

PAPER • OPEN ACCESS

## Poverty incidence identification of cities and municipalities using convolutional Neural Network as applied to satellite imagery

To cite this article: J Mesina *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **482** 012044

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

# Poverty incidence identification of cities and municipalities using convolutional Neural Network as applied to satellite imagery

**J Mesina<sup>1</sup>, J Isanan<sup>1</sup> and C Maderazo<sup>1</sup>**

<sup>1</sup>Department of Computer and Information Sciences, University of San Carlos, Talamban, Cebu City, Cebu 6000, Philippines

joshuaisanan@gmail.com

**Abstract.** Poverty remains to be a hindrance to national growth for many developing countries. This presents a problem in a country's resource management and urban planning of its people. Furthermore, there is a big gap when it comes to data collection and monitoring in developing countries such that the government fails to prioritize this area of responsibility. The authors aim to bridge this gap by classifying wealth through the use of satellite images and emerging technology, particularly Convolutional Neural Networks. The researchers' goal is to test whether AlexNet, a Convolutional Neural Network Architecture, can identify the poverty levels of municipalities in the Philippines, by estimating their poverty incidences, based on satellite images. With this research, it can measure a broader set of society growth indicators. Indicators such as the material of people's houses, the size of their houses, their living conditions, access to roads and also how highly urbanized an area is, are considered important factors to indicate poverty. The results shows an accuracy of 0.84 and average precision-recall of 0.86 and 0.84, respectively. This study will give deeper insight to the government as to which areas to improve on in certain sectors given that a poverty incidence is high in a certain municipality.

## 1. Introduction

An essential building block of a state are its people. The problem is that some countries lack the awareness of the state of their people in order to respond and act accordingly. Although a developing country, it still does not fully cater to the entirety of the poor. One reason for this is that the government is not aware of the state they are living in. This research was intended to test AlexNet, a Convolutional Neural Network Architecture, using satellite imagery in order to reduce manual information collection on poverty incidences with the help of a Neural Network. Data preparation started with gathering satellite images of municipalities in Cebu and other provinces. Convolutional Neural Network (CNN) is utilized to train the gathered images that will extract the vital features used to identify and predict poverty incidence of specific municipality. The data consists of RGB satellite images obtained from Google Earth. The model developed by the researchers did not take into account rich areas that were abundant in mountains since they were lacking during the gathering of data. An example for this would be privately owned lots and subdivisions located in mountainous areas that have few but wealthy residents. The time at which the images were made were not taken into account and these areas may contain obstructions, blur or noise.



## 2. Review of Related Literature

### 2.1. Basis for Poverty

For a study on predicting poverty by [1], a basis they had used for economic estimation was the Global Multidimensional Poverty Index also known as MPI. Their source of features and data were not conventional as opposed to other studies wherein they used satellite imagery alone. Instead, the researchers used environmental data which was connected to food security, economic activity, accessibility to facilities and call data records. Their empirical results prove the above average accuracy when using data that differ greatly, yielded a Pearson Correlation of 0.91. They then attempted to predict poverty in context of health, education, and standards of living. Surprisingly their Pearson Correlation showed that they are good indicators for poverty prediction with Pearson Correlations through 0.84-0.86.

### 2.2. Poverty Prediction

A 2016 study conducted by [2] used artificial intelligence, particularly Convolutional Neural Networks and high-resolution satellite imagery to predict the economic well-being of 5 African countries. They found the idea to use satellite features instead of local data collected because collection of data manually is expensive and time-costly. Their study was able to predict 75 percent of the variation in local-level outcomes. This supports the current research greatly because of the high probability of its success and because according to [3], the Philippines suffers a higher poverty incidence than the general population which is estimated at 21.6 percent in the year 2015. [4] proposed a study to use Convolutional Neural Networks but had a different approach when it came to the features used in satellite imagery. They used both object and texture features extracted from satellite imagery. They took into account features such as the number and density of buildings using the existence of shadow area, which is used as a model for building height, the number of cars within those villages, the length and density of roads, paths and roof material to name a few. Another team of researchers [5], have also taken to using satellite imagery as basis for the prediction of poverty except they used publicly available, freely distributed satellite imagery. Their study differed from previous ones because they made use of multi-spectral satellite imagery.

## 3. Methodology

### 3.1. Data Gathering and Pre-processing

Data on poverty was gathered from the local office of the National Economic and Development Authority. The gathered data comprises the Poverty Incidences of the different municipalities of Cebu in the year 2012. Poverty is formally defined by [1] in Equation 1 where Q represents the number of families or individuals with per capita annual income or expenditure less than the per capita poverty threshold divided by total number of families or individuals. These incidences indicate the standard of living in a country. Satellite images of municipalities were retrieved from Google Earth for training. Municipalities were divided using the extracted coordinates of their boundaries using GADM database of Global Administrative Areas which provided a list of coordinates that form a polygon to map out each municipality for a given country.

$$P = (Q / n) * 100 \quad (1)$$

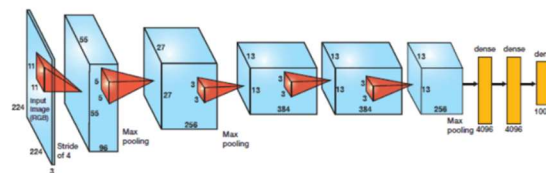
**3.1.1. Data Treatment.** The images were re sized to 227 by 227 pixels. The image data sets were cleaned by padding anything outside of the boundaries black.

**3.1.2. Data Augmentation.** After the images were cleaned, the images were mirrored or inverted which was a technique to gain more image data sets. 2 sets of data were obtained, and each image within the sets were rotated by 1 degree. They rotated each image until they reached 360 rotations to gain more

data. In total, the researchers gained 72,000 images, enough for a substantial study to be conducted. They then split this data into 70 percent for training and 30 percent for testing.

### 3.2. Network Architecture

Similar to previous studies [2], [6], the researchers used Convolutional Neural Networks to classify objects using the municipalities extracted. AlexNet was used for the network architecture. The reason as to why AlexNet was the chosen network architecture is because AlexNet reduces over-fitting by using a dropout layer that is applied to every neuron. Another reason as to why AlexNet was the chosen architecture was because of how AlexNet uses Rectified Linear Unit over Sigmoid which lessens the possibility of the occurrence of the vanishing gradient problem. A large amount of data may contain overfitting which Rectified Linear Unit can prevent. Rectified Linear Unit provides a way for simplifying activation functions which is currently popular for deep neural network [7].



**Figure 1.** AlexNet CNN Diagram

Caffe, a deep learning framework that included support for AlexNet, was used. The epoch count was modified to 440 in the CNN to fit the nature of the data to avoid over training. 100 images were gathered which was augmented to 72,000 images. The 72,000 images were divided into 50,396 images for training and 21,604 images for testing. Each image would be labeled after its respective class which in our case is 0 for 0-10 poverty incidence, 1 for 11-20, 2 for 21-30, 3 for 31-40 and 4 for 41-beyond. The reason as to why the researchers grouped together municipalities with poverty incidences beyond 40 is because of the lack of municipalities beyond that poverty incidence. Per iteration during the training, 50 images were propagated through the network. The loss function that was used by the researchers was Softmax with Loss. The reason why softmax is useful is because it converts the output of the last layer in the neural network into what is essentially a probability distribution. The trained model was used to classify each image in the test set of the data set. To measure the performance of the trained model, the researchers calculated the Accuracy, Recall, Precision and the F1-Score of the classifier. In the words of [8], Recall in statistics, is the ability of a model to find all the relevant cases within a data set. Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives. While recall expresses the ability to find all relevant instances in a data set, precision expresses the proportion of the data points our model says was relevant, actually were relevant [8]. The F1-Score takes into account both Precision and Recall. According to [9], F1 score is used if one needs to convey the balance between Precision and Recall. It is also important to use F1-Score if there is an uneven class distribution.

## 4. Results and Discussions

Initially, 50 images were extracted which were then augmented to 36,000. The initial dataset showed signs of under-fitting because of the lack of data. 50 more images of municipalities were extracted which totaled to a 100. This produced an augmented total of 72,000 images. The data was divided into 70 percent for training (50,396 images) and 30 percent for testing (21,604 images). The first idea was to make each unique poverty incidence a specific class and train the model but at an early stage of the training the model showed signs of underfitting because all of the classes only contained 720 or 1,440 images. This was not substantial enough data for features to be detected by the CNN. A new approach

was attempted in creating the classes, which was to use two classes only. The classifier only classifies an image as above or below a certain poverty incidence threshold. The mean poverty incidence of 24 was taken from the data set. The model was trained with binary classification as its specification. It was discovered that the model didn't learn because there were too many features in both classes to conclude whether a certain municipality in an image was either above or below the poverty incidence threshold. Finally, the researchers tried grouping together poverty incidences by range of 10 such that the classes were 0-10, 11-20, 21-30, 31-40, 41-beyond for poverty incidence classification and also adjusted the parameters by increasing the number of epochs for better results. This experimentation resulted in the loss value growing due to over-training but was continuously modified resulting in an epoch count of 440 with a loss value of 0.30. The trained model was used to classify the test set and was able to reach an accuracy of 84%, classifying 18,149 images correctly out of 21,604. In Table 3 there is a sample of experience results by the CNN. The statistical tests of the classifier were calculated with recall, precision and F1 Score and came up with these results which is shown on Table 1.

