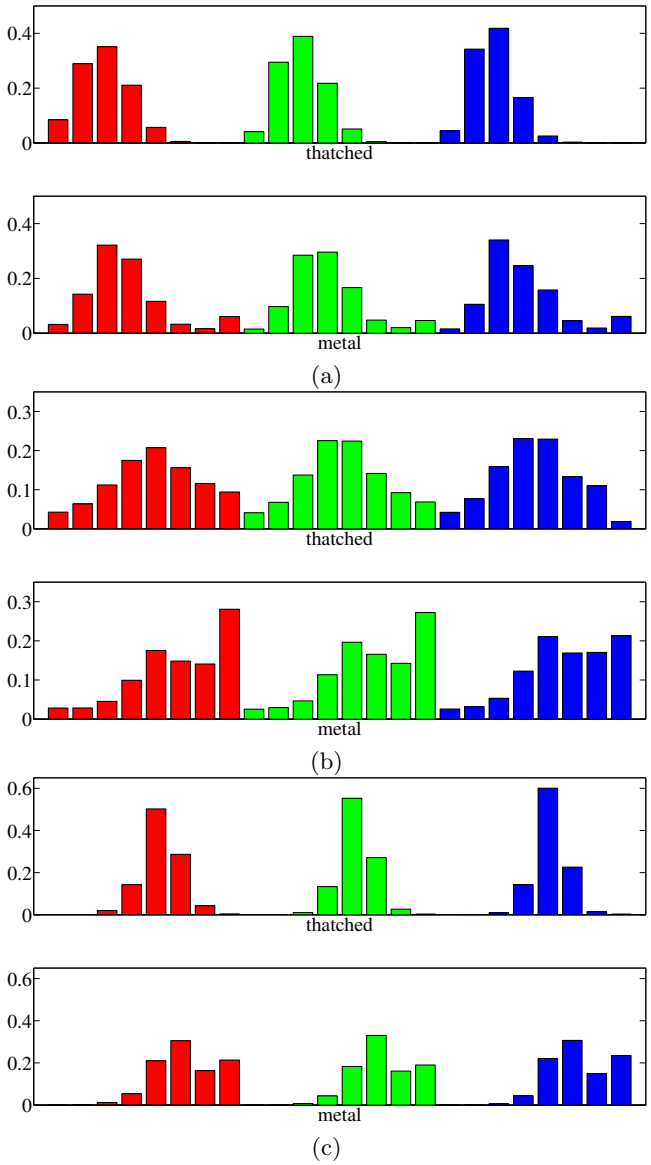

Figure 9: Average color histogram by roof type in 15×15 pixel patches of training set in (a) wet, (b) dry, and (c) hazy images.

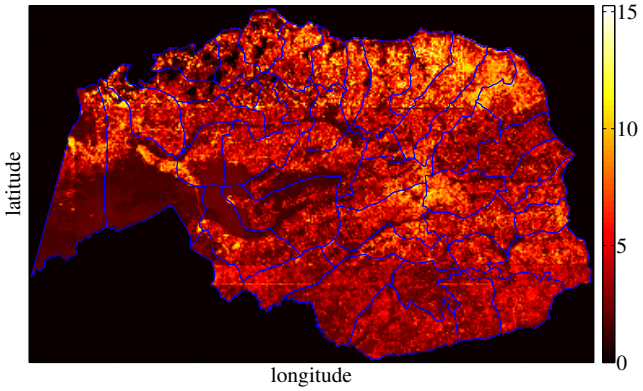

Figure 10: Heat map of number of total estimated roofs per 400×400 pixel image in the region of interest.

