ISSN: 2515-8260 Volume 06, Issue 01, 2019

PREDICTION OF POPULATION DENSITY & POVERTY RATE USING MOSAICSWITH SATELLITE IMAGERY

B. SATYANARAYANA, Principal & Professor, Department of CSE

CMR Institute of Technology, Kandlakoya (V), Medchal Road, Hyderabad :

bsat777@gmail.com

Abstract

The project involves combination of Machine Learning along with Satellite Imagery (SIML) which is having potential for addressing major global problems by remotely accessing socio-economic and meteorological conditions in data poor areas, although SIML's resource requirements will limits its access and utilization. Combining Machine Learning along with Satellite Imagery (SIML) is enabling better characterizations for population densities and poverty Rates. Further, this ML (Machine Learning) Application proves to be a path which is effective to convert such huge amount of unformed image data into formed assess of conditions of ground.

Keywords—ML (Machine Learning), Satellite Imagery, Image encoding, CNN, MOSAIKS, Regression, SIML

1. INTRODUCTION

The capital requirements to group SIML technologies, although, restrict their obtainability and practice. Satellite based dimensions are especially underutilized in less income circumstances, Where the high-tech capability to enactment SIML maybe less, but those dimensions would probably disclose the utmost advantage. For instance, government authorities in less income context might want to grasp local water pollution, unlawful fields usage and also mass relocations. However, SIML, residue mainly not able to reach the possible end user, as present-day paths need the crucial resource comprehensive endeavour, compelling an integration of particular task proficiency, Remote-Sensing and skill of Engineering, accessing the images, customization including attuning knowledgeable ML(MachineLearning) architectures, and huge computational assets.

To detach obstacles, a new path to SIML is required which empower the non-experts by gaining the level of development performance by not using the particularised computational assets or even by advancing the complicated procedure for prediction. This single performed task sceptic encoding which converts the satellite images into features(vector of variables) empower this path by decoupling users from expensive handling of images. While anterior work has pursued an unsupervised method to convert these satellite images into a data of single set of features to accomplish competitive-performance with methods of deep learning including a mixture of tasks to scale them globally.

Our path permits familiar sources of images to change into compacted sets (features) which analysers can acquire, where individuals solve the tasks which are heterogeneous, this will separate the further people (users) to avoid the expensive steps like procuring, handling, image storage, and to process the images. Immensity of these results will increase in size by enlarging SIML-user-community and also the global imagery data scale. This system of

ISSN: 2515-8260 Volume 06, Issue 01, 2019

SIML naming "Multi-task Observation using Satellite Imagery and Kitchen Sinks" (MOSAIKS), will make SIML more approachable and broad-based by dividing the process into 2 different steps: a single performed "Featurization step" by transforming the satellite images into a short-vector representation (satellite image → a), and the "Regression step" which will grasp the particular task quantity by mapping the sets (features) to the outputs of taken (given) task (a→b). For every satellite-image this featurization process will be done only once to get one set (outputs), where these sets will be used further to solve various dissimilar functions by repetition of this regression process through numerous individualistic users.

A. Objectives

1. Enabling the users with basic resources to easily predict the Population Density and Poverty Rate by using only Satellite images.

B. Problem Statements

Many biggest global challenges like managing poverty rate & population density etc., are necessary for nation's ecosystem development. Planet scaled which are based on ground, these systems are generally monitored at restrain cost, satellite image represents an alternative for gathering comprehensive data globally. For data poor regions it is the difficulty is more to analyze and predict population density. This project mainly looks into the problem to predict the assets of some small regions like population density in a single time period, By considering the inputs which are only the day-time satellite images with high resolution.

2. RELATED WORK

Abhishek P and others used various machine learning techniques for the prediction of land coverage from the images obtained from a satellite. For obtaining features of the input from satellite image they have used a time series technique called NDVI. They have executed their work mostly using Python and they have observed that K Nearest Neighbor algorithm is the most accurate technique for their work. [1].

Shailesh P and others proposed an approach for predicting and calculating the poverty rate in the rural areas in India. The Google Static Maps API to extract images for the villages from the determined geocodes. Their process consists of two steps, in the firststep they tried to train multi-task fully convolutional model and in the next step they tried to predict poverty rates. [2]

Stefanos G and others developed an implementation of RandomForest and they called it as Geographical Random Forest to predict and calculate population density with Remote Sensing data(High Resolution). They observed that their proposed method is better tomodel the heterogeneous remote sensing data. [3].

Fariha S and others proposed a method which involves mapping(agricultural) and monitoring in the area of Habiganj for the prediction of growth of crop and also yield. They obtained high resolution images from Landsat-8 to monitor the areaof Habiganj and those images are processed and obtained indices which were correlated to growth and crop yield. For predicting the yield they have used time-sereies techniques such as Long Short-Term Memory and Autoregressive Integrated Moving Average. [4].

Kuldeep C and others tried to predict the types of land coverage. For this purpose they have used raw images from Sentinel2A Satellite and also they have used classification techniques such as Support Vector Machine, Random Forest and other classification techniques, they also tried to obtain the accuracy of these classification techniques and they have observed that Random Forest classification technique (95.67% accuracy) provides betterresults than others. [5].

M. Das's proposed paper consists of two parts one is preprocessing of data and another is autonomous learning or prediction. They have used an online prediction model which is an autonomous prediction model 'OPAL' .Recurrent Neural Network was used in OPAL technique for enabling the online prediction. [6].

Tianjun W and others tried to perform population mapping, here they have used residential geo objects and treated their model as spatial prediction model for performing mapping predictions and they have used machine learning methods with satellite images that have high resolution. They have observed that their model provided better results for finding distributions of population. [7].

B. A. Gebreegziabher's paper tried to solve the supply chain crisis by obtaining road pavement quality information, they have obtained imagery from Sentinel2A Satellite. Several deep learning techniques were used to obtain road pavement information.He had observed that deep learning techniques such as U-Net and IoU provide better accuracy than other machine learning techniques such as Random Forest. [8].

M.S. Andreano and others tried to address and solve the digital divide between people .The main cell phone parameters they have taken into account are cost of service and cell phone adoption. They have observed that image recognition approach (predicting 40% data variance) is better than baseline approach (predicting 20% data variance).Among different approaches they have observed that CNN method provides 41% data variance.[9].

R. Jarry and others observed that to make proper prediction of poverty rate they need reliable poverty indicators. They also concluded that by only taking nighttime satellite imagery is not a good approach for poverty prediction so they decided to use CNN approach where they obtain different features from satellite images and they perform regression techniques to obtain better results for poverty prediction. [10].

Nischal, K. N et al : We can predict future poverty, population growth, gross domestic product (GDP), and forest cover in India by using night-time satellite imagery and census data. This study presents a technique for doing just that. There are a number of techniques we employ to fill in the blanks in census data, including predictive model-based methods, multivariate regression analysis, and the ARIMA model. We use the ARIMA model and the regression model to verify the accuracy of census data when predicting. We provide a method for obtaining economic and timely information on poverty, which may be used to establish monetary policy, foreign aid, or to direct other types of assistance.[11]

Upadhyay, A.,et al : One of the most critical components of advanced image processing is remote sensing, which is used to gather geospatial data from satellite images. The monitoring of land use areas is an essential aspect of remote sensing's application in monitoring various resources on the earth's surface. There are a variety of strategies and procedures that may be used to classify satellite images based on various themes. Land use and land cover mapping using the LISS-III satellite image will be discussed in this study. Pixel-based categorization of LISS-III satellite images has been used to categorise images of water, forest, mangroves and human development in various geographic regions. After evaluating the first training data set, several linear regressions are performed to this dataset to see how the optimal training data set affects classification accuracy. Analyzing classification accuracy requires calculating the confusion matrix and the Kappa coefficient. Without training data, ANN classifiers have an accuracy of 90.19 % and a Kappa Coefficient of 0.8533, but when training data is evaluated using multiple linear regression, ANN classifiers' accuracy increases to 98.74 percent and a Kappa Coefficient of 0.9766, which is a remarkable improvement in accuracy and performance. The suggested strategy improves the classification accuracy of ANN classifiers for LISS-III satellite image categorization.[12]

Luo, H.,et al :Forest fire area can't be accurately predicted using the Multivariate Linear Regression Model due to an error in interpreting the linear connection between independent and dependent variables. In order to minimise multicollinearity, model the data, and predict the fire area, the article makes use of the ridge regression model. Unstable standardised ridge regression coefficients or stable coefficients with tiny absolute values are first eliminated. For the Support Vector Machine model, the remaining characteristics serve as new input values. It is then possible to collect classification results from the new dataset, which has been partitioned into training data and test data. Finally, the model's correctness is examined in light of the results. The results of the tests show that the approach is capable of accurately predicting the locations of fires.[13]

Pandey, P.,et al :Improved contrast may be achieved by expanding the dynamic range of intensity levels used by Conversion functions, which in turn aids in the enhancement of contrast. To alter picture quality and borders, several conversion methods use local significant material. Here, an effective approach for enhancing images' contrast has been developed using an efficient combination of universal and local conversion functions that retains the image's intensity and fine details. The universal conversion function maintains all of the image's contrast and intensity settings. Satellite pictures that are poorly concentrated or deteriorated may be preprocessed using this paper's intuitive technique, which consists of a series of autonomous processing steps that reduce the effects of blur and noise. Finally, each stage's output is compared using PSNR parameters to determine the overall quality.[14]

Yuan, Z.,et al :In the field of intelligent video surveillance, an important research topic is the automatic assessment of population density. With the present state of big data, it is difficult to keep up with the design elements required by traditional methodologies. The use of deep learning to video surveillance is also an enticing trend as artificial intelligence approaches such as this one gain popularity. The need for a multilayer convolutional neural network (MCNN) arises as a result of the drawbacks of conventional manual feature extraction and the shortcomings of a single-layer convolutional neural network (CNN). The features of CNN learning images will not be affected by changes in head size, such as the penetration effect, in this article. In other words, even if we don't know the input map's viewpoint, we can use adaptive kernel to properly estimate the density of the population. The population density map is created by combining the graphs of each layer's characteristic graphs. According to our findings from our studies, this network topology is capable of providing a more precise estimate of the population size.[15]

ISSN: 2515-8260 Volume 06, Issue 01, 2019

Suhail, M.,et al : It is critical to accurately estimate k in ridge regression analysis. Thereare a variety of ways to estimate this kind of parameter. Some of these approaches have been examined in this paper, and novel estimators based on the generalised ridge regression approach have also been developed. Based on the MSE criteria, a simulation study has been conducted to assess the performance of the suggested estimators. If the suggested estimators perform well under particular situations, they outperform LSE and other popular current estimators. Finally, a numerical case was examined, and the results therein seem to corroborate the simulation's predictions.[16]

Lin, C. Y.,et al :Chlorophyll concentration (CHLS) was estimated using multispectral and hyperspectral data acquired from the WorldView multispectral camera. The multispectral effective chlorophyll indicators (MECIb and MECIc) generated an estimate of roughly 38 percent PRMSE, which is almost four times more than the value of the measure (1-R 2) in CHLS-MECI models. Multiple regression models produced using the standard least squares technique and the ridge method were both afflicted by similar levels of uncertainty in their estimates of tree leaf chlorophyll. In contrast to prior models that used just one NIR band, a ridge regression model based on the reflectance of both NIR bands was able to minimise CHLS estimate uncertainty greatly. The ridge model obtained RMSE=0.35 mg/g and PRMSE=22%, demonstrating its ability to increase estimate accuracy. The ridge model was shown to have a 65 percent decrease in the PRMSE of fresh leaves compared to CHLS-MECI models. However, the CHLS assessment of water-stressed leaves did not improve.[17]

Li, T.,et al : Satellite image road recognition is a hotly debated problem in the world of image processing. Based on convolutional network findings, we provide a new road extraction and tracking strategy that improves road identification. Connected-tube marked point process (MPP) model and a post-tracking technique are used in the suggested approach.Our approach for detecting roads in remotely-sensed photos is shown using data from the Massachusetts roads dataset.[18]

La, Y.,et al : Data from the Census, satellite imagery, and Land-Use/Landcover (LULC) may be used to map and describe urban expansion. In this work, we utilized NTL to evaluate urban development patterns in Adelaide, Australia and Tokyo, Japan, and then studied the link between NTL, LULC, and population census data for assessing urban growth in the two cities. Following geographical correlation research, it was shown that both Adelaide ($r = 0.90$) & Tokyo ($r = 0.81$) had a substantial positive association between urban/built-up area and population density. Furthermore, a multiple linear regression model was used to estimate Tokyo's population density using NTL imaging and urban/built-up data, with a correlation value of $R^2 = 0.80$. Urban population increase may be predicted using a mix of NTL and LULC data.[19]

Jing, R.,et al : Multi-scale SLIC-based image segmentation is first applied to multi- temporal pictures in order to obtain segmented objects while retaining as much edge information as feasible. CNN architecture is utilized to build a change map, and then a SCAE feature-based classification technique is employed to generate "from" and "to" change data. Finally, the identification of changes is improved using Bayesian information criteria. SLIC image segmentation has an impact on the consistency of change regions; CNN features have an effect on the integrity of change regions; and SCAE features have an effect on the performance of support vector machine (SVM) classifiers, according to this study's tests. Features derived from the structures also improve the capacity to retrieve information from the earth. Other techniques of change detection are outclassed by this one in a side-by-side comparison.[20]

Ayush, K.,et al : Governments and humanitarian groups must be able to accurately evaluate poverty at the local level in order to monitor progress in improving livelihoods and distributing limited resources. Even while computer vision developments in satellite images have improved accuracy in predicting poverty, they do not provide characteristics that can be interpreted by policymakers, making it difficult for practitioners to embrace these techniques because of need alone. By applying object detectors to high resolution (30cm) satellite photos, we establish an interpretable computational framework for effectively predicting poverty at the local level. Predicting Ugandan village poverty using weighted counts of items yields a r2 of 0.539, a 31% improvement over previous (and less interpretable) benchmarks. Ablation and feature importance demonstrate apparent links between object count and poverty forecasts. Our findings show that performance does not have to be sacrificed for interpretability, at least in this critical area.[21]

Ferreira, B.,et al The numerous Earth Observation methodologies and their relevance to the UN Sustainable Development Goals are presented and explored in this study. Sustainable Development and its aims are reviewed, followed by Earth Observation methodologies pertinent to this subject, with specific emphasis on Machine Learning methodsand algorithms and their potential and capabilities to promote the attainment of Sustainable Development Goals.. With Earth Observation's cost-effectiveness and information richness, it can play an important role in monitoring the Sustainable Development Goals. Data extraction and synthesis are critical to the effectiveness of Machine Learning in Earth Observation data processing, despite the fact that it has been widely used. To find the most important aspects of Sustainable Development, it is necessary to conduct a thorough and comprehensive study of the data that is now accessible. All in all, this study sheds light on how Earth Observation and Machine Learning may be used to help Sustainable Development in nations, and how they can be used to uncover relationships between them. In order to achieve the Sustainable Development Goals, new methodologies and techniques, including Machine Learning, are highly suggested because of the significance and expanding quantity of data provided by Earth Observation.[22]

Stratoulias, D.,et al : After 25 years of declassification of the Corona reconnaissance satellite mission, a historic collection of high-resolution panchromatic photographs from the Cold War period is now openly accessible to the public Bulgaria has seen a decline in the population and a decline in agricultural output at the same time. Based on the analysis of a 1968 Corona photograph, we want to map the villages surrounding the Bulgarian city of Plovdiv. The findings are compared to data from current Sentinel-2B and Landsat-8 photos. The capacity of the Corona image to detect settlements and the possible application of textural analysis in the context of land use and land cover mapping of historical photos are discussed in this study. Here we show how an early satellite picture might be used for feature extraction using textural analysis, as well as an overview of the state-of-the-art in textural analysis.[23]

Tsvetkovskaya, I. I.,et al :A new generation of remote sensing tools and technologies may be used to high-resolution satellite pictures collected through space radio communications. According to the research presented in this article, the electromagnetic spectrum can be mapped using ground-based, aerial, and space-based cameras. Artificial spacecraft and technologies for collecting space data are discussed in this article. Other examples include how fake Earth satellite data may be processed using convolutional neural networks (CNN). Using data, CNN is able to automatically learn the best functions depending on the information.[24]

ISSN: 2515-8260 Volume 06, Issue 01, 2019

Jarry, R.,et al To accurately estimate poverty using satellite photos, it's important to have accurate and precise poverty indices. There have been a number of approaches lately presented to deal with this issue. A proxy (e.g., nighttime light) is used to supplement limited data in most current techniques. A CNN is built and trained on a large number of pictures, and it is then used to extract features from those images. In the end, a regression model is built using pairs of retrieved feature vectors and poverty labels.[25]

Zhang, S.,et al By adding the L2 regularisation component to the equations, ridge regression, a common regression technique in machine learning, may effectively handle the problem of solutions with some aspects irreversible in multiple linear regression. Multiple linear regression also has its own set of drawbacks, such as long calculation times, sparse, unstructured data, and so on. The use of t-distribution parameters in the glowworm swarm optimization approach has been recommended in this study, which has a higher computing efficiency in achieving the optimal solution and can effectively avoid the solution from sliding into the local minimum trap, making it more efficient. [26]

Keswani, M.,et al Satellites in constant orbit around the planet keep tabs on the planet's surface all the time. As a consequence, massive amounts of satellite picture data are generated. This data is becoming more detailed in terms of both geographical and temporal resolution. Satellite photography is useful for a variety of tasks, including land cover categorization. Deforestation, desertification, and water shortages may all be detected using data gleaned from land cover categorization to help with resource planning. The categorization of land cover is also useful in determining changes in the land cover through time. Pixel-wise data is provided by most satellites and high-end sensors. There are billions of pixels created even for a tiny research area because of high-resolution photographs. For this reason, it is necessary to strike a compromise between runtime and accuracy. In this research, we compared the Landsat-8 multi-spectral and temporal Landsat-8 satellite data classification models using the Multi CNN model to the standard Landsat-8 classification models.

