

# A correlation analysis study on satellite image nightlight features and development of Africa regional economy

Xizhi Li<sup>1\*</sup>

<sup>1</sup>The Affiliated High School of South China Normal University, Guangdong Province, China

**Abstract**—Nightlight intensity has become an important factor of measuring the wealthiness of a country or an area, it mostly relies on extracting the feature from the satellite images and it becomes a dominate factor that determines the economic development. This study used CNN based deep learning model to extract the light intensity feature from the satellite images and associate it with additional survey information. CNN has been widely used for image feature extraction. Then, the study combined the survey data and light intensity feature together and conducted comprehensive experiments on different regression models that using different regularizations and optimization approaches. The paper studied the influence of regularization and optimization approaches to the model. Through the feature selection, hyper-parameter tuning, and model evaluation, the study can select the best model. This paper compares different linear regression models. They utilize different regularization and optimization. The experiment results indicate that Lasso regression model is the best model.

## 1 Introduction

Economic growth of countries can represent a lot of things: development of the countries, national power and so on. A country's economic growth also determines whether the country is developed or developing, appealing to investors to invest or not, and whether United Nations (UN) needs to fund it. Furthermore, the data collected from the country also determines what possible policy should UN use to fund or country may carry out.

It is hard to do economic survey in an entire country. In most of developing countries, according to department of Computer Science in Stanford University, data on key measures of economic development are still scarce though the quantity and quality of economic data are improving. For instance, according to World Bank data, there are only 39 of 59 African countries that conducted surveys nationally from 2000 to 2010 [1]. This phenomenon can be also observed through Demographic and Health Surveys (DHS): nationally-representative household surveys and a common measure of wealth. For the same 11-year period, 20 of 59 African countries had no DHS and 19 only had done one [2]. The problem does not only exist in those poor developing countries but also exist in strong developing countries like China. Since 1949, China only has done population census for six times. Even for those developed countries, economic growth surveys are hard to be done due to its huge financial and man-powered burden. It costs hundreds of billions of U.S. dollars to measure every target of the United Nations Sustainable Development Goals in every country over a 15-year period [3].

Moreover, the result can also be misled by traditional survey. Morten Jerven, in his book "Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It" first mentioned that in previous GDP estimation, economic activities worth about 13 billion dollars were missed. Then, he also pointed out that in Africa, many countries have a delay in up-dating information. Though China may not have this problem, it also has some officers extravagating their local GDP and cause to misconception [4]. There is an urging need for both African countries and China to find a more accurate way in getting more accurate result which cannot be altered as the traditional surveys do.

Modeling is using a math formula as basis and alter parameter to find the model revealing the best outcome. Comparing to traditional survey, modeling is better in two ways. Firstly, it can provide a model that be used to provide prediction of poverty in a faster way. While a traditional household survey may need to spend a year to complete and time span between each survey is long, a model can provide results with little labor and continuously provide current analysis. Secondly, it is also may provide a more accurate analysis for regions that are in extreme poverty or chaos. This is because those regions do not have enough resource to start a traditional survey or have a high-quality survey.

Since the difficulty for people to get cheap and accurate result, people start to seek for better methods to access result in a cheaper and more accurate way in field of big data. There are many applications in big data. For instance, Joshua Blumenstock et al collected data from mobile phones to predict wealth index of a country.[5] Among all studies based on big data, Neal Jean et al use

\*lixz.alex2019@gdhfi.com

big data of satellite images from 5 poor African Countries, extract features by convolutional neural network and start a new method in finding reasons of poverty and regions' development. In their research, however, have two short comings. First, it does not include developed region in its analysis and focus on one country in East Africa and 4 countries staying close in West Africa. Second, they directly use ridge regression model to train and analyze. This leaves us further space to explore the correlation between satellite images and economic growth.

The paper aims to propose a case study of using satellites images features to analyze regions' economic development. We use Convolutional Neural Network to extract light intensity features and combine them with the traditional surveys. Then we use methods from Economy and Econometrics to find correlation between images and economic growth. Through the obtained dataset from 30 African countries light intensity images and Demographical and Health Surveys, we have 3 hypotheses: 1) To what degree geographical factors like coast lines and mainland that affect economic development; 2) We separate data's countries into different regions: southern, eastern and western Africa and observe from a relatively macro view; 3) We compare and contrast countries that have similar and different DHS results to find out similarities and differences. This model's innovative part results in providing a faster and cheaper way for countries to survey its economic development and extending data source into 11 countries from western, central, eastern, and southern parts in Africa.

### 1.1 Related work

In 2016, Neal Jean et al have published a study about predicting poverty by combining satellite images and machine learning on Science. They use surveys and satellite images from five African countries: Nigeria, Tanzania, Uganda, Malawi, and Rwanda. They primarily use Convolutional Neural Network (CNN) to extract images' light intensity features and ridge regression to train model. CNN has been greatly used in the computer vision task: image and video recognition, image classification, and so on. The CNN contains convolution layer and pooling layer, which aims to extract the features from the original image [6]. To extract the light intensity feature from the satellite images we adopt Michael Xie et al.'s approach [7]. They use transfer learning method to solve the problem of challenges in extracting satellite images. They provide a way for people to use CNN to identify images' features that can explain up to 75% of the change in local-level economy result. However, their experiment has some limitations as well. In all these applications, a group of scientists use CNN and satellite images of 5 poor African Countries and start a new method in finding reasons of poverty. In their research, however, does not include developed countries and focus on one country in East Africa and 4 countries staying close in West Africa.

Kumar Ayush et al focus on doing in a research that helps improve the interpretability of poverty map that apply object detection in satellite images so that

policymakers can better understand the poverty map [8]. They use linear regression, ridge regression and lasso regression to train model and Pearson's  $r^2$  in measurement. They achieve 0.539 Pearson's  $r^2$  in prediction of village level poverty in Uganda. However, it only includes Uganda in its data.

Burak Uz Kent et al conduct a research trying to add more training data in satellite interpretation's pretrained model part [9]. They combine the satellite images and crowdsourced annotation at that area from Wikipedia articles. By doing so, their pretraining strategy can boost pre-trained model's performance by up to 4.5% on Image Net in F1 score. Their limitation is that their boost is only a little higher Intersection-over-Union scores than that on Image Net which may cause because different pre-training datasets contain different levels of level tasks.

Machine learning becomes more and more popular in economic modeling. In Giovanni Cicceri et al. work, they proposed Nonlinear Autoregressive with exogenous (NARX) model to better predict the influential factors that could be possible cause the economic recession. [10]. The development of the agriculture can also be modeled using machine learning techniques. Hugo Storm et al. detailedly explain the limitations of current econometric and simulation model toolbox in applied economics and the advantages of machine learning models [11]. Recent study is trying to find the different measurements on the development of the wealth of a certain region. In David Stifel's work, he adds the features: household assets, consumption, expenditure, and price data into the analysis model [12]. However, they ignored the effect of the regularization and the optimization approaches to the model. In addition, the preprocessing of the data can also influence the model performance.

In this paper we proposed 5 different machine learning models and the study the effect of different regularization and optimization approach and we also introduce the night-light intensity feature into the model.

### 1.2 Light Intensity feature extraction

In extracting light intensity, we use International Standards Organization (ISO) sensitivity in light intensity. Ranges of ISO is from 25 to 6400 and the lower the number, the less sensitive image sensor is. If the number grows stronger, there will be more noise on the picture. ISO sensitivity plays a vital role in acquiring and crafting digital images like satellite images. ISO sensitivity method is used to observe the light intensity of different countries at night. The luminance is calculated as following:

$$luminance = \frac{I_0}{H} = \frac{I_0}{t} \quad (1)$$

In the formula shown above,  $I$  is the raw brightness extracted from the raw image,  $t$  is assumed to be 0.9.

$I(\text{patch luminance}) = \text{light box} * luminance * 10(\text{-patch intensity})$ . Patch intensity means the density equation of observed data points on the area of country. The light box illuminance is from one of factor in  $t$  record: lightness.  $T$  record is from our data source which includes the raw data on lightness.

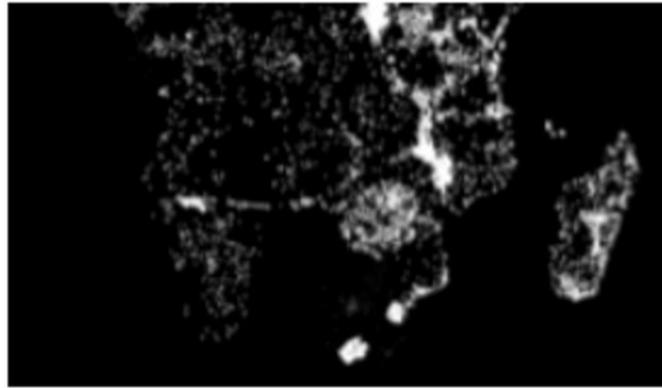


Fig 1. raw image of light brightness at night

### 1.3 About our dataset

Our dataset is consisted two parts: the CNN part and traditional part.

In CNN part, the data is mined using deep learning skills and process features. The raw satellite images of African countries are from Google Earth Engine. Then, we use ISO sensitivity method to extract light intensity. The traditional part is from Domestic Household Surveys in different African countries. These two parts are combined and become our data base.

Here are all original categories of data: svid, wealth pooled, wealthpooled5country, wealth, iso3, hv000, year, country, region, iso3n, households, LATNUM, LONGNUM, URBAN\_RURA. Svid means survey id, and an id may represent a country's survey in a certain year. Wealth-pooled means the household density in a country in this area and it may show the impact of over-population

towards poverty. The concept of wealthpooled5country is like wealth pooled but it selects 5 countries household's density. Iso3 and hv000 are different kinds of methods in calculating light brightness in different region. Iso3N shows the calculation of the result. Region means whether this area belong to middle, western, eastern, or southern Africa. URBAN\_RURA represents whether this area is urban or rural.

However, some of variables are dropped or changed into dummy variables. Svid is dropped because the id of a survey has no direct correlation with wealth. Iso, country, region, urban and rural are changed into dummy variables since they are not a continuous variable but categorical variables. Thus, it is necessary to change them into categorical variables and we use dummy variables to achieve this purpose.

### 1.4 Feature visualization

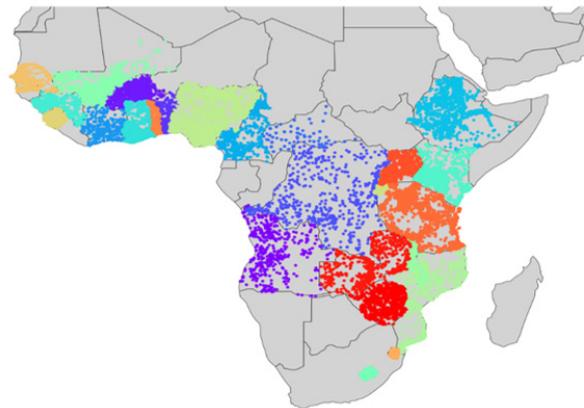


Fig 2. An illustration of geographical locations on data samples

The world map is imported as a background image which helps viewers better identify each point belongingness to which country. Color grey represents countries and color black stands countries edge.

From the map above, we can see that there are a lot of data points and their distribution is quite diverse. There are regions along coastline, regions located in mainland and regions on coun-tries' boarder.

The purpose of presenting this image is to better show where our data is collected.

### 1.5 Dummy Variables

In order to select the best model, some of factors in the data need to be change into dummy variables. A dummy variable is in binary format, which only has the value of 0 or 1, indicates yes or no.

## 2 Methods

The coefficients of the linear model can best represent the relationship between independent variables and dependent

variable. We assume that there is a linear relationship between the extracted features and wealth condition. The multivariate linear equation of a linear model is:  $y = \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \dots + \beta_m * x_m + \varepsilon$ , where  $\beta = \{\beta_1, \beta_2, \beta_3, \dots, \beta_m\}$  are coefficients;  $\{x_1, x_2, x_3, \dots, x_m\}$  are selected features for the model input and  $\varepsilon$  is the biased term or the error term. Then, we compare the different regularization approaches and optimization approach for finding the best model, subject to the minimum loss.

### 2.1 Ordinary Least Square (OLS) linear Regression

OLS has been greatly used in economic modeling and study the relationship between the dependent variable and independent variables [13]. Its most advantage is the interpretability and simplicity. We decided to choose the OLS model as the baseline. OLS regression model is the fundamental linear regression.

$$Loss(\ell) = \sum_{i=1}^n (y_i - \varepsilon - \sum_{j=1}^m \beta_j * x_i)^2 \quad (2)$$

Its loss function is measured by the average sum of residual of all observed data samples in the training, where  $n$  is the number of samples in the training set and  $m$  is the number of features. Then our goal is to minimize the loss during the training. This can be easily done by take the derivative with respect to the matrix  $\beta$  and solve the equation by setting it to 0.

### 2.2 Lasso Regression

Lasso Regression is a special linear regression by adding the regularization approach. Regularization has been greatly used in machine learning area. The overfitting is a biggest challenge when we try to train a model. Regularization can be the solution for this challenge by adding the penalty terms to the loss function. Lasso Regression utilizes the L1-regularization approach, which was first introduced in 1986 by Fadil Santosa and William W. Symes [14]. The loss function of the lasso regression can be expressed as:

$$Loss(\ell) = \sum_{i=1}^n (y_i - \varepsilon - \sum_{j=1}^m \beta_j * x_i)^2 - \frac{\lambda}{n} \sum_{j=1}^m |\beta_j| \quad (3)$$

where Lasso regression is like OLS loss function, but its L1 regularization.  $\lambda \sum_{j=1}^m |\beta_j|$  is L1 penalty.  $\beta_j$  represents one coefficient.  $\lambda$  is the hyperparameter which controls the effect of penalty term. The most important for L1-regularization is the ‘sparsity’ of the output coefficients. During the training, for those unimportant features, it can quickly set the coefficients to be 0, so it is used a process of feature selection and it can also be used for prevent overfitting in some degree.

### 2.3 Ridge Regression

Ridge regression is like Lasso regression, but it uses a different regularization technique, which we called it l2-regularization [15]. Comparing with the l1-regularization, it is mainly used for prevent the overfitting. The training

process of model usually tries to let the coefficients to be small, and eventually construct a model, which all coefficients are small. Small coefficients allow the model to be robust for different dataset. Therefore, l2-regularization term is usually can be called “shrinkage term”. The cost function of ridge regression is:

$$Loss(\ell) = \sum_{i=1}^n (y_i - \varepsilon - \sum_{j=1}^m \beta_j * x_i)^2 - \frac{\lambda}{2n} \sum_{j=1}^m \beta_j^2 \quad (4)$$

where  $\beta_j$  represents coefficient of one input feature.  $\lambda$  is the hyperparameter, which controls the effect of l2-regularization. Overall, regularization is an effective approach to avoid the problem of overfitting and ensure the generalization of the model by explicitly controls the model’s complexity.

### 2.4 Elastic Net

To avoid the limit of single l1 and l2 regularization and obtain a more flexible regularization term, Hui Zhou and Trevor Hastie proposed a new regulation approach, which combines l1 and l2 regularizations—elastic net [16]. The cost function is:

$$Loss(\ell) = \sum_{i=1}^n \left( y_i - \varepsilon - \sum_{j=1}^m \beta_j * x_i \right)^2 + \frac{\lambda_1}{n} \sum_{j=1}^m |\beta_j| + \frac{\lambda_2}{2n} \sum_{j=1}^m (\beta_j)^2 \quad (5)$$

where  $\lambda_1, \lambda_2$  are the two hyperparameter for controlling the effect of l1 and l2 regularization. This hybrid regularization provides a more a flexible way. Actually, we can clearly see that it becomes a l1 loss when  $\lambda_2$  is 0 and becomes l2 loss when  $\lambda_1$  is 0.

### 2.5 Stochastic Gradient Decent Regression (SGD regression)

SGD regression utilizes the stochastic gradient decent approach on the loss function. Stochastic Gradient Decent approach is commonly used in the optimization of deep learning. The gradient descent approach is to compute the gradient on the loss function that is:  $\beta \leftarrow \beta - \eta * \nabla(loss(\ell))$  for each training instance. Here,  $\eta$  is the learning rate that is also a hyperparameter. In order to better compare the difference between different models and control confounding variables in the experiment, we still decide the ordinary squared loss. Moreover, the SGD regression also supports the different regularization approaches. Since, the data preprocess increases the dimension of the input features, we decide to use the l1 loss by our hypothesis. Then, the optimization process illustrate in mathematical way can be expressed as following:

$$\beta \leftarrow \beta - \eta \left[ \alpha \frac{\partial R_{l1}(\beta)}{\partial \beta} + \frac{\partial loss(\ell)}{\partial \beta} \right] \quad (6)$$

$\alpha$  is the  $l_1$  loss ratio, where we set it to be 1. From the formula, we can see that the SGD training is an iterative process that gradually adapts the  $\beta$ , so we can think it as a stochastic approximation of gradient descent optimization. The advantage of stochastic process can

solve the problem of ‘local minimum’ and reach the ‘global minimum’ in the training process.

OLS regression will be the baseline for the experiment. As for lasso, ridge, and elastic-net regression, they use the different regularization approaches to improve the model performance and avoid the over-fitting problem. SGD regression, however, uses a different optimization technique and still uses the L1-regularization. These techniques introduce new parameters into the model. Then, we conduct a naive based hyper-parameter tuning to find the best parameter. The more details will be discussed in the Experiment section.

