International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

Prediction of Housing Price and Forest Cover Using Mosaics with Uncertain Satellite Imagery

[1]**Manam Pavan Sai**, Computer Science and Engineering VR Siddhartha Engineering College Vijayawada, India pavansaimanam@gmail.com
[2]**V. Amaleswara Rao**, Computer Science and Engineering VR Siddhartha Engineering College Vijayawada, India chintuamalesh@gmail.com
[3]**K.Suvarna Vani,** Computer Science and Engineering VR Siddhartha Engineering College Vijayawada, India suvarnavanik@gmail.com
[4]**P. Poul**, Computer Science and Engineering VR Siddhartha Engineering College Vijayawada, India poulpolimetla99@gmail.com

Abstract

The growing world is more expensive to estimate land use, road length, and forest cover using a plant-scaled ground monitoring system. Satellite imaging contains a significant amount of detailed uncertain information. Combining this with machine learning aids in the organization of these data and the estimation of each variable separately. The resources necessary to deploy Machine learning technologies for Remote sensing images, on the other hand, restrict their reach ability and application. Based on satellite observations which are notably underutilised in impoverished nations, while practical competence to implement SIML might be restricted. Encoded forms of images are shared across tasks, and they will be calculated and sent to an infinite number of researchers who can achieve top-tier SIML performance by training a regression analysis onto the actual data. By separating the duties, the proposed SIML solution, MOSAIKS, shapes SIML approachable and global. A Featurization stage turns remote sensing data into concise vector representations, and a regression step makes it possible to learn the correlations which are specific to its particular task which link the obtained characteristics to the set of uncertain data.

Keywords: Satellite imagery, machine learning, image encoding, CNN, MOSAIKS, Regression, SIML.

1. INTRODUCTION

To solve comprehensive universal concerns including controlling anthropogenic climate change, demographic changes, ecological change, or sustainable growth, many experts, decision-executives demand complete access to accurate, enormous observations for multiple factors simultaneously. ML is proven to be a powerful tool for converting huge quantity of disorganized imagery data into organised estimators. For example, the use of satellite imagery combined with machine learning (SIML) has allowed for more precise identification of tree cover, usage of land, poverty rates, and population density which is improving research and decision-making. A task is a term used to describe the process of predicting a single variable. The expanding need of estimate based on SIML is evidenced by huge count among private sector providers specialised in estimating either one or few projects. From the other side, the resources necessary to develop techniques of SIML constrain its application and to use. In impoverished nations, where practical ability to use SIML might be restricted although where these measures are probable to be helpful, satellite measurements are of little utility. Government agencies in impoverished communities, for example, can be curious in local river pollution, illicit land usage, or mass migration. SIML, on the other hand, is pretty much entirely out of reach for this and many other user groups, as current approaches necessitate a large-scale, organisation with substantial resources associated with the combination of domain-specific expertise, spatial analysis and engineering expert knowledge, image access, intricate architecture customization, and enormous machine learning computational resources.

Multiple researchers and decision-makers (hence referred to as "users") need access to reliable large-scale observations of many variables simultaneously when tackling global challenges such as climate change, population migrations, ecological modifications or economic development. With over 700 satellites in orbit, satellite photography is a viable alternative to large-scale ground-based surveillance systems for this purpose. These enormous volumes of unstructured photo data are also being translated into organised ground condition evaluations using machine learning. Using satellite imaging and machine learning, for example, we've been able to increase forest coverage while decreasing poverty and population density (SIML). A single variable prediction is what we mean by a "task."

As seen by the growing number of private service providers that specialise in one or more of these professions, SIML-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

based estimations are becoming more essential. Even yet, SIML's resource requirements make it difficult for the general public to utilise. While SIML may be more difficult to implement in low-income nations, satellite-based metrics are nevertheless underutilised despite their potential benefit. Some governments may be interested in learning more about river pollution or illegal land use in their region. As things stand now, the only way that SIML can be made available to these and other potential users of the system is through an extremely resource-intensive enterprise that combines task-specific domain knowledge with remote sensing and engineering expertise, access to imagery, customization and tuning of sophisticated machine learning architectures, and a large amount of computing power.

A new method to SIML that doesn't need the employment of specialised computer resources or the creation of a sophisticated prediction mechanism is needed to simultaneously remove these constraints. Encoding techniques that convert each satellite picture into a vector of variables (thus, features) might make this method practicable by minimising the need for users to change expensive photographs. SIML issues may benefit from an unsupervised encoding method, especially when compared to deep-learning SIML algorithms that utilise techniques initially designed for natural pictures. Complex methods are needed to deal with the lack of consistency in many essential elements in natural photography, such as topic or camera viewpoint, which may be superfluous when using satellite images to learn.

Attempts to encode satellite pictures employing unsupervised techniques have been made in the past, but no one collection of characteristics has been shown to compete with DL systems across a variability of applications and scale internationally.

