Satellite Image Analytics, Land Change and Food Security

Sunandan Chakraborty[1,2], Zohaib Jabbar[2], Lakshminarayanan Subramanian[1,2], Yaw Nyarko[2]
[1]Department of Computer Science
[2]Center for Technology and Economic Development, NYU Abu Dhabi
New York University, New York, USA
{sunandan, lakshmi}@cs.nyu.edu, {zohaib.jabbar, yn1}@nyu.edu

ABSTRACT

Changing patterns and reduction in agricultural land are among the fundamental problems that impacts food security in developing regions like India. Rapid economic growth coupled with increasing populations and changes in climatic patterns are among the main factors impacting availability of agricultural land. On a macroscopic scale, due to the lack of good quality data, governments do not have a complete and clear cut picture of changes in land usage patterns. In this paper, we present the design of a satellite image analytics engine that we use to perform a detailed analysis of changes in agricultural land patterns over a 13-year time period (2000-2012) in West Bengal, India, traditionally considered one of the most fertile areas in the world. Our satellite analytics engine can perform a fine-grained analysis of macro-granular satellite images (elevation of 11 km) and classify small portions of land in each image into different categories: agricultural, developed, forest and water bodies. Our analytics engine can analyze temporal changes in land patterns and compute the percentage of change in land under each category. Based on detailed food production data gathered in collaboration with the bureau of statistics of West Bengal, we analyze the correlations between changes in agricultural land patterns and corresponding changes in food production (normalized by change in yield patterns). Our tool can be used at varying levels of spatial granularities ranging from macroscopic analysis at a state level to fine-grained analysis at sub-district levels. This analytics tool is targeted for government and non-governmental policy makers to analyze land pattern changes and correlate them with food security metrics.

1. INTRODUCTION

Agricultural land availability is undergoing dramatic changes across the globe. This phenomenon is more rampant in the developing world where rapid economic growth and increasing population is resulting in unplanned development. The loss of arable area is estimated to be 1-21% in South America and around 18% in Africa [36]. The major causes identified for this decline are: (1) rapid urbanization of these countries including industrialization; (2) the migration of farmers to cities resulting in the sale of farmland for non-agricultural development, a trend that has escalated in the past few years due to rising real estate prices. Loss of arable land has a direct impact on food security. Most developing regions are also predominantly agrarian economies and changes in arable land can significantly impact food production and availability. Apart from urbanization and industrialization, changes in climatic patterns and other environmental factors are also resulting in degradation of farmlands and eventual disappearance. There are reports that Sahara desert is expanding southwards at an alarming rate[9]. Rising sea levels are increasing salinity of soil and decreasing productivity of land [18]. In countries like India, unpredictable monsoon is also harming production and land quality.

Loss of arable land is also a well documented phenomena in developed regions around the world, especially in North America and Europe. Unlike developing regions, land usage is well-documented in most developed regions at fine-grained granularities. For example, from 1982 to 2007, more than 23 million acres of agricultural land was converted to developed land in the USA[4][5] , with each state losing significant areas of farmland. Similarly, Germany lost five million acres of its Utilized Agricultural Area (SAU) between 1960 and 2010, a decline of almost 11 percent. France has about 50 percent of its land used for agricultural activities in 2010 and it is also declining [1]. Most developed countries have traditionally maintained detailed electronic records to monitor change in land patterns over 5-10 decades. In contrast, such fine-grained data is often found lacking in developing regions.

In this paper, we propose the design of an automated satellite image analytics tool that can leverage publicly available satellite image data sources to provide a fine-grained longitudinal analysis of changes in land pattern in a given region. Our goal is to design a data analytics system that can understand the longitudinal relationship between the changes in agricultural land pattern in a given small geographic area and its corresponding impact on food production. This paper is specifically contextualized for the region of West Bengal, traditionally considered one of the most fertile areas in the world being in the delta of the Gangetic plains. We used a corpus of satellite images gathered from Google Earth, which maintains an updated repository of satellite images along with an archive of older images across the globe. Based on detailed food production data gathered in collaboration with the bureau of statistics of West Bengal, we computed the correlations between the changes in agricultural land patterns and in food production in these years, at the district level of West Bengal.

The key building block of our analytics tool is a satellite image analysis engine that can analyze potentially noisy satellite images and provide fine-grained classification of regions within each image into different categories such as: arable land, water body, developed land, forest etc. Given historical data about the same lo-

Figure 1: Image of a location in Angola taken in 2003

Figure 2: Image of a location in Angola taken in 2011

cation, the image analysis engine can provide a detailed analysis of land pattern changes. Figure 1 is an example of a land in Angola in 2003 and a developed version of the same land in 2011 is shown in Figure 2. Our engine can detect such changes at different location granularities (small region, district, state level etc.). In the case of West Bengal, we obtained data over a 13 year time period from 2000-2012 and could track land evolution over this entire time period. We correlate this land change pattern with food production data over the same time period gathered by the Bureau of Statistics in the government. This tool can be helpful to policymakers to monitor the changes in the land pattern and take appropriate steps if any drastic changes are noticed.

