Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assessment of Machine Learning Algorithms for Detection of Forest Degradation in the Remote Sensing Data of Omo Forest Reserve

B. A. Babalogbon

maltracksuite@gmail.com

National Space Research and Development Agency

O. O. Awotoye

Obafemi Awolowo University

O. J. Matthew

Obafemi Awolowo University

P. C. Odoh

University of Ilesa

D. E. Ekpa

Obafemi Awolowo University

Research Article

Keywords: Forest degradation, Remote sensing, Machine learning, Omo Forest Reserve, Support Vector Classifier, Deforestation monitoring

Posted Date: July 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10395750/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study explores the application of machine learning (ML) algorithms for the automated detection of forest degradation using remote sensing data. Five machine learning models: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), and Support Vector Classifier (SVC) were trained and evaluated using multi-temporal satellite imagery and vegetation indices such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). A reference dataset was generated through remote sensing image interpretation and field validation to train and validate these models. The results revealed that all models performed well in distinguishing degraded and undisturbed forest areas, with accuracy scores ranging from 92% to 99%. Among the models, SVC demonstrated the highest performance, achieving 99% accuracy, followed by DT (98%) and RF (97%). The Receiver Operating Characteristic (ROC) Curve analysis further confirmed the superiority of these models in detecting forest degradation patterns. These findings highlight the effectiveness of ML-based remote sensing analysis as a reliable and scalable tool for monitoring forest health. Thus, it recommends the adoption of ML-driven forest assessment frameworks in Nigeria's conservation policies to enhance sustainable forest management.

Introduction

The forest ecosystems play a crucial role in maintaining environmental stability by regulating climate, mitigating pollution, regulating floods, preventing soil erosion and supporting biodiversity. The Food and Agriculture Organization (FAO) recommends a forest cover of 20–30% for every country globally (FAO, 2020). Deforestation is one of the threatening global development challenges and environmental problems facing the world and Nigeria today. Rapid deforestation poses significant threats to ecological balance, particularly in a region like Nigeria, where forests are under constant pressure from agriculture, logging, and urban expansion.

Nigeria has a declining forest cover, currently estimated to be 9.9% of its total land area, equivalent to approximately 9.2 million hectares Over the past few decades, Nigeria has witnessed rapid deforestation, with an annual deforestation rate of 3.7%, one of the highest in the world (Sambe et al., 2018). The drivers of deforestation in Nigeria include population growth, poverty, agricultural expansion, and excessive logging for timber and fuelwood (Oyediji & Adenika, 2022). Omo Forest Reserve, one of the most vital conservative areas in Nigeria is increasingly affected by these activities, necessitating efficient monitoring and management strategies.

Traditional methods of forest monitoring, such as field surveys, are time-consuming, labour-intensive, costly, and often impractical in large and inaccessible forested areas. Today, technological developments allow the monitoring and processing of large data. Thus, Nigeria must adopt modern technologies such as Remote Sensing, Geographic Information Systems (GIS) with machine learning (ML) algorithms for automated detection of forest degradation and enhanced management. This study

is aimed at assessing and comparing different machine learning algorithms for detecting forest degradation.

Methodology

Study area

Omo Forest Reserve (OFR) was gazetted in 1925 as part of the old Shasha Forest Reserve of southwestern Nigeria. It is located between latitudes 6^{o} 35' to 7^{o} 05' N and longitudes 4^{o} 19'to 4^{o} 40' E in Ijebu East and North Local Government Areas of Ogun State, South-western Nigeria (Fig. 1). The reserve covers an area of about 1,305km^2 (Amusa, 2015). The reserve has a tropical climate characterised by rainy and dry season (Larinde *et al.*, 2011; Adedeji et al., 2015).

Machine Learning and Generation of machine learning reference datasets

A total of nine variables/indicators/parameters was used for developing the ML models. These include the number of relevant bands in the Landsat remote sensing dataset (namely: red, green, blue, near infrared (NIR), shortwave infrared (SWIR), and thermal bands), the Biomass, Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) datasets. Undisturbed and degraded parts of the forest in the study area were identified through remote sensing image interpretation and fieldwork. Random points data covering the study area was generated in a GIS software, and subsequently overlaid on the time series and change analyses maps of Biomass, EVI and NDVI to identify and assign the forest degradation states for each of the points. Careful attention was also paid to data points that retained or changed their forest states throughout the period of 1985 to 2023 (1985, 2002, 2019 and 2023). Google Earth Engine and fieldwork were used to confirm and verify the forest states associated with each of these points (Talukdar et al., 2020). These point data served as reference for characterizing, modelling and validating the spectral behaviour of the identified forest states in the study area.