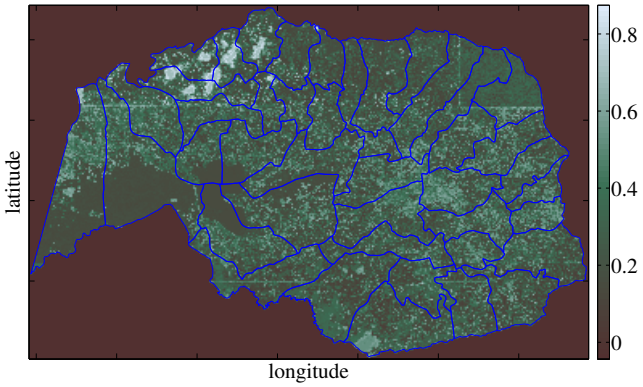

Figure 11: Heat map of proportion of roofs that are metal in the region of interest.

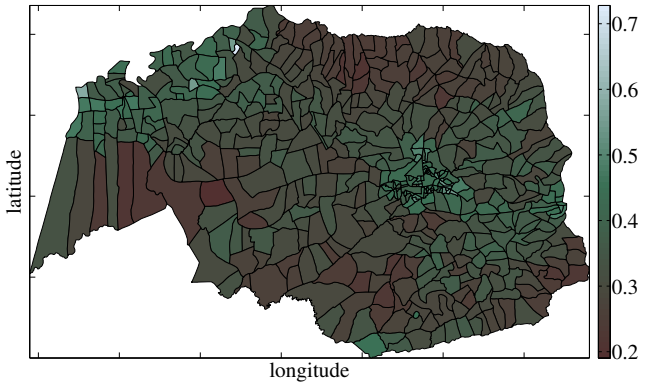

Figure 12: Estimated proportion of metal roofs in villages in the region of interest.

Table 3: Village Rankings (Names Obfuscated)

Ranking	Village Name	Proportion Metal
1	C14C	0.19039
2	C107	0.21016
3	C12E	0.21664
4	V10F	0.22197
5	V13D	0.22446
6	L180	0.22547
7	L111	0.22733
8	L106	0.22876
9	L193	0.22877
10	C14D	0.22916
11	L106	0.22934
12	C170	0.22986

with local knowledge. One small circularly-shaped village in the south central part of the region of interest is specifically delineated as a separate village by government authorities precisely bacause it is a small enclave with lower poverty; the estimation results bear this out: the predicted metal proportion is much lower than the surroundings.

6. IMPACT

In early 2014, GiveDirectly piloted the algorithm to identify approximately 50 villages in western Kenya for its largest campaign to date, which will move US$ 4 million in cash transfers. The approach immediately resulted in two major improvements: reduced staff time and reduced risk of inconsistent results.

Under the previous model, one staff member could collect requisite data on five villages per day. The 507 villages that the algorithm assessed would have required 101 person days to complete manually. This represents a cost savings of approximately US$ 4,000 in wages, which is equal to the value of cash transfers for four households (or twenty individuals) and improves overall operating efficiency by approximately ten basis points (defined as the fraction of total funds delivered directly to the poor). Reducing the person-hours required also allows implementation to move more quickly by completing the assessment in a matter of minutes rather than days, which modestly increases leverage of fixed costs like infrastructure and management time. These efficiencies will accumulate as the algorithm is applied on an ongoing basis.

The previous model also posed a risk of inconsistent results because it relied on multiple individuals, each liable to make different human errors. Although the historical error rate is unknown, it is reasonable to conclude that the algorithm improves outcomes by introducing a consistent approach to the entire task.

7. CONCLUSION

In this work, we have applied data mining for social good by developing a remote sensing approach for targeting unconditional cash transfers to the extremely poor. The algorithms and other system components we have developed have been deployed as decision support systems for planning GiveDirectly's operations. The impact of this work includes tangible benefits for the NGO including operational efficiency in cost and time, and reduction of inconsistency risk. This project also demonstrates how enthusiastic crowds can be when the goal is social good.

The machine learning and image processing components used in this work are fairly standard, but are composed in a novel way. Especially novel is the construction of sieve-like features from template matching, which was possible because of properties of thatched and metal roofs, but may find application in other settings as well. The steps taken to develop the system, its architecture, and even the results, can be applied more generally to uplift humanity beyond just targeting for unconditional cash transfers because fine-scale estimates of poverty and other social indicators in rural villages have myriad uses.