Milojevic-Dupont, N.,et al The full potential of artificial intelligence and machine learning for climate change mitigation has yet to be realised. For climate change mitigation, we perform a comprehensive assessment of applied machine learning works including remote sensing, transportation and buildings. Research in this area has been going on for over a decade, and it's just going to become bigger. Data and machine learning approaches are enabling climate solution research to surpass general suggestions and deliver policy options that are tailored for unique settings yet scalable for global mitigation potentials on a local and regional level. In order to accelerate, improve, and change urban infrastructure provision, we propose a meta-algorithmic architecture and framework for applying machine learning to optimise urban planning.

Huang, X.,et al With the fast development of remote sensing tools, we have access toa wide range of information about the ground, which can be combined with new deep learning systems that can extract hidden features and patterns. First, this study compares the performance of common deep learning models in predicting population distribution from remote sensing photos, and investigates the influence of adjacent effect and possible systematic biases in population estimation. A mapping between Sentinel-2 image patches and the LandScan population grid's corresponding population count allows us to train four common deep learning architectures end-to-end: VGG, ResNet, Xception, and DenseNet. Under all chosen nearby situations, DenseNet surpasses the other three models, while VGG suffers the poorest outcomes. As for the adjacent impact, contrary to previous research, our findings show that increasing neighbouring sizes reduces population estimation performance across all four models in each of the evaluation measures. All chosen deep learning models tend to overestimate sparsely populated picture patches and underestimate highly populated image patches, independent of nearby sizes. This is an important and universal bias to note. This article's methodological, experimental, and contextual information should be useful for a broad variety of future investigations that use remote sensing images to estimate population distribution.

Gebreegziabher, B. A. Keeping the world's food supply secure is made more difficult by losses that occur throughout the supply chain. Agricultural output has increased dramatically in recent years, but the infrastructure required to handle it has lagged behind. After harvesting agricultural crops, this inadequacy leads in post-harvest losses. It is common in underdeveloped nations for post-harvest losses to impede agricultural food supply systems. Because transportation is a critical component in the post-harvest chain, these losses have a significant impact. Even in developing countries, where road transportation is the primary mode of transportation, the quality of transportation services is critical. To guarantee quality, it is necessary to keep an eye on, maintain, and repair roads. Due to a shortage of funds, these activities are not carried out on a regular basis. Agricultural items such as tomatoes are damaged during transportation because of the widespread use of low-quality roads. These issues may be addressed by using spatial road quality information, according to this research. More significantly, making this information easily accessible in resource-constrained places like Sub-Saharan Africa is critical. To that end, this study looked at the possibility of utilising machine learning approaches to map road pavement condition using publicly available optical satellite photos. Using reference data gathered for a corridor stretching from Accra (Ghana) to Ouagadougou (Burkina Faso) utilising crowdsensing technologies, shallow and deep learning models were created to extract road condition information from Sentinel-2 satellite photos. Using such a data source for easy access to road pavement quality information yielded good results. U-Net, a deep learning model, outperformed the shallow ML alternative, random forest, with an F1-score of 37.93% and an IoU of 32.40%. Because of the inherent data imbalance, it is impossible to compare the results of this work with those of a typical segmentation task. However, the findings were comparable to similar road extraction programmes that used Sentinel-2 photos. For the purpose of evaluating the relative usefulness of Sentinel-2 photography in this endeavour, the researchers also compared it to data from Planet's imaging satellite. In the pixel-wise categorization of road pavement condition, Sentinel-2 photographs outperformed Planet ones, according to the findings (RPQ).

3. PROPOSED WORK

Figure 1 depicts the Satellite images are included with various potential functions. For instance, 9 satellite-images consisting of population density and poverty rate from the India are taken, both having distinct identified features which are observable in each and every image. Later, N satellite-images will be converted using RCF's into k-dimensional feature vector which is highly descriptive before labels are well-known. After computing these features, they are kept in a table form known as matrix X. It will be useful for various functions excluding re-computation. The people (users) having a new function (z), can mix their self-taken labels to given features and can train them on their own. Linear prediction and total MOSAIKS sample consisting of features will give SIML assess (labels) at any locations.

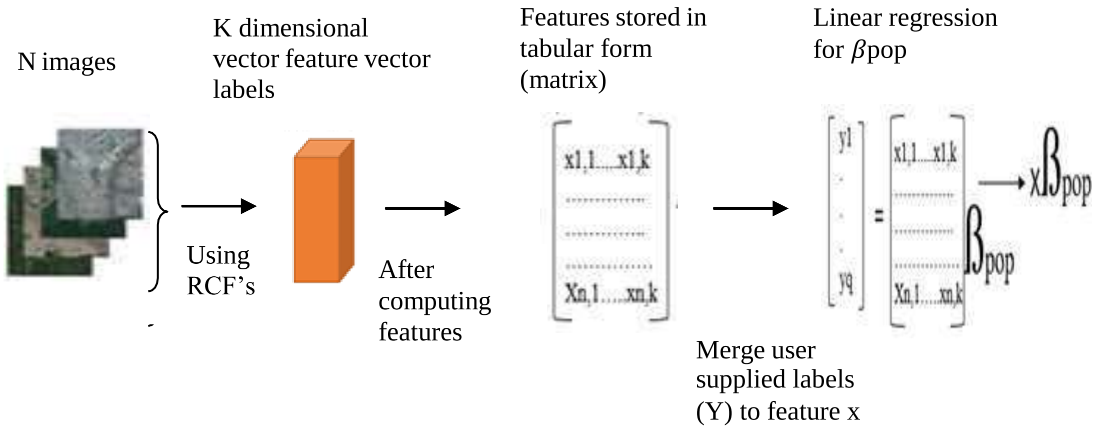

Figure. 1. Proposed Model Block Diagram

A. Featurization Step

This step is to convert satellite images into vector-representations(images→x). The victory of generalizability depends on how these images are converted into features. The layout of featurization function improves on the conceptually grounded-ML(MachineLearning) idea random Kitchen-sinks, where it will be put into images by forming Random-Convolutional-Features (RCF's). RCF's will capture a pliable measuring of similarity among sub-images of mixed set of images by not considering the circumstantial and any other details. Regression process here tends features 'q' on a point to predict 'r'. This can be non-linear functioning of images as shown in fig 2.