Development of trained machine learning models

Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), and Naïve Bayes (NB) were those considered for this study because they have been identified as common and mature machine learning models that are suitable for classification tasks (Maxwell et al., 2018). The scikit-learn toolbox (http://scikit-learn.org/), deployed to the Python programming language by Pedregosa et al., (2011), was used to develop these machine learning models for the detection of degraded forest regions in the study area. The parameters considered for developing these models were the individual bands of the remote sensing data that were related to the field-based biomass estimation. Eighty percent (80%) of the reference data representing both forest states was used in the training and export of the several machine-learning models, while the remaining twenty percent (20%) of the

reference datasets was used for the validation of the models. The implementation, training and validation of these machine-learning models were performed using the Python programming language.

Validations and selection of most suitable machine learning model

The accuracy of developed machine-learning models was validated using the remaining 20% of the reference dataset generated. Statistical indicators were interpreted to determine the level of accuracy of each of the developed machine learning models, and subsequently determined the most suitable ML model for detecting and monitoring forest degradation in the study area.

Results and Discussion

Machine Learning Models for Forest Degradation Detection in Omo Forest Reserve

Performance indices for each of the trained and validated machine learning model are shown in Table 1 (Logistic Regression, Decision Tree, and Radom Forest, Naïve Bayes and Support Vector Classifier). Precision indicates the proportion of true positive predictions among all positive predictions made by the model. It focuses on the model's ability to avoid false positives. Recall measures the proportion of true positive predictions among all actual positive instances in the dataset. It highlights the model's ability to detect all relevant instances (minimize false negatives). F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance that considers both false positives and false negatives. Accuracy measures the proportion of correct predictions out of the total predictions made. It provides an overall view of how well the model performed across all classes.

The Logistic Regression model showed high precision with perfect precision (1.00) for the "Disturbed" class but had a slightly lower recall of 0.85, leading to an F1-Score of 0.92. For the "Undisturbed" class, it achieved a precision of 0.86 and a perfect recall of 1.00, again resulting in an F1-Score of 0.92. The overall accuracy of this model was 0.92, with macro and weighted averages of 0.92. The Decision Tree model performed exceptionally well with both the "Disturbed" and "Undisturbed" classes, achieving identical scores: a precision, recall, and F1-Score of 0.98 for both classes. This consistency resulted in an overall accuracy of 0.98 and macro and weighted averages of 0.98 across all metrics, indicating a highly reliable model.

Table 1
Trained Machine Learning Models and their Performance Indices!

Machine Learning Models	Performance Indices				
	Class	**Precision**	**Recall**	**F1-Score**	**Support**
Logistic Regression	Disturbed	1	0.85	0.92	518
	Undisturbed	0.86	1	0.92	484
	Accuracy			0.92	1002
	Macro average	0.93	0.92	0.92	1002
	Weighted average	0.93	0.92	0.92	1002
Decision Tree	Disturbed	0.98	0.98	0.98	518
	Undisturbed	0.98	0.98	0.98	484
	Accuracy			0.98	1002
	Macro average	0.98	0.98	0.98	1002
	Weighted average	0.98	0.98	0.98	1002
Random Forest	Disturbed	0.97	0.98	0.97	518
	Undisturbed	0.98	0.97	0.97	484
	Accuracy			0.97	1002
	Macro average	0.97	0.97	0.97	1002
	Weighted average	0.97	0.97	0.97	1002
Naïve Bayes	Disturbed	0.97	0.89	0.93	518
	Undisturbed	0.89	0.98	0.93	484
	Accuracy			0.93	1002
	Macro average	0.93	0.93	0.93	1002
	Weighted average	0.93	0.93	0.93	1002
Support Vector Classifier	Disturbed	1	0.99	0.99	518
	Undisturbed	0.99	1	0.99	484
	Accuracy			0.99	1002
	Macro average	0.99	0.99	0.99	1002
	Weighted average	0.99	0.99	0.99	1002

The Random Forest model also demonstrated robust performance, with a precision of 0.97 and a recall of 0.98 for the "Disturbed" class, and a precision of 0.98 and a recall of 0.97 for the "Undisturbed" class. This balance led to F1-Scores of 0.97 for both classes and an overall accuracy of 0.97, with macro and weighted averages of 0.97.

The Naïve Bayes model, while effective, showed slightly lower performance compared to the previous models. It had a precision of 0.97 and a recall of 0.89 for the "Disturbed" class, and a precision of 0.89 and a recall of 0.98 for the "Undisturbed" class, resulting in F1-Scores of 0.93 for both classes. The overall accuracy was 0.93, with macro and weighted averages also at 0.93. Lastly, the Support Vector Classifier (SVC) emerged as the best performer, achieving perfect precision (1.00) for the "Disturbed" class and nearly perfect recall (0.99), leading to an F1-Score of 0.99. For the "Undisturbed" class, it achieved a precision of 0.99 and a perfect recall of 1.00, also resulting in an F1-Score of 0.99. The model's overall accuracy was the highest at 0.99, with macro and weighted averages of 0.99, indicating its superior ability to balance precision and recall across both classes.

Figures 2 shows the ROC AUC score for each of the five ML. The Receiver Operating Characteristic Area Under the Curve (ROC AUC) graphs and scores provide a clear indication of the discriminatory power and effectiveness of each machine learning model in distinguishing between classes (Disturbed and Undisturbed) based on forest degradation detection from remotely sensed satellite imagery by plotting the true positive rate against the false positive rate across different thresholds. Logistic Regression achieved a perfect ROC AUC of 1.00, indicating flawless discrimination between the classes without errors. The Decision Tree model followed closely with a high ROC AUC of 0.98, demonstrating strong performance in classification.

Random Forest also achieved a perfect ROC AUC of 1.00, leveraging its ensemble approach to enhance classification accuracy. Naïve Bayes, despite its probabilistic assumptions, performed exceptionally well with an ROC AUC of 0.99, indicating robust discrimination ability. Support Vector Classifier (SVC) mirrored Logistic Regression and Random Forest with a perfect ROC AUC of 1.00, highlighting its efficacy in separating the classes based on the support vectors chosen

Figure 3 shows the Confusion matrix for each of the five ML models. The confusion matrix shows the frequency of each of the True Positive (TP), False Negative (FN), False Positive (FP), and True negative (TN) of a classification model in a matrix table layout, depicting the number of correct predictions (TP and TN) and errors (FP and FN) made per class (disturbed and undisturbed). True Positives (TP) are instances correctly predicted as positive (e.g., correctly identified as "Disturbed" forests), False Positives (FP) are instances incorrectly predicted as positive (predicted "Disturbed" but actually "Undisturbed"), True Negatives (TN) are instances correctly predicted as negative (e.g., correctly identified as "Undisturbed" forests), while False Negatives (FN) are instances incorrectly predicted as negative (predicted "Undisturbed" but actually "Disturbed"). The TP, FN, FP and TN for the Logistic Regression model are 438, 80, 0, and 484 respectively, while those for the Decision Tree model are 510, 8, 11, and 473 respectively. The TP, FN, FP and TN for the Random Forest model are 507, 11, 15, and 469

respectively, and those for the Naïve Bayes model are 459, 59, 12, and 472 respectively. Finally, the TP, FN, FP and TN or the Support Vector Classifier model are 511, 7, 1, and 483 respectively. All models showed strong performance, the Support Vector Classifier stood out with the highest accuracy and balanced metrics, closely followed by the Decision Tree and Random Forest models.

The performance metrics of the trained and evaluated machine learning models in this work, which are Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, and Support Vector Classifier models, show an overall accuracy of 0.92, 0.98, 0.97, 0.93, and 0.99, respectively, with Support Vector Classifier emerging as the best performing model in classification tasks for detecting forest degradation. This is in accord with the findings of Solórzano and Gao, (2022) and Cui et al., (2023). Similarly, Loiola et al., (2023) also shows that Support Vector Classifier is a good machine learning model for classification tasks.

Conclusion and Recommendations

The combined application of artificial intelligence and satellite remote sensing technologies for detection and monitoring forest degradation in Omo Forest Reserve proved to be accurate and cost-effective, with the Support Vector Classifier model achieving an accuracy of 99% as the best-performing machine learning model. This study recommends the use of ML-driven for assessment of forest in Nigeria in order to achieve sustainable forest management.

Declarations

Author Contribution

B. A. , O. O. and O. J. conceptualized the work. B. A. executed the methodology. O. O. and O. J. supervised the work. B. A. , P. C., and D. E. wrote the manuscript. All authors reviewed the manuscript.

Funding Declaration

This research was not grant funded.

Ethics Declaration

Not applicable.

References

1. Adedeji OH, Tope-Ajayi OO, Abegunde OL. Assessing and Predicting Changes in the Status of Gambari Forest Reserve, Nigeria Using Remote Sensing and GIS Techniques. J Geographic Inform Syst. 2015;7:301–18.

2. Amusa TO. Trade and income generation potentials of African bush pepper (Piper guineense, Schum. & Thonn) in tropical lowland rainforests of Southwestern Nigeria. Nat Faune. 2015;29(2):59–63.
3. Cui L, Wang L, Su J, Song Z, Li X. (2023). Classification and identification of degraded alpine m eadows based on machine learning techniques. 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL), 263–267.
4. Larinde SL, Olasupo OO. (2011). 'SocioEconomic Importance of Fuelwood Production in Gambari Forest Reserve Area, Oyo State, Nigeria. J Agric Social Res (JASR), 11(1).
5. Loiola TM, Fantinel RA, dos Santos FD, de Bastos F, Schuh MS, Fernandes P, de Oliveira BAS, Pereira RS. Use of Machine Learning Algorithms in the Classification of Forest Species. Volume 46. Anuário Do Instituto de Geociências; 2023. pp. 1–8.
6. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: An applied review. Int J Remote Sens. 2018;39(9):2784–817.
7. Oyediji OT, Adenika OA. Forest degradation and deforestation in Nigeria; poverty link. Int J Multidisciplinary Res Anal. 2022;5(10):2837–44.
8. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
9. Sambe LN, Adeofun CO, Dachung G. The economic and ecological effects of deforestation on the Nigerian environment. Asian J Adv Res Rep. 2018;1(2):1–25.
10. Solórzano JV, Gao Y. Forest disturbance detection with seasonal and trend model components and machine learning algorithms. Remote Sens. 2022;14(3):803.
11. Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. Remote Sens. 2020;12(7):1135.

Figures

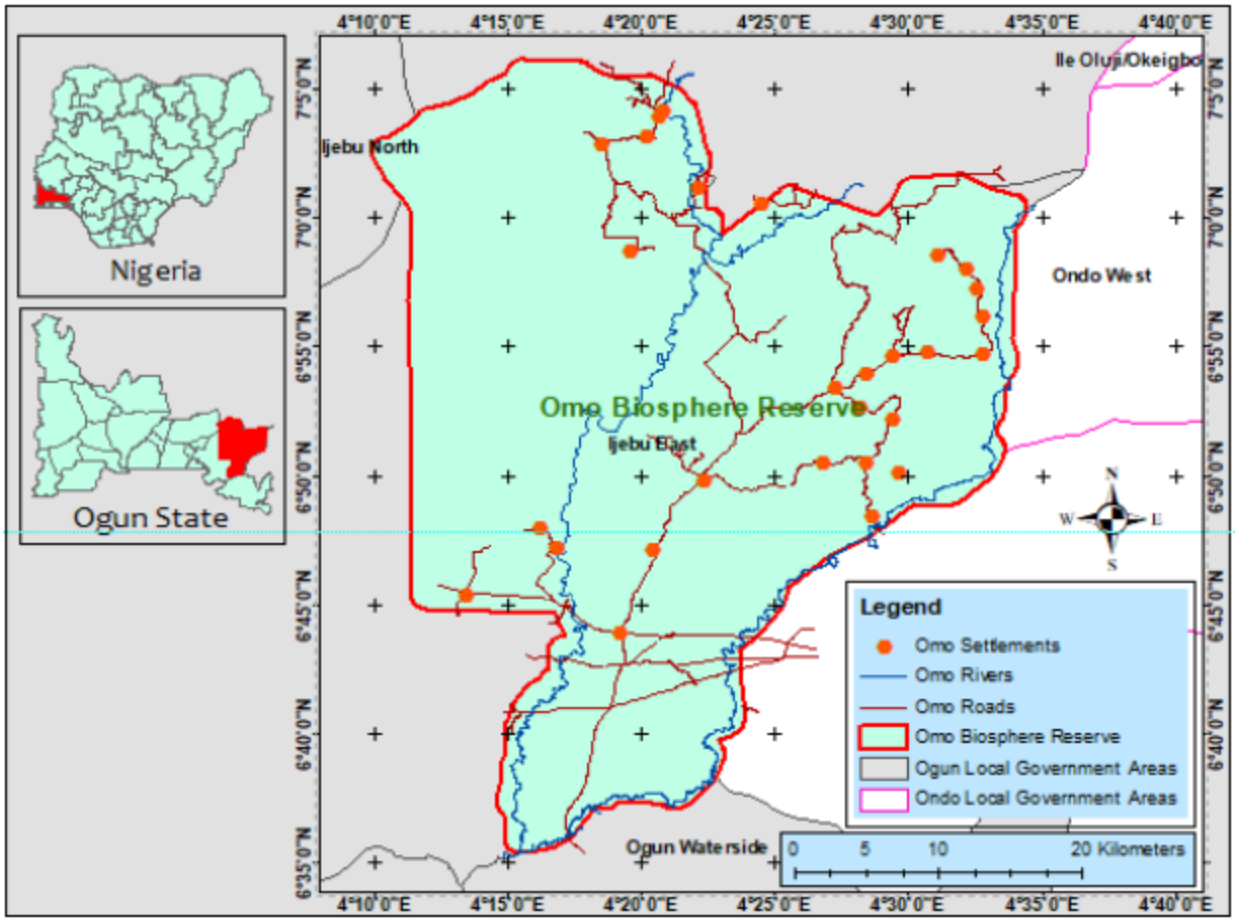


Figure 1

Map of Omo Forest Reserve, Southwest Nigeria

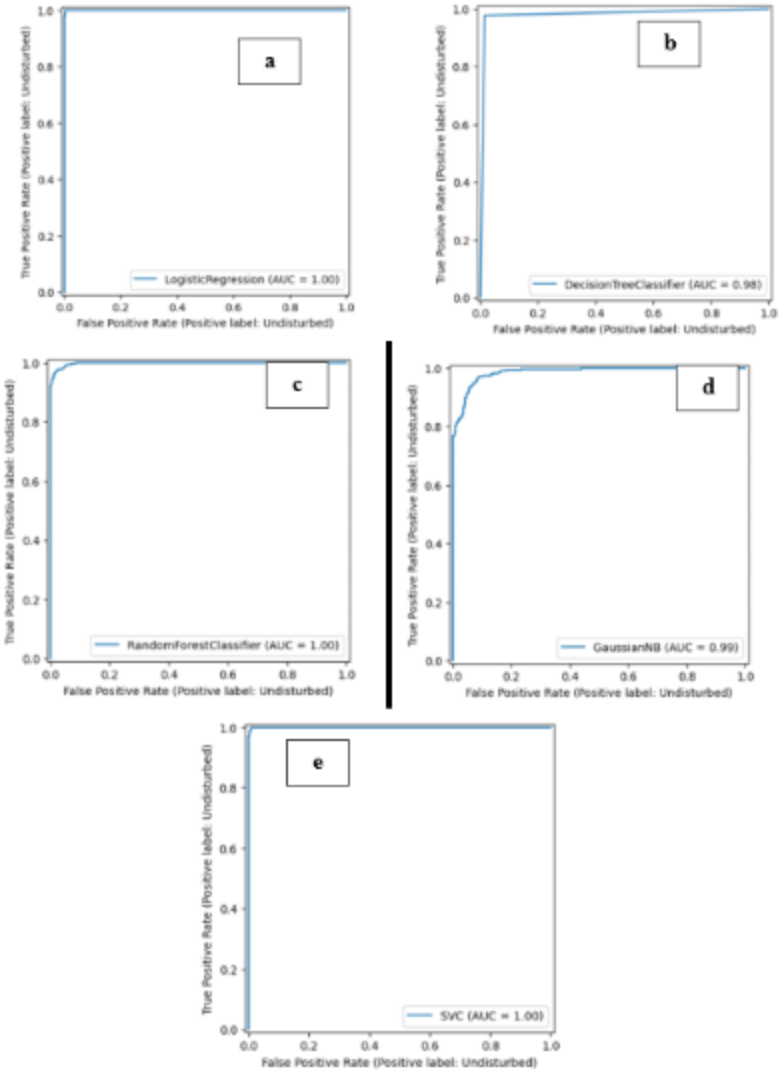


Figure 2

Area Under Curve Graphs and Scores for: (a) Logistic Regression, (b) Decision Tree, (c) Random