LAND POP ANALYZER

Yawar Ali
Department of Computer Science and Engineering(AIML) RD Engineering College info@rdec.in

***Abstract* - The issue of population density in a particular area has been delved into by this paper, and various factors that contribute to the phenomenon are examined.The research finds that population density has a significant impact on land use patterns, with areas of high population density experiencing more intense land use and greater pressure on natural resources. A comprehensive solution that employs statistical and geographic tools, such as SVM (Support Vector Machine) for land classification, Edge Detection with the Laplacian of Gaussian filter for calculating unused land area, and Linear Regression for predicting population density is proposed by the paper to tackle this problem. Additionally, QGIS (Quantum Geographic Information System) and its related plugins like HCMGIS and SCP are utilized by the paper to export graphical maps and satellite images for further analysis.**

KEYWORDS - Population density, SVM, Edge detection, Prewitt method, Mul- tiple linear regression, Land classific- ation.

I. INTRODUCTION:

India's population density has been steadily rising in the last 100 years. It has been changed from 234.4 in 1980 to 431.11 in 2022 [1]. The predicted population density by 2040 is expected to be 540 per km^2. The majority of the population in India, roughly 70% of the total, is seen to live[2]. The population density in states like Maharashtra is higher than the national average. State populations are less dense in places like the seven sisters of India in comparison to the average population density across the country. Currently, Uttar Pradesh has the largest population of 23.32 Cr, followed by Maharashtra (12.54 Cr), Bihar (12.49 Cr), West Bengal (9.86 Cr), and Madhya Pradesh including Tamil Nadu (7.66 Cr) according to various newspaper sources. Approximately half of the nation's population is seen to reside in these five states.

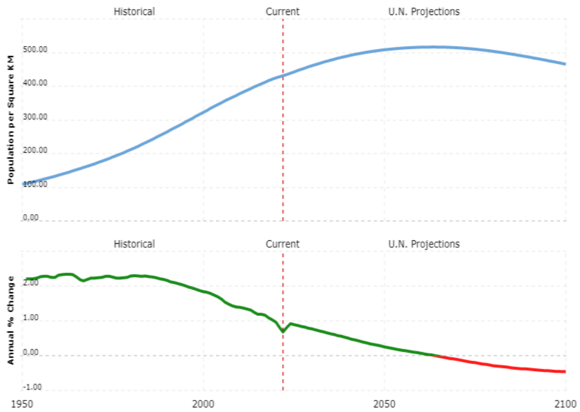

Fig. 1. **Population trend from 1950 to 2050**

A lot of land in India is unused, whether it has been acquired by governments for establishing large and medium-sized public sector industries or purchased by citizens as an investment but not used. This leads to issues such as resource waste and unrealized potential of fundamental community resources, as well as misuse of land. Other issues include frozen ground, vacant private property, abandoned industrial wastelands, and purposefully uncultivated farmland.

ISSUES FACED:

I. Wastage of resources: Many perfectly useful items may be wasted and products made with an inherent antiquatedness in the culture where an exceptional quality of life entails the consumption of new commodities very hastily. Excessive production of factory waste, scrap, and refuse contributes to pollution, and industries. Machinery replaces labour to make production go swift resulting in the wastage of human resources.

II. Unrealized potential of community resources: The unrealized potential of community resources may be seen in small mining villages, where other natural resources such as undeveloped land, untapped water supplies, or unmanaged natural forests with significant lumber are frequently overlooked and left underdeveloped. Since the property is typically owned and maintained by outside organizations, it is not always recognized as an untapped source of economic potential, leaving its resources unused and the community deprived of its economic advantages. Legal processes for any new endeavours are difficult, drawn-out, and complex, and competing for irrigation needs from farms downstream lead to an abandonment of lake recreation plans. Any new industry will require not simply land and water power, but also ancillary services.

III. Brownfields: The legacy of large cities is the presence of abandoned industrial structures constructed in the 19th and early 20th centuries on extensive tracts of valuable urban land. Action may be stymied by the heated argument over restoration vs destruction. Some view the dilapidated or rusted factories as useless, while social historians and certain architects desire to preserve these locations as examples of industrial legacy.

This paper explores the relationship between the density of a population and how the land is utilized. This study aims to understand how population density impacts land use patterns and how land use, in turn, affects population density. The research will examine both urban and rural areas, and the discovering will have implications for land use planning and urban development. The paper will utilize data from a variety of sources, including census data and satellite imagery, to gain a comprehensive understanding of the relationship between population density and land use. However, this explration will contribute to a better understanding of how human activity affects the environment and how land use planning can be used to promote sustainable development.

II. RELATED WORK:

I. A framework by combining multi-source social sensing data and remote sensing images is proposed in a paper by Wenliang Li. The study focuses on the land-use patterns of New York City and employs the use of the random forest method. Results reveal an overall accuracy of 77.31% in the level I classification and 66.53% in the level II classification [3].

II. The authors of the "Detection of Urban and Environmental Changes via Remote Sensing" paper(Karim Ennouri, Slim Smaoui, and Mohamed Ali Triki) present an examination of changes in land cover and the associated potential impacts on factors such as soil depletion, amplified run-off, water balance, and climate change [4].

III. Adel Shalaby and Hossam S. Khedr, in their paper "Remote Sensing and GIS for Land Use/Land Cover Change Detection in Dakhla Oasis" have employed the use of three multispectral satellite images: Landsat TM (1988), Landsat TM (2003), and Sentinel 2 (2018), to analyse land use change in Dakhla Oasis in the Western Desert (Egypt). Results show a significant increase in the total areas of agricultural land and built-up areas, while the water bodies slightly increased in the period under examination [5].

In their paper, H.Z.M Shafri and F.S.H Ramle have already conducted a comparison between The comparison between various land classification algorithms to determine which one of them gives the best accuracy has been done by Syam Kakrala [13].

IV. A comparative study of various edge detection operators and their suitability for specific models has already been conducted by various researchers [15, 16, 17, 18].

V. Lei Yang, Xiaoyu Wu, Dewei Zhao, Hui Li, and Jun Zhai have presented an enhanced version of the Prewitt algorithm for a noised image in their research [19].

The land has been classified using GIS, SVM, decision trees, CNN, and other machine learning techniques by all previous studies, but population densities have not been concurrently calculated by any of them to locate the predicted population density and vacant land to relocate the growing population to prevent under- and over-utilizing the available space. However, in our model, QGIS is used only for fetching the latest satellite images and the formation of bands. Furthermore, the SVM algorithm is utilized for the classification of land into water bodies, agricultural areas, forested areas, vacant land, and built-up areas, which provides maximum efficiency. The area of vacant land is calculated through edge detection, which determines the exact empty area for population migration.

III. OVERVIEW:

A variety of purposes, including farming, growing forests, grazing animals, mining, setting up industries, and building homes, roads, trains, etc. are utilised for and by land. The sustained growth and affluence of any nation are necessary for the right and sensible use of the land. The kind of terrain, fertility, depth, water retention ability, mineral content that is accessible, and transportation options, among other factors, rely on land usage.

Any area of land that has been cleared for development but has not yet undergone any further construction or industrial usage is referred to as unused land. This can include bare soil land, forested land, or agricultural land. In metropolitan contexts, bare earth is considered a crucial component of the landscape. If we saw the National Institute of Hydrology, 51.09% of the land in India is cultivated, 21.81% is covered by forest, and 3.92% is used for grazing. 12.34% of the land is made up of built-up areas and undeveloped land. Uncultivated waste makes up around 5.17% of the overall land area and can be used for agriculture. The remaining 4.67% of the land is made up of other forms. (Source: Kundra, 1999)

A thorough analysis of the land and its categorization is done in the paper. The goal is to classify the land into several categories, including those for cultivation, industry, water bodies, open space, and forested areas, as well as to determine the extent of bare soil land. To estimate possible population movement around the region, attention is also being paid to the area's existing population.