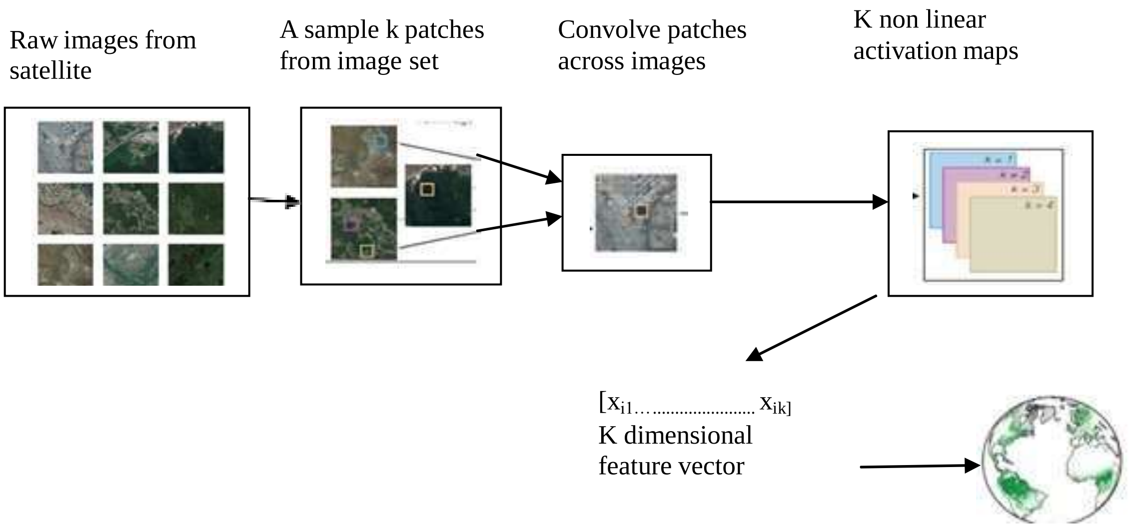

Figure. 2. Process of Featurization step

B. Regression Step

For every image this process will be done once to get a single set (outputs), where these sets will be used further to solve various dissimilar functions by repetition of this regression process through numerous individualistic users. Here we used Ridge Regression. In this

ISSN: 2515-8260 Volume 06, Issue 01, 2019

Ridge Regression, it is of the form Y=XC+e. Here Y is the user contributed labels i.e., labels (population density) for 1 to q locations. X is the K- dimensional feature vector. Every user solving a linear regression(single) for C. The linear forecasting by βs and total MOSAIKS sample of features ‘q’, will give SIML assess (labels) at any locations as shown in fig 3.

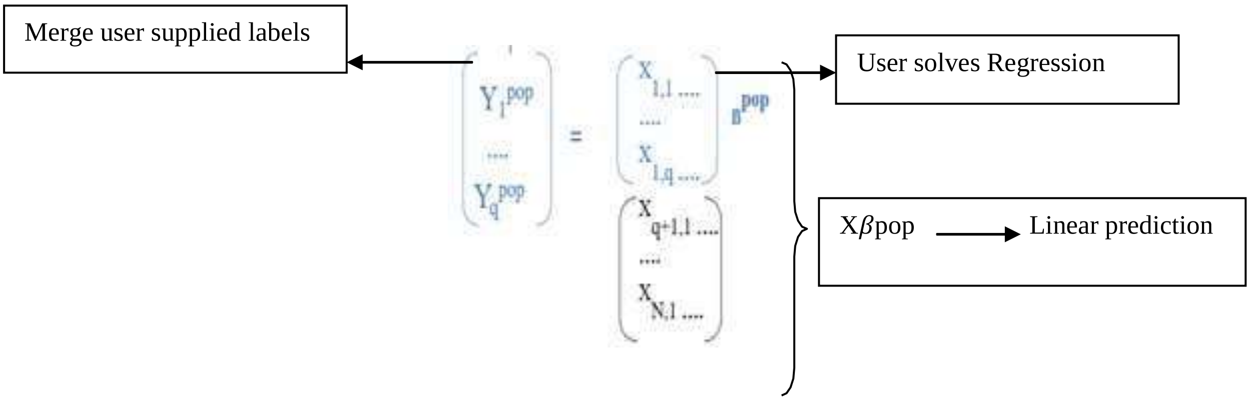

Figure. 3. Process of Regression Step

C. MULTI TASK PERFOMANCE OF MOSAIKS

100,000 daytime images were taken and by featurizing the images by running them via MOSAIKS feature extraction method which produces 8,192 features and stored. Here got K= 8,912 features per image which then repeating regression process by using ridge regression which is cross validated, for every function to easily forecast the population density and poverty rate. This can also be extended for elevation predictions by considering only the matrix of features (X) which is generated. The tasks are selected on a basis of exploration and diverseness. By using the same procedure, the predictions for each task are generated.

4. ALGORITHM

The Prediction procedure is a two step process i.e. Featurization and Ridge Regression. The overall Approach can be defined with a simple algorithm as follows:

Merge X, Y

RidgeRegression(X,Y)

Predict Y

Where,

X is the Feature matrix obtained in Featurization Step

Y is the User supplied labels, Here the User supplied labels are Population Density and Poverty Rate.