Based on our detailed analysis of satellite data and food production data from 2000-2012, we present the following key results in this paper:

- In the state of West Bengal agricultural land area has declined by 2%.
- The decline in agricultural land area has demonstrated positive correlations with rice/wheat production and agriculture productivity index.
- The loss in arable land was higher in districts closer to urban and industrial zones.

In this paper, we present the satellite analytics tool and the data generated by it to analyze food production and security issues. However, this analytics tool can be used in other contexts as well. This tool is a general purpose satellite image analytics tool whose end-goal is to process satellite images from any part of the world and generate data about distribution of land patterns and how the pattern has changed over the years in that region. The generated data can be used in any relevant analysis.

2. RELATED WORK

There are two different classes of related works for the work presented in this paper. There are many works which have focused on detecting objects, patterns from images. These techniques have been or can be applied to satellite images. On the other hand there are various other works which have used satellite images and information extracted from them to infer different metrics and build novel applications.

We start with novel applications using satellite images and then discuss some techniques used in such applications. There are quite a few applications targeted at disease detection using satellite images. Ceccato et al [11] used remote sensing technology to monitor malaria risk. They used high resolution images of Landsat data to monitor features on earth' surface. They monitored features, such as, surface water, vegetation, temperature, humidity which has a direct impact on outbreaks of disease such as malaria. Safi et al [8] has a similar approach where they predicting malaria risk using remote sensing data. They used neural networks and general linear model to combine precipitation, temperature and vegetation index obtained from remote sensed data to build a predictive model for malaria cases. They ran this model for 23 provinces in Afghanistan and showed good prediction accuracy. Another example of applying image analysis in public health is the work by Kelly et al [21]. They used an object based image classification approach to identify objects within the satellite images that can help in understanding various public health issues. Another similar work can be seen in the paper describing an early warning system for malaria [32]. Famine detection and prediction is another common application that have been targeted using satellite imagery. In a slightly earlier work, Hutchinson [20] developed an early warning system for famine in sub-Saharan Africa. This work demonstrated better efficiency and accuracy of using satellite data in building such system compared to datasets from ground sources. In a recent work a similar approach and data was used to predict famine Uganda [28]. In this work, two main parameters used for famine detection – vegetation index and rainfall – were derived from multi-spectral satellite images to characterize crop yield. Related to famine, there are numerous works that have focused on food security [19][17][22]. Huber et al [19] has argued that early warning systems for food security can be vastly improved with the help of satellite data. Their claim is that frequent temporal data on water stress and surface temperature can help in detecting drought early, thus, reducing their impact on the society.

Deforestation and other environment related issues can also be well understood from satellite data. Vibrans et al [33] created a land database of a state in Brazil to determine the remaining land area covered by Atlantic rain forest. A specific deforestation case was covered by Rahman et al [30], where they looked the disappearance of mangrove forests, making way for aquaculture. They used MODES data from NASA's Terra and Aqua satellites to study the Mahakam delta mangrove forest in Indonesia between the years 2000 and 2010. Another work focusing on environmental issue using satellite data is by Diouf et al [13]. There goal was to detect the concentration of Saharan dusts and sea surface chlorophyll from satellite data.

There are other works which have analyzed satellite data to infer various metrics. Example of such an application is using night-time satellite images to infer poverty in a region [26]. Abelson et al [6] used satellite image data to identify extremely poor regions using roof top of buildings as a proxy of poverty. This work has been to identify poor regions for better cash flow using mobile platforms. They have shown that the intensity of artificial lights during night time, as captured in night-time satellite images, has strong correlation with GDP of that region. Dahmani et al [12] used satellite images to build cartographic database of slums to facilitate better planning and policies.

On the other hand, there have been many works that has developed novel techniques for processing satellite images or digital images in general that can be applied to satellite images as well. Object detection [25][34] has been a very popular method used to extract information from satellite images [6]. A typical feature of satellite images is they are heterogeneous and more continuous in nature. Within a small area there can be different objects placed one after the other. Hence, segmentation becomes an important preprocessing step before any kind of information extraction task. There are works which have specifically targeted satellite images for segmentation [10] [29]. There are numerous other works focusing on image segmentation [15] [14] or superpixel extraction [16] [24] [31][7][35] that can be applied to satellite image analysis as well .

Based on past works it can be understood that information extracted from satellite images can be very useful to infer many useful parameters, particularly in public health, environmental issues and public policy. Such a technique is even more essential when ground source data – particularly for developing or underdeveloped regions – are error-prone, irregular and difficult to obtain at regular intervals.

3. LAND PATTERN ANALYSIS AND FOOD SECURITY

In this section, we provide a brief context and motivation for the problem addressed in this paper. We begin by stating the importance of the problem before providing specifics of the data analysis study presented in the paper.

Context: Food security is emerging as one of the biggest problems that human populations may face in the upcoming century due to growing populations, urbanization and reduction in arable land. Reduction in arable land is a relatively hard task to reverse and food statistics around the world clearly indicate that agricultural land patterns are on the decline, and in recent times the rate has accelerated.

The problem of decrease in agricultural land is particularly an important question for developing countries which are predominantly agrarian economies. Given unprecedented population growth, coupled with only a relatively modest growth in agricultural yields, any minor changes in land can have catastrophic consequences on the food security for the population in such countries. While western countries have relied on large scale imports, developing countries do not have the economic horsepower to rely on imports for tackling food deficiencies.

A critical metric that needs to be monitored for food security in agrarian economies is agricultural land availability. While developed nations are known to have much more detailed records on land change patterns, the relative documentation of such data in developing regions is often lacking due to the lack of fine grained data. In many developing regions, data is not frequently updated or not complete. Policy decision-making is often done based on stale or incomplete data and an important phenomena such as disappearance of arable land often go unnoticed.

Even if policy makers are aware of the decline in arable land, the second problem is the lack of knowledge of the underlying causes behind this disappearance. On many occasions, land acquisition for development is happening illegally [3]. Local authorities are unaware of this change and at what rate it is happening. Apart from this, there might be reasons for which land is losing productivity [27] due to bad agricultural practices or environmental factors. The exact transformation of agricultural land to what other types can provide valuable clues to the solution of the problem. For all these factors, land pattern statistics at regular intervals are essential.

3.1 Satellite Images

Advantage of using satellite images is that they can give a clear picture of the state of the land. Over the years, the quality of these images have greatly improved providing rich information about the surface area of different regions. Processing these images can reveal the present status of a region. Moreover, having access to historical images, short and long term changes in a region can be easily tracked. How historical satellite images can detect changes is evident from Figure 1 and 2. This images are taken from Angola in the year 2003 and 2011 respectively. In the 2003 it shows a green patch of land, probably used for agriculture, with marks of being converted into developed land. Within 8 years the entire green land has turned into an urban area. Thus, the sattelite images can have an important impact in tracking changes in the land pattern.

In addition, satellite images can provide an account of of what type of changes happened in a region. A piece of agricultural land disappearing might not always be a very useful observation to tackle the problem. On the other hand, identifying the exact change can provide more insights. An existing agricultural land can change to some other land type – acquired for urban or industrial development. Alternatively, due to changing climatic conditions or bad agricultural practices, the land's quality has deteriorated and hence, it is turning into a barren land. Being able to track these changes, a policy maker can address the situation properly. In the first case, the solution is to deal with illegal and forcible acquisitions and in the second case come up with sustainable solution to protect existing farmlands. In conclusion, these steps can protect agricultural lands, moreover, ensure food security and stabilize food prices for long term sustainability.

3.2 West Bengal: A Case Study

In this paper, we contextualize our study of the agricultural land availability problem for the state of West Bengal in India. West Bengal lies in the north Indian Gangetic Plains, as a result fertile alluvial soil is abundant in the state. Hence, agriculture is the predominant driving force of the economy of the state. Table 1 shows some facts of the state. The reason we chose West Bengal is due its relatively smaller area and its dominance in agriculture. Based on the data presented in Table 1 it is clear that that the economy of the state is heavily dependent upon agriculture. Agriculture contributes to about 24% [2] of the state's domestic product and with such a high percentage of the labor force engaged in agriculture, it is apparent any decline in agriculture will have a drastic effect on the state, its economy. This will impact the food security of the state as well. One of the goal of this paper is to use the tool to estimate land pattern statistics of West Bengal and observe what changes have occurred in the last 10 years.

In this paper, we try to analyze the agricultural land availability problem by building a satellite image analysis tool that can classify satellite images into various categories. In addition, it can monitor the changes over time and report *what is changing* and *how it is changing*. Subsequently, we use the data produced by the tool to compare changing land patterns with officially collected government data on food production at a district granularity.

4. SATELLITE IMAGE DATA

We created our dataset from the satellite image repository of Google Earth (GE). GE has freely available satellite images from across the world, including an archive of historical images. To de-

Table 1: West Bengal at a glance

Area	88,752 sq km
Population (2011)	91,347,736
Population density (2011)	1000/sq. km
Rural population (2011)	72%
Area under agriculture (2012)	5,666,000 h.a.
Labor force engaged in agriculture	67%

velop and test our image analytics tool, we gathered satellite images from different regions across India and different countries across Africa. We collected around 8800 images from 7 different countries in Africa and 5100 images from 2 different states in India. All these images were captured at an elevation of 11 KM, to make sure they are not too blurry (if captured from too close to the ground) or missing details (if captured from a very high elevation). The time range of these images is between 2000 and 2012. A typical image is shown in Figure 1.

One of the key challenges in processing the satellite image is the variability in the quality of images. Google Earth images do not have consistent quality. Particularly, for older images the quality is quite poor and blurry. Second, the precision of images presented by Google Earth can historically vary. Third, in many images the land area is not visible in some parts due to completely blanked out images or due to cloud cover. Finally, the background color across images is not consistent which makes it difficult to classify land into different categories. For this study, we ignored such noisy images and built the model assuming the images are clear. For future work, we intend to build a more robust model to tackle such noisy images.

5. SATELLITE IMAGE ANALYSIS TOOL

The image analysis tool is aimed at classifying satellite images into 4 different categories. To design the classifier we used Convolutional Neural Network (CNN) [23]. The network takes an image as an input and has 4 output nodes for each class to describe the image. The categories are - ***Arable, Tree-covered, Water body, Developed.***

We experimented with a variety of standard methods and feature sets to build the tool. In order to evaluate each method, we split the data into a training and a cross validation set. We set aside 10% of the images for cross validation from each individual category. So, our training set is the combined set of 90% of the images from 7 different African countries and 2 Indian states.

5.1 Preprocessing

A single frame in a satellite image covers a large portion of land and they typically have more than one category of land. Fig 3 show a typical satellite image from our dataset and this image has agricultural land, tree-covered areas, as well as water bodies. Thus, every image needed to be broken into segments or superpixels, where each superpixel represents a homogeneous area, potentially belonging to only one of the categories mentioned above. There are many related works that involve segmenting images and extracting superpixels [16] [24] [7][35], we chose Markov Random Field based approach [31] to extract the superpixels. In this method the conditional distribution of a pixel is determined by the 8 neighboring pixels. A graphical model is constructed where every node is a pixel and edges are drawn between adjacent pixels. The edge potentials define how similar they are or whether they belong to the same superpixel or not. Based on some annotated superpixels, the model learned the edge potentials and based on the learned model, the superpixels were extracted from the larger image. Figure 3 shows the resulting superpixels after segmentation. Here, each superpixel represents a uniform region within the image. We treat these superpixels as an atomic unit of the images and are fed into the network for classification.

5.2 Image Classification

Our dataset essentially consists of segmented superpixels extracted from satellite images. After the segmentation process and extracting the superpixels, it is assumed that the superpixels have one and only one type of land type. Thus, each superpixel can be classified into one category (i.e. arable, tree-covered, water body or developed). We experimented with different feature sets and classification methods to classify these superpixels into one of these categories. The purpose of these experiments were to identify the best performing method to achieve our ultimate goal of building a satellite image classification tool and get an accurate estimate of land patterns over a geographical region. We employed two classification methods – support vector machine based method using a suite of feature sets and a convolutional neural network based model with implicit feature extraction module. In the rest of the section, we provide a brief overview of the methods used.

Support Vector Machines (SVM) are designed for binary classification. However, in our case, we are classifying the images into 4 classes by combining several binary classifiers using one vs rest method. We used several image features to experiment with identify the best performing model. Our features include standard image features, such as, color histograms, texture etc. We started with greyscale features, as many of the satellite images have non-standard color codes. That is, in many images, agricultural land and water bodies are depicted with similar colors. To test this intuition, we used RGB color histograms as well. We also considered the texture of an image, as many images across categories had similar color distributions, we used a couple of texture based features, such Grey Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) features sets. All these different types of features were extracted from the superpixels and fed into an SVM classifier to classify each superpixel into a category. The detailed results from these experiments are discussed in Section 5.4.

The second type of model used to build this tool was Convolutional Neural Networks (CNN). CNNs are multi-layered artificial neural networks which incorporate both unsupervised feature extraction and classification. A CNN consists of a series of convolutional and pooling layers that perform feature extraction followed by one or more fully connected layers for the classification. The inputs of a unit in a convolutional layer come from just a small rectangular subset of units of the previous layer. In addition, the nodes of a convolutional layer are grouped in feature maps sharing the same weights. The inputs of each feature map are tiled in such a way that correspond to overlapping regions of the previous layer making the aforementioned procedure equivalent to convolution while the shared weights within each map correspond to the kernels . The output of convolution passes through an activation function that produces nonlinearities in an element-wise fashion. A pooling layer follows which subsamples the previous layer by aggregating small rectangular subsets of values. Max or mean pooling is applied replacing the input values with the maximum or the mean value, respectively. A number of fully connected layers follow with the last one having a number of units equal to the number of classes. This part of the network performs the supervised classification and takes as input the values of the last pooling layer which constitute the feature set. For training the CNN a gradient descent method is applied using back propagation.

However, the input superpixels (Section 5.1) are non-uniform in terms of dimension and neural networks are usually designed for fixed size input. Thus, non-overlapping patches of size 32 are extracted from the inside of each superpixel image. In order to increase the amount of training data and prevent over- fitting we artificially augment the training patch dataset by using label-preserving transformations such as flip and rotation as well as the combinations of the two. In total, 16 transformations are used. Then, we calculate the mean over the training image patches and subtract it from all the patches of the dataset so the CNN takes as input mean centered RGB pixel values.

Using the created super-pixel dataset we train a deep CNN with a six layer architecture. The network has four convolutional layers with 5×5 kernels; the first three layers have 32 kernels while the last has 64, producing equal number of feature maps. Each convolutional layer is followed by a pooling layer with 3×3 pooling regions and stride equal to two; the first one outputs the maximum value out of each pooling region while the following three use the average. The last two layers of the network are fully connected with 128 and 7 units, respectively. The output of each hidden neuron was set to zero with a probability p forcing the network to learn more robust features for the description of the input regardless of the inactive neurons. Here, the dropout probability p is set to 0.5. The softmax function is used so as to normalize the outputs of the last layer so each output is between zero and one and they all sum up to one. This way, the output values represent a categorical probability distribution so a cross-entropy loss function is used to calculate the error used by gradient descent training. Finally, as far as the weight learning is concerned, a schema with a decay of the learning rate along with a momentum coefficient was used. The base learning rate is set to 0.001 with an exponential decay policy and the momentum is set to 0.9. The CNN model learns and optimizes the filters in each layer through the back propagation mechanism. These learned filters extract important features that uniquely represent the input image of a homogeneous segment of a satellite image.