Using a desktop computer's linear regression to predict ground conditions over several tasks may provide a lot of information in remote sensing satellite images. It is our goal to utilise just high-resolution daylight satellite images to estimate the features of tiny areas (such as the average home price) for a particular time period. Google Static Maps API embedding's may be generalised across jobs because of their ubiquitous availability at fine resolution, geo-rectification, and pre-processing to reduce cloud opacity. In addition to SIML applications, more data sources might theoretically be used. We take an entirely new strategy to dealing with the particular challenges and opportunities of SIML applications, one that is simple but very effective. As opposed to leading deep neural networks, satellite pictures, shot from a fixed distance, viewing angle, record repeating patterns, objects, allowing us to make significant computing gains in model training and testing. In addition, researchers have routinely used the same photos to answer a variety of unrelated problems. Using our method, researchers may access a centralised collection of characteristics that can be used to solve a variety of different problems. Aside from saving time and money now, customers won't have to go through the tedious process of creating their own images and editing them in the future. As the number of SIML users and the volume of global imaging data grows, so does the scale of the benefits that come with it. Differentiation in SIML applications from existing designs. As opposed to leading deep neural networks, satellite pictures, shot from a fixed distance, viewing angle, record repeating patterns, objects, allowing us to make significant computing gains in model training and testing. In addition, researchers have routinely used the same photos to answer a variety of unrelated problems. Using our method, researchers may access a centralised collection of characteristics that can be used to solve a variety of different problems. Aside from saving time and money now, customers won't have to go through the tedious process of creating their own images and editing them in the future. As the number of SIML users grows and the volume of global imaging data rises at a rate of more than 80 terabytes per day, the magnitude of the advantages that will be gained will grow as well.

2. LITERATURE SURVEY

In this proposed research prediction of land usage from satellite data is done using ML approaches. Time-series normalised difference vegetation index is used to collect input features from satellite images (NDVI). The work was done entirely in Python, and the KNN method was used to achieve the highest level of accuracy. In this research, a mixture Habiganj monitoring and mapping of agriculture, as well as crop growth and production prediction, is presented. Landsat-8 photos of Habiganj with multi-spectral bands have been prepared, and satellite image indicators linked to agricultural yield and production have been retrieved. Habiganj's crop yield is projected using existing parameters, and the datasets of future values are forecasted using 2 types of time-series data analysis models for improved accuracy (ARIMA and LSTM).In this paper, classifier like SVM, Decision Tree, RF, NB, an effort was made to better detect land cover categories from Sentinel2A data (CART) are used. Performance measures computed in the paper also validate the findings obtained by the used models. According to the findings, the random forest classifier surpasses different classification techniques with a 95.67 percent accuracy.

This proposed system, The Geographical Random Forest (GRF) is indeed a localized Random Forest execution (RF) to forecast density of population using Remote Sensing with Extremely High Resolution such as VHHRS data. As a result,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

the GRF technique is proposed as a viable fact-finding and illustrative technique for modelling spatially heterogeneous remotely sensed relationships. This study suggests a two-step approach for using satellite images to anticipate poverty in India's rural areas. To extract images for the villages from the determined geocodes, we used the Google Static Maps API. Training a multi-task fully convolutional model first, followed by training the network to predict income levels. Residential geo-objects are used as fundamental levelling units in this research, and the problem is formalised as geographical forecasting model applying HSR satellite-based imagery and multi-source geo-spatial data with algorithms of Machine-learning (ML).

The Land cover analysis using fundamental pixel-based features extracted from much more sophisticated Ultra Spectral imagery is the subject of this study. Second, for pixel-based land cover analysis, an exploration of parametric and non-parametric machine learning techniques. They employed SPOT-5 aerial photographs with a range of nearly 2.64m for an experimental investigation. We choose Maximum Likelihood Estimator (MLE), Support Vector Machine and Neural Networks from the techniques of machine learning collection (ANN). Higher performance of these algorithms in pattern recognition tasks led to their selection. Scant Flora, Sugar Beet, Urban Areas, Water, Roads, Tobacco and Rough Terrain are seven types in which the feature space is divided. Using satellite photos and machine learning, anticipate air heavy metal contamination. Satellite pictures from the Google Earth Engine platform and sampling data from the UNECE International Cooperative Program (ICP) Vegetation Data Management System were used to train the model. The pollution and satellite pictures were correlated using the KNN technique for data modelling. To obtain some indexes, researchers collect and evaluate samples. A sampling is rarely done for objective purposes, and the size of the sampling grid might be quite large. Modelling may be a good option in this circumstance. Our plan is to train a specific statistical model using real-world data on heavy metal concentrations and indexes obtained from satellite photos. In this proposed work, to identify modifications in tropical forests during the 29-year period (1987–2015) images from remote sensing satellites are used. To reclaim multispectral data that has been lost, they first suggest a spatiotemporal inpainting mechanism because the original data is badly inadequate and cluttered with artefacts. The spatial filling procedure uses data from surrounding transient instances, after that, sparse encoding-based restoration. The goal of change identification is formulated like regional classification task. An region with several resolutions is mapped out to provide a collection of candidate bounding box suggestions that may include potential change zones.