IV. ADVANTAGES:

I. This paper's output categorizes land into several categories: industrial, agricultural, bare soil, water bodies, and forested land.
II. We can avoid the over-exploitation of resources in a particular region.
III. The quantity of pollution and global warming may also reduce if the population is distributed equitably over the whole nation.
IV. Problems with water and energy are being experienced more in urban cities like Delhi and Mumbai than in less developed places like Meerut and Muzaffarnagar. As one of the major cities, Delhi is quite populous and has a lot of traffic and pollution-producing automobiles. Also, Delhi is overcrowded as a result of the development of new structures. Construction site dust, including cement, wood dust, and other types, spreads widely and mixes with other microscopic particles in the atmosphere. These issues can also be fixed.
V. The vacant property can be further observed and put to some use.
VI. It is possible to address the issue of population build-up in a certain state, locale, or region. Further, the government can work on issues that keep the land unoccupied or prevent population migration.

V. CHALLENGES:

I. Quality assurance was the first and foremost challenge.
II. Fear of failure because, despite having a clear plan, the struggle was to locate the required resources.
III. To create the groundwork for the paper, consulting several research papers, case studies, common statistical data, demographic statistics from recent years, and different geologically connected works was a task since the majority of the data was dated and old. 2009, 2011, and 2014 were the most often cited years.
IV. Lack of clarity since the paper is made up of multiple distinct issues, each of which might stand alone as a significant paper. Since this paper involves machine learning (ML), the key worry was completing things expertly, accurately, and on schedule.
V. It was difficult to select the ideal satellite image with the least amount of cloud cover and the one with all the necessary geographical areas for categorization.

VI. METHODOLOGY:

The proposed methodology utilizes a user-friendly platform that allows for easy input of geographic regions of interest. The user has the option to either select the region directly from a map by clicking on the upper left and lower right corners to obtain the coordinates or to manually enter the coordinates of the upper left and lower right corners of the region. Once the coordinates are obtained, the model automatically retrieves the most recent Landsat 8 satellite images for the chosen area using the QGIS API. The user can then select a specific satellite image from the available options for further analysis.

The model then employs a machine learning technique, specifically Support Vector Machine (SVM), to classify the area based on the various bands of the selected satellite image [21]. Edge detection is used to specifically calculate the amount of unused land within the region. Additionally, the model employs linear regression to forecast population density for the chosen region.

The final result of the model provides the user with the present and predicted population densities as well as the amount of unused land in square kilometres. This information can be used to estimate potential population migration patterns throughout the area and inform decision-making in land use planning and management.

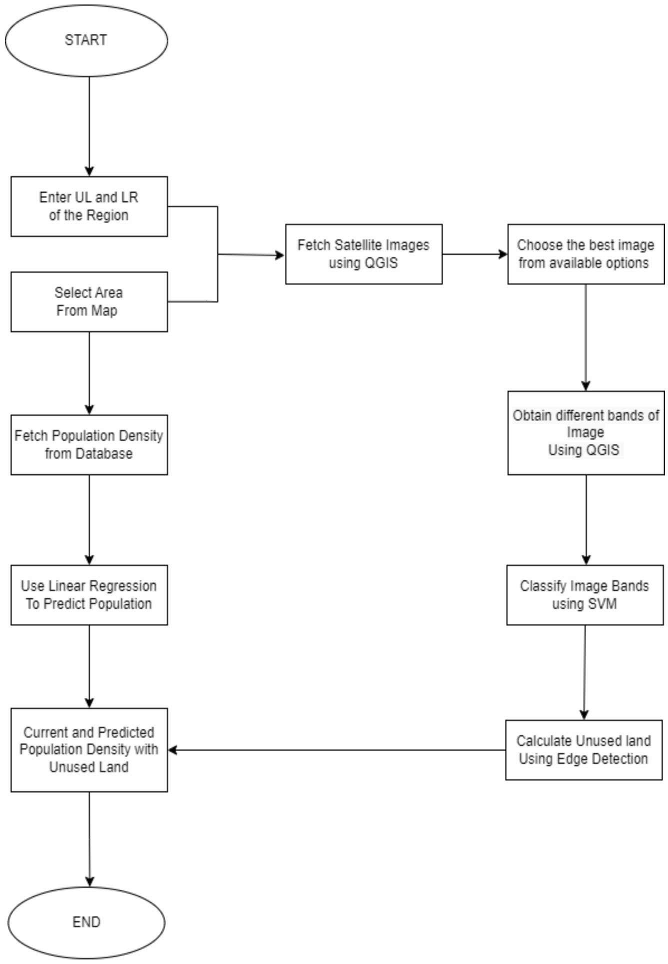

Fig. 2. Flowchart of the model

Terminologies used:

1. QGIS: QGIS (Quantum Geographic Information System) and its numerous plugins are used to fetch satellite images and access their bands.