However, given the characteristics of the dataset, images belonging to same category can have vastly different feature values. For example, the color of an arable land can have different shades of green as well as in many cases shades of brown. Using greyscale versions of the image cannot capture this variation. Hence, we experimented with different feature sets and methods to empirically find the optimum solution to this problem.

5.3 Estimation of Land Pattern

Finally, the trained model can be applied to the satellite images collected from a region across time to get an estimate of land patterns and changing trends in the pattern in that region. To estimate the land pattern of a given region, we collect satellite images covering the entire region based on th elatitude-longitude coordinates bounding the region. We repeat this image collection scheme for all the years for which we would like to have the estimate. For a given year, each superpixel α_i is fed into the trained model and classify it to obtain the category k, which best describes the land type α_i belongs to. This information of individual superpixels can be aggregated to obtain the land pattern statistics of a region. In other words, the land pattern statistics can be computed for a $year$ as,

$$\theta_k^{year} = \frac{\sum_i \alpha_{ik}^{year}}{\sum \alpha_{[1:K]}^{year}}$$

where, α_{ik}^{year} is the total number of superpixels in the image segments i which are assigned to the class k. Monitoring θ_k^{year} for different years can give an estimation of how land area under k has changed over the year.

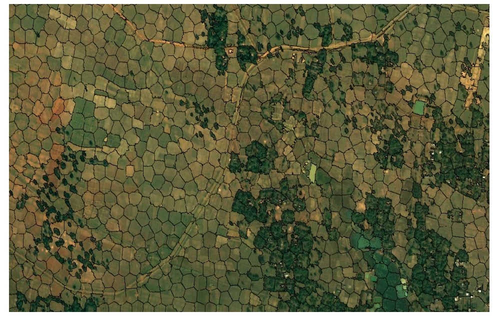

Figure 3: Segmented Image

5.4 Performance

Our dataset contains around 10,000 images across 7 years. Extraction of superpixels resulted in an average 500 segments per image. For the development of tool, we randomly selected 8000 superpixels across all years to train the model and another 2000 as a cross validation set. These images and superpixels were manually annotated to create a labeled training and cross-validation set to develop the tool. We experimented with different kinds of feature-set and computed the accuracy on this cross-validation set. The accuracy is computed as the percentage of the total number of cases where a superpixel was assigned to the correct cluster compared to the entire cross validation set.

We compared the CNN based method to several other popular feature sets used in image classification. Our experiments showed that the CNN based method had the best performance in the cross-validation set and we chose to use it for the land analysis study. Below we present a brief description of all these experiments.

Our first approach was based on using greyscale color histograms as the features. We ran SVM multi-class classification algorithm on these features. This method gave very low accuracy of 38.28%. This is due to the fact that the satellite images had very similar colors for many different objects. Moreover, in greyscale the objects looked alike and color histograms could not discriminate against different types of objects (e.g. tree cover, open fields, buildings). Next we tried a similar approach but with RGB color information separately instead of converting them to greyscale. This increased the performance to an accuracy of 51.04%.

Different experiments based on color histograms could only provide a maximum accuracy of 51.04%. A closer look at the color distribution of the images reveal that there are not much difference between the histograms across different categories. On the other hand, texture of the images are very different for different land types. Irrespective of colors, farmlands tend to have a smoother texture compared to other categories. We added Grey Level Co-occurrence Matrix (GLCM) based features to include the texture of the images. This method increased the accuracy to 70.93%. Based on this observation, we experimented with other texture based feature Local Binary Pattern (LBP) and the color variant of this feature set, RGB-LBP. Although, with pure LBP the accuracy dropped to 68.35% but the RGB variant showed improvement with an accuracy of 76.93%. In all these experiments we used SVM based multi-class classifier to classify the images. Finally, the CNN based method described in the previous section showed the best performance with an accuracy of 89.41%. Table 2 summarizes the per-

Table 2: Accuracy of the tool for different features used

Feature-set	Accuracy
Greyscale color histogram	38.28%
RGB color histogram	51.04%
Texture (GLCM)	70.93%
Local Binary Pattern (LBP)	68.35%
LBP + RGB	76.93%
CNN	**89.41%**

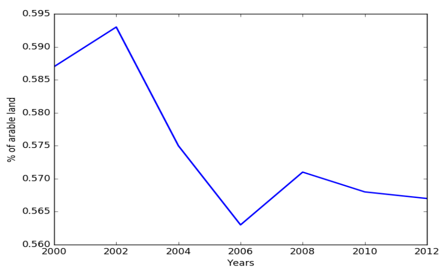

Figure 4: Percentage of arable land per year

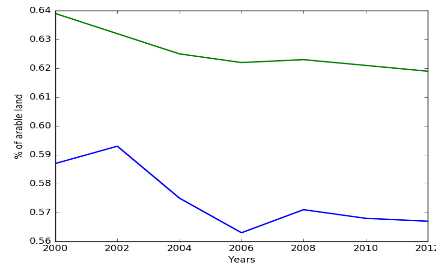

Figure 5: Comparison between official percentage cropped area (upper) and estimation from the satellite images analysis tool (lower)

Table 3: Comparison between arable land and food production

Index	Correlation coefficient	p-value
Net rice production	0.58	0.012*
Net wheat production	0.61	0.041*
Agricultural area index	0.68	0.023*
Agricultural production index	0.72	0 .030*

formance of the satellite image analysis tool based on different features. We performed the rest of the analysis with the CNN based image analysis tool, given the best performance of the model.

6. LAND PATTERN ANALYSIS OF WEST BENGAL

After the development of the tool, we applied it to the satellite images collected from the Indian state of West Bengal. Our goal was to analyze the variation of agricultural land in that state across several years. We collected satellite images covering entire area of the state for 7 years between 2000 and 2012. The tool was applied on this data to estimate the percentage of arable land in all these years and observe the change in percentage across years. For a year, the total number segments under arable land category was aggregated to estimate the percentage of arable land for that year.

For the case study, we collected the images from West Bengal from the GE repository for 7 different years, starting from 2000 till 2012 at an interval of 2 years. For each year we had 7,505 images covering the entire land area of the southern part of the state. Each image captured were rectangular in shape with a fixed height and width. All the images were captured at an elevation of 11 KM,

The estimated percentage of arable land as computed by our tool is shown if Figure 4. We see that there has been a decline in agricultural land in the state between these year. Although, the percentage slightly rose between 2000 and 2002, according to the estimate there has been a drop of 2.0% between 2000 and 2012.

We wanted to compare our findings with the food production statistics published by the *Bureau of Applied Economics and Statistics* affiliated to the *Government of West Bengal* [2]. The bureau publishes various statistics about the state in their annual Economic Review journal. The different food production related metrics we collected are - *Land area under rice, Land area under wheat, Net rice production, Net wheat production, Net cropped area, Agricultural area index, Agricultural production index.*
Our goal was to validate our findings as well understand the implications of reduction in arable land and food production from these data. To validate the estimate given by the tool we compared the result with the cropped area published in the official report. The comparison of the official and our computed values are shown in Figure 5. Although our findings do not exactly match with the official figures, we see that the trends in both the plots have similarity.

To understand the implication of decline in arable land, we compared our findings with the food production data published by the government. We computed the correlation coefficients to see how arable land area can affect food production. The results are summarized in Table 3. The results indicate that in a region reduction in arable land has a positive correlation with food production.

Apart from looking at the entire state as a whole, we have also analyzed different districts of the state separately. In India, districts are the second level administrative boundaries in each state. The state of West Bengal is divided into 19 districts, including the urban district of Kolkata, which is also the state's capital. In this paper we have focused on 12 districts in the southern part of the state, excluding the district of Kolkata. Figure 6 shows the map of West Bengal and district boundaries of the region considered in this study. The southern part of the state is part of the large Gangetic plain and conditions are well suited for agriculture.

Our satellite images were labeled with the latitude-longitude coordinates of the location from where they were extracted. Using a GIS database we identified the district from where the image was extracted. Then for each such cluster we computed the year-wise share of each land type and produced a district level data of land pattern and its changes. The percentage change in agricultural land for these districts is summarized in Table 4. The figure in the table shows the change between year 2012 and 2000, as produced by our tool.

We see that in some districts the disappearance of agricultural land is higher than others. Interestingly, the districts with higher rate of disappearance, such as, Howrah, Hooghly are close to the urban district of Kolkata. Due to increasing population of the city, more land from the neighboring districts are being taken up for urban development. Similar arguments can be applied to relatively high rate of disappearance of agricultural land in the Burdwan district. Burdwan is a very populous district with industrial towns such as Asansol and Durgapur. On the other hand, in some dis-

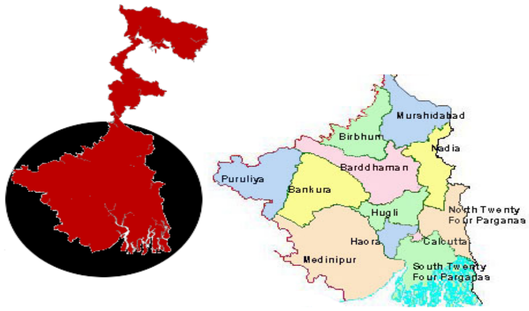

Figure 6: Map of West Bengal and district boundaries of southern part of the state

Table 4: District-wise change in agricultural land between 2000 and 2012

District	Change in agricultural land (%)
Bankura	6.10
Burdwan	-1.44
Birbhum	-2.36
Midnapore[1]	0.23
Howrah	-1.76
Hooghly	-5.30
24 Parganas (North)	-0.36
24 Parganas (South)	-2.11
Nadia	2.09
Murshidabad	-2.20
Purulia	1.48

[1] Midnapore district was split into 2 districts in 2002. Due to unavailability of separate data for earlier years we have considered them as one aggregating their data wherever needed

tricts, such as, Purulia and Bankura where we observe a increase in agricultural land. These districts are in the western part of the state and industrialization initiatives are limited in that region. Also, the population density is relatively lower.