Panda, A., et al [2018] In this study, different machine learning approaches, such as the closest neighbour algorithm, DT, SVM, RF, & NB classifier, were employed to estimate land cover from satellite images. Time-series normalised difference vegetation index is used to gather the input characteristics from satellite images (NDVI). An impenetrable forest is the only one of the six class classes to provide an impermeable output. Synthetic Minority Oversampling (SMOTE) was utilised to ensure that the data in each class was evenly distributed. Python has been used for all of the work. k-NN is the most accurate method.[1]

Shahrin, F., et al [2020] Most of Bangladesh's people work in agriculture, which accounts for a large portion of the country's total employment. However, agricultural yields are unreliable and farming infrastructure is inefficient, which has a negative impact on food security. Because of its unusual geography, Habiganj was chosen as a research region for its sensitivity to flooding and drought. In this study, agricultural mapping & surveillance in Habiganj is paired with crop growth and output forecasting. Remote sensing indicators relating to crop growth and yield are calculated using Landsat-8 multispectral images of Habiganj. You may use K-means or Mask R-CNN methods in Python or Mat lab to estimate crop growth. Two ML approaches and two time series analysis models are then used to forecast the agricultural output of Habiganj, as well as its dataset's future values. Analytical analyses and predictions may help monitor agricultural growth dynamics and detect early symptoms of crop yield decline. To evaluate which environment and model is most suited for this study, two platforms, algorithms, and time series analysis are compared.[2]

Chaurasia, K., et al [2020] Satellite technology has advanced to the point that it is now feasible to collect satellite photographs of almost any corner of the planet with ease and frequency. Data from satellites offers a wealth of information that may be used to a number of beneficial ends. Land cover identification in a region by hand may be a time-consuming and demanding process, though. Sentinel-2A imagery has been used in this publication to better classify landcover types based on common classifiers such as random forest and SVM (CART). The performance measures of the models employed in the paper are also utilised to validate the findings. As a result of this study, the random forest classifier surpasses all other classification algorithms with a 95.67 percent success rate. Large datasets may be processed automatically, decreasing the need for labour-intensive human tagging.[3]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

Georganos, S., et al [2019] Geo-graphical RF is a local version of RF that may be used to forecast population density using VHHRS data. The 2013 Dakar census population density is used as an independent variable, while the fractions of the 3 distinct built-up kinds in each neighbourhood are used as explanatory features. By calibrating GRF to a suitable geographic scale, we may increase the accuracy of our estimations by including geographical variability into the data. It is also possible to plot the GRF findings, showing the performance of local sub-models and other noteworthy geographical changes. In the end, GRF may be used to investigate and explain distant sensing interactions that are geographically diverse.[4]

Pandey, S. M., et al [2018] Economic activity may be studied using high-resolution daylight satellite photography. These photos provide a comprehensive view of a wide region while also providing a close-up view of certain communities. Current approaches, on the other hand, only use photos on a single geographic level. Deep learning is used to aggregate qualities observed at many geographic levels to forecast economic indicators. Ordinal regression is used to estimate the hyperlocal economy in tiny towns. As a second stage, the interconnectedness of hyperlocal economies is summarised to derive district-level properties By using hyperlocal and district-level data, the model calculates district-level economic indicators. When it comes to forecasting important metrics like population, buying power, and energy consumption, our novel multi-level learning model beats established strong baselines. For example, when learned features from one nation are tested on data from Malaysia, Thailand and Vietnam, the model is able to generalise to other countries. Multi-level models have consequences for assessing inequality, which is a critical first step in policy and social science studies on poverty and inequality.[5]

Wu, T., et al [2020] For future scientific decision-making, reliable specialization of socioeconomic data is critical in understanding the distribution of human social development status over time and place. Traditional specialization of social economy macroeconomic data is used to focus this study on population mapping. When it comes to population forecasting, typical grid-based mapping or administrative divisional mapping, such as townships, might fall short on both pattern and accuracy. Machine-learning (ML) methods and high spatial resolution satellite remote sensing photos and multisource geospatial data are utilised to formalise the process of mapping this region utilising residential geo-objects. Analysis of population density factors such as residential geoobject area, building existence index, terrain slope, night-light intensity, POI density and road network density from Internet electronic maps and location factors such as road and river distances can be done using Random Forests or other ML algorithms. These include: As a result of the nonlinear regression relationship that was discovered, weights of disaggregation for each unit were calculated, and this information was then utilised to generate a population density distribution at the size of individual home geo-objects. It has been shown in experiments conducted across a county area that the approach is capable of creating more accurate and fine-grained spatial patterns of population distribution than the old deterministic methods. As a consequence, the methodological framework may be proposed for expansion to additional spatialization sectors of socioeconomic data in order to further profit from the optimization of mapping findings.[6]

Ali, K., et al [2016] Categorizing remote sensing images has long been a focus for those in the field, since classification findings provide the foundation for a wide range of environmental and socially beneficial uses. In the classification process, selecting an effective image classification technique is a key component. The study examines remote sensing picture classification supervised learning algorithms. This study's major goal is to classify (LC) Land Cover and usage. Machine learning techniques, such as supervised learning, are used in this project. It's examined how remote sensing images are categorised using pixel-based supervised classification algorithms. The data was labelled with labels in order to perform the investigation. Pixel-based supervised classification was shown to work best with a support vector machine out of the five methods tested (i.e. maximum likelihood estimation, minimum distance classifier, principal component analysis, is clustering and support vector machine). [7]

Uzhinskiy, A., et al [2018] Air pollution is a major problem in Europe and Asia. 92.00% of the world's population, according to the WHO, lives in places with pollution levels that are too high to be considered healthy. Regional and worldwide efforts to protect the environment are many. In the end, they are all trying to figure out what the present state of affairs is and how it is changing. Typically, researchers gather samples and analyse them in order to come up with certain indices. It's very uncommon for a sample grid to be exceedingly large, due to natural factors. Modelling might be a good option in this circumstance.