### 3 Experiment

We develop the five models under Python 3.6 and Scikit-learn (sklearn). Sklearn which is built on NumPy, Scipy and Matplotlib is an efficient tool for predictive data analysis. It is capable to provide functions like regression, dimensionality reduction and clustering.

We evaluate the model performance by two evaluation metrics: 1. R-squared and 2. MSE (Mean Squared Error):

$$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

R-squared is a statistical approach for measuring the proportion of the variance for the degree of explanation of dependent variable by the independent variable. In above formula,  $\hat{y}_i$  is the model predicted value.  $y_i$  represents the actual value.  $\bar{y}$  means the average of the observed value.

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

In this formula, n represents the number of data points.  $\hat{y}_i$  represents the predicted value returned by model.  $y_i$  is the observed sample’s value.

#### 3.1 Model selection

We increase dimensions of the original dataset by adding dummy variables. By doing this, our original 14 variables increase into 72 variables. At the same time, the dataset becomes sparser than the previous one.

As mentioned in our method part, L1 penalty in lasso regression can quickly change those unimportant features into 0 during the training. Because of this, lasso regression is more suitable to our sparse dataset.

Comparing to L1-regularization, ridge regression’s L2-regularization leads to all coefficients to be small. Since the dataset already has 72 variables, it will lead coefficients smaller. In SDG model, we use L1-regularization, but the SDG model is more fitted in a continuous dataset instead of a dataset with a lot of categorical features.

For all of reasons above, we indicate that lasso regression is the optimal model.

#### 3.2 Hyperparameter tuning

Hyperparameter tuning is to find the optimal hyperparameter for certain model. The optimal

hyperparameter ensures that the model minimizes its loss function based on given data. It is important because we need to get each model best pre-trained then compare them to find the optimal model.

In tuning part, we use Mean Squared Error (MSE), gold standard, to find the optimal hyperparameter. MSE works as an evaluation matrix.

The best hyperparameter for each model is shown in following table.

**Table1.** Optimized hyperparameter for 4 models

Model	Optimal hyperparameters
Lasso	0.0001149
Ridge	12.3284
Elastic net	(L1: 0.22, L2: 0.0)
SDG	0.3342

The table shows four models’ corresponding optimal hyperparameters. Since OLS does not have a hyperparameter, there is no need to show it in the table. The graphs of four models that show MSE changes as hyperparameter changes in are shown in appendix. With all models get their best combination of hyperparameters, we can use R-Squared method and MSE to find the best model.

### 4 Analysis

Our results are presented in the following table.

**Table2.** A table summary of MSE and R-Squared results for five different proposed models

Models	MSE	R-Squared
SDG	0.3344248561320945	0.4794704858405102
Elastic Net	0.333972917842489	0.4808387986203442
Ridge Regression	0.3340009532629111	0.4808969698824157
Lasso Regression	<b>0.3339721121421821</b>	<b>0.4809417946441902</b>
OLS	0.3340380205212713	0.4808393600883569

Based on table II, we can clearly see that lasso regression model has lowest MSE and highest R-Squared. This proves that lasso regression is the optimal model for our dataset.

#### 4.1 Results

The main part of model is shown as following:

$$\text{Wealth} = 1.579657 \cdot 10^{-3} \cdot \text{year} + 1.48680799 \cdot 10^{-4} \cdot \text{iso3n} + -2.90026294 \cdot 10^{-3} \cdot \text{households} + -2.68290928 \cdot 10^{-2} \cdot \text{LATNUM} + -8.86857605 \cdot 10^{-3} \cdot \text{LONGNUM} + \dots + -4.75360325 \cdot 10^{-1} \cdot \text{rural} + 7.85720199 \cdot 10^{-1} \cdot \text{urban}$$

(more coefficients are shown in appendix)

The final linear regression model can be integrated as following: '1 unit increase in the iso3n is corresponding to 0.000149 unit increase in local wealth.' From the model, we can testify our hypothesis on the night-light feature. The more brightness of the country, the higher economic development it has. Moreover, we further prove the feasibility in using light intensity to analyze and predict region's development in different regions by using a larger dataset.

1 unit increase in the year is corresponding to 0.00158 unit increase in local wealth. Then we can conclude that the development of Africa economy is still a slow process. Based on our background research there are many other confounding variables could be reasons such as religion, conflicts between tribes, diseases such as Ebola, unstable social environment etc.

