Detection and Classification of Mahogany Tree Species via Satellite Imagery, Google Earth Engine, and Deep Learning

Joy Roy, Muhammad Anwarul Azim, Abu Nowshed Chy, and Mohammad Khairul Islam
Department of Computer Science and Engineering,
University of Chittagong,
Chittagong-4331, Bangladesh

ABSTRACT
The mahogany tree is widely used in various industries worldwide for its valuable timber. In response to the increasing scarcity of mahogany wood, particularly in Bangladesh, this study presents an automated process for detecting and classifying mahogany tree species. Focusing on the University of Chittagong, Bangladesh region where multi-polygons delineating mahogany trees and other land cover types were generated using Google Earth Engine. Sentinel-2 satellite imagery from 2019 and 2020 provided spectral band wavelength data, enabling the creation of datasets with two classes: mahogany tree and non-mahogany tree cover types. To enhance model performance and accuracy, the four satellite datasets removed irrelevant columns and eliminated 7,661 duplicate entries, resulting in 2,339 unique entries. The refined dataset comprises 12 spectral band features and one class attribute, which significantly improves the classification accuracy. Supervised classification employed a multilayer perceptron deep neural network with three hidden layers. The model achieved a training accuracy of 99.76% and testing accuracy of 96.58%, with a precision of 98.36%, recall of 99.05%, and F1-Score of 97.6%. The optimal performance was observed with the Spectral band wavelength dataset of Satellite Surface Reflectance 2019 images using the Adam optimizer. This research contributes to advancing automated methods for mahogany tree detection, facilitating conservation efforts for tree species, and informing resource management practices in Bangladesh and beyond.

Keywords
Mahogany Tree Species, Google Earth Engine, Spectral band wavelength, Supervised Classification, Multilayer perceptron, Deep Neural Network

1. INTRODUCTION
Mahogany (Swietenia spp.) is a semi-deciduous, medium-sized tree 30-35 meters long. It has a short but solid foundation, a wide canopy, plenty of thick trees, and dense shade. While several assorted trees are mentioned in Mahogany from around the world, the name refers to the Swietenia family species [1]. Swietenia mahagoni, West Indian mahogany, is all the more suitably alluded to as Mahogany by its logical name. Mahogany, known for its aromatic hardwood used in furniture production, is a tropical tree [2], [3]. A bright, dappled shadow is cast on the ground below by Mahogany, rendering it a perfect shade tree for landscapes with enough space to flourish [3]. Mahogany plants growing along streets on sidewalks, or even in private gardens, form a lovely overhead canopy that can be enjoyed by anyone [3]. It is temporarily deciduous and in spring it loses its leaves. Because of its color and longevity, wood is prized in the lumber industry for fine cabinets and furniture. It is possible to have smooth, interlocked, uneven, or wavy grain. Whitish or yellowish is the sapwood. When split, the heartwood color is reddish, pinkish, or yellowish and grows to a dark reddish-brown.

Mahogany trees are prized for their durable timber, used in various applications from furniture to musical instruments, particularly in countries like Bangladesh where the wood is highly valued. However, due to its increasing scarcity and high cost, there's a pressing need to protect and identify mahogany tree species in different regions.

In this research, the mahogany tree is considered as a sample to develop the solution method of detecting and classifying tree species. In other words, the proposed solution approach can be extended to other tree species using satellite imagery. Traditional methods of species classification, such as remote sensing and satellite imagery, are costly and time-consuming, posing challenges for surveillance efforts in tropical landscapes [4].

By delivering a cloud computing framework for earth observation data processing, Google Earth Engine (GEE) brings open access geospatial analysis one phase forward. It integrates a publicly available catalog which, from its creation in 1972 to the present day, consists of an almost complete collection of Landsat imagery, with a large-scale computing facility configured for parallel geospatial data processing. In a notable forest resources mapping exercise, to describe tree cover and eventual tree cover loss and gain over 2012, a total of 20 tera pixels of satellite imagery were processed on GEE, utilizing one million CPU-core hours on 10000 computers in parallel [6]. This process was finished on GEE in a matter of days, but it would have taken 15 years to finish on a single machine [7]. GEE also offers a cloud-based system for real-time supervised learning on large datasets, using hand-drawn inputs to rapidly classify ground cover types online, revolutionizing and accelerating remote sensing processes.

This study proposes a methodology to address the aforementioned issue by utilizing Google Earth Engine. Randomly chosen training points represented as polygons are employed to acquire spectral band wavelength data for various satellites, enabling the classification of mahogany tree species. This approach leverages multilayer perceptron (MLP) deep neural networks (DNN) for accurate classification

2. RELATED WORK
The remote-sensing community has long drawn the interest of remote-sensing studies based on image classification since the classification findings are the basis for many socioeconomic and ecological applications [8]. The first important step for an effective classification for a particular reason is the collection

of appropriate sensor data [9] [10]. Different selection techniques could be used, such as single image, seed, and polygon, but they affect the results of classification, specifically for classifications of image data of high spatial resolution [11].

The related work of this thesis has been divided into two categories.

- Lidar image-based tree species identification Research
- Satellite image-based tree species identification Research

2.1 Lidar image-based tree species identification

In both botanical taxonomy and computer vision, plant image recognition has been an interdisciplinary subject. The first dataset of plant photographs taken by cell phones in the natural scene is presented, containing 10000 photos of 100 species of ornamental plants on the campus of the Beijing Forestry University [12]. For large-scale plant classification in natural ecosystems, a 26-layer deep learning model consisting of eight residual basic components is planned. The proposed model achieves a 91.78 percent identification score on the BJFU100 dataset, showing that a useful approach for smart forestry is deep learning [12]. Specific tree species classification based on lidar using Convolutional Neural Network uses terrestrial lidar since in the dark forest it can achieve high-resolution point clouds [13]. The classification of individual urban tree species using very High Airborne Multi-Spectral imagery using spatial resolution profiles longitudinal, suggests an approach to improving the classification of individual tree species using longitudinal profiles from airborne imagery with a very high spatial resolution [14].

2.2 Satellite image-based tree species identification

Detection on Landsat images with GEE of industrial oil palm plantations is most relevant to this paper. This Research demonstrates the use of GEE for the identification in Tripa, Aceh, Indonesia of commercial oil palm plantations [15]. Using separate spectral bands (RGB, NIR, SWIR, TIR, all bands), ground cover classification of the picture of Landsat 8 images was carried out, defining the following groups of ground cover: immature oil palm, mature oil palm, non-oil palm, forest, water, and clouds [15]. Using all bands for the classification of land cover, the total precision and Kappa coefficient were the best, followed by RGB, SWIR, TIR, and NIR. Trees for regression and grouping (Classification And Regression Tree) and random forest tree algorithms provided categorized land cover maps with higher overall accuracies and coefficients of Kappa than the algorithm for minimum distance (MD). GEE can be further evolved into an open and low-cost instrument for autonomous bodies to track and control the growth of tropical oil palm plantations [15]. This oil palm detection research was done using a pixel-based spectral band wavelength dataset which was trained and classified in GEE. However, this research accuracy is poor due to using 30-meter resolution Landsat-8 images and many duplications of spectral band wavelength. This research does not use any neural networks to improve accuracy.

3. LOCATION AREA

The experimental site is situated within the University of Chittagong (22.4716° N, 91.7877° E), approximately 22 kilometers north of Chittagong, Bangladesh. Spanning around 2110 acres in Fatehpur Union, Hathazari Upazila, Chittagong, the university is enveloped by a diverse landscape of evergreen forests, hills, cropland, and ponds. Over the years, the university administration has systematically planted various tree species, including the highly prized mahogany tree, renowned for its value, aesthetics, fragrance, and resilience to strong winds. Initially planted along a 2km stretch from zero point to gate no. 1, the mahogany trees have flourished, forming distinct patches across the campus. Motivated by this proliferation, numerous additional mahogany trees have been planted in different locations within the university grounds. Thus, the University of Chittagong has been chosen as the focal point for a case study aimed at detecting and classifying mahogany tree species using deep neural networks

4. METHODOLOGY

4.1 Dataset generation using the Google Earth Engine

Google Earth Engine is a cloud-computing platform for processing satellite picturing and geospatial datasets with planetary-scale inspection. The unlimited access GEE portal offers access to (1) petabytes of freely accessible remote sensing data and other fully prepared explorer web software products; (2) parallel high-speed computing and machine learning algorithms utilizing Google's computing resources; and (3) a repository of Application Programming Interfaces (APIs) that support common coding languages, such as JavaScript and Python, with development environments. All of these key features enable users to powerfully explore, interpret, and simulate geospatial big data without having access to supercomputers or advanced skills in coding. For interactive data and algorithm creation, the accessible and user-friendly front end provides a convenient atmosphere. Users get the ability also to incorporate and customize their data and collections while doing all the processing using Google's cloud tools. The application areas, ranging from woodland and vegetation studies to fields of medicine such as malaria, were very varied. The most commonly used dataset was Landsat; it is the largest portion of the GEE data portal, with data accessible for use and download from the first through the latest Landsat collection.

4.1.1 Creation of Multi-polygon and Random GPS Coordinates:

To identify mahogany tree species, the process began by delineating multi-polygons representing various land cover types—mahogany tree areas, forests excluding mahogany trees, water bodies, croplands, and buildings—within the satellite imagery of the study area using GEE. Data on mahogany tree locations and attributes were gathered from local sources and the Institute of Forestry and Environmental Sciences at the University of Chittagong. Using Google satellite imagery, 24 polygons were manually created for mahogany trees, 22 for non-mahogany forests, 15 for water bodies, 15 for croplands, and 25 for buildings

Fig 1: Polygons Representing Non-Mahogany and Mahogany Trees in Google Earth Engine (GEE)

Fig 2: Polygon of building and water in Google Earth Engine

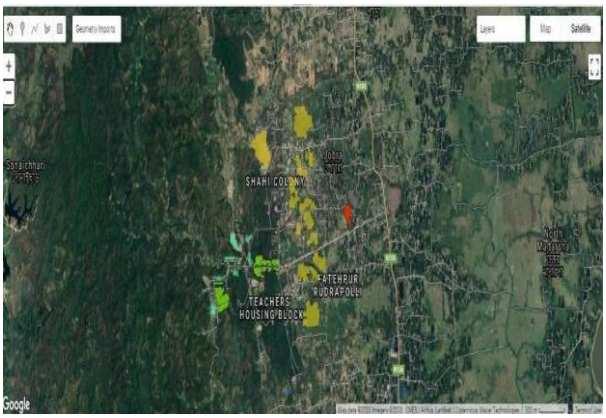

Fig 3: Polygon of Satellite view of the different polygon data in Google Earth Engine

Each polygon was assigned a distinct color scheme: gray for mahogany tree areas, green for non-mahogany forests, blue for water bodies, yellow for croplands, and cyan for buildings. Subsequently, 5000 random GPS coordinates (Longitude and Latitude) were generated within the mahogany tree polygons and stored as variables using the GEE editor. Next, 2000 random GPS coordinates were systematically generated from the non-mahogany forest polygon, followed by 1000 coordinates each from the cropland, buildings, and water polygons. These coordinates were stored as variables using the Google Earth Engine editor. Assigning class names, the 5000 random GPS coordinates within mahogany tree areas were labeled as "mahogany," corresponding to class level "0". For the remaining coordinates—2000 from non-mahogany forests, 1000 from croplands, 1000 from water bodies, and 1000 from buildings—they were grouped under the class name "not a mahogany tree," representing class level "1".

4.1.2 Generate and Export Spectral Band Wavelength Data:

Two satellite images were utilized to retrieve spectral band wavelengths corresponding to the generated random GPS coordinates.

- Sentinel-2 Satellite Surface Reflectance (SR) [17]
- Sentinel-2 Satellite Top of Atmosphere (TOA)[17]

The search for Satellite Surface Reflectance (SR) images began from May 1st, 2019 to July 30th, 2019 in the GEE data catalog. Specifically, an image with a 20% cloud cover was selected, utilizing cloud masking techniques facilitated by the QA60 spectral band of the SR Satellite and bit masking.

Table 1. Level 2A Surface Reflectance spectral bands

Feature	Band Name	Band Resolution	Band Description
1	B1 Band	60 m	Aerosols Band
2	B2 Band	10 m	Blue Band
3	B3 Band	10 m	Green Band
4	B4 Band	10 m	Red Band
5	B5 Band	20 m	Red Edge 1 Band
6	B6 Band	20 m	Red Edge 2 Band
7	B7 Band	20 m	Red Edge 3 Band
8	B8 Band	10 m	NIR Band
9	B8A Band	20 m	Red Edge 4 Band
10	B9 Band	60 m	Water Vapor Band
11	B11 Band	60 m	SWIR 1 Band
12	B12 Band	60 m	SWIR 2 Band

The SR dataset encompasses 13 distinct spectral bands with variable resolutions ranging from 10 to 60 meters. The 12 spectral bands listed in Table 1 below are also utilized in the Satellite Top of Atmosphere (TOA) dataset. From these images, wavelengths of the 12 spectral bands were retrieved for various random geographic locations generated through GEE. These wavelengths were stored as variables representing the input vectors of the dataset. Subsequently, the band wavelengths were exported to Google Drive, saving the dataset in CSV format for further analysis. Following a similar procedure, TOA images from May 1st, 2019 to July 30th, 2019 were searched for, selecting images with the least cloud cover. These images comprise 13 spectral bands and three QA bands for cloud masking, providing data for analysis. Similarly, wavelengths of the 12 spectral bands mentioned earlier were fetched for different geographic locations and stored as variables representing input vectors. The resulting dataset was then exported to Google Drive. Utilizing the same methodology, two additional spectral band wavelength datasets were created: one from Satellite SR images spanning May 1st, 2020 to July 30th, 2020, and another from TOA images during the same period. Finally, employing Deep Neural Networks, the performance of these four different datasets was compared to identify mahogany tree species.

4.2 Data Preprocessing

In this subsection, data preprocessing techniques are delineated, encompassing two distinct approaches outlined below: In the preceding section, four distinct datasets were produced, namely:

- ✓ SR Satellite spectral band wavelength of the 2019 dataset
- ✓ SR Satellite spectral band wavelength of 2020 dataset
- ✓ TOA Satellite spectral band wavelength of 2019 dataset
- ✓ TOA Satellite spectral band wavelength of 2020 dataset

After generation using Google Earth Engine, each of the four datasets comprises 15 feature attributes or input vectors. Among the 15 feature attributes, 12 represent the spectral band wavelengths of different satellites. The "Forest" feature attribute serves as the class attribute, with values of "0" denoting mahogany trees and "1" indicating non-mahogany trees. The ".geo" attribute, autogenerated by GEE, holds null values and doesn't impact the model. Similarly, the "system: index" attribute, also generated by GEE, contains dataset serial numbers. For enhanced model performance, preprocessing steps begin with noise elimination. The "geo" column, solely containing null values and irrelevant to the deep learning model, is removed. As for the "system: index" attribute, used for assigning serial numbers to the dataset, its correlations are explored using Principal Component Analysis (PCA). PCA reveals that the "system: index" feature is independent of other attributes, and no significant correlations exist among the remaining features. Consequently, the "system: index" feature is removed from all four datasets iteratively. After removing the unnecessary columns "system: index" and ".geo" from the four datasets, a total of 13 feature attributes remain, including 12 spectral bands and 1 class attribute. These spectral bands are captured at resolutions of 10 meters, 20 meters, and 60 meters per pixel. Given the varying resolutions, numerous duplicate entries are expected within the datasets, especially in areas with different pixel resolutions. To address this, the datasets are imported into Google Colab for duplicate and redundant row detection. Using Python libraries, redundant data is identified and removed. This process identifies 7661 redundant entries across the datasets, significantly enhancing performance. Post-removal, there are 2339 unique entries remaining. Among these, 488 belong to the mahogany class (class level "0"), with the remaining classified as non-mahogany (class level "1"). This duplicate data detection and deletion process is applied iteratively across all four datasets, ensuring data integrity and enhancing model accuracy

4.3 Multilayer Perceptron

Artificial Neural Networks, particularly the multilayer perceptron (MLP), emulate the brain's microstructure and are crucial for various Artificial Intelligence tasks. MLPs use backpropagation for supervised learning and feature multiple layers with nonlinear activation functions, allowing them to handle complex data. In an MLP, perceptrons in the input layer transmit outputs to those in the hidden layer, which then relay them to the output layer. With multiple weights per signal, MLPs construct elaborate structures that facilitate advanced data processing. Deep neural networks, such as a three-layer MLP, form the basis of deep learning, which enables computers to learn and comprehend data hierarchically. This approach allows systems to derive insights from experience without explicit human intervention in defining data features. By employing multiple processing layers, deep learning facilitates the acquisition of diverse data representations across various levels of abstraction.

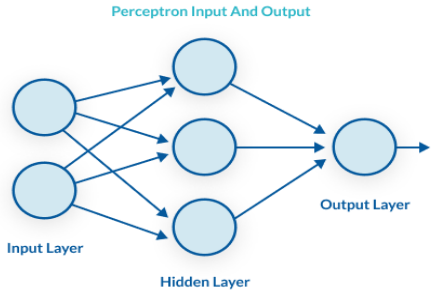

Fig 4: Multilayer Perceptron Diagram

4.4 The process flow for supervised classification using MLP DNN

Utilizing Google Earth Engine (GEE), a comprehensive dataset for Mahogany tree species identification was constructed. This dataset included various geographical and environmental features. The dataset underwent preprocessing steps, which involved feature reduction to eliminate redundant and irrelevant features and duplