Combined Optical and SAR Remote Sensing for LULC Mapping of Imphal Valley Using Machine Learning Algorithm

Priyanka Gupta and Dericks Shukla

December 11, 2022

Combined Optical and SAR remote sensing for LULC mapping of Imphal valley using Machine Learning Algorithm

Priyanka Gupta
Ph.D. Scholar, School of Civil and Environmental Engineering,
Indian Institute of Technology (IIT)-Mandi,
Himachal Pradesh, India,
d20062@students.iitmandi.ac.in

*Dericks Praise Shukla**
Associate Professor, School of Civil and Environmental Engineering,
Indian Institute of Technology (IIT)-Mandi,
Himachal Pradesh, India,
dericks@iitmandi.ac.in

***Abstract*—**
Imphal Valley (India) is the part of United Nations Educational, Scientific and Cultural Organization (UNESCO) biodiversity hotspots in India where the Imphal River and Loktak Lake form their lifelines. This region is dynamically changing at an unprecedented pace and the effects of such Land use change need to be quantified accurately. In this work, we used both optical satellite data and microwave (Synthetic Aperture Radar (SAR)) data to prepare the Land Use Land Cover (LULC) map of the area for 2015 and 2021. We used Landsat-8 and Sentinel 1A satellite data for 2015 and 2021. We compared our results with optical-derived LULC as well as earlier published literature. We have classified the area into seven classes namely waterbody, Phumdis, settlement, horticulture farm, paddy farm, forest, and bare land, and used the Random Forest Machine learning (ML) algorithm for image classification. We observed a significant decrease in waterbody, Phumdis, and forest areas with an increase in the settlement, paddy, and horticulture farm between 2015 and 2021. We observed that settlement and water bodies were better classified when we used both the data rather than optical data only. The k-fold overall accuracy (OA) and Kappa coefficient (KC) for integrated data was 87% and 0.83 for 2015 and OA-90%, KC-0.86 for 2021, where k=10. These values were lower for only Landsat-8 derived LULC.

Keywords— Synthetic Aperture Radar **(**SAR); Land Use Land Cover (LULC)Mapping; Loktak Lake; Machine Learning (ML) algorithm; Google Earth Engine (GEE).

I. Introduction

One of the primary drivers of environmental change worldwide is Land Use Land Cover (LULC) [1]. The changes in LULC happen due to improper use of the limited land resource by the ever-increasing population for agricultural and other purposes. This impacts the environment and all the ecosystem badly, which in turn acts as an important factor for LULC change. Thus, this forms a circular pattern and a positive feedback loop. To address these issues, it becomes inevitable to prepare accurate maps and monitor LULC change for sustainable and holistic planning at regional and national levels.

The greater accessibility of data from numerous, complementary remote sensing sensors and sources has allowed research to exploit these data for creating better LULC maps. Since the beginning of the last decade, Synthetic Aperture Radar (SAR) data is becoming readily and easily accessible. So larger number of users have started using SAR for research, practical and commercial purposes. Pixel-based classification approach on SAR data have produced LULC maps with acceptable accuracy but they are still less used by the remote sensing community [2]. In recent years there has been a shift from the statistical approach to a more flexible and powerful Machine Learning (ML) approach where various algorithms and applications have been developed for geospatial applications [3]. Among these ML algorithms, Random Forest (RF), Support Vector Machine (SVM), and KNN have shown better, more accurate, and enhanced performance for LULC mapping studies [2; 4; 5]. RF has been used with both optical and SAR data, making the classification better for all categories [6].

Combining separate data sets enables the use of bands from various electromagnetic spectrum regions [7]. Integrating Optical and SAR helps in availing the benefits of both data types [8]. Each sensor has the ability in better classifying certain land cover types over others. SAR data alone is more efficient in delineating built-up areas whereas the optical data is inept in classifying the natural vegetation [7; 9]. Hence the combination of both types of data may reduce the problem of misclassification [10].