Using satellite pictures and machine learning, we have attempted to forecast the concentration of heavy metals in the atmosphere. Satellite pictures from Google Earth Engine and sample data from the UNECE International Cooperative Program (ICP) Vegetation Data Management System were used to train the model. Sb for Norway and Mn for Serbia were correctly predicted by our model.[8]

Khan, S. H., et al [2017] there are a number of reasons why monitoring land cover change is critical to regional resource management and catastrophe response. Images taken by satellites over a period of 29 years are analysed in this article to determine how forest cover has changed over time (1987-2015). In order to recover the lost surface reflectance information from the original data, we first construct a spatiotemporal inpainting process. A sparse encoding-based reconstruction is used to fill in the gaps in the temporal data that are present in the surrounding temporal occurrences. Change detection is approached as an issue of area categorization. It is possible to obtain a list of prospective change regions using the target area's multiresolution profile (MRP). We use a deep neural network to learn area representations automatically, rather than depending on handcrafted attributes. By labelling the candidate set of proposals, we can detect forest changes, predict their start and end timeframes. We have a 91.6 percent average patch classification rate and an onset/offset prediction error on average of 4.9 months using our method, which would be a 16 percent improvement and a five-month decrease in error over a strong baseline. Our findings show that the suggested forest change detection method may be applied to new areas by analysing the identified changes in unlabeled images.[9]

Nischal, K. N., et al [2015] there is a correlation between light intensity in photos and state-level poverty, population, gross domestic product (GDP), and forest cover in this article based on night-time satellite photograph of India & census data. Using the predictive model-based approach, multivariate regression analysis, and the ARIMA model, we are able to fill in the gaps left by the lack of data. Census data accuracy may be verified and predictions made using ARIMA and regression models. Poverty data may be obtained economically and quickly, allowing policymakers, relief organisations, and others to make better decisions.[10]

Upadhyay, A., et al [2017] this refers to the physical land types that are present in a certain location. These include flora, aquatic bodies, urban areas, roadways, and more. There is an urgent need to assess whether or not the effect of recent environmental and land-use planning reforms is on the correct track, given the fast shifts in land-use patterns. This is why remote sensing technology may play a critical role in assessing the current landscape by analysing Land Use and Land Cover (L ULC) categorization maps based on satellite photos of the area. Using classification maps created by ML categorization algorithms on a specific geographic area around Nirma University, Ahmadabad, India, this research provides a unique way for change detection analysis. [11]

Luo, H., et al [2017] 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. Data are modelled, and a fire's likely path is predicted using an approach called ridge regression. To begin with, certain variables are selected and eliminated that have unstable or tiny absolute values in standardised ridge regression coefficients or stable coefficients. For the Support Vector Machine model, the remaining characteristics will be used as input values in a new dataset. Classification results may be produced from the new dataset once it has been divided into two sets: a training set and a test set. 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.[12]

Pandey, P., et al [2017] Conversion functions are able to take use of an image's universal data content, which increases the dynamic range of intensity levels that may be utilised. Several conversion techniques use locally important material to affect image quality and limits. For example, a plotting tool that uses both universal and local conversion algorithms may successfully boost contrast while keeping the intensity and fine features of an image. All of the image's contrast and intensity settings are preserved throughout the universal conversion process. Satellite images may be pre-processed using a series of independent processes that convert noise, boost contrast, and improve the quality of the picture while also decreasing the impacts of blur and noise. Once each step has been compared, PSNR parameters are utilised to evaluate the results.[13]

Yuan, Z., et al [2018] in intelligent video surveillance, population density estimate is an important research field. It is difficult to keep up with today's huge data since traditional approaches need the construction of features manually. Deep learning, a type of artificial intelligence, is on the rise, and its use in video surveillance is becoming more and more common. Therefore, a multilayer convolutional neural network (MCNN) is proposed in response to the disadvantages of conventional manual feature extraction and the deficiencies 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. For example, using an adaptive kernel, we may properly determine population density even when we do not know the input map's viewpoint. In order to create a population density map, the graphs of each layer are

ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

combined. Research shows that this network topology is more accurate in estimating population sizes.[14]

The authors of [15]-[17] have analysed, configured, and employed the most recent variant of CNN termed as Capsule Network for leaf retrieval. The authors in [18] presented a CNN for Devanagari handwritten text recognition.

Kibria, B. G. [2003] The estimate of the ridge parameter k is a significant issue in ridge regression analysis. It is possible to estimate this parameter in a variety of ways. The generalised ridge regression strategy has been used to develop various novel estimators, which have been discussed in this article simulation research was carried out to estimate the performance of recommended estimators according to the mean squares estimators (MSE) criterion. According to results of the simulation research, the suggested estimators outperform least squares estimators (LSE) and other widely used current estimators under specific situations. It was also possible to verify the simulation results using a real-world numerical example.[19]

3. PROPOSED WORK

MOSAIKS' design lends itself to two additional characteristics: the capacity to fuse combines information from various sensors with photo forecasts and assign sub-image scale locations to image-scale predictions [20] [21]. SIML predictions can be improved with the help of available satellites that have a variety of characteristics (for example, sample timing and wavelength). As the step of regression in the attributes is linear, MOSAIKS' design enables for incorporation of information seamlessly from various satellites [22][23].

