Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Spl Issue [5] 2022 : 398-406

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
REVIEW ARTICLE **OPEN ACCESS**

Supervised Classification– An Overview of Machine Learning in Agricultural Practices

Susanta Kundu*[1], Vinod Kumar[2]
1,2 Department of Computer Science and Engineering,
Shree Guru Gobind Singh Tricentenary University – Gurugram, India

ABSTRACT
In recent years, machine learning algorithms and various areas of agriculture have aligned very closely with the availability of flexible programming tools (like Python and R) in terms of code simplicity, well-developed libraries, support from the open-source community, access to the high-end computing platform, more accessible access to public domain data. The process involves adequate studies to evaluate ML algorithms' performance in various agricultural application domains. The paper reviews applications of data science in agriculture involving AI/ML in a few representative areas like the dynamism of Land use and Land cover (LULC), food balance in agricultural exports; forest vegetation; groundwater resources; and soil moisture. The applications considered classification and prediction techniques in supervised ML algorithms, like Random Forest (RF), Extreme Gradient Boosting (EGB), Support Vector Machine (SVM), Decision Tree (DT), Normal Bayes (NB), and Long Short-Term Memory (LSTM), and compared related classification performance algorithms. The outcome is a comparative reference for researchers in these domains.
KEYWORDS DEM, *Food balance, Groundwater, LULC, Soil moisture*

Received 22.10.2022 Revised 23.11.2022 Accepted 20.12.2022

INTRODUCTION
Data science is a multidisciplinary approach to extracting information from large and voluminous data collected or created from relevant applications. The process involves preparing data for analysis, processing advanced research to reveal patterns, presenting the results, and enabling stakeholders to draw informed conclusions[1].With the availability of faster computing platforms and well-developed algorithms, Machine Learning offers various approaches to extracting information from remote sensing imageries [2].
For data analysis, data has to undergo cleaning and aggregation before processing using appropriate implementation tools. The analysis involves algorithms to deploy AI/ML/DL models to find information from data and transform identified patterns into predictions in support of decision-making processes. Scientifically designed tests validate the accuracy of predictions. The patterns and trending results are communicated to the target audience through data visualization tools [2].
AI/ML emerged well with high-performance computing platforms supporting data-intensive, multidisciplinary agricultural practices for yield prediction, crop management, weed detection, disease detection, species recognition, crop quality, and many more application areas[3].

MATERIAL ND METHODS
To strengthen the review process, literature surveys are essential for analytical studies. The process was substantiated by an advanced search on practice areas like LULC, food security, forest vegetation, groundwater, and soil moisture.
For journal search, a combination of Machine Learning AND (Practice Area) was considered. Each Practice area (refer to table 1) was tried one at a time. The search outcome, generated from the J gate plus database of global e-journals is captured in Table 1.

Table 1. Journal Search Outcome from Jgate database

Sr #	Practice Area	All	Full Text	Text with ML performance
1	LULC	134	77	44
2	Food security	249	141	49
3	Forest vegetation	12	5	1
4	Groundwater	444	188	78
5	Soil moisture	391	217	85
	Total	1230	628	257

Out of 1230 relevant documents listed from the journal search, 628 were full texts. The count was reduced to 257 when ML performance was added (AND).The available literature list was further shortlisted to28for analysis and performance accuracy in machine learning algorithms.
These are open access journals having relevance to the application areas under consideration.

SUPERVISED MODELS

Machine learning (ML) techniques are a promising alternative to process remote sensing data and are applied in many data processing and analysis tasks [2]. These techniques are classified primarily into supervised, and unsupervised learning models[4], that can benefit the end users like the farmers, in minimizing losses in the farming through recommendations and insights about the crops, if shared in time. ML algorithms are used in many applications, with suitable performance metrics appropriate for each domain[3]. The ML methods allow establishing non-parametric and nonlinear relationships between parameters (dependent and dependent variables), resulting in overall performance improvement when compared with conventional linear models [5].
A computing platform facilitates the classification and regression analysis for supervised models, clustering, and association for unsupervised models. The analytical studies in the time domain, and behavioral patterns during variance findings, log, and factor studies, originate from the artificial neural network (ANN) [2].
Unlike other areas of data science applications, the modeling process has several overlapping operations applicable in agricultural management steps. Representative steps in these agricultural domains are capturing raw data from appropriate sources; data preparation through cleaning, deduplication, and reformatting; data analysis for prediction and regression. The findings are finalized in a report or chart through data visualizations, easing the decision-makers to understand the impact [6].
The two functions involved in these processes are classification and regression. Classification predicts the object classes of unknown labels. It derives a model defining and differentiating data classes. The derived model is subjected to training with data of known class labels. Regression Analysis is for prediction, focused on the identification of distribution trends based on available data. The core transformations can be estimated by the supervised classification[7]. Supervised learning models require input at human intervention, with a feedback provision to enhance the prediction accuracy.
Supervised learning approaches are simple and widely used in data classification processes [8] of agricultural applications.Multi-spectral satellite images are used to classify land into a built-up area, bare soil, water body, vegetation, and wetland. Such classification help to conduct change detection analysis or time series analysis [9].