There are several avenues for future research. First, several improvements to the algorithm are possible. In this work, each image patch estimation is treated independently; Markov random field models connecting neighboring image patches, including thin-plate and thin-membrane priors often used in remote sensing, could be used to improve the estimation results [15]. Moreover, the feature engineering can be further enhanced using techniques beyond the very basic template matching and color histograms used herein. Alternative supervised learning beyond random forests could also yield improved performance.

The first deployment has been to prioritize villages in support of a human decision maker. Additional future work includes collecting data on all other considerations in village targeting and using them to fully automate the process. In this line of work, operational costs could be accounted for in the learning algorithms, cf. [14, 13], and schedules and routes for campaigns could also be constructed in a data-driven way.

One area of improvement for future campaigns relates to the fact that not all roofs seen in satellite imagery correspond to separate households. As discussed earlier, in this work, we count each structure within a compound separately. However, in the part of the world in which the NGO currently operates, it is common for kitchens to be separate structures and for there to be sleeping houses that dependent sons live in once they reach puberty, both of which should not be counted separately from the main house.

8. ACKNOWLEDGMENTS

The authors thank Jake Porway, Brian Dalessandro, Dave Goodsmith, Shubha Bala, Peter Darche, Julia Marden, Craig Barowsky, the rest of the DataKind family, Carolina Toth, Piali Mukhopadhyay, and the rest of the GiveDirectly staff for their assistance.

9. REFERENCES

[1] T. Bah. *Inkscape: Guide to a Vector Drawing Program*. Pearson, Boston, MA, 2010.

[2] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.

[3] R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley, Chichester, United Kingdom, 2009.

[4] R. DuPlain. *Instant Flask Web Development*. Packt Publishing, Birmingham, United Kingdom, 2013.

[5] A. C. Gamage-Mendis, R. Carter, C. Mendis, A. P. De Zoysa, P. R. Herath, and K. N. Mendis. Clustering of malaria infections within an endemic population: Risk of malaria associated with the type of housing construction. *Am. J. Trop. Med. Hyg.*, 45(1):77–85, July 1991.

[6] T. Gevers and A. W. M. Smeulders. PicToSeek: Combining color and shape invariant features for image retrieval. *IEEE Trans. Image Process.*, 9(1):102–119, Jan. 2000.

[7] J. Haushofer and J. Shapiro. Household response to income changes: Evidence from an unconditional cash transfer program in Kenya. Working paper, Abdul Latif Jameel Poverty Action Lab, Cambridge, MA, Nov. 2013.

[8] N. Hughes and S. Lonie. M-PESA: Mobile money for the "unbanked". *Innovations*, 2(1–2):63–81, Winter/Spring 2007.

[9] Z. J. Liu, J. Wang, and W. P. Liu. Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic Hough transform. In *Proc. IEEE Geosci. Remote Sens. Symp.*, pages 2250–2253, Seoul, Korea, July 2005.

[10] W. Mason and D. J. Watts. Financial incentives and the "performance of crowds". In *Proc. ACM SIGKDD Workshop Human Comp.*, pages 77–85, Paris, France, June 2009.

[11] J. A. dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falcão. Multiscale classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, 50(10):3764–3775, Oct. 2012.

[12] J. de la Torre. Organising geo-temporal data with CartoDB, an open source database on the cloud. In *Proc. Biodiversity Informatics Horizons*, Rome, Italy, Sept. 2013.

[13] T. Tulabandhula and C. Rudin. Machine learning with operational costs. *J. Mach. Learn. Res.*, 14:1989–2028, July 2013.

[14] T. Tulabandhula, C. Rudin, and P. Jaillet. The machine learning and traveling repairman problem. In *Proc. Int. Conf. Alg. Decision Theory*, pages 262–276, Piscataway, NJ, Oct. 2011.

[15] A. S. Willsky. Multiresolution Markov models for signal and image processing. *Proc. IEEE*, 90(8):1396–1458, Aug. 2002.