Similar to the state-level data, we tried to find the relationship between arable land and food production in the districts as well. We did this comparison between the land share and rice production in these districts. We chose rice was because it is the most produced crop and the most consumed food in this region. The result of this comparison is shown in Figure 7. The figure shows a positive relationship between agricultural land (computed by the tool) and rice production in these districts. Both the values have been normalized to a score between 0 and 1. The scatter plot shows an approximate linear relationship between land area under agriculture and rice production. Hence, even in the districts we see that land area is changing over time and this change is showing positive relationship with food production.

7. DISCUSSION

In this paper, we presented a framework to estimate food production by identifying land pattern distribution in a region. We have demonstrated that the estimation of agricultural land in a region using our tool correlates strongly with food production indices in that region. The state of the art methods of measuring these metrics are

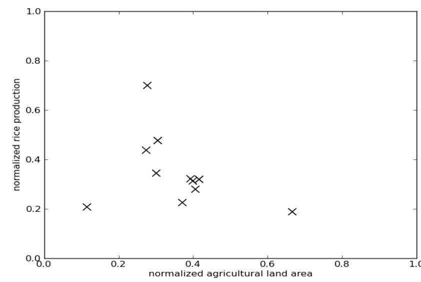

Figure 7: Relationship between agricultural land area and rice production for different districts (2012)

done manually by visiting the site, hence they are, cumbersome, time consuming and prone to errors. With our solution, reasonably accurate estimation of the same parameters can be done much faster. With this improvement, our tool can be an important tool for government employees and policy-makers, who can have faster access to such data, leading to better decision-making.

Based on the early findings from the study presented in the paper, we are in the process of collaboration with Government of West Bengal by deploying our tool at their units, offering our findings and in exchange including their expertise and additional data to improve the system's functionalities. We are also planning to deploy our tool in Ghana at the Ministry of Food and Agriculture (MoFA) field offices. MoFA field workers can greatly benefit from this tool by obtaining data about the area under their supervision. There is a shortage of field workers recruited by MoFA and thus, each worker has a huge area to supervise. Monitoring the land, which normally done manually by visiting the area in person, is slow and cumbersome. This tool can provide a positive impact on their operations by reducing time and increasing accuracy in their data. A new feature can further benefit them, if the tool can access updated satellite images of the region at a more frequent intervals.

The main component of the study is the satellite image analysis tool. The performance of the tool is greatly dependent upon the underlying data on which the model is trained and applied. In this study, we have used a freely available repository of satellite images from Google Earth. Although this data is free and easily accessible, the quality is poor. Re-training the model with high resolution satellite images can vastly improve the performance of the system and quality of the data produced. A major emphasis in the future direction of this work will be to acquire better data and train an improved model.

Apart from producing agriculture related information, the increased impact of the tool can be realized by applying it to other applications. This tool can be used – at its present form – to monitor water bodies. This can help in two different applications - firstly, detecting unlawful filling of ponds and lakes for development, and secondly, monitor the impact of rising water levels in coastal areas or disappearance of land due to river erosion. Generally, this tool can be a useful apparatus for policy makers and law-enforcement agencies to detect illegal constructions on protected or vulnerable land, provided that it has access to supporting data.

8. CONCLUSION

This paper presents and evaluates a hypothesis that changing land pattern can affect food production in a region. In order to evaluate the hypothesis, we built a tool that processes satellite images to estimate the distribution of land patterns in a region, computes the changing proportion in arable land and show that these changes correlate with food production indices in the region. The satellite image analysis tool introduced in the paper, processes Google Earth satellite images to classify land area in a region into 4 classes. The image classification engine of the tool demonstrated an accuracy of around 89%. Further improvement in the accuracy can greatly increase the tool's usability and reliability. So, one direction of the future work is to improve the system performance specifically for satellite images used and goals aimed in this work. Another approach in the future versions of this work would be to incorporate datasets from other sources, with less noise, better coverage and more frequent updates. In future, we would also like to focus on other regions, apart from West Bengal and build nationwide maps of land patterns. In addition, highlighting regions where there has been drastic changes in this pattern over the recent past.

9. ACKNOWLEDGEMENTS

We thank the NYU Abu Dhabi Research Institute and the Center for Technology and Economic Development (CTED) in NYU Abu Dhabi for supporting Sunandan Chakraborty and Lakshminarayanan Subramanian on this project.

10. REFERENCES

[1] `http://www.momagri.org/UK/agriculture-s-key-figures/Every-single-day-over-550-acres-of-\agricultural-land-the-equivalent-of-\\four-average-size-farms\-are-disappearing-in-France_1057.html`. [Online; accessed Feb-2016].

[2] Economic review and statistical appendix, department of statistics and programme implementation, government of west bengal, 2000-2012.

[3] Land lost, singur farmer said no to compensation, commits suicide. `http://archive.indianexpress.com/news/land-lost-singur-farmer-said-no-to-\\compensation-commits-suicide/31930/`, 2007. [Online; accessed Feb-2016].

[4] Yale environment 360: Most productive u.s. farmland disappearing at fastest rate. `http://e360.yale.edu/digest/most-productive-us-farmland-disappearing\-at-fastest-rate-report-says/2627/`, 2010. [Online; accessed Feb-2016].