A 1 unit increase in the household is corresponding to 0.00290 unit decrease in local wealth. Based on our background research, due to factors like local custom, social environment and stereotypes, the living standard per capital and local economic development decreases as the household increases. Thus, family members spend more time in surviving instead of learning more skills. This finally leads to an endless poverty circle.

## 5 Conclusion

In this paper, we apply a novel approach by adding extra features on traditional features which utilizes the state-of-art image processing deep learning model. Through extracting night lightness in Africa's different regions from satellite images, we find a correlation between light intensity and wealth. The regions with higher light intensity are more developed. Moreover, we also take traditional factors like longitudes, latitudes, and household into consideration in our re-search. From the aspect of big data, we use different kinds of regularizations and evaluations methods to model the optimal correlation. The best model in our research is lasso regression model, but we find that different ways of pre-processing data will lead to different outcomes. For instance, without dummy variables, the best model is ridge regression instead of lasso regression. Different regularizations will lead to different outcomes as well.

Our model has two significant points. First, it validates our hypothesis that light intensity has a correlation with wealth and can be used to analyze region's economic development. Secondly, it also reveals problems some African countries have in improving their economic growth: the household variable reveals the stereotype and social environment problems that some African countries have.

These African countries' governors may get two useful results. Firstly, they can use our light intensity model to

analyze a region's development. Secondly, they need to pay efforts in helping their people have fewer stereotypes.

## 6 Future work

In the future, we want to expand our dataset by adding more regional economy data collected from China and explore the effectiveness of our model in analyzing China's economic development. We are also interested in using more advanced models to improve our current analysis. For example, deep learning is a cutting-edge area in machine learning, especially for the neural network. We want to explore the deep learning techniques in the analysis of economy development such like feed-forward neural network.

## References

1. World Bank. 2018. Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO
2. United Nations, "A World That Counts: Mobilising the Data Revolution for Sustainable Development" (2014)
3. United Nations. The Sustainable Development Goals Report 2020.
4. M. Jerven, "Benefits and costs of the data for development targets for the Post-2015 Development Agenda," Data for Development Assessment Paper Working Paper, September (Copenhagen Consensus Center, Copenhagen, 2014).
5. Blumenstock, J., et al. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science*, vol. 350, no. 6264, 2015, pp. 1073–1076., doi:10.1126/science.aac4420.
6. Jean, N., et al. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science*, vol. 353, no. 6301, 2016, pp. 790–794., doi:10.1126/science.aaf7894.
7. Xie, Michael. "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping." *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, 2016
8. Ayush, Kumar, et al. "Generating Interpretable Poverty Maps Using Object Detection in Satellite Images." *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020, doi:10.24963/ijcai.2020/608.
9. Uzkent, Burak, et al. "Learning to Interpret Satellite Images Using Wikipedia." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019, doi:10.24963/ijcai.2019/502.
10. Cicceri, Giovanni, et al. "A Machine Learning Approach to Forecast Economic Recessions—An Italian Case Study." *Mathematics*, vol. 8, no. 2, 2020, p. 241., doi:10.3390/math8020241.

11. Storm, Hugo, et al. "Machine Learning in Agricultural and Applied Economics." *European Review of Agricultural Economics*, vol. 47, no. 3, 2019, pp. 849–892., doi:10.1093/erae/jbz033.
12. Sahn, David E., and David Stifel. "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data." *Review of Income and Wealth*, vol. 49, no. 4, 2003, pp. 463–489., doi:10.1111/j.0034-6586.2003.00100.x.
13. Goldberger, Arthur S. (1964). "Classical Linear Regression". *Econometric Theory*. New York: John Wiley & Sons. pp. 158. ISBN 0-471-31101-4.
14. Santosa, Fadil; Symes, William W. (1986). "Linear inversion of band-limited reflection seismograms". *SIAM Journal on Scientific and Statistical Computing*. SIAM. 7 (4): 1307–1330. doi:10.1137/0907087.
15. Gruber, Marvin (1998). *Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators*. Boca Raton: CRC Press. pp. 7–15. ISBN 0-8247-0156-9
16. Zou, Hui; Hastie, Trevor (2005). "Regularization and Variable Selection via the Elastic Net". *Journal of the Royal Statistical Society, Series B*. 67 (2): 301–320. CiteSeerX 10.1.1.124.4696. doi:10.1111/j.1467-9868.2005.00503.x