Loktak Lake of Imphal valley is a Ramsar site and is one of the largest freshwater lakes in South Asia. It is also one of the two sites listed in the Montreux record. It is famous for floating islands called Phumdis, which is a heterogeneous mass of soil, vegetation, and organic matter. One such phumdi is home to Keibul Lamjao National Park, the only floating park in the world. Also, this valley shelters more than two-thirds of Manipur's population and hence is densely populated. The LULC pattern of the state is changing frequently. As this is an intermontane valley lying close to the tropics so it receives heavy rainfall (>200 cm annual rainfall) during most months of the year. Hence, it is cloudy most of the year, so getting cloud-free satellite images for LULC classification is very difficult. Hence to overcome the problem of atmospheric issues, SAR data from Sentinel 1A is used, which operates in the microwave part of the electromagnetic spectrum that could penetrate the clouds. [4].

There was no comparable analysis being conducted using both optical and SAR images for LULC mapping using ML algorithm in this rich biodiverse area before this work. Thus, the primary goal of the study is to show the applicability of the ML algorithm to derive accurate LULC maps by using both optical and SAR datasets. Additionally, accurate mapping of human habitation and vegetation with a focus on Phumdis and their change in 5 years is studied in this work using Google Earth Engine (GEE). The results are interesting and could be used for better management practices.

II. Study Area

Imphal valley is an intermontane valley located in the lesser Himalayan region of North East India (figure 1), in the Manipur state of India between the 23˚N to 25˚N latitudes and 93˚E to 94˚E longitudes[11]. This valley covers an area of around 1864.44 sq. km. and is interspersed by many small rivers originating from the surrounding hills and flowing from North to South. The most important one among them is the Imphal river and its tributaries are Iril, Thoubal, Khuga, and Sekmai.[12]. It has a subtropical monsoon-type climate where the maximum summer temperature lies between 32°C – 34°C and winter temperature goes to a minimum of 1°C - 2°C. Manipur state has a total population of about 2,855,794 as per the 2011 census and the Imphal valley holds around 60% of this population.

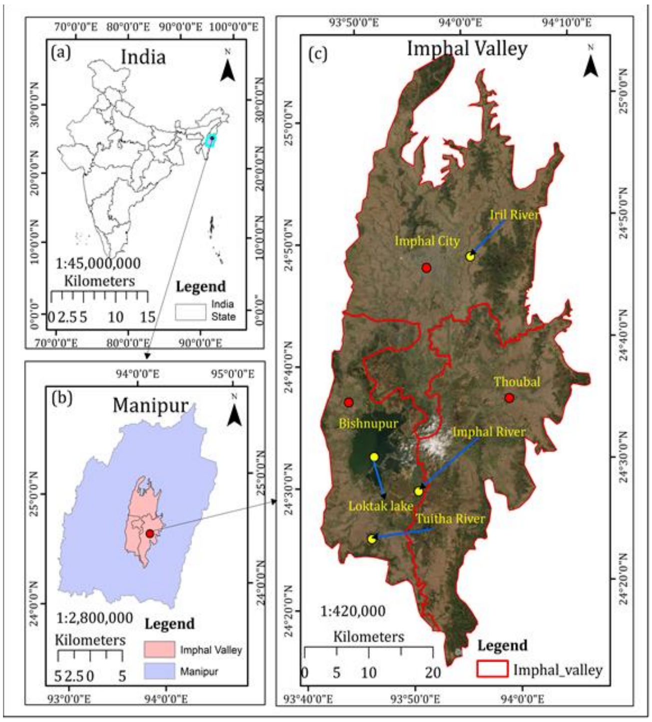

Figure 1: Study area map of Imphal Valley.

III. Data Used

In this study, we have used Landsat 8 (L8) surface reflectance Tier 1 dataset for March to April for both the years i.e 2015 and 2021, for LULC mapping using Band B1 to B7 and NDVI (Normalized difference vegetation index) of the study area. Additionally, VV and VH products of Sentinel 1A ortho-corrected SAR Data which includes S1 Ground Range Detected (GRD) scenes processed using Sentinel-1 Toolbox in combination with Landsat 8 data is used for LULC mapping. In all Bands B1-B7, NDVI, VV, and VH are used for LULC mapping using Random Forest (RF) classifier. This work is carried out on in Google Earth Engine platform and analyzed using ArcMap 10.5.

IV. Methodology

The methodology used in the study is shown in the flowchart in figure 2. For the first step of pre-processing, L8 surface reflectance data is taken, filtered by date, and clipped by a region of interest. Band 1 of aerosol was removed from further analysis and so the Blue band was renamed as B1, Green as B2, Red as B3, and so on. Cloud masking is then used to obtain cloud-free images. The NDVI is then calculated for each scene and its mean value is taken. After that, a composite is prepared using the bands from B1 to B7 and NDVI. Similarly, Sentinel 1A data is filtered for the date, and we choose VV and VH bands that are mosaicked, and then to reduce speckle in the image. We have applied a morphological mean filter to each band of an image using a circle kernel type with a smoothing radius of 50 meters to remove speckle noise. After that, we created a collection made up of all L8 and Sentinel 1A data bands. In this collection, data training is carried out using samples from 7 different classes namely: waterbody, Phumdis, settlement, horticulture farm, paddy farm, forest, and bare land. We applied Random Forest (RF) [5, 6] classifier for the classification of the image to prepare the LULC map of the study area. The samples are divided into training (80%) and testing (20%) and that was used to obtain the overall accuracy and Kappa coefficient of classification using K-fold accuracy estimation where k=10.

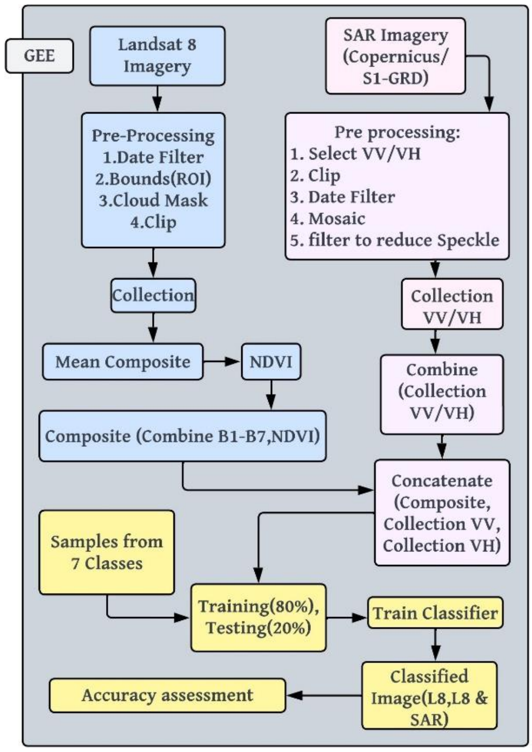

Figure 2: Methodology used in the study.

V. Results And Discussion

Results depict a significant change in the study area. The land use land cover classes for the study area can be seen in Figure 3, and Figure 4. Table 1 shows how much area each class takes up in the study area. K fold (k=10) overall accuracy using optical data (Landsat 8) for the years 2015 and 2021 is 84.61% and 89.24%, respectively, with a Kappa coefficient (KC) of 0.80 and 0.86. However, utilizing integrated SAR and optical data, the k-fold (k=10) overall accuracy (OA) for 2015 and 2021 were increased to 87% and 90%, respectively, with KC values of 0.83 and 0.86. Using both L8 and SAR, it is evident that between 2015 and 2021, there is an increase in settlement, horticulture farms, and paddy farms, while there is a decline in waterbodies, phumdis, and forests. Hence it reflects land cover is continuously decreasing and land use increasing, so proper management of land resources is required to save the environment from changes that are happening [13] in Imphal Valley impacting the health of Loktak Lake and Keibul Lamjao National Park.

The results showed that waterbodies occupies around 7% of the total area of the region in 2015 while it got decreased by around 2.5% in 2021 occupying 4.07% of the total area using L8 data. Similar results are obtained from combined data as well, i.e. ~2.30% of change between 2015 and 2021.

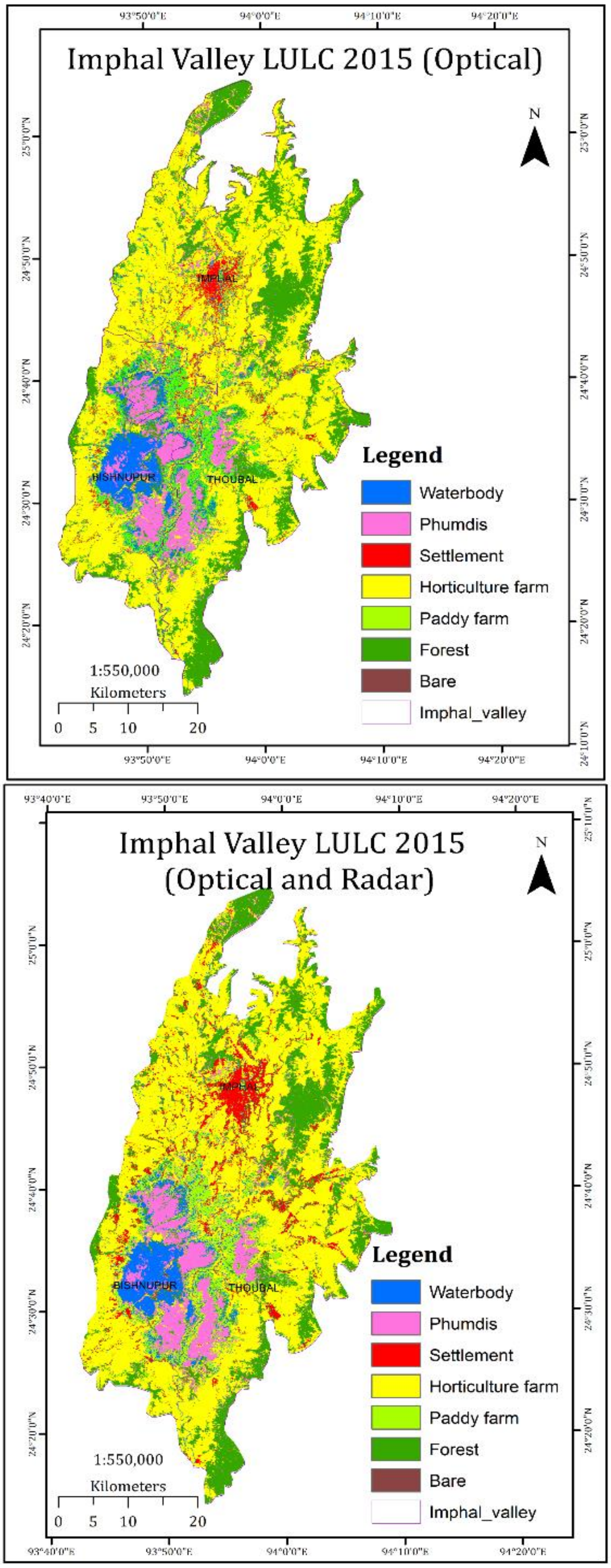

Figure 3: LULC map of Imphal Valley for 2015 using L-8 data (top) and integrated L-8 and Sentinel-1A data (bottom).

Anand and Oinam [11], applied maximum likelihood algorithm using Landsat 8 data and overserved that the waterbodies occupy 1.68% of the total area. Farming practices specifically paddy farms increased by around 7% between 2015 and 2021 while horticulture farms remained nearly the same with the least percentage change (table 1). Due to the construction of the Ithai barrage in 1984 [14], which transforms the Loktak into a reservoir for a hydroelectric project, the agricultural land in the lake expands [15]. Forest cover in the area decreased by around 5% in the study area which can be seen in table 1. There is an increase of 3.49% in settlement area between 2015-2021 when mapped using L8 data only but this increase is 1.79% when mapped using combined data. Settlements have higher backscatter in SAR data, so they are mapped correctly when using SAR data. But with only optical data, there is overestimation, hence higher change is observed for L8 data only.

Phumdis register a decrease of around 2% may due to the removal of floating herbaceous wetlands from the central area and converting them to Athaphums from Phumdis for the fishing purpose [11,12,14]. A naturally changing marsh land was turned into a reservoir by the barrage, which resulted in the inundation of nearby areas, the extinction of migratory fishes, the diminution and deterioration of habitat in national parks, and a decline in water quality.

TABLE I. CLASS STATISTICS USING L8 AND BOTH (L8+SENTINEL 1A) DATA

Class statistics - L8							
	Classes	**2015 (area in sq. km.)**	**2021 (area in sq. km.)**	**Change (sq. km.)**	**2015%**	**2021%**	**Change %**
1	Water	141.36	86.20	-55.15	6.67	4.07	-2.60
2	Phumdis	189.60	138.64	-50.95	8.94	6.54	-2.40
3	Settlement	71.36	145.33	73.97	3.37	6.85	3.49
4	Horticulture farm	1112.94	1122.01	9.07	52.49	52.92	0.43
5	Paddy	192.53	312.57	140.03	9.08	15.69	6.60
6	Forest	411.57	290.48	-121.08	19.41	13.70	-5.71
7	Bare	0.90	5.01	4.10	0.04	0.24	0.19

Class statistics – L8+Sentinel 1A							
	Classes	**2015 (area in sq. km.)**	**2021 (area in sq. km.)**	**Change (sq. km.)**	**2015%**	**2021%**	**Change %**
1	Water	133.81	85.12	-48.69	6.31	4.01	-2.30
2	Phumdis	185.93	157.30	-28.63	8.77	7.42	-1.35
3	Settlement	134.52	172.51	37.99	6.34	8.14	1.79
4	Horticulture farm	1093.51	1098.79	5.28	51.57	51.82	0.25
5	Paddy	160.75	309.11	148.36	7.58	14.58	7.00
6	Forest	410.37	294.78	-115.61	19.36	13.90	-5.45
7	Bare	1.33	2.62	1.28	0.06	0.12	0.06

Environmentalists are extremely concerned about the removal of phumdis (herbaceous wetlands), which are home to the endangered Rucervus eldii and several other animal species. In the last ten years, the basin's climate has seen a dramatic transformation. Settlement increased (table 1) by around 2-3% which can also be seen clearly in Figures 3 and 4. According to the 2011 Manipur population census, the human population in the basin has grown significantly, placing great strain on the lake's natural resources and the livelihood of the locals. In particular, sectoral developmental planning within the basin has put the ecosystem services and biodiversity of the Loktak Lake complex under stress. Projects to enhance water resources for irrigation, flood prevention, and hydropower generation have resulted in changes to hydrological regimes that have a significant impact on the wetlands' processes, functions, and characteristics [14, 16].

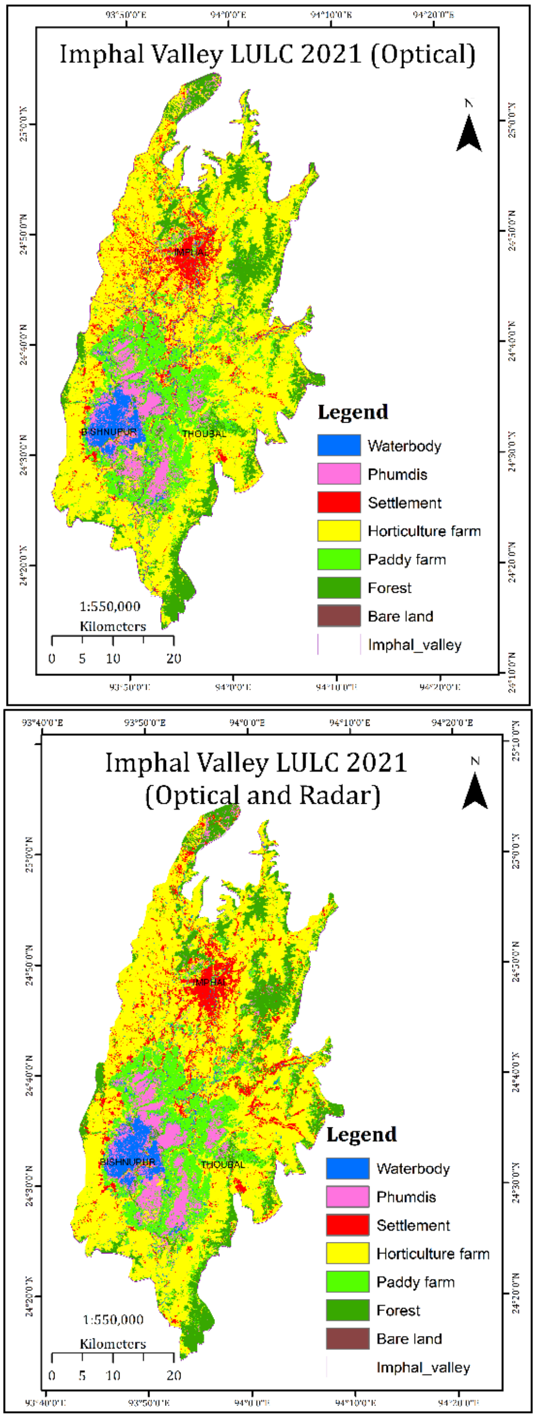

Figure 4: LULC map of Imphal Valley for 2021 using L-8 data (top) and integrated L-8 and Sentinel-1A data (bottom).

VI. Conclusion

In this study, we have used Random Forest, a ML algorithm to prepare the updated LULC map of the Imphal valley, home to Loktak Lake and Keibul Lamjao National Park. We have classified the area in 7 classes with the help of on combined optical and SAR data of 2015 and 2021 on GEE platform. Out of them waterbody, Phumdis and forest classes are found to be decreasing while settlement, farming is found to be increasing from 2015 to 2022. This change is mainly seen in Phumdis where settlement is increasing and it is causing a major concern. We found that combined data of L8 and Sentinel 1A performs better in the classification of Phumdis and settlement area due to higher backscattering from urban areas.

References

[1] Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*(7720), 639-643.

[2] Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, *115*(5), 1145-1161.

[3] Waske, B., & van der Linden, S. (2008). Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(5), 1457-1466.

[4] Niraj, K. C., Gupta, S. K., & Shukla, D. P. (2021). Kotrupi landslide deformation study in non-urban area using DInSAR and MTInSAR techniques on Sentinel-1 SAR data. *Advances in Space Research*.

[5] Stefanski, J., Mack, B., & Waske, B. (2013). Optimization of object-based image analysis with random forests for land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(6), 2492-2504.

[6] Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., ... & Jin, X. (2017). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological indicators*, *73*, 105-117.

[7] Haack, B. N., Herold, N. D., & Bechdol, M. A. (2000). Radar and optical data integration for land-use/land-cover mapping. *Photogrammetric Engineering and Remote Sensing*, *66*(6), 709-716.

[8] Nuthammachot, N., & Stratoulias, D. (2017). Use of SAR and optical satellite data for land use and land cover classification in the Songkhla Lake Basin, Thailand. *International Journal of Applied Engineering Research*, *12*(24), 14358-14364.

[9] Nicolau, A. P., Flores-Anderson, A., Griffin, R., Herndon, K., & Meyer, F. J. (2021). Assessing SAR C-band data to effectively distinguish modified land uses in a heavily disturbed Amazon forest. *International Journal of Applied Earth Observation and Geoinformation*, *94*, 102214.

[10] De Beyer, L. H. (2015). *Integrated use of polarimetric Synthetic Aperture Radar (SAR) and optical image data for land cover mapping using an object-based approach* (Doctoral dissertation, Stellenbosch: Stellenbosch University).

[11] Anand, V., & Oinam, B. (2020). Future land use land cover prediction with special emphasis on urbanization and wetlands. *Remote Sensing Letters*, *11*(3), 225-234.

[12] Kumari, M., Ghosh, S., & Victoria, K. (2018). Spatio-Temporal Analysis Of Phumdi Proliferation Around Lotak Lake Of Manipur Using Geospatial Technology. In *19th ESRI India User Conference*.

[13] Lambin, E. F., & Geist, H. J. (Eds.). (2008). *Land-use and land-cover change: local processes and global impacts*. Springer Science & Business Media.

[14] LDA and WISA. 2006. "Loktak Protection Act 2006. Imphal, Manipur, India." https://www.loktaklake.org/docs/Loktak%20Lake%20Protection%20Act%202006.pdf

[15] Kangabam, R. D., Selvaraj, M., & Govindaraju, M. (2019). Assessment of land use land cover changes in Loktak Lake in Indo-Burma Biodiversity Hotspot using geospatial techniques. *The Egyptian Journal of Remote Sensing and Space Science*, *22*(2), 137-143.

[16] Meitei, I. L., & Singh, A. M. Fuzzy Logic Based Site Suitability Assessment For Cold Storage Construction At Western Villages Of Loktak Lake, Manipur