1. The Featurization step includes Grid Creation, Feature Extraction and Label Creation.
2. The Regression Step mainly focuses on two steps i.e. loading Feature matrix X and Running Regression.
   - Loading Feature Matrix

```
X = {}
latlons = {}
X["POP"], latlons["POP"] = io.get_X_latlon(c, "POP")
X["POV"], latlons["POV"] = io.get_X_latlon(c, "POV")
```

Where,
POP is the labels of Population density
POV is the labels of Poverty Rate

- Regression

```
Regression()
{
subset_n = slice(None)
subset_feat = slice(None)
solver = solve.ridge_regression
(
        this_X,
        this_X_test,
        this_Y,
        this_Y_test,
        this_latlons,
        this_latlons_test,
    ) = parse.merge_dropna_transform_split_train_test(
        c, label, X[sampling_type], latlons[sampling_type]
    )
    this_X = this_X[subset_n, subset_feat]
    this_X_test = this_X_test[:, subset_feat]
    this_Y = this_Y[subset_n]
    this_latlons = this_latlons[subset_n]
kfold_results = solve.kfold_solve(
        this_X,
        this_Y,
        solve_function=solver,
        num_folds=c.ml_model["n_folds"],
        return_model=True,
        return_preds=True,
        svd_solve=False,
        clip_bounds=bounds,
    )
    preds = np.vstack([solve.y_to_matrix(i) for i in best_preds.squeeze()]).squeeze()
    truth = np.vstack(
        [solve.y_to_matrix(i) for i in kfold_results["y_true_test"].squeeze()]
    ).squeeze()
ll = this_latlons[
        np.hstack([test for train, test in kfold_results["cv"].split(this_latlons)])
    ]

    data = {
        "truth": truth,
        "preds": preds,
        "lon": ll[:, 1],
```

```
        "lat": ll[:, 0],
        "best_lambda": best_lambda,
    }
with open(save_path_validation, "wb") as f:
        pickle.dump(data, f)
    results_dict = r2_score(truth, preds)
}
```

5. RESULTS

The above screenshot is taken after the prediction of population density analyzing the final estimates. The Map is presented with different colors on the basis of density scale which is ranging from 0-4. When mouse hovers it displays the density, density scale and id of that particular state in India.

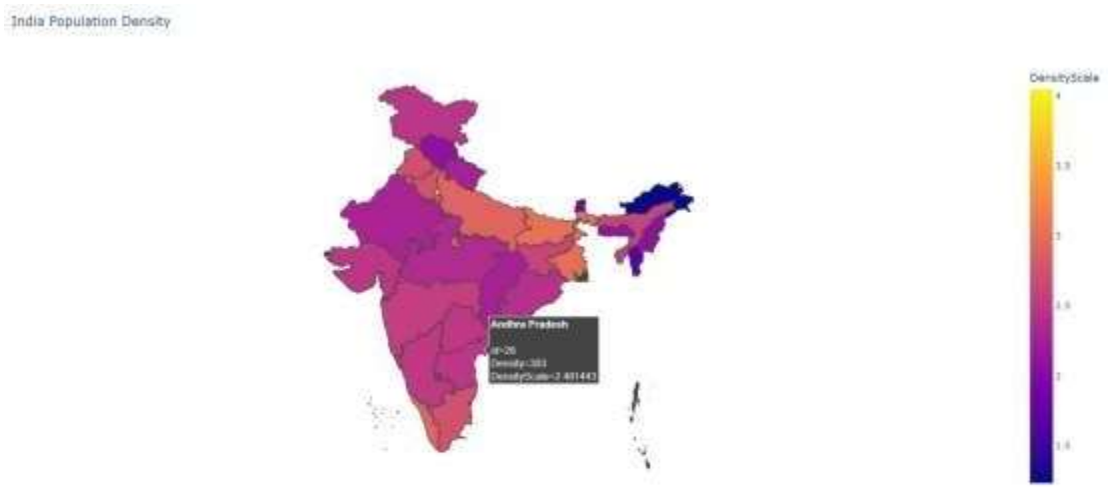

Figure. 4. Prediction of Population Density for Indian maps

ISSN: 2515-8260 Volume 06, Issue 01, 2019

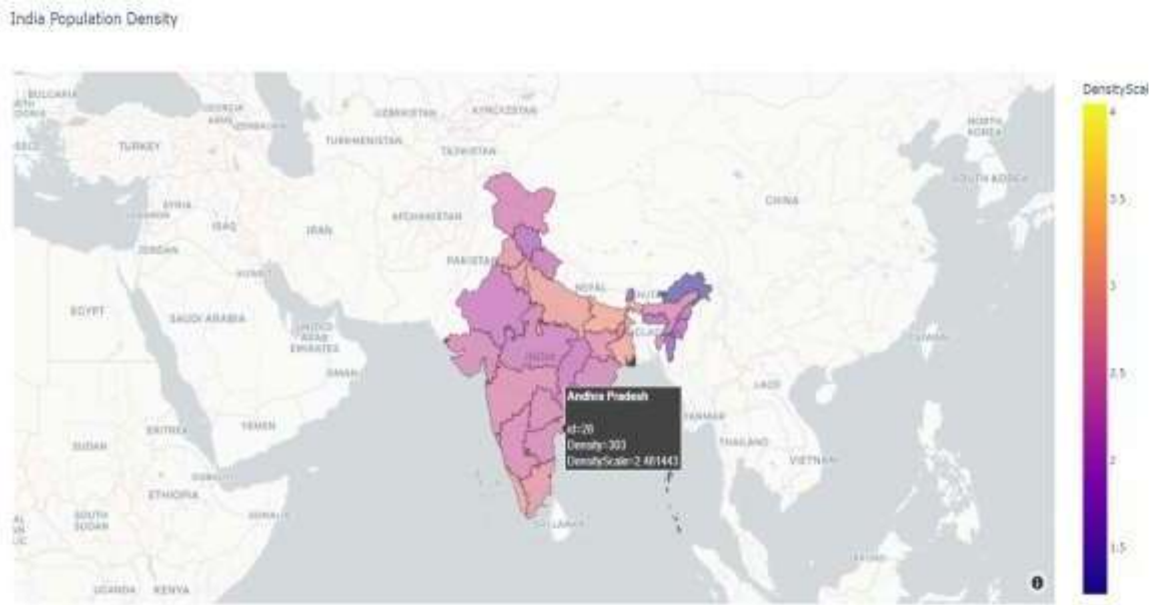

Figure. 5. Prediction of Population Density for Indian maps representing in world maps

The above figure represents the prediction of population density for Indian maps which is plotted in world maps. Similarly, It also shows the density, density scale and id of each state. The density is obtained by dividing population with area. The density scale ranges from 0-4.