2. API: Application Programming Interface, or API, is the acronym for the word. The connection between two applications and the sharing of information and communication between them is handled by the software. The connection of two applications or devices is enabled by API to aid information exchange. The interface is what the other software components use. The standards or documentation created to explain the construction of such connections are known as API specifications.

3. Database: An organised collection of structured data, often known as a database, is generally kept digitally inside the system. As the database, MySQL is utilized to hold the dossier on population density from 1990 to 2021.

4. Satellite images: A range of tasks, including transmitting telecommunications signals, gathering data for tactical objectives, or forecasting weather patterns, are performed by artificial satellites in orbit. Many aspects of reality are examined by the paper through the photos that are gathered by satellites.

ALGORITHMS:

I. Support vector machine: Support Vector Machine (SVM) is a supervised ML algorithm that is used for regression as well as classification.

ALGORITHM OF SVM:

1. The necessary libraries, including NumPy, sklearn, and matplotlib, were imported.
2. The satellite imagery dataset, including both the image data and the corresponding land cover labels, was loaded.
3. Pre-processing steps, including resampling, normalization, and feature extraction, were performed on the image data.
4. The dataset was divided into training and testing sets.
5. An SVM classifier was trained on the training set using a radial basis function (RBF) kernel.
6. The classifier was tested on the te
7. sting set and its performance was evaluated using metrics such as accuracy, precision, and recall.
8. The trained classifier was applied to classify the entire image, creating a land cover map.
9. The results were visualized using matplotlib.

CODE:

```
import numpy as np

from sklearn import svm

from sklearn.metrics import accuracy_score

from sklearn.model_selection import train_test_split

# Load Data set

Z, y = load_dataset()

# Preprocessing

Z = preprocess(Z)

# Split data into training and testing sets

Z_train, Z_test, y_train, y_test = train_test_split(Z, y, test_size=0.2)

# Train SVM classifier

a = svm.SVC(kernel='rbf') where a is the classifier

a.fit(Z_train, y_train)

# Test classifier on the test set

y_pred = a.predict(Z_test)

# Evaluate performance

accy = accuracy_score(y_test, y_pred)

print("Accuracy:", accy)

# Classify the entire image

result = a.predict(Z)
```

II. Edge detection: It is a method of image processing that locates the edges of objects in pictures. It operates by looking for changes in brightness.

Laplacian of gaussian: Also known as Marr Hildreth Operator is a gaussian based operator that takes an image's 2nd derivative using the Laplacian. When the change in the grey level appears abrupt, assistance is rendered by the technique. It operates using the zero-crossing approach, where a maximum level is determined by where the 2nd-order derivative passes zero. It's referred to as an edge location. The sharp edges are found by the Laplacian operator while the noise is minimized by the Gaussian operator.

ALGORITHM OF PREWITT METHOD:

1. The image is recommended to be loaded into the algorithm as the first step.
2. If the image is not already in grayscale, it should be converted to grayscale.
3. The Prewitt edge detection filter should be applied to the image by convolving it with the corresponding Prewitt kernel.
4. In the 4th step, a binary image is created where all pixels with an intensity greater than a certain value are set to 1, and all other pixels are set to 0 by thresholding the filtered image.
5. Further, the number of pixels with a value of 1 is determined by iterating through all the pixels in the binary image.
6. Finally, the total area of the edges in the image is found by multiplying the count of edge pixels by the area of each pixel.

CODE:

```
import numpy as np

# Create the Prewitt kernel for the y-direction

prewitt_iy = np.array([[-1, -1, -1], [0, 0, 0], [1, 1, 1]])

# Load the image

image = …

# Apply the Prewitt operator to the image

filtered_image = np.abs(np.convolve(image, prewitt_iy, mode='same'))
```

In the above code, the prewitt_iy variable holds the Prewitt kernel for the y-direction, which is a 3x3 matrix. The convolve function is used to convolve the image with the kernel, and the mode parameter is set to 'same' to ensure that the output image has the same size as the input image. The np.abs function is

used to take the absolute value of the filtered image, as the Prewitt operator can produce negative values.

III. Multiple Linear Regression: This statistical method can be used to predict a response (dependent) variable based on multiple predictors (independent) variables. The general form of the equation for multiple linear regression is:

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta n*xn$$

where y is the response variable, x1, x2, ..., xn are the predictor variables, and β0, β1, β2, ..., βn are the coefficients to be estimated. The goal is to find the values of the coefficients that minimize the sum of the squared errors between the predicted values and the actual values of the response variable.

CODE:

```
from sklearn.linear_model import LinearRegression

import numpy as np
```

```
# Prepare the data

Z = np.array(data[['x1', 'x2', 'x3', ..., 'xn']])
# predictor variables

y = np.array(data['y']) # response variable

# Create the linear regression model

model = LinearRegression()

# Fit the model to the data

model.fit(Z, y)

# Print the coefficients

print('Intercept:', model.intercept_)

print('Coefficients:', model.coeff_)

# Predict the response for a new set of predictors

Z_new = np.array([[z1_new, z2_new, z3_new, ..., zn_new]])

y_pred = model.predict(Z_new)
```

To check accuracy, sklearn.metrics library is used

CODE:

```
from sklearn.metrics import mean_squared_error, r2_score

# Predict the response for the data

y_pred = model.predict(Z)

# Calculate the mean squared error and the R-squared score

mser = mean_squared_error(y, y_pred)

r2 = r2_score(y, y_pred)

print('Mean squared error:', mser)

print('R-squared score:', r2)
```

VII. IMPLEMENTATION:

Case study:

The bulk of Telangana's cities is included in the case study. Telangana, India's 29th state, was created on June 2, 2014, and is home to a population of 3,50,03,674 people living in an area of 1,12,077 square kilometres. The state was previously a part of Hyderabad, which united with the state of Andhra to create Andhra Pradesh (AP) as a state from 17-09-1948 till 1-11-1956 [22]. The UL (77.530993514541,19.955743223192) and LR (80.480867711167,16.676502301880) coordinates are entered manually.

Fig. 3. Map depicting the area of Telangana

Fig. 4. Map of Telangana

The screen displayed the satellite images that were accessible from 2021-01-01 to 2021-03-31 among which one was chosen.

The B1 to B11 bands are fetched automatically.

