MULTISPECTRAL SATELLITE IMAGE SEGMENTATION AND CLASSIFICATION OF LAND COVER AREA USING LINEAR REGRESSION OVER RANDOM FOREST WITH IMPROVED ACCURACY

V. Sathwik Sai[1], R. Kesavan[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: Spatial and spectral satellite image segmentation and land cover classification utilising Novel linear regression versus random forest with higher accuracy Linear regression outperform random forest in terms of accuracy

Materials and Methods: Multispectral Satellite Image Segmentation using Linear Regression (N=10) and Random Forest (N=10) with the split size of training and testing dataset 60% and 40% using G-power setting parameters: (α=0.05 and power=0.85) respectively

Results: Linear Regression with Accuracy 80.04 % is more Accurate than the Random Forest with Accuracy 74.07% and attained the significance value 0.053 (Two tailed, $p>0.05$)

Conclusion: The Linear Regression model is significantly better than the Random Forest for multispectral satellite Novel Image Segmentation.

Keywords: Land Cover Area, Linear Regression, Random Forest, Novel Image Segmentation, Satellite, Accuracy.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602105.
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode - 602105.

1. Introduction

The availability of new earth observation satellites, such as Sentinel 2, has sparked interest in their precision agriculture potential (Proffitt 2006). The influence of inter row space and vine geometry on the assessment of non-continuous crops, such as grapevines, may pose issues (Thwal, Ishikawa, and Watanabe 2019)**)**. Satellites are used to examine the qualities of plants. This article depicts people's perspectives and informs them about their alternatives for action and the expenses associated with them (Borra, Thanki, and Dey 2019). To this end, explicit spatial layout awareness varies not just in connection to major visual and ocular-motor parameters, but also in proportion to the costs of executing specified actions (Acharya, Yang, and Lee 2016). Although explicit consciousness is changeable in this way, visually guided behaviours that are directed at one's immediate surroundings are not. Applications of this study are providing a base map for graphical reference, assisting planners, engineers, extracting mineral deposits with remote sensing based spectral analysis, disaster mitigation planning and recovery, agriculture development (Bengtsson, Nordin, and Pedersen 1994).

There were many distinct performances of Linear Regression and Random Forest simple. Around 188 related papers were found in IEEE Xplore and 197 were found in the ScienceDirect database. Many Python libraries were utilised in the development, including Keras, which included a net for Multispectral Satellite Novel Image Segmentation, and TensorFlow, which was created by Google and is used to build deep learning neural networks by performing algorithms (Berhane et al. 2018). Describes the different numerical methods used to create a credible land cover map. The Land cover area intricacy necessitates the use of a variety of data, including Landsat satellite pictures, digital elevation models, digital orthophotos. Linear Regression is Compared over Random Forest architecture. The proposed method achieves a significant improvement in performance and efficiency (Kulkarni and Rege 2021).

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). Presenting ways for determining correctness. Comparing three nonparametric machine-learning methods in this work. The study disadvantages are as follows: Image de-striping, Local cloudiness, limited temporal and geographical resolution, and image gaps make vegetation classification a difficult process in land cover area. The aim was to forecast Satellite novel Image Segmentation using Novel Linear Regression, which delivers the highest accuracy rate when compared to the Random Forest Algorithm ((Acharya, Yang, and Lee 2016).

2. Materials and Methods

The study setting of the proposed work was conducted in the DBMS Laboratory, Saveetha School of Engineering in guidance with faculty. To perform this research two groups were taken. Group 1 is the Linear Regression and group 2 is Random Forest**.** The Sample size was calculated using clinical analysis by keeping G power fixed with 80%, 440 sample sizes estimated per group, totally 880, 93% confidence, pretest power 80%, and enrolment ratio 1 and the maximum accepted error is fixed as 0.05. The dependent variables are their location and proximity to other data and independent variables are generally on the Passive sensors collecting radiation. In this study, the accuracy of two classifiers Linear Regression and Random Forest was compared.

The two groups that used Linear Regression and Random Forest algorithms were performed by taking the dataset containing 10 columns and 20 rows. The dataset was split into training and testing parts accordingly using a test size of 0.2. The first group in this paper is the linear regression algorithm which performs classification by forming groups of every different class in the data. Random Forest classifier takes k groups as input size and tries to do classification with the k groups. Significance value p = 0.053 in Table 4. The proposed work is designed and implemented with the help of google colab software. The platform to assess deep learning was Windows 10 OS. The Hardware configuration was an Intel corei7 processor with a RAM size of 8GB. The system sort used was 64-bit. For the implementation of code, the python programming language was used. As for code execution, the dataset is worked behind to perform an output process for accuracy.