IMPLEMENTATION TOOLS

The benefits of data science in agricultural practices explore data-driven algorithms to build, train, and evaluate the machine learning models with agriculture-related data. The choice of computational platform and the programming language play a pivotal role in ML applications. There are several choices available in this regard to meet the requirements easily, and provide AI algorithms commonly used and discussed [10].
In its implementation, Python, as a popular programming language, has many appropriate libraries contributed by the data science community. The popular and significant python libraries consist of NumPy for scientific computing, Pandas for data analysis, Skit-learn, and Tensor Flow for machine learning and deep learning applications. Other Python packages allow suitable development and production environments for agricultural application-specific domains.
The statistical data analysis summarizes the key characteristics discovering attributes and trends with visual methods. Data exploration helps assess quality, quantity, and first-level features by data interpretation, visualizing in plot, histogram, or chart for researchers to study and analyze further. ML algorithms are chosen extensively for classifying satellite images to map the relevant Earth surface.
Among the popular ML algorithms, Support Vector Machine (SVM) and Random Forest (RF) emerges with improved accurate results [9].

The advantages and disadvantages of commonly used ML methods are discussed based on the widely adopted practices put by researchers for data classification and regression.

Table 2. Application, highlights of ML algorithms

Application	Methodology	Highlights	Reference
Predicted species distribution	RF	Improves prediction accuracy	[11]
	SVM	Works well with limited experimental data	
Prediction of carbon and energy flux	ANN	Effective with traditional models to reduce uncertainty predictions	[11]
		Has excellent data mining capacity	
Hazard assessment and prediction	ANN	Can deal with data of higher dimensions	
		Adaptability to mapping nonlinearity	
	SVM	Efficient in solving classification areas	
Forest management	ANN	Decent forecasting volume of wood, the height of trees, biomass, etc.	
	RF	Potential in hybrid modeling RF and spatial interpolation	

SVM is easier to model nonlinear decision boundaries with options to choose from many kernels, allowing robust fitments. It avoids over fitting, particularly for higher-dimensional data spaces. As a feature, it can train models with fewer but meaningful pixels with the ability to accommodate limited information [4]. SVM is memory-intensive and tricky to tune by selecting the appropriate kernel. It will result ina poor extrapolation model in case of inconsistent prior data, and model dependency on previous support vector records.
Decision Trees (DT) are part of SVM having decision nodes, and data split continuously based on certain criteria. The tree provides non-linear associations with parameters without affecting the tree performance. It is a tree-shaped structure, representing a set of decisions, generating rules for the classification of a dataset. With lesser effort in its data preparation, Tree methods are easier to understand and interpret for analytics. It is database size-independent and usable as an extension to a large database. The classification and Regression Tree have a lesser impact on outliers in its output.
Random Forest under the SVM learning algorithm is used for both classification and regression. It runs by creating a host of decision trees during training. For classification criteria, the output of the Random Forest (RF) is the class selected by most trees. Random Forests are noted toreduce the overfitting risks effectively.
ANN is used in learningmodels for bothcomplex and non-linear data with high fault tolerance toward noisy data and the ability to parallel processes. ANNs cannot identify the relative importance and effects of individual environmental variables [4].

AGRICULTURAL PRACTICES

With the advent of analytical solutions, monitoring the farms, soil, plant activities, etc. are now easier to manage with additional clarity developed in agricultural systems. The enormous data generated from various wireless sensors used in the agricultural domain canbe processed to discover patterns paving paths for the farmers to make logical decisions in enhancing farm production and improved quality[10].
For analysis of agricultural practices, field data collected from agricultural farms are classified by SVM to classify elements like the type of crop identification. Such classification and linear regression results in higher class accuracies, typically of the order 88–99%. Comparing two entities, such as crop qualities; by paired t-tests often differentiates known and predicted elements[12]. Differences between known and predicted mean are usually measured by root mean square error (RMSE). The results of studies and previous research depicted emerging practices and estimations of various parameters for effective monitoring and verifying agricultural management practices over spatial distribution (over larger areas)with temporal changes(distributed over longer durations)[13].
The following subsections cater to a few representative application areas illustrating data science usages under the agricultural scope of emerging digital technologies in agriculture[14].

The dynamism of LULC and food balance

Human interventions in land usage for progress and growth is a continuous process ever since the inception of human civilization. Expansions in farming land for societal causes have seen significant

demographic transformation due to agricultural development. Agricultural trade plays a crucial role in farmers' incomes and associated people engaged in the food supply chain. It reduces food imbalance globally and allows wider choice in buyer merchandise, having grown significantly over the last two decades between 2001 and 2019. It adds global value to agricultural and food processing chains spread over several countries by linking agro-food sectors to the economy across the globe [15].
Precise month-on-month predictions of agricultural exports from a country are crucial to facilitate import-export planning and domestic consumption, resulting in a balance between production and marketing [16]. Researchers worked on predicting agricultural exports in the time domain using long short-term memory (LSTM). The research process explored Purchasing Managers' Index (PMI) of a few industries using LSTM to predict agricultural export trends. The results included keyword vectors in the PMI of the insurance and finance domain, that impact the prediction of agricultural export results. The agribusiness operators and policymakers can assess and regulate domestic and foreign production and sales on such prediction results [16].
Land cover is the scope of associating physical areas and using items on the earth's surface. Land use refers to social and economic causes associated with the overall human development process. The dynamism in LULC, added by the human interventions, has multiple bearingsin demographic transformation and related consequences on agricultural production [14].
Lesser vegetation leads to soilsusceptibility towardssoil erosion and degradation due to reduced water retention capability. Supervised classification models of spectral images obtained from satellite images can identify built-up areas, bare soil, vegetation, wetland, water body, etc. along with change detection analysis in time series. Results and analysis of such built-up area and related dynamics over a few decades in time frame can reveal the undergoing urban expansion processes [13].
The LULC assessed by multispectral (MS) satellite images produces land classification [17].In a study from south-central Sweden, the researchers collected stratified random samples from Sentinel-2 images of mixed-use areas. The samples, segmented into training and evaluation sets in a 70: 30 ratio for data modeling had an overall accuracy in SVM (0.758 ± 0.017), EGB (0.751 ± 0.017), RF (0.739 ± 0.018), and DL (0.733 ± 0.0023) [11].
Mapping vegetation is a type of LULC classification using MS data from large areas captured by remote sensing satellite images. Remote sensing as a prevailing tool acquires information about an object or phenomenon by measuring emitted and reflected radiation. It facilitates digital images, with an environment to perform multiple image operations. Advanced image classification techniques can map vegetation and land uses like urban land, forest, and water [18].
A study over Dehradun analyzed spatial distributions of LULC [19] in and around the city area using Landsat-8 and Sentinel 2 images. The algorithms used in the study were RF, SVM, Classification and Regression Tree (CART) [20], and Gradient Tree Boosting (GTB) [21], for both the satellite images. The outcome compared the performance of supervised ML algorithms on image classification.The accuracy measurement of the algorithms followed a confusion matrix and the Kappa calculation method. The accuracy measurement in CART was 89.24% to 93.52% and Kappa coefficient 85.64% to 91.36%; RF had 91.45% to 95.86% accuracy and Kappa coefficient 88.59% to 94.48%, GTB produced 87.71% to 95.33% accuracy and Kappa coefficient 83.58% to 93.37%. In comparison, SVM had an accuracy measurement of 73.54% to 84.96% and Kappa coefficient 76.28% to 79.99% [19] using Sentinel-2 and Landsat-8 data.
LULC analysis helps to predict land-use changes over a period. The present studies indicate a probable saturation between food security and environmental conservation during the next few decades, with an equilibrium setting on agricultural production and urban expansionthrough land resource conservation [22].

Mapping Forest Vegetation

Mapping forest vegetation is an essentialprocess to provide data on the surficial changes of specified areas[23]. Different ML algorithms offer to extract information from the surfaces using multi-spectral remote sensing footprints of these areas. Studies on ML applications for mapping forest vegetation present a framework of environmental planning with relevant data models[4]. Performance of ML algorithms like RF, SVM, DT (DecisionTree), and NB (Normal Bayes) evaluates these models. The outcome models validated by DT have the highest levels of accuracy in related activities [24]./ / /
Maps of trees in forests are required to estimate terrestrial carbon resources, forest records, and predict tree mortality after the wildfire. Statistical methods and modified RF are used for analytics and classification. Treesare mapped in a forest using biophysical landscape characteristics with data collected from 30 × 30 m spatial resolution to produce tree-level maps. The results for forest cover were within the class agreement of 79%, forest height of 96%, and vegetation group of 92% [25].