**Table 1.** Results of the Testing.

Classes	Precision	Recall	F1-Score	Support
Avg/total	0.86	0.84	0.84	21,604
0-10	0.91	0.95	0.93	6,265
11-20	0.85	0.73	0.78	3,457
21-30	0.63	0.92	0.75	3,241
31-40	0.88	0.85	0.86	5,184
41-beyond	0.98	0.67	0.79	3,457






#### 4.1. Class 0-10

The statistical test for this class performed remarkably well in terms of all the statistical tests. In fact, it has one of the highest scores in each test compared to all the other classes. The kinds of images in this class consisted mostly of densely populated areas such as the Metro Manila area. There were more buildings and structures present in the images and contained very little barren land.

#### 4.2. Class 11-20

Compared to 0-10 class, this class did not perform as high but its results are still considered above average. The pattern of the images that were identified in this class had buildings and structures but not as much as 0-10. The patches of green and empty land were still overwhelmed by the number of buildings.

**Table 2.** Samples of Images and Respective Actual Poverty Incidence.



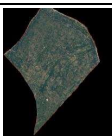


Image					
Poverty Incidence Label	0-10	11-20	21-30	31-40	41-beyond

#### 4.3. Class 21-30

The statistical tests for this class has substantial scores but something important to note is that, compared to the rest in this precision test, the classifier regarded many of the images as truly part of the 21-30 range but it also regarded many other images that are not in this class as part of this poverty incidence

range. This could also be described as a high amount of true positives but an even higher amount of false positives. Images that were labelled with that class would usually consist of mountainous regions combined with very few amounts of buildings and sometimes no buildings at all. The cases where there were almost no buildings at all in the image allowed the model to classify it as 31-40 or 41-beyond. In cases where there were a few buildings present in the image, the model performed well and classified it correctly to the 21-30 class.

**Table 3.** Results Sample.

Image					
Poverty Incidence Label	20.7	27.5	27.9	9.7	16.6
Predicted Class/Range	41-beyond	31-40	41-beyond	41-beyond <sup>a</sup>	41-beyond <sup>a</sup>

<sup>a</sup> These are incorrectly classified images because mountainous regions take up majority of the area

#### 4.4. Class 31-40

The model was able to classify precision and recall well for this class as it has picked up features such as municipalities that have really large mountainous regions while having small amounts of buildings visible within the satellite imagery. Compared to the Class 41-beyond which contained mostly no buildings present within the image.

#### 4.5. Class 41-beyond

Almost all the actual poverty incidences within this class were what the classifier predicted correctly because of its high precision but many of the actual poverty incidences within this class were deemed as false or one could say this class contained high false negatives. The 41-beyond class has the highest precision among the classes, but this class suffers the lowest recall. The reason why the recall of this class is very low is because of the presence of mountainous regions in the image that were classified as 41-beyond. This meant that the model did not yet find features that could possibly classify wealthy areas that satisfied this criteria. This caused the amount of False Negatives to increase because most of the mountainous municipalities would be classified as 41-beyond which affected the model's recall.

### 5. Conclusion

A major problem of society today is the manual and costly labor of data gathering. The researchers' goal was to test if the use of a certain CNN architecture namely AlexNet could estimate the poverty levels within a certain country accurately by breaking it down into manageable images. From the results, the most significant reason why the system could not obtain a higher estimation rate was because the researchers lacked the data set for wealthy areas with low poverty incidences that contained mountainous regions within them. Advice for further studies would be to increase the variability of the data set since it lacked the accuracy for it to be reliable for government use and most probably use another method for coming up with results since the output comes in the form of brackets. It would be best if there was a specific predicted output for a given expected poverty level and then perform regression tests on it.

### References

- [1] Pokhriyal N and Jacques D C 2017 Combining disparate data sources for improved poverty

- prediction and mapping *Proc. of the National Academy of Sciences* vol **114** no **46**
- [2] Jean N, Burke M, Xie M, Davis W M, Lobell D B and Ermon S 2016 Combining satellite imagery and machine learning to predict poverty *Science* vol **353** pp 790-4
  - [3] Anonymous 2018 *Philippine Statistics Authority Glossary of Terms* <https://psa.gov.ph/poverty-press-releases/glossary>
  - [4] Engstrom R, Hersh J and Newhouse D 2016 Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-Being and Geographic Targeting
  - [5] Perez A, Yeh C, Azzari G, Burke M Lobell D and Ermon S 2017 Poverty prediction with public landsat 7 satellite imagery and machine learning *Preprint* arXiv:1711.03654
  - [6] Wang W, Yang S, He Z, Wang M, Zhang J and Zhang W 2018 Urban perception of commercial activeness from satellite images and streetscapes *Companion of the The Web Conf. 2018 on The Web Conf. 2018* pp 647-54
  - [7] Ramachandran P, Zoph B and Le Q V 2018 Searching for activation functions *Under review as a Conf. paper at ICLR 2018*
  - [8] Koehrsen W 2018 *Beyond Accuracy: Precision and Recall* <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>
  - [9] Shung K P 2018 *Accuracy, Precision, Recall or F1?* <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>