A *Featurization Step*

This phase is used to convert uncertain satellite imagery (images x) into concise vector representations [24][25]. The way images are represented as characteristics determine how generalizable they are. The featurization function is based on MOSAIKS, which we utilise to construct RCFs from satellite images and is theoretically supported [26][27]. RCFs quantify the measure of correlation among each sub-image in each and every set of pictures without utilising task-specific information or contextual information [28][29]. As shown in the diagram, The properties of x are then used by MOSAIKS as an overfitted means of determining any y, that might be a image objects non-linear function [30][31].

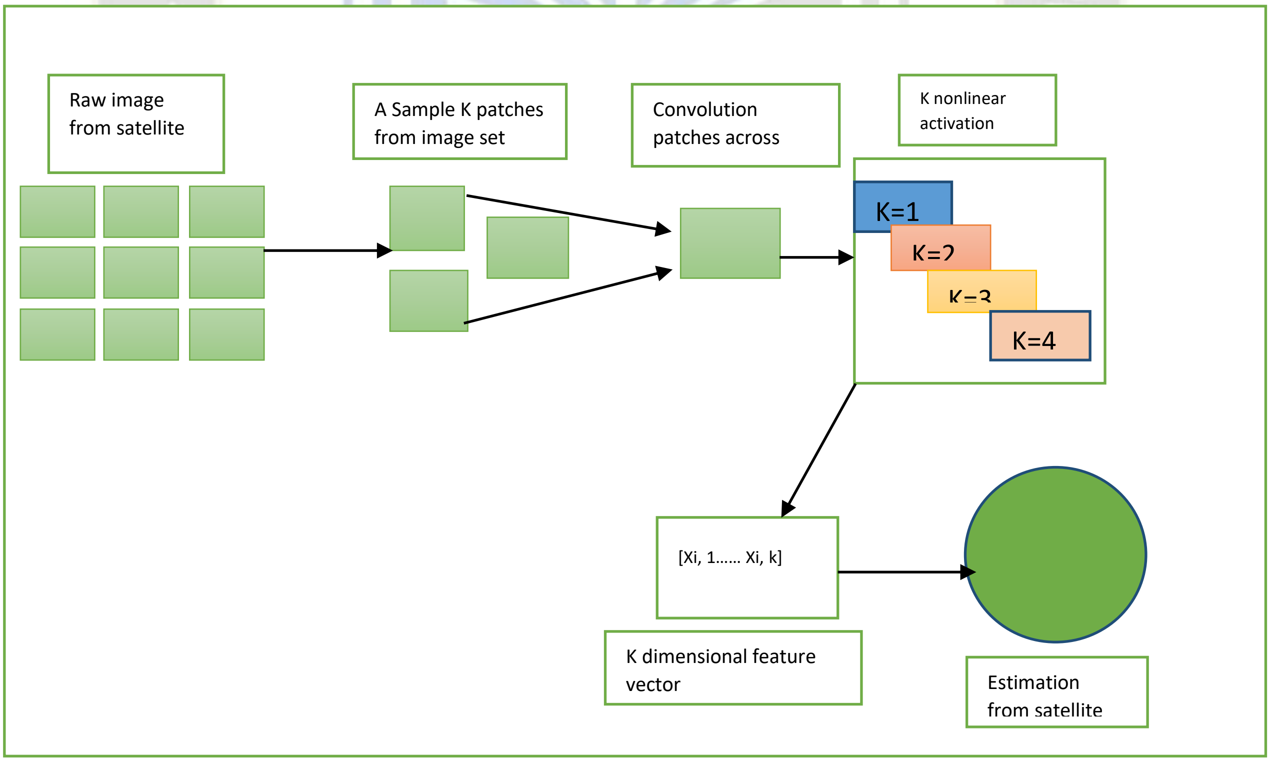

Figure 1: Process flow diagram

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

The above figure 1 represents about the process of featurization.

B *Regression Step*

The unsupervised featurization stage can be performed once for each image, resulting in a single collection of results which can be utilized to resolve a variety of problems by numerous independent users applying the regression step repeatedly. Ridge Regression was employed in this case. It is of the form Y=XB+e in this Ridge Regression. For 1 to q locations, Y is the user-supplied labels, such as population density, poverty, and elevation. The K-dimensional feature vector is denoted by X [32]. Then, for B, each user runs a single linear regression. The whole MOSAIKS feature set as well as linear prediction with e. After that, X generates the for all locations, SIML forecasts regarding label elements [33][34].

C *Multi Task Performance of Mosaics*

We begin by futurizing the photos by running them via MOSAIKS' feature extraction method, which produces 8,192 features for each image. We then repeat to perform a cross-validated ridge regression for every job to forecast forest cover ($R2 = 0.91$) , road length ($R2 = 00.54$), & average house price ($R2 = 00.53$) using generated feature matrices (X) in the regression procedure [31]-[33]. The figure represents the prediction of Road Length in the US continent with 1km x 1km resolution daytime images.

Algorithm:

```
Featurization:
Gridcreation();
Featurization();
Regression();
Gridcreation()
{
    latmin = 25
    latmax = 50
    lonmin = -125
    lonmax = -66
   gridvals <- makegrid(zoom, pixels, lonmin, lonmax, latmin, latmax)
   latVals <- gridvals[[2]]
   lonVals <- gridvals[[1]]
  save(file.path(data_dir, "int/grids", paste0(filename, ".npz")), lon = lonVals, lat = latVals, zoom = zoom, pixels = pixels)
}
Featurization()
{
subgrid_files = Path(c.grid_dir).glob("[!grid_]*.npz")
area = grid_name_lst[0]
sample = grid_name_lst[3]
image_folder = base_image_dir / f"{area}_{sample}"
outfpath = Path(c.features_dir) / f"{image_folder.name}.pkl"
featurize_and_save(image_folder, out_fpath, c);
}

featurize_and_save(image_folder, out_fpath, c)
{
X_lift, names, net = featurize(image_folder, c)
   latlon = np.array([i.split("_")[:2] for i in names], dtype=np.float64)
   lon = latlon[:, 1]
   lat = latlon[:, 0]
   zoom_level, n_pixels = [int(i) for i in names[0].split("_")[2:4]]
   ij = spatial.ll_to_ij(
       lon,
       lat,
       c.grid_dir,
       c.grid["area"],
       zoom_level,
       n_pixels,
   )
   ij = ij.astype(str)
   ids = np.char.add(np.char.add(ij[:, 0], ","), ij[:, 1])
}
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"],
```

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

```
        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)
}
```

4. IMPLEMENTATION

The below figures 2 shows the bar plots of the labels forest cover in each states of India and across each districts of Andhra Pradesh. Where X-axis represents states and Y-axis represents forest cover Area.

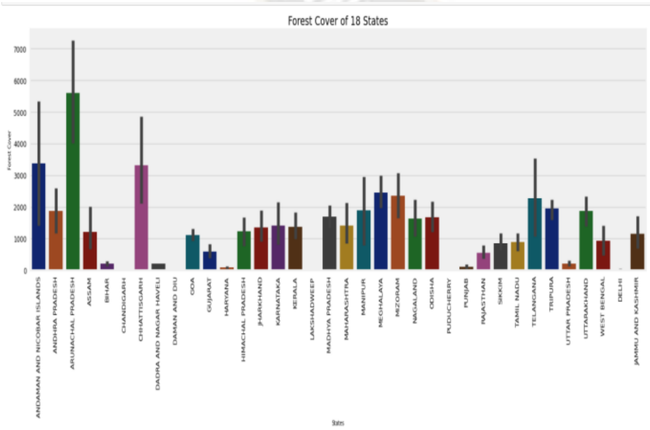

Figure 2 Representation of Forest Cover in different states across India.

The lowest forest cover is in the state of Haryana followed by Punjab while the highest forest cover is in the state of Arunachal Pradesh. The above data is taken from 2015 forest cover dataset.

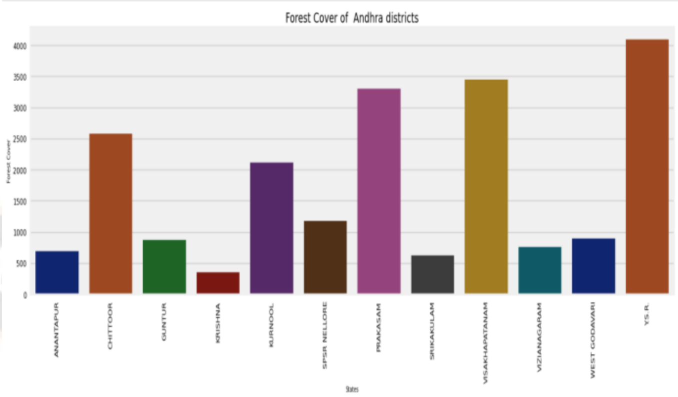

Figure 3: Representing Forest Cover in different districts of Andhra Pradesh

The lowest forest cover is in the district of Krishna while highest forest cover is in the district of Y.S.R Kadapa show in figure 3.

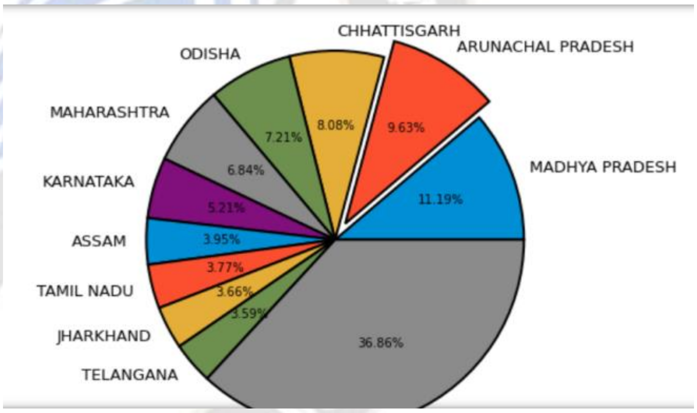

Figure 4: Representing Total share of top 10 states in India

The highest forest cover share is contributes by Madhya Pradesh followed by Arunachal Pradesh show in figure 4.

The below figure 5 represents the scatter plot of labels Housing price in each states of India.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

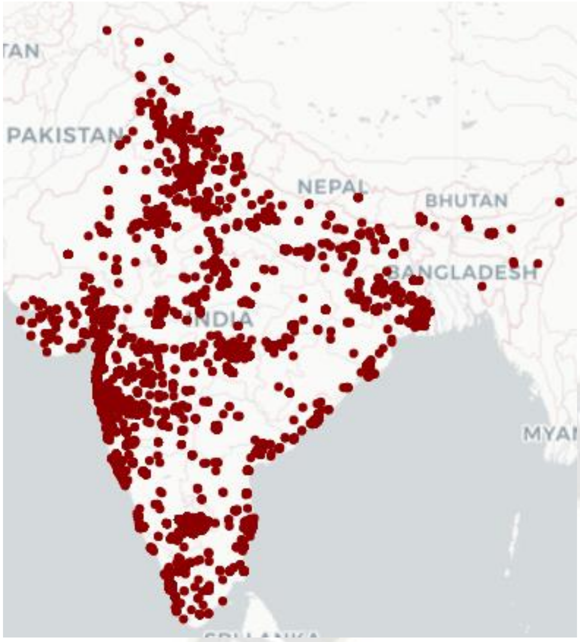

Figure 5: Representing Housing Price in different states in India

5. ADVANTAGES

These particular systems have following advantages:

i. The photographing of the ground surface is a constant operation. Having a duration of 4 days.
ii. As a result, the most appropriate image was selected.
iii. The formalities of aerial photography and flight planning are omitted in this case.
iv. The cost of using satellite images is far less than that of using aerial photographs.

6. CONCLUSION

The MOSAIKS platform as a whole, which includes linear prediction and featurization. It can be thought as a double layered Convolutional Neural Network with a massive private intermediate layer created with filters which are not trained, or a computationally practical kernel ridge regression approximation for a fully convolutional network. A Featurization stage turns remote sensing uncertain data into concise vector representations, and a regression step makes it possible to learn the correlations which are specific to its particular task which link the obtained characteristics to the set of uncertain data.

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] Shahrin, F., Zahin, L., Rahman, R., Hossain, A. J., Kaf, A. H., & Azad, A. A. M. (2020, December). Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. In 2020 11th International Conference on Electrical and Computer Engineering (ICECE) (pp. 21-24). IEEE.

[3] Chaurasia, K., Neeraj, B., Burle, D., & Mishra, V. K. (2020). Topographical Feature Extraction Using Machine Learning Techniques from Sentinel-2A Imagery. In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium (pp. 1659-1662). IEEE.

[4] Georganos, S., Grippa, T., Gadiaga, A., Vanhuysse, S., Kalogirou, S., Lennert, M., & Linard, C. (2019, May). An application of geographical random forests for population estimation in Dakar, Senegal using very-high-resolution satellite imagery. In 2019 Joint Urban Remote Sensing Event (JURSE) (pp. 1-4). IEEE.

[5] 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.

[6] Monir, M. B., & Mohamed, A. A. (2022). ENERGY AWARE ROUTING FOR WIRELESS SENSOR NETWORKS. International Journal of Communication Networks and Information Security (IJCNIS), 14(1).

[7] Wu, T., Luo, J., Dong, W., Gao, L., Hu, X., Wu, Z., ... & Liu, J. (2020). 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.

[8] Ali, K., Rahman, H. U., Khan, R., Siddiqui, M. Y., Najeeb, F., Amin, S., ... & Chaudhry, N. J. (2016). Land usage analysis: a machine learning approach. International Journal of Computer Applications, 141(12), 23-28.

[9] Uzhinskiy, A., Ososkov, G., Goncharov, P., & Frontsyeva, M. (2018). Combining satellite imagery and machine learning to predict atmospheric heavy metal contamination. Advisory committee, 145.

[10] Khan, S. H., He, X., Porikli, F., & Bennamoun, M. (2017). Forest change detection in incomplete satellite images with deep neural networks. IEEE Transactions on Geoscience and Remote Sensing, 55(9), 5407-5423.

[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.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 8
DOI: https://doi.org/10.17762/ijritcc.v10i8.5666
***Article Received**: 10 June 2022 **Revised**: 15 July 2022 **Accepted**: 30 July 2022 **Publication**: 31 August 2022*

[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] Sandeep Pande and Manna Sheela Rani Chetty, "Analysis of Capsule Network (Capsnet) Architectures and Applications", Journal of Advanced Research in Dynamical and Control Systems, Vol. 10, No. 10, pp. 2765-2771, 2018.

[17] Sandeep Pande and Manna Sheela Rani Chetty, "Bezier Curve Based Medicinal Leaf Classification using Capsule Network", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 6, pp. 2735-2742, 2019.

[18] Sally Fouad Shady. (2021). Approaches to Teaching a Biomaterials Laboratory Course Online. Journal of Online Engineering Education, 12(1), 01–05. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/43

[19] Pande S.D., Chetty M.S.R. (2021) Fast Medicinal Leaf Retrieval Using CapsNet. In: Bhattacharyya S., Nayak J., Prakash K.B., Naik B., Abraham A. (eds) International Conference on Intelligent and Smart Computing in Data Analytics. Advances in Intelligent Systems and Computing, vol 1312.

[20] Sandeep Dwarkanath Pande, Pramod Pandurang Jadhav, Rahul Joshi, Amol Dattatray Sawant, Vaibhav Muddebihalkar, Suresh Rathod, Madhuri Navnath Gurav, Soumitra Das, "Digitization of handwritten Devanagari text using CNN transfer learning – A better customer service support", Neuroscience Informatics, Volume 2, Issue 3, 2022.

[21] Kibria, B. G. (2003). Performance of some new ridge regression estimators. Communications in Statistics-Simulation and Computation, 32(2), 419-435.

[22] Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064

[23] Lin, C. Y., & Lin, C. (2019, July). Using ridge regression method to reduce estimation uncertainty in chlorophyll models based on worldview multispectral data. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 1777-1780). IEEE.

[24] Chandrakar, I. (2022). Improved Technique for Preserving Privacy while Mining Real Time Big Data. International Journal of Communication Networks and Information Security (IJCNIS), 14(1).

[25] Li, T., Comer, M., & Zerubia, J. (2019, September). Feature extraction and tracking of CNN segmentations for improved road detection from satellite imagery. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 2641-2645). IEEE.

[26] Fakharian, E., Nabovati, E., Farzandipour, M., Akbari, H., & Saeedi, S. (2021). Diagnosis of Mechanical Low Back PainUsing a Fuzzy Logic-Based Approach. International Journal of Intelligent Systems and Applications in Engineering, 9(3), 116–120. https://doi.org/10.18201/ijisae.2021.239

[27] La, Y., Bagan, H., & Takeuchi, W. (2019, July). Explore urban population distribution using nighttime lights, land-use/land-cover and population census data. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 1554-1557). IEEE.

[28] Jing, R., Gong, Z., & Guan, H. (2020). Land cover change detection with VHR satellite imagery based on multi-scale SLIC-CNN and SCAE features. IEEE Access, 8, 228070-228087.

[29] Ayush, K., Uzkent, B., Burke, M., Lobell, D., & Ermon, S. (2020). Generating interpretable poverty maps using object detection in satellite images. arXiv preprint arXiv:2002.01612.

[30] Ferreira, B., Iten, M., & Silva, R. G. (2020). Monitoring sustainable development by means of earth observation data and machine learning: a review. Environmental Sciences Europe, 32(1), 1-17.

[31] Stratoulias, D., & Kabadayı, M. E. (2020). Land Cover Feature Extraction from Corona Spy Satellite Images during the Cold WAR-1968. In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium (pp. 6069-6072). IEEE.

[32] Tsvetkovskaya, I. I., Tekutieva, N. V., Prokofeva, E. N., & Vostrikov, A. V. (2020, March). Methods of Obtaining Geospatial Data Using Satellite Communications and Their Processing Using Convolutional Neural Networks. In 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT) (pp. 1-5). IEEE.

[33] Jarry, R., Chaumont, M., Berti-Équille, L., & Subsol, G. (2021, January). Assessment of CNN-based Methods for Poverty Estimation from Satellite Images. In International Conference on Pattern Recognition (pp. 550-565). Springer, Cham.

[34] Garigipati, R. K., Raghu, K., & Saikumar, K. (2022). Detection and Identification of Employee Attrition Using a Machine Learning Algorithm. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 120-131). IGI Global.

[35] Mythreya, S., Murthy, A. S. D., Saikumar, K., & Rajesh, V. (2022). Prediction and Prevention of Malicious URL Using ML and LR Techniques for Network Security: Machine Learning. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 302-315). IGI Global.

[36] Saikumar, K., Rajesh, V., Babu, B.S. (2022). Heart disease detection based on feature fusion technique with augmented classification using deep learning technology. Traitement du Signal, Vol. 39, No. 1, pp. 31-42. https://doi.org/10.18280/ts.390104

[37] Kailasam, S., Achanta, S.D.M., Rama Koteswara Rao, P., Vatambeti, R., Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. International Journal of Intelligent Computing and Cybernetics, 15(2), pp. 184–197

[38] Saikumar, K. (2020). RajeshV. Coronary blockage of artery for Heart diagnosis with DT Artificial Intelligence Algorithm. Int J Res Pharma Sci, 11(1), 471-479.

[39] Raju K., Chinna Rao B., Saikumar K., Lakshman Pratap N. (2022) An Optimal Hybrid Solution to Local and Global Facial Recognition Through Machine Learning. In: Kumar P., Obaid A.J., Cengiz K., Khanna A., Balas V.E. (eds) A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems. Intelligent Systems Reference Library, vol 210. Springer, Cham. https://doi.org/10.1007/978-3-030-76653-5_11

[40] Uçak, K. (2021). Runge-Kutta MIMO NARMA-L2 Controller. International Journal of Intelligent Systems and Applications in Engineering, 9(2), 38–47. https://doi.org/10.18201/ijisae.2021.232