[5] Incredible shrinking farmland. `http://grist.org/sustainable-farming/2011-11-07-incredible-shrinking-farmland/`, 2011. [Online; accessed Feb-2016].

[6] B. Abelson, K. R. Varshney, and J. Sun. Targeting direct cash transfers to the extremely poor. KDD '14, 2014.

[7] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11), Nov. 2012.

[8] F. Adimi, R. P. Soebiyanto, N. Safi, and R. Kiang. Towards malaria risk prediction in afghanistan using remote sensing. *Malaria Journal*, 9(125), May 2010.

[9] BBC. Desertification - expansion of the sahara desert. `http://www.bbc.co.uk/programmes/p00xbj4w`. [Online; accessed Feb-2016].

[10] A. Bretto, A. Ducournau, and S. Rital. A hypergraph reduction algorithm for joint segmentation and classification of satellite image content. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 6419. 2010.

[11] P. Ceccato, S. Connor, I. Jeanne, and M. Thomson. Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. *Parassitologia*, 47(1), '05.

[12] R. Dahmani, A. Fora, and A. Sbihi. Monitoring the proliferation of slums through gis and satellite image processing in the rural common of sidi taibi. *I. J. of Engineering Research and Development*, 10(8), Aug 2014.

[13] D. Dioufa, A. Nianga, J. Brajardb, M. Creponb, and S. Thiriab. Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method. *Remote Sensing of Environment*, 130, Mar 2013.

[14] A. Ducournau and A. Bretto. Random walks in directed hypergraphs and application to semi-supervised image segmentation. *Computer Vision and Image Understanding*, 120:91–102, Mar 2014.

[15] A. Ducournau, A. Bretto, S. Rital, and B. Laget. A reductive approach to hypergraph clustering: An application to image segmentation. *Pattern Recognition*, 45(7), July 2012.

[16] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *IEEE Intl. Conf. on Computer Vision*, 2009.

[17] J. Hielkema and F. L. Snijders. Operational use of environmental satellite remote sensing and satellite communications technology for global food security and locust control by fao: The artemis and diana systems. *Acta Astronautica*, 32(9), Sep 1994.

[18] M. Hossain. Global warming induced sea level rise on soil, land and crop production loss in bangladesh. In *World Congress of Soil Science, Soil Solutions for a Changing World*, 2010.

[19] S. Huber, R. Fensholt, K. Rasmussen, S. Proud, and A. Anyamba. Improving early warning systems for food security in africa with geostationary earth observation data. *IOP Conference Series: Earth and Environmental Science*, 6:–, 2009.

[20] C. F. Hutchinson. Uses of satellite data for famine early warning in sub- saharan africa. *Intl. J. of Remote Sensing*, 12(6), 1991.

[21] M. Kelly, S. Blanchard, E. Kersten, and K. Koy. Terrestrial remotely sensed imagery in support of public health: New avenues of research using object-based image analysis. *Remote Sensing*, 3(11), 2011.

[22] B. W. Khan MM, Mock NB. Composite indicators for famine early warning systems. *Disasters*, 16(3), Sep 1992.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 1998.

[24] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *CVPR, 2011*, 2011.

[25] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *Pattern Analysis and Machine Intelligence*, 23(4), Apr 2001.

[26] A. M. Noor, V. A. Alegana, P. W. Gething, A. J. Tatem, and R. W. Snow. Using remotely sensed night-time light as a proxy for poverty in africa. *Population Health Metrics*, 6(5), 2008.

[27] T. Philpott. New research: synthetic nitrogen destroys soil carbon, undermines soil health. `http://grist.org/article/2010-02-23-new-research-synthetic-nitrogen\\-destroys-soil-carbon-undermines/`, 2010. [Online; accessed Feb-2016].

[28] J. A. Quinn, W. Okori, and A. Gidudu. Increased-specificity famine prediction using satellite observation data. ACM DEV '10, 2010.

[29] K. R., S. R., and S. N. Satellite image resolution enhancement using contrast limited adaptive histogram equalization. *Digital Image Processing*, 6(3), 2014.

[30] A. F. Rahman, D. Dragoni, K. Didan, A. Barreto-Munoz, and J. A. Hutabarat. Detecting large scale conversion of mangroves to aquaculture with change point and mixed-pixel analyses of high-fidelity modis data. *Remote Sensing of Environment*, 130, 2013.

[31] A. Schick, M. Bauml, and R. Stiefelhagen. Improving foreground segmentations with probabilistic superpixel markov random fields. In *CVPRW, 2012*, 2012.

[32] C. S. Thomson MC. The development of malaria early warning systems for africa. *Trends Parasitol*, 17(9):438–445, Sep 2001.

[33] A. C. Vibrans, R. E. McRoberts, P. Moser, and A. L. Nicoletti. Using satellite image-based maps and ground inventory data to estimate the area of the remaining atlantic forest in the brazilian state of santa catarina. *Remote Sensing of Environment*, 130, 2013.

[34] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR 2001*, 2001.

[35] S. Wang, H. Lu, F. Yang, and M.-H. Yang. Superpixel tracking. In *ICCV, 2011*, Nov 2011.

[36] X. Zhang and X. Cai. Climate change impacts on global agricultural land availability. *Environmental Research Letters*, 6(1), 2011.