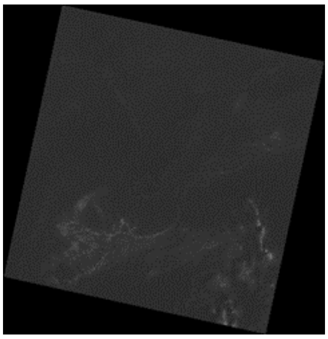

Fig. 5. B1

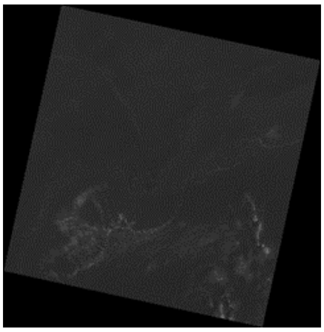

Fig. 6. B2

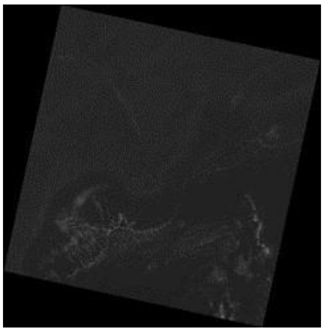

Fig. 7. B3

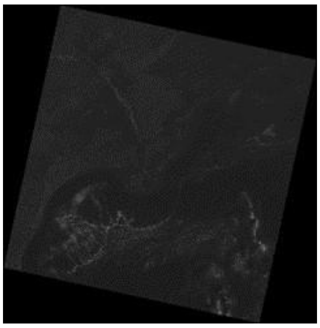

Fig. 8. B4

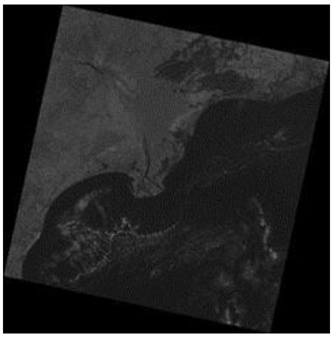

Fig. 9. B5

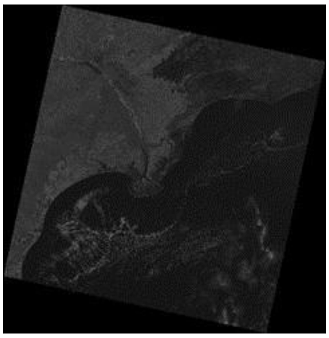

Fig. 10. B6

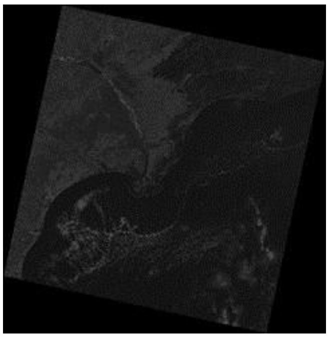

Fig.11. B7

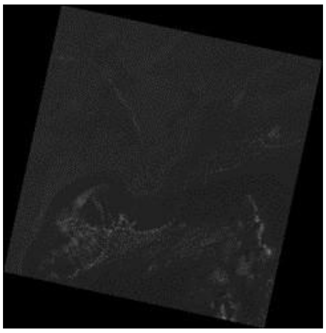

Fig.12. B8

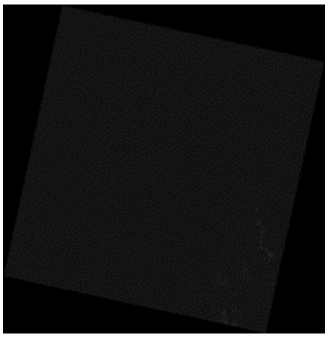

Fig. 13. B9

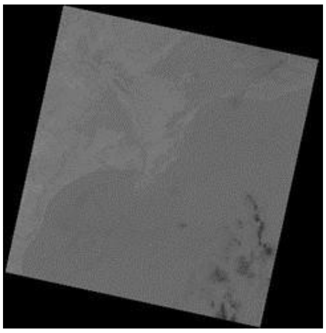

Fig. 14. B10

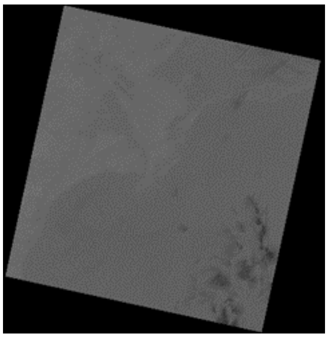

Fig. 15. B11

WAVELENGTH	BANDS
0.433-0.453	*Coastal/Aerosol*
0.450-0.515	*Blue*
0.525-0.600	*Green*
0.630-0.680	*Red*
0.845-0.885	*Near infrared*
1.560-1.660	*Shortwave infrared 1*
2.100-2.300	*Shortwave infrared 2*

TABLE I. Wavelengths for particular bands

COMPOSITE NAME	BANDS
Natural colour	4 3 2
False colour(urban)	7 6 4
Colour infrared(vegetation)	5 4 3
Agriculture	6 5 2
Healthy vegetation	5 6 2
Land/Water	5 6 4
Shortwave infrared	7 5 4
Vegetation analysis	6 5 4

TABLE II. Spectral band designations for Landsat 8

Fig. 16. Natural satellite image (4-3-2)

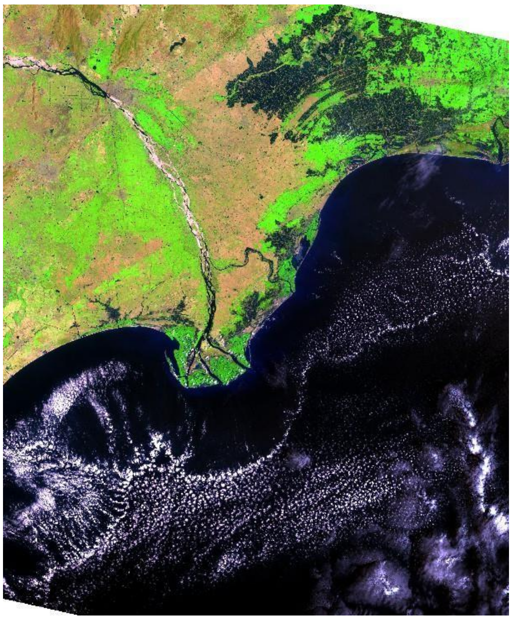

Fig. 17. Satellite image depicting agricultural area, bare soil, and water bodies (6-5-4)

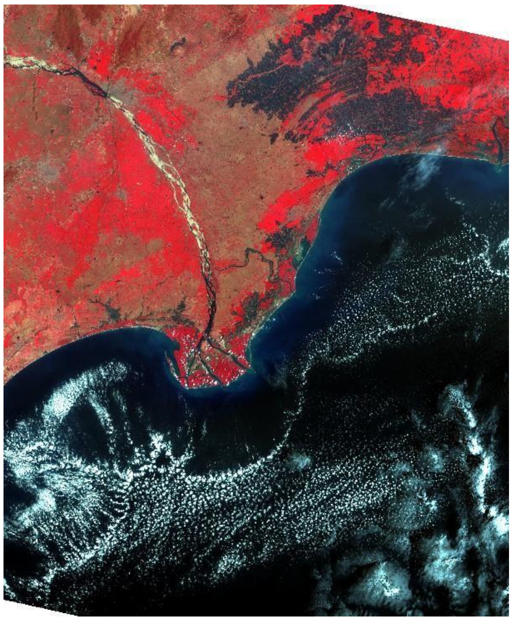

Fig. 18. Satellite image depicting forested areas and built-up areas (5-4-3)

SVM is used for classification, and the resultant accuracy is 98%.

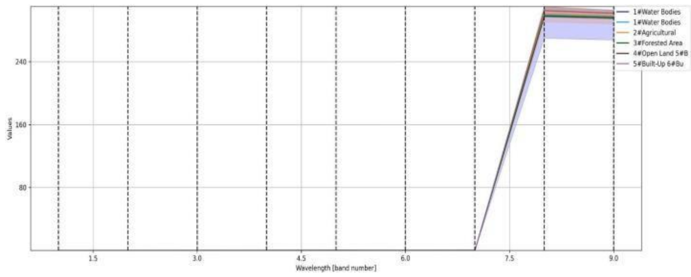

Fig. 19. Spectral signature plot

Spectral signature plot: It is used to display the values of the spectral signature in relation to wavelength.

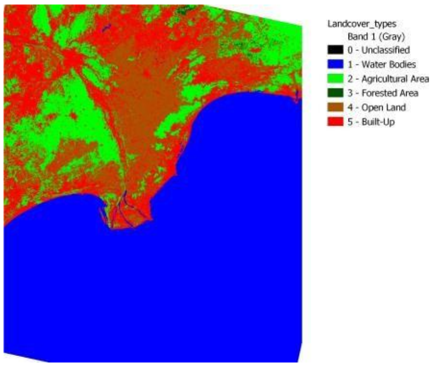

Fig. 20. Classified map after applying unsupervised learning

The population density of Telangana is shown as 312 person/sq. km and predicted to be 40.9 million, or 364 person/sq. km, by 2041, indicating a projected growth of 7.4 percent. The overall amount of unused land found is 2.76 crore acres, with 1.65 acres being arable, and the remainder being forest, arid, and unusable terrain.

RESULT: Using SVM, the land is classified into 5 classes. Land cover types taken into consideration are water bodies, agricultural areas, forested areas, open land, and built-up areas. The accuracy given by the algorithm is 98%.

Classes	Precision	Recall	f1-score	support
0	1.00	1.00	1.00	16222
1	1.00	1.00	1.00	23570
2	1.00	1.00	1.00	6095
3	1.00	1.00	1.00	16790
4	1.00	1.00	1.00	13545
5	1.00	1.00	1.00	9066
accuracy			1.00	85288
Macro avg	1.00	1.00	1.00	85288
Weighted avg	1.00	1.00	1.00	85288

Fig. 21. Classification Report

Additionally, 2.29 crore acres of unused land have been computed using the edge detection (Prewitt method) which was expected to be 2.73 crore acres with an accuracy of 84%. By 2041, the population density is anticipated to be 364 people/square kilometre, up from 312 people/square kilometre in 2021, which is still fewer than the current national average population density of 431.11 people/ square kilometre. Taking the data into consideration we can suggest that around 5,379,696 more people can be migrated to the state increasing the current population density of Telangana to 360 person/sq. km.

VIII. CONCLUSION:

In conclusion, the proposed methodology in this paper utilizes a user-friendly interface to analyse population density and unused land in a specific geographic region in India. The model employs a combination of satellite imagery, supervised machine learning techniques, edge detection, and linear regression to classify and forecast population density, and calculate the amount of unused land. The final result provides the user with present and predicted population densities and the amount of unused land in square kilometres. This information can be used to estimate potential population migration patterns and inform decision-making in land use planning and management.

The generated output data enables the normalization of discrepancies between lower and greater density regions of India by dividing the population evenly across the country. Another benefit is realized through the identification of geographical regions that require greater attention and resources. Furthermore, the anticipated land can be utilized in disaster situations as well as to divide the population equitably as some natural disasters cause long-term harm, making it impossible for the local people to survive for the time being.

The Future scope of this study is to continue to refine and improve the model through the incorporation of additional data sources and machine learning techniques, and to apply the model to other regions in India to better understand the dynamics of population density and land use across the country.

XI. REFERENCES:

[1] Indrajit Dey and Sujit Kumar De, "A simple way to study the global population growth," Man and Life, ISSN-0972-4109, vol. 41, pp. 171-180, June 2015.

[2] Shahab Fazal and Md Kaikubad Ali, "Rural – Urban transition in India: A perspective of future urbanization," ANNALS of the National Association of Geographers India, vol. 38, pp. 343-365, February 2019.

[3] Wenliang Li, "Mapping urban land use by combining multi-source social sensing data and remote sensing images," Earth Science Informatics 14, pp. 1537-1545, May 2021.

[4] Karim Ennouri, Slim Smaoui and Mohamed Ali Triki, "Detection of Urban and Environmental Changes via Remote Sensing," Circular Economy and Sustainability 1, pp. 1423-1437, March 2021.

[5] Adel Shalaby and Hossam S.Khedr, "Remote Sensing and GIS for Land Use/Land Cover Change Detection in Dakhla Oasis," Earth and Environmental Sciences Library, March 2021.

[6] Pandian Mangan, "Land use and land cover change detection using remote sensing and GIS in parts of Coimbatore and Tiruppur districts, Tamil Nadu, India," International Journal of Remote Sensing & Geoscience, vol. 3, January 2014.

[7] Sandipan Das, A. Bhattacharya and S. Mali, "Study on Urban Land Suitability Assessment Using Remote Sensing and GIS: A Case Study of Khairagarh, in Chhattisgarh," International Journal of Computer Applications, vol. 74, pp. 20-26, July 2013.

[8] Ujjwal Maulik and Debasis Chakraborty, "Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques," IEEE Geoscience and Remote Sensing Magazine, vol. 5, pp. 33-52, March 2017.

[9] Giorgos Mountrakis, Jungho Im and Caesar Ogole, "Support vector machines in remote sensing: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, pp. 247-259, May 2011.

[10] Mahesh Pal and Paul M.Mather, "Support vector machines for classification in remote sensing," International Journal of Remote Sensing, vol. 26, pp. 1007-1011 March 2005.

[11] G. Mountrakis, J. Im, Caesar Ogole, "Support vector machines in remote sensing: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, pp. 247-259, May 2011.

[12] H.Z.M. Shafri and F.S.H. Ramle, "A Comparison of Support Vector Machine and Decision Tree Classifications Using Satellite Data of Langkawi Island," Information Technology Journal, *pp. 64-70, 2009.*

[13] Syam Kakarla, " Land Cover Classification of Satellite Imagery using Python," Towards Data Science, March 2021.

[14] Hela ElMannai, Monia Hamdi and Abeer AlGarni, "Enhanced Support Vector Machine Applied to Land-Use Classification," Communications in Computer and Information Science, vol 1097, pp. 236-244, December 2019.