The rapid advancement of remote sensing technology has expanded the choice of imagery sources in agricultural practices. These images when used to map vegetation needs to align with different processes, considerations, and techniques. These image sources have varied properties in their spatial, spectral, temporal, and radioactive characteristics, making them viable to use for appropriate vegetation and mapping purposes. For vegetation classification, vegetation mapping at a species or community level forms the initial steps. It follows the vegetation type correlations (communities or species) with noticeable spectral characteristics from remote sensed (RS) imageries identified. Such spectral classes formed from the RS images are translated into the vegetation types using image interpretation and processing.

Ground Water Resources

Groundwater recharge is essential for agricultural fields, grasslands, and forests[26]. An assessment of measurable groundwater recharge is necessary, considering inevitable land cover changes by natural disasters[27]. Researchers adopted MS satellite images and Geographic information system (GIS) data to classify groundwater areas. The maximum likelihood classifier of a supervised model produced a kappa coefficient of 0.88 and an accuracy of 91.7%[28]. The topsoil loss due to landslides impacts groundwater recharge and will not improve fast in normal and natural conditions[29]. It takes nearly 100 years for the groundwater recharge to regain normalcy[30]. Efforts should be made to overcome the loss due to natural calamities.

The groundwater availability is influenced by multiple parameters like soil, rainfall, drainage density, and LULC. A study explored potential groundwater zones in India using geospatial techniques for Ranchi district, Jharkhand (India), using DEM, Landsat 8 images. The study identified priority areas to develop strategies for sustainable groundwater development [31].

Soil Moisture and Nutrients

For many years, soil moisture has been crucial in agriculture, climatology, and hydrology. In the present-day context, precision farming depends on the trend and pattern of soil moisture content at the root zone of the concerned area[32].

Researchers evaluated soil moisture at 0–40 cm depth with a time-series soil moisture data of approximately 40 years[33]. Descriptive and inferential statistics applied to GIS data analysis resulted in the mean soil moisture of the specified study area. The soil moisture has an increasing trend in some seasons with changes in climatic conditions. It also follows a time-dependent (temporal) pattern that is not statistically significant. However, the spatial variation was statistically significant for different soil classes. The study revealed apotential link between GIS and remote sensing resources to assess global environmental changes [32].

Soil moisture is a significant climate variable also[34]. In a study in China, researchers explored soil moisture at 0–10 cm depth from 1948 to 2014 with the adjoining air temperature utilizing statistical analysis. The results indicated a significant decrease in yearly soil moisture ($p < 0.01$) in multiple study areas. The reduction in soil moisture trend was seasonal. It was lower in summer and winter than in spring. It also revealed that soil moisture at 0–10 cm negatively correlated with air temperature 2m above the ground in certain areas with higher temperatures. The combined results may improve the understanding of climatic changes and variations in regional soil moisture [35], although the climate change effects on agricultural productivity and vegetation are not uniform over time and space.

ML techniques help to create comprehensive models to evaluate soil nutrient content. The evaluation methods for soil nutrients, based on SVM and ANN can create accurate soil nutrient content estimation, that has higher impacts on forest regeneration and plant growth process[36]. The performance evaluated by these models indicates higher model efficiencies reflected in MSPE, ME, and RMSE with smaller values.

Soil Organic Carbon Stock

The land-use changes show that converting forest land and wetland to cropland caused SOCS loss from the soil. Such changes challenge our ecosystem, adversely impacting SOCS[37].The predictability of SOCS changes requires technological advancements in collecting and processing satellite-based remote sensing data[38]. It also involves soil data from different regional and global databases in published documents or databases to supplement SOCS predictability analysis[39].Soil samples collected from study plot areas are primary field data elements tagged with positional information (latitude, longitude)for geo-referencing. Under standard laboratory procedures, the soil samples were examined for SOC-related tests using total oxidized carbon and wet oxidation measurement as per standard laboratory procedures [40].

Remote Sensing data from satellite images supplements estimating SoC Stock (SOCS)trained models. The performance accuracy of predictive statistical models, such as random forest, partial least square

regression, etc., can be likened to assessing the SOC content. Predicting soil organic carbon in spatial distribution for land management relates to carbon emission and soil health [41]. Performance comparison of these models suggests that the predictability is not entirely dependent on spectral behavior[42]. SOC prediction using a Convolutional Neural Network (CNN) is encouraging compared with Random Forest (RF) with multiple environmental factors. The results show improved accuracy of CNN with added land surface varieties and other environment variables. CNN had higher prediction performance than RF regardless of the variables added, indicating CNN is a good soil mapping approach in a regional setup [43].
SOCS is one primary dependent variable for SOC estimation, with local climate, terrain conditions, and spectral parameters independent variables, depending on the study scenario[39].
Prediction accuracy depends on various factors, like soil condition, variability of input characteristics, surface area with varying moisture, roughness due to vegetation and crop, and atmospheric conditions during image acquisition. [44]
The precision of prediction models based on data from multispectral bands of different satellites was recorded as RPD between 1.4 to 3.1 and RPIQ between 1.8 to 2.1 [18].Prediction accuracy is obtained by ML/DL model or techniques or approaches bythe Ratio of Performance to Deviation (RPD), coefficient of determination (R2), and Root Mean Square Error (RMSE) [45].

RESULTS

LULC performance results captured in Table 3 relateto applications in LULC landscape classification and mapping vegetation using Landsat-8 and Sentinel-2 data for a matrix view.

Table 3. Performance Accuracy Matrix

Training Model	Accuracy	Accuracy check by	RS Satellite	Reference
SVM	0.758± 0.017	Confusion Matrix	Sentinel-2	[11]
EGB	0.751± 0.017	Confusion Matrix	Sentinel-2	[11]
RF	0.739± 0.018	Confusion Matrix	Sentinel-2	[11]
DL	0.733± 0.0023	Confusion Matrix	Sentinel-2	[11]
CART	0.8924	Confusion Matrix	Sentinel-2	[19]
RF	0.9145	Confusion Matrix	Sentinel-2	[19]
GTB	0.8771	Confusion Matrix	Sentinel-2	[19]
SVM	0.7354	Confusion Matrix	Sentinel-2	[19]
CART	0.9352	Confusion Matrix	Landsat-8	[19]
RF	0.9586	Confusion Matrix	Landsat-8	[19]
GTB	0.9533	Confusion Matrix	Landsat-8	[19]
SVM	0.8496	Confusion Matrix	Landsat-8	[19]
CART	0.8564	Kappa coefficient	Sentinel-2	[19]
RF	0.8859	Kappa coefficient	Sentinel-2	[19]
GTB	0.8358	Kappa coefficient	Sentinel-2	[19]
SVM	0.7628	Kappa coefficient	Sentinel-2	[19]
CART	0.9136	Kappa coefficient	Landsat-8	[19]
RF	0.9448	Kappa coefficient	Landsat-8	[19]
GTB	0.9337	Kappa coefficient	Landsat-8	[19]
SVM	0.7999	Kappa coefficient	Landsat-8	[19]

A confusion matrix describes the performance of a classification model for which True-False values are known. The matrix is organized as shown in Table 3 below.

Table 4. Confusion Matrix

		Actual Values	
		True (T)	False (F)
Predicted values	True (T)	TT	TF
	False (F)	FT	FF

In a confusion matrix, the overall accuracy is (All Correct / All) = (TT + FF) / (TT + TF + FT + FF).
With Sentinel-2 data, SVM performed better in LULC when compared with EGB RF, and DL.Kappa coefficient estimates the degree of agreement between two data sets collected on different instances. In the case of performance results of various algorithms, Landsat-8 and Sentinel-2 data were used as two separate instances. Landsat-8 had better results for LULC performance.

DISCUSSION

Data science has several unsolved challenges relating to data inadequacy, developing reasoning skills, and understanding architectures with solutions. Next-level challenges are modeling physical phenomena using heterogeneous data and developing algorithms relating to spatial, temporal, and spectral data.
Although the initial and available RS framework caters to most research tools for analysis, more work is needed in analyzing applications. It requires skills in signal processing, electromagnetics, and multisensory systems. These all add to a genuinely multidisciplinary approach [2].
ML techniques are useful in identifying important features and choosing appropriate environment variables. It also improves the model's accuracy to predict ecological phenomena. The decision-tree-based random forest is an efficient machine-learning model tested in many studies. These decision tree models are efficient in managing numerous variables when compared with other parametric models to address multiple collinear ties[44].
When classification and regression tasks are compared, the regression algorithm is noted as more accurate and performed better than classification algorithms [18].