Forest, (d) Naïve Bayes, and (e) Support Vector Classifier models.

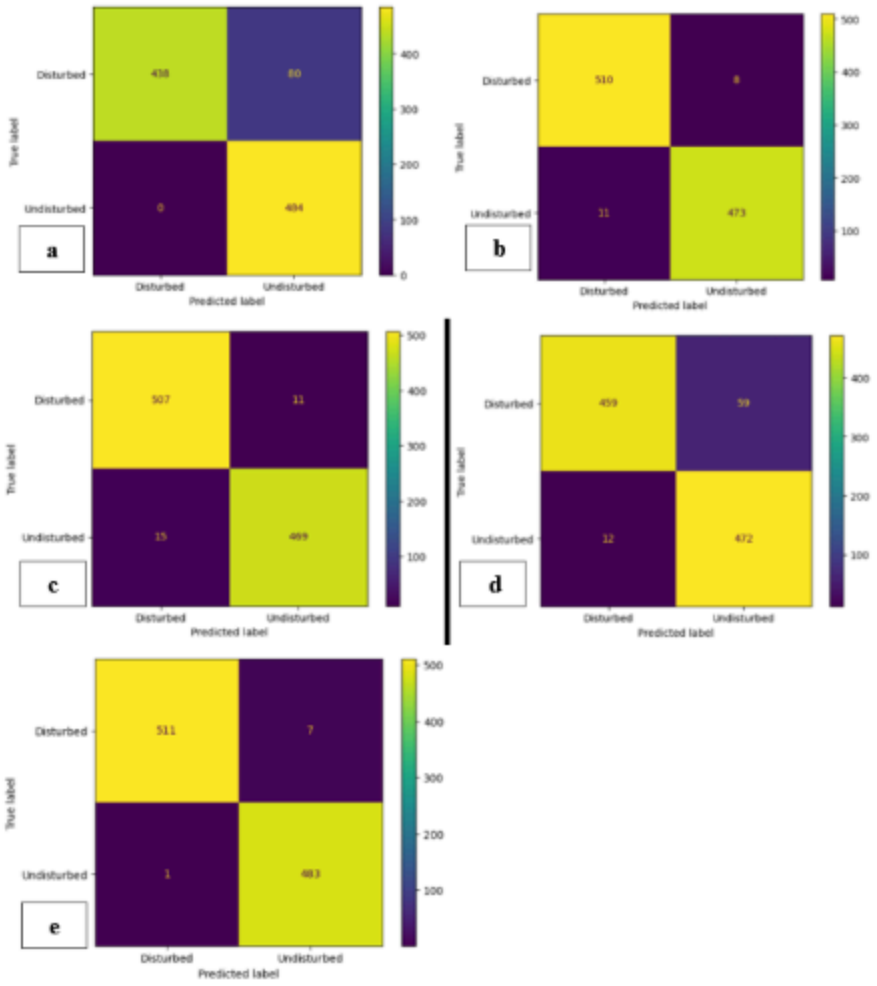


Figure 3

Confusion Matrix for: (a) Logistic Regression, (b) Decision Tree,

(c) Random Forest, (d) Naïve Bayes, and (e) Support Vector Classifier models.