Linear Regression

In insights, straight relapse is a direct way to deal with demonstrating the connection between a scalar reaction and at least one informative factor (otherwise called reliant and autonomous factors). The instance of one informative variable is called straightforward direct relapse; for multiple, the cycle is called different straight relapses over the land cover area. This term is unmistakable from multivariate straight relapse, where different associated subordinate factors are anticipated, rather than a solitary scalar variable.

Pseudocode for Linear Regression

INPUT: Training data D, number of epochs e, learning rate n.
OUTPUT: Classifier accuracy
Ensure: Weights w_0, w_1w1,.. w_k
Step 1: Initialise weights wo, w_1,... w_k from standard normal distribution with
Zero mean and standard deviation σ
Step 2: for epoch in 1... e do
Step 3: for each (x, y) ∈ D in random order do
Step 4: if $\hat{y} \leftarrow \omega 0 + \sum_{i=1}^{k} \omega_i \chi_i$
Step 5: if ($\hat{y}$ > 1 and y = 1) or ($\hat{y}$ < −1 and y = −1) then
Step 6: continue
Step 7: wo ← wo - η 2($\hat{y}$ — y) -
Step 8: for i in 1... k do
Step 9: $w_i \leftarrow w_i - \eta\, 2(\hat{y} - y)$
Step 10: end for
Step 11: end for
Step 12: return wo, w_1,... w_k

Random Forest Algorithm

Choice trees are a famous strategy for different AI assignments. Tree learning comes nearest to meeting the prerequisites for filling in as an off-the-rack technique for information mining, since it is invariant under scaling and different changes of element values, is strong to the incorporation of insignificant highlights, and delivers inspectable models. Notwithstanding, they are only occasionally precise. Specifically, trees that become extremely profound will generally advance exceptionally unpredictable examples: they overfit their preparatthe ion sets in land, cover area, for example have low predisposition, yet extremely high difference. Arbitrary timberlands are an approach to averaging numerous profound choice trees, prepared on various pieces of a similar preparation set, determined to diminish the difference. This comes to the detriment of a little expansion in the inclination and some deficiency of interpretability, yet for the most part significantly helps the exhibition in the last model.

Pseudocode for Random Forest

INPUT: Training data D, number of epochs e, learning rate n.
OUTPUT: Classifier accuracy
Step 1:To generate c classifiers:
For i = 1 to c do
Step 2:Randomly sample the training data D with replacement to produce Di
Step 3:Create a root node. Ni containing Di
Step 4:Call BuildTree(Ni)
Step 5:End for
Step 6:If N contains instances of only one class then return
Else
Step 7:Randomly select x% of the possible splitting features in N
Step 8:Select the feature F with the highest information gain to split on
Step 9:Create f child nodes of N, N1,......,Nf
For i = 1 to f do
Step 10:Set the contents of Ni to Di
Fi
Step 11:Call Build Tree(Ni)
Step 12:End for
Step 13:End if

Statistical Analysis

The statistical analysis is done using IBM's SPSS statistical analysis tool with version 26. Independent Sample T-test analysis was performed by using the Machine learning models and evaluated the quality of the study. In SPSS the dataset is prepared using the 10 samples from each of the algorithms and the total samples is 20. Group id is given 1 for Linear Regression and 2 for Random Forest.

3. Results

The group statistical analysis on the two groups shows: Linear Regression has more mean accuracy than Random Forest and the standard error mean is slightly less than Linear Regression. The Linear Regression algorithm scored an accuracy of 80.04% and Random Forest has scored 74.07% as shown in Table 4. The accuracies are recorded by testing the algorithms with 10 different sample sizes and the average accuracy is calculated for each algorithm. Figure 1 represents the bar chart of accuracies with standard deviation error is plotted for both the algorithms. The Mean value of Linear Regression is better when compared with the Random Forest with a standard deviation of **1.**514121 and 2.01753 respectively. Table 4 shows the Independent sample T-test data of Random Forest and Linear Regression with the significance value obtained is 0.053 (Two-tailed, p>0.05).

4. Discussion

From the results of this study, Linear Regression is proved to be having better accuracy than the Random Forest algorithm over Novel

Image Segmentation in Table 1. Linear Regression has an accuracy of 80.04% whereas Random Forest has an accuracy of 74.07%. In Table 2, the group statistical analysis on the two groups shows that Random Forest has more mean accuracy than Random Forest and the standard error mean including standard deviation mean is slightly less in Table 3. These involve various combinations of the proposed scheme's constituent components, specifically, rough-set initialization (Jenicka 2021)).

Similar paper provides an overview of state-of-the-art computer vision algorithms, particularly, and deep learning models based on the models trained on the first partition of the dataset their application to land use classification using satellite imaging data is presented in this study effort land cover area, A study by (Johnson et al. 2003) gives context for the Special Issue on Micro Strategy and Strategizing's origins, ideas, and papers. For opposite urban hydrological investigations, detection using remote sensing photography is critical using Novel Image Segmentation. Urban hydrology is a developing scientific field that enables us to improve and manage urban water systems in order to address environmental challenges created by increasing urbanisation.

The study examines the computing power requirements per unit of area. Their imaging action is easily repeatable. The Limitations of this study are Local cloudiness, low temporal, and Signal reception can be spotty at times. satellites are their unstable signal. (Borra, Thanki, and Dey 2019) and gaps on the image create a complex task for vegetation classification. Future scope enables far broader coverage, and because all information is digital in land cover area,, it can be easily linked with software (Sozzi et al. 2019). Cloud cover has little effect on results in some circumstances.

5. Conclusions

In this research work, the prediction of the accuracy percentage of Multispectral Satellite Image Segmentation using Linear Regression to have enhanced accuracy 80.40%. When compared to the Random Forest 74.07% shown in Fig. 1. Accuracy estimation for various Satellite Image Segmentation has been successfully calculated for the Images. The main focus was on the algorithmic substance of various attention processes, as well as a summary of how they are used. Conclude that we have succeeded in creating a Machine learning model that is a major improvement above all other Multispectral Satellite Image Segmentation Previously available. Accurate descriptions of accurate calculations for each Image can be done using this model.

DECLARATIONS

Conflicts of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author VSS was involved in data collection, data analysis, and manuscript writing. Author RK was involved in conceptualization, data validation, and critical reviews of manuscripts.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organisations for providing financial support that enabled us to complete the study.

1.Best Enlist, Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Acharya, Tri, In Yang, and Dong Lee. 2016. “Land Cover Classification Using a KOMPSAT-3A Multi-Spectral Satellite Image.” Applied Sciences. https://doi.org/10.3390/app6110371.

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. “Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield.” Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Bengtsson, E., B. Nordin, and F. Pedersen. 1994. “MUSE--a New Tool for Interactive Image Analysis and Segmentation Based on Multivariate Statistics.” Computer Methods and Programs in Biomedicine 42 (3): 181–200.

Berhane, Tedros M., Charles R. Lane, Qiusheng Wu, Oleg A. Anenkhonov, Victor V. Chepinoga, Bradley C. Autrey, and Hongxing Liu. 2018. “Comparing Pixel- and Object-Based Approaches in Effectively Classifying Wetland-Dominated Landscapes.” Remote Sensing 10 (1): 46.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Borra, Surekha, Rohit Thanki, and Nilanjan Dey. 2019. Satellite Image Analysis: Clustering and Classification. Springer.

Jenicka, S. 2021. Land Cover Classification of Remotely Sensed Images: A Textural Approach. Springer Nature.

Johnson, L. F., D. E. Roczen, S. K. Youkhana, R. R. Nemani, and D. F. Bosch. 2003. "Mapping Vineyard Leaf Area with Multispectral Satellite Imagery." Computers and Electronics in Agriculture. https://doi.org/10.1016/s0168-1699(02)00106-0.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Kulkarni, Samadhan C., and Priti P. Rege. 2021. "Application of Taguchi Method to Improve Land Use Land Cover Classification Using PCA-DWT-Based SAR-Multispectral Image Fusion." Journal of Applied Remote Sensing. https://doi.org/10.1117/1.jrs.15.014509.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Proffitt, Tony. 2006. Precision Viticulture: A New Era in Vineyard Management and Wine Production.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Sozzi, M., A. Kayad, D. Giora, L. Sartori, and F. Marinello. 2019. "Cost-Effectiveness and Performance of Optical Satellites Constellation for Precision Agriculture." Precision Agriculture '19. https://doi.org/10.3920/978-90-8686-888-9_62.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Thwal, Nyein Soe, Takaaki Ishikawa, and Hiroshi Watanabe. 2019. "Land Cover Classification and Change Detection Analysis of Multispectral Satellite Images Using Machine Learning." Image and Signal Processing for Remote Sensing XXV. https://doi.org/10.1117/12.2532988.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." International Journal of Biological Macromolecules 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

TABLES AND FIGURES

Table 1. Group, Accuracy, and Loss value uses 8 columns with 8 width data for Multispectral Satellite Image Segmentation

SI.NO	Name	Type	Width	Decimal	Columns	Measure	Role
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input
3	Loss	Numeric	8	2	8	Scale	Input

Table 2. Accuracy and Loss Analysis of Linear Regression and Random Forest.

S.No	GROUPS	ACCURACY	LOSS
1	Linear Regression	**80.04**	**12.30**
		79.12	**18.64**
		76.26	**16.11**
		80.00	**19.53**
		74.65	**22.00**
		71.65	**13.00**
		74.00	**11.87**
		73.00	**25.75**
		80.97	**24.75**

		70.00	**12.45**
2	Random Forest	**74.07**	**25.75**
		69.65	**24.75**
		69.00	**12.45**
		71.45	**23.45**
		74.00	**12.45**
		73.09	**11.25**
		68.45	**15.12**
		74.11	**16.23**
		70.00	**15.25**
		70.45	**11.03**

Table 3. Group Statistical Analysis of Linear Regression and Random Forest. Mean, Standard Deviation and Standard Error Mean are obtained for 10 samples. Linear Regression has higher mean accuracy and lower mean loss when compared to Random Forest.

	GROUP	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
ACCURACY	Naive Bayes	**10**	**75.9690**	**38.9529**	**1.23180**
	Linear Regression.	**10**	**71.4870**	**23.2110**	**.73400**
LOSS	Naive Bayes	**10**	**15.8450**	**44.0643**	**1.39343**
	Linear Regression.	**10**	**16.7730**	**57.2529**	**1.81050**

Table 4. Independent Sample T-test: is insignificantly Naive Bayes better than Random Forest with p value 0.053 (Two tailed, p>0.05).

		F	Sig.	t	df	Sig (2-tailed)	Mean Diffencene	Std.Error difference	Lower	Upper
ACCURACY	**Equal variances assumed**	**1.170**	**.053**	**5.665**	**18**	**.000**	**8.60100**	**1.60100**	**5.54060**	**14.15140**
	Equal Variances not assumed			**5.665**	**16.893**	**.000**	**8.60100**	**1.69468**	**4.02380**	**14.17820**
LOSS	**Equal variances assumed**	**.745**	**.053**	**-406**	**18**	**689**	**-52800**	**2.28463**	**-3.72784**	**1.87184**
	Equal Variances not assumed			**-406**	**16893**	**690**	**-82800**	**2.28463**	**-3.75049**	**2.89449**

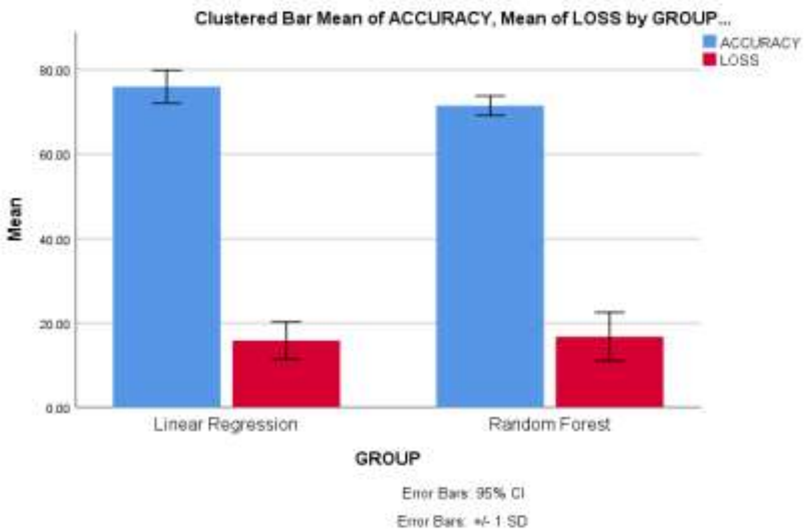

Fig. 1. Represents the mean accuracy of the software effort estimation for Linear Regression and Random Forest. The Linear Regression obtained 80.40% accuracy and the Random Forest obtained 74.07% accuracy. The Linear Regression achieved better than Random Forest. Mean Accuracy with +/- 1 SD.