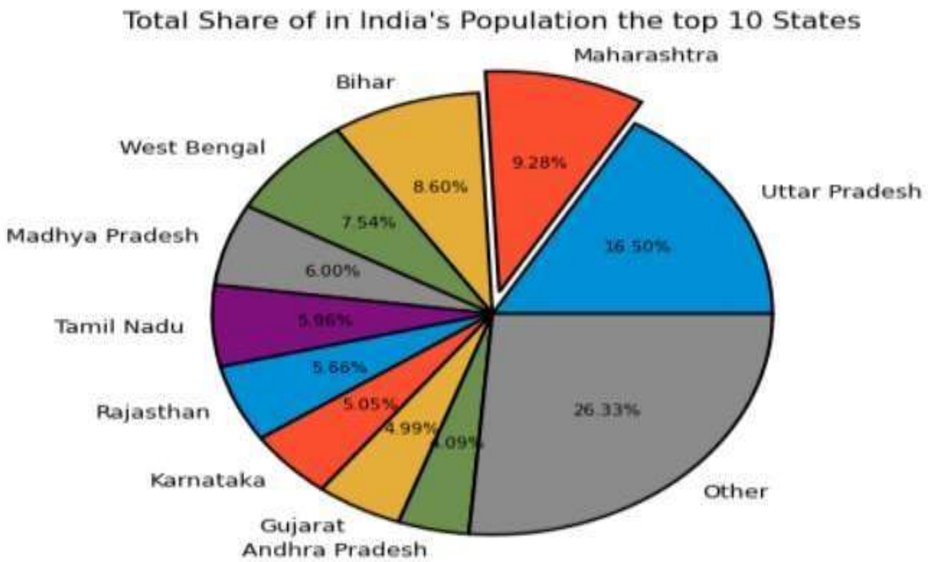

Figure. 6. Pie Chart representation of India's Population

The above Figure represents the India's population in top 10 states which indicates Uttar Pradesh has Highest Population density. It also replicates the Highest and Lowest Population Density states.

ISSN: 2515-8260 Volume 06, Issue 01, 2019

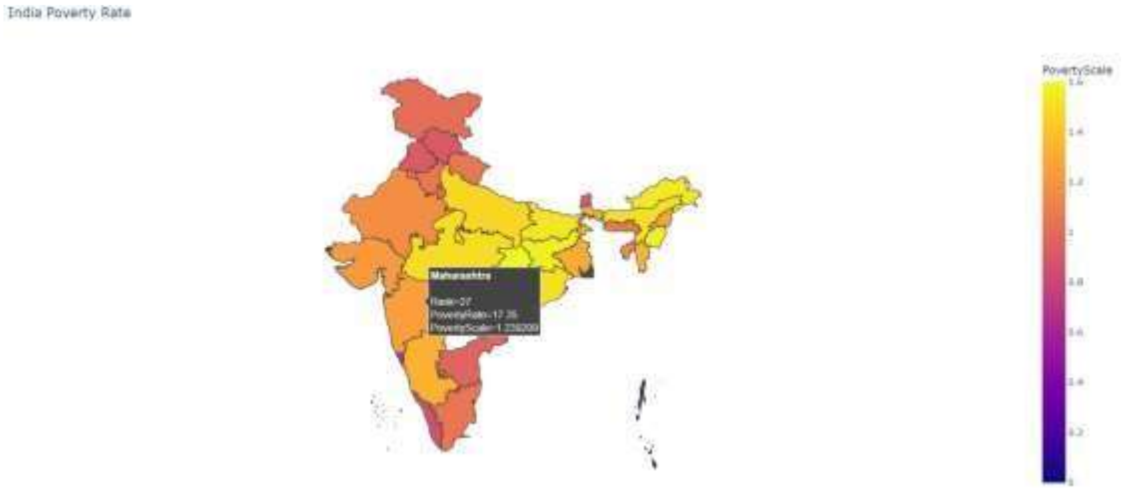

Figure. 7. Poverty Rate of Indian states

The above screenshot represents the predictions of Poverty rate of Indian states. The Poverty scale ranges from 0-1.6. When mouse hovers to particular state then Rank, Poverty Rate and Poverty Scale of that state will appear

Figure. 8. Poverty Rate of Indian states represented in world map

The above figure represents the prediction of population density for Indian maps which is plotted in world maps. Similarly, It also shows the density, density scale and id of each state. The density is obtained by dividing population with area. The density scale ranges from 0-4.

6. Conclusion

The total MOSAIKS manifesto, holding Linear prediction and Featurization, holds up likeness to few related paths. Especially, this may be elucidated as reckoning practicable estimation to kernel-ridge-regression for complete convolution between the image and the kernel. or, on the other hand, the two layered CNN with a phenomenal broad layer which is hidden having filters which are untrained. As MOSAIKS has been inspired from enabling theorize and adept SIML forecast. It will be fulfilled through nesting images that want to be descriptive (models which are trained on unity basis attain large skill on various labels) as well as efficient (skill accomplished by taking moderately less dimensions). The path of nesting depends on concept of random kitchen sinks, which is a process to generate features which qualify the Linear estimation of chosen tasks, which are in a manner. It is the use of polynomial features, distinct Fourier changes to estimate the functions basically, the 1-dimension functions. The user puts in these features on linear regression, to recognize the linear outputs on a basis, vector key to predict particular set of labels. By high dimension inputs, the satellite images we consider, shows tentatively that arbitrarily chosen subspace of basis often function as total basis for prediction problems. This project can also be extended for Predicting Poverty Rates and Elevation.

REFERENCES

[1] Panda, A., Singh, A., Kumar, K., Kumar, A., & Swetapadma, A. (2018, April). Land cover prediction from satellite imagery using machine learning techniques. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1403-1407). IEEE.

[2] Pandey, S. M., Agarwal, T., & Krishnan, N. C. (2018, April). Multi-task deep learning for predicting poverty from satellite images. In Thirty-Second AAAI Conference on Artificial Intelligence.

[3] Georganos, S., Grippa, T., Gadiaga, A., Vanhuysse, S., Kalogirou, S., Lennert, M., & Linard, C. An application of geographical random forests for population estimation in Dakar, Senegal using very-high-resolution satellite imagery. In Joint Urban Remote Sensing Event (JURSE) (pp. 1-4:2018). IEEE.

[4] Shahrin, F., Zahin, L., Rahman, R., Hossain, A. J., Kaf, A. H., & Azad, A. A. M. Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. 11th International Conference on Electrical and Computer Engineering (ICECE) (pp. 21- 24:2018). IEEE.

[5] Chaurasia, K., Neeraj, B., Burle, D., & Mishra, V. K. Topographical Feature Extraction Using Machine Learning Techniques from Sentinel-2A Imagery. In IGARSS IEEE International Geoscience and Remote Sensing Symposium (pp. 1659-1662:2018). IEEE.

[6] Das, M. Online prediction of derived remote sensing image time series: an autonomous machine learning approach. In IEEE International Geoscience and Remote Sensing Symposium (pp. 1496-1499). IEEE.

[7] Wu, T., Luo, J., Dong, W., Gao, L., Hu, X., Wu, Z., ... & Liu, J. Disaggregating county-level census data for population mapping using residential geo-objects with multisource geo-spatial data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1189-1205: 2018.

[8] Gebreegziabher, B. A. Mapping road pavement quality from optical satellite imagery using machine learning (Master's thesis, University of Twente) : 2017.

[9] Oughton, E. J., & Mathur, J. Predicting cell phone adoption metrics using machine learning and satellite imagery. Telematics and Informatics, 62, 101622: 2018.

[10] Jarry, R., Chaumont, M., Berti-Équille, L., & Subsol, G. Assessment of CNN-based Methods for Poverty Estimation from Satellite Images. In International Conference on Pattern Recognition (pp. 550-565). Springer, Cham-2018.

[11] Nischal, K. N., Radhakrishnan, R., Mehta, S., & Chandani, S. (2015, March). Correlating night-time satellite images with poverty and other census data of India and estimating future trends. In Proceedings of the Second ACM IKDD Conference on Data Sciences (pp. 75-79).

[12] Upadhyay, A., Upadhyay, A., & Maurya, S. (2017, September). Regression and artificial neural network based improved classification of LISS-III satellite image. In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) (pp. 917-921). IEEE.

[13] Luo, H., & Liu, Y. (2017, November). A prediction method based on improved ridge regression. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp. 596-599). IEEE.

[14] Pandey, P., Dewangan, K. K., & Dewangan, D. K. (2017, April). Enhancing the quality of satellite images by preprocessing and contrast enhancement. In 2017 international conference on communication and signal processing (ICCSP) (pp. 0056-0060). IEEE.

[15] Yuan, Z., & Tao, C. (2018, November). Estimation population density built on multilayer convolutional neural network. In 2018 5th International Conference on Systems and Informatics (ICSAI) (pp. 424-428). IEEE.

[16] Suhail, M., & Chand, S. Performance of some new ridge regression estimators. In 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS) (pp. 1-4). IEEE-2018.

[17] Lin, C. Y., & Lin, C. Using ridge regression method to reduce estimation uncertainty in chlorophyll models based on worldview multispectral data. In IEEE International Geoscience and Remote Sensing Symposium (pp. 1777-1780: 2018). IEEE.

[18] Li, T., Comer, M., & Zerubia, J. Feature extraction and tracking of CNN segmentations for improved road detection from satellite imagery. In 2018 IEEE International Conference on Image Processing (ICIP) (pp. 2641-2645). IEEE.

[19] La, Y., Bagan, H., & Takeuchi, W. Explore urban population distribution using nighttime lights, land-use/land-cover and population census data.In IGARSS IEEE International Geoscience and Remote Sensing Symposium (pp. 1554-1557: 2018).

[20] Jing, R., Gong, Z., & Guan, H. Land cover change detection with VHR satellite imagery based on multi-scale SLIC-CNN and SCAE features. IEEE Access, 8, 228070-228087 -2017.

[21] Ayush, K., Uzkent, B., Burke, M., Lobell, D., & Ermon, S. Generating interpretable poverty maps using object detection in satellite images. arXiv preprint arXiv:2002.01612- 2016..

[22] Ferreira, B., Iten, M., & Silva, R. G. Monitoring sustainable development by means of earth observation data and machine learning: a review. Environmental Sciences Europe, 32(1), 1-17 : 2018.

[23] Stratoulias, D., & Kabadayı, M. E. Land Cover Feature Extraction from Corona Spy Satellite Images during the Cold WAR-1968. In IEEE International Geoscience and Remote Sensing Symposium (pp. 6069-6072: 2019). Tsvetkovskaya, I. I., Tekutieva, N. V., Prokofeva, E. N., & Vostrikov, A. V. Methods of Obtaining Geospatial Data Using Satellite Communications and Their Processing Using Convolutional Neural Networks. In Moscow Workshop on Electronic and Networking Technologies (MWENT) (pp. 1-5). IEEE-2017.

[24] Jarry, R., Chaumont, M., Berti-Équille, L., & Subsol, G. Assessment of CNN-based Methods for Poverty Estimation from Satellite Images. In International Conference on Pattern Recognition (pp. 550-565: 2016). Springer, Cham.