CONCLUSION

In agricultural practices, data available from multiple sources are analyzed to arrive at a decision. Data plays a significant role in the process of classification and related regression accuracy. Similar data from different sources may increase the scope of ambiguity while analyzing agricultural applications of measured accuracy. The results from different case studies indicate dynamism in measured accuracies in classification and regression. The attributes can be due to prolonged change processes in agricultural farm data, the time to validate the model depends onnatural processes like seasonal crop growth, crop removal/cutting, etc.
The prospective models predict data and advancement, using machine learning as a statistical tool. It may not be a cure-all solution. Researchers are studying various areas and factors to reinforce results through a systematic understanding, avoiding unauthentic relations. The results of such studies indicate the possibility to predict the ecological phenomena, while simultaneously improving the accuracy of such models [44] with automated methodologies derived from the most appropriate environmental variables.

ACKNOWLEDGMENT

The authors acknowledge the support from the SGT University for writing the paper.

Conflicts of interest: The authors declare no conflicts of interest.

REFERENCES

1. C. Ley and S. P. A. Bordas, “What makes Data Science different? A discussion involving Statistics2.0 and Computational Sciences,” *Int J Data Sci Anal*, vol. 6, no. 3, pp. 167–175, Nov. 2018, doi: 10.1007/s41060-017-0090-x.
2. J. E. Ball, D. T. Anderson, and C. S. Chan, “Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community,” *J. Appl. Remote Sens*, vol. 11, no. 04, p. 1, Sep. 2017, doi: 10.1117/1.JRS.11.042609.
3. K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine Learning in Agriculture: A Review,” *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/s18082674.
4. Z. Liu, C. Peng, T. Work, J.-N. Candau, A. DesRochers, and D. Kneeshaw, “Application of machine-learning methods in forest ecology: recent progress and future challenges,” *Environ. Rev.*, vol. 26, no. 4, pp. 339–350, Dec. 2018, doi: 10.1139/er-2018-0034.

5. P. Feng, B. Wang, D. L. Liu, and Q. Yu, "Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia," *Agricultural Systems*, vol. 173, pp. 303–316, Jul. 2019, doi: 10.1016/j.agsy.2019.03.015.
6. K. K. Kurt, H. T. Atay, M. A. Cicek, S. B. Karaca, and S. Turkeli, "The Importance of Multidisciplinary in Data Science: Application of the Method in Health Sector to Telecommunication Sector," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Turkey, Sep. 2019, pp. 1–4. doi: 10.1109/IDAP.2019.8875917.
7. S. Paul, "Change detection and Future Change Prediction in Habra I and II block Using Re-mote Sensing and GIS – A case Study," *International Journal of Engineering and Geosciences*, Sep. 2021, doi: 10.26833/ijeg.975222.
8. R. Saravanan and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, Jun. 2018, pp. 945–949. doi: 10.1109/ICCONS.2018.8663155.
9. C. Avci, M. Budak, N. Yağmur, and F. Balçik, "Comparison Between Random Forest and Support Vector Machine Algorithms for LULC Classification," *International Journal of Engineering and Geosciences*, Nov. 2021, doi: 10.26833/ijeg.987605.
10. C. N. Vanitha, N. Archana, and R. Sowmiya, "Agriculture Analysis Using Data Mining And Machine Learning Techniques," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2019, pp. 984–990. doi: 10.1109/ICACCS.2019.8728382.
11. A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.
12. R. S. Bricklemyer, R. L. Lawrence, P. R. Miller, and N. Battogtokh, "Monitoring and verifying agricultural practices related to soil carbon sequestration with satellite imagery," *Agriculture, Ecosystems & Environment*, vol. 118, no. 1–4, pp. 201–210, Jan. 2007, doi: 10.1016/j.agee.2006.05.017.
13. R. Jimoh, Y. Afonja, C. Albert, and N. Amoo, "Spatio-Temporal Urban Expansion Analysıs in a Growing City of Oyo Town, Oyo State, Nigeria Using Remote Sensing And Geographic Information System (GIS) Tools," *International Journal of Environment and Geoinformatics*, vol. 5, no. 2, pp. 104–113, Aug. 2018, doi: 10.30897/ijegeo.354627.
14. H. S. James Jr, "Introduction to the Handbook on the Human Impact of Agriculture," in *Handbook on the Human Impact of Agriculture*, Edward Elgar Publishing, 2021, pp. 1–11. doi: 10.4337/9781839101748.00008.
15. "Monitoring the changing landscape of agricultural markets and trade," *oecd*, [Online]. Available: https://www.oecd.org/agriculture/topics/agricultural-trade/
16. J.-L. Xu and Y.-L. Hsu, "Analysis of agricultural exports based on deep learning and text mining," *J Supercomput*, vol. 78, no. 8, pp. 10876–10892, May 2022, doi: 10.1007/s11227-021-04238-w.
17. H. T. Abd El-Hamid, W. Caiyong, Z. Yun, and A. M. El-Zeiny, "Environmental hazards of land use/land cover dynamics using data observation techniques and GIS: case study of Guyuan watershed, China," *Arab J Geosci*, vol. 15, no. 5, p. 426, Mar. 2022, doi: 10.1007/s12517-022-09655-6.
18. I. Azpiroz, N. Oses, M. Quartulli, I. G. Olaizola, D. Guidotti, and S. Marchi, "Comparison of Climate Reanalysis and Remote-Sensing Data for Predicting Olive Phenology through Machine-Learning Methods," *Remote Sensing*, vol. 13, no. 6, p. 1224, Mar. 2021, doi: 10.3390/rs13061224.
19. A. Srivastava, S. Bharadwaj, R. Dubey, V. B. Sharma, and S. Biswas, "MAPPING VEGETATION AND MEASURING THE PERFORMANCE OF MACHINE LEARNING ALGORITHM IN LULC CLASSIFICATION IN THE LARGE AREA USING SENTINEL-2 AND LANDSAT-8 DATASETS OF DEHRADUN AS A TEST CASE," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B3-2022, pp. 529–535, May 2022, doi: 10.5194/isprs-archives-XLIII-B3-2022-529-2022.
20. W. Loh, "Classification and regression trees," *WIREs Data Mining Knowl Discov*, vol. 1, no. 1, pp. 14–23, Jan. 2011, doi: 10.1002/widm.8.
21. B. M. Sreedhara, A. P. Patil, J. Pushparaj, G. Kuntoji, and S. R. Naganna, "Application of gradient tree boosting regressor for the prediction of scour depth around bridge piers," *Journal of Hydroinformatics*, vol. 23, no. 4, pp. 849–863, Jul. 2021, doi: 10.2166/hydro.2021.011.
22. L. Zhen *et al.*, "Future land use and food security scenarios for the Guyuan district of remote western China," *iForest*, vol. 7, no. 6, pp. 372–384, Dec. 2014, doi: 10.3832/ifor1170-007.
23. G. Reyes-Palomeque, J. M. Dupuy, C. A. Portillo-Quintero, J. L. Andrade, F. J. Tun-Dzul, and J. L. Hernández-Stefanoni, "Mapping forest age and characterizing vegetation structure and species composition in tropical dry forests," *Ecological Indicators*, vol. 120, p. 106955, Jan. 2021, doi: 10.1016/j.ecolind.2020.106955.
24. D. E. G. Furuya *et al.*, "A Machine Learning Approach for Mapping Forest Vegetation in Riparian Zones in an Atlantic Biome Environment Using Sentinel-2 Imagery," *Remote Sensing*, vol. 12, no. 24, p. 4086, Dec. 2020, doi: 10.3390/rs12244086.
25. K. L. Riley, I. C. Grenfell, and M. A. Finney, "Mapping forest vegetation for the western United States using modified random forests imputation of FIA forest plots," *Ecosphere*, vol. 7, no. 10, Oct. 2016, doi: 10.1002/ecs2.1472.
26. Y. Ouyang, W. Jin, J. M. Grace, S. E. Obalum, W. C. Zipperer, and X. Huang, "Estimating impact of forest land on groundwater recharge in a humid subtropical watershed of the Lower Mississippi River Alluvial Valley," *Journal of Hydrology: Regional Studies*, vol. 26, p. 100631, Dec. 2019, doi: 10.1016/j.ejrh.2019.100631.
27. S. O. Owuor *et al.*, "Groundwater recharge rates and surface runoff response to land use and land cover changes in semi-arid environments," *Ecol Process*, vol. 5, no. 1, p. 16, Dec. 2016, doi: 10.1186/s13717-016-0060-6.

28. H. Amano and Y. Iwasaki, “Estimation of Groundwater Recharge in Kumamoto Area, Japan in 2016 by Mapping Land Cover Using GIS Data and SPOT 6/7 Satellite Images,” *Sustainability*, vol. 14, no. 1, p. 545, Jan. 2022, doi: 10.3390/su14010545.
29. A. Kumawat, D. Yadav, K. Samadharmam, and I. Rashmi, “Soil and Water Conservation Measures for Agricultural Sustainability,” in *Soil Moisture Importance*, R. Swaroop Meena and R. Datta, Eds. IntechOpen, 2021. doi: 10.5772/intechopen.92895.
30. M. F. P. Bierkens and Y. Wada, “Non-renewable groundwater use and groundwater depletion: a review,” *Environ. Res. Lett.*, vol. 14, no. 6, p. 063002, Jun. 2019, doi: 10.1088/1748-9326/ab1a5f.
31. M. Kumar *et al.*, “GIS-based multi-criteria approach to delineate groundwater prospect zone and its sensitivity analysis,” *Appl Water Sci*, vol. 12, no. 4, p. 71, Apr. 2022, doi: 10.1007/s13201-022-01585-8.
32. O. A. Ibanga, O. F. Idehen, and M. G. Omonigho, “Spatiotemporal variability of soil moisture under different soil groups in Etsako West Local Government Area, Edo State, Nigeria,” *Journal of the Saudi Society of Agricultural Sciences*, vol. 21, no. 2, pp. 125–147, Feb. 2022, doi: 10.1016/j.jssas.2021.07.006.
33. W.-Z. Zeng *et al.*, “Estimating Root Zone Moisture from Surface Soil Using Limited Data,” *Ecological Chemistry and Engineering S*, vol. 24, no. 4, pp. 501–516, Dec. 2017, doi: 10.1515/eces-2017-0033.
34. H. Griesbauer, S. C. DeLong, B. Rogers, and V. Foord, “Growth Sensitivity to Climate Varies with Soil Moisture Regime in Spruce–Fir Forests in Central British Columbia,” *Trees*, vol. 35, no. 2, pp. 649–669, Apr. 2021, doi: 10.1007/s00468-020-02066-8.
35. G. Han *et al.*, “Temporal and Spatial Variation of Soil Moisture and Its Possible Impact on Regional Air Temperature in China,” *Water*, vol. 12, no. 6, p. 1807, Jun. 2020, doi: 10.3390/w12061807.
36. C. Wu, Y. Chen, X. Hong, Z. Liu, and C. Peng, “Evaluating soil nutrients of Dacrydium pectinatum in China using machine learning techniques,” *For. Ecosyst.*, vol. 7, no. 1, p. 30, Dec. 2020, doi: 10.1186/s40663-020-00232-5.
37. A. Suleymanov, I. Gabbasova, R. Suleymanov, E. Abakumov, V. Polyakov, and P. Liebelt, “Mapping soil organic carbon under erosion processes using remote sensing,” *HunGeoBull*, vol. 70, no. 1, pp. 49–64, Apr. 2021, doi: 10.15201/hungeobull.70.1.4.
38. L. Poggio *et al.*, “SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty,” *SOIL*, vol. 7, no. 1, pp. 217–240, Jun. 2021, doi: 10.5194/soil-7-217-2021.
39. E. G. Jobbágy and R. B. Jackson, “THE VERTICAL DISTRIBUTION OF SOIL ORGANIC CARBON AND ITS RELATION TO CLIMATE AND VEGETATION,” *Ecological Applications*, vol. 10, no. 2, pp. 423–436, Apr. 2000, doi: 10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.
40. D. Žížala, R. Minařík, and T. Zádorová, “Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions,” *Remote Sensing*, vol. 11, no. 24, p. 2947, Dec. 2019, doi: 10.3390/rs11242947.
41. M. Ließ, J. Schmidt, and B. Glaser, “Improving the Spatial Prediction of Soil Organic Carbon Stocks in a Complex Tropical Mountain Landscape by Methodological Specifications in Machine Learning Approaches,” *PLoS ONE*, vol. 11, no. 4, p. e0153673, Apr. 2016, doi: 10.1371/journal.pone.0153673.
42. E. Ben-Dor *et al.*, “Using Imaging Spectroscopy to study soil properties,” *Remote Sensing of Environment*, vol. 113, pp. S38–S55, Sep. 2009, doi: 10.1016/j.rse.2008.09.019.
43. L. Yang, Y. Cai, L. Zhang, M. Guo, A. Li, and C. Zhou, “A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102428, Oct. 2021, doi: 10.1016/j.jag.2021.102428.
44. P. Holloway, D. Kudenko, and J. R. Bell, “Dynamic selection of environmental variables to improve the prediction of aphid phenology: A machine learning approach,” *Ecological Indicators*, vol. 88, pp. 512–521, May 2018, doi: 10.1016/j.ecolind.2017.10.032.

CITATION OF THIS ARTICLE

Susanta Kundu, Vinod Kumar: Supervised Classification– An Overview of Machine Learning in Agricultural Practices. Bull. Env.Pharmacol. Life Sci., Spl Issue [5]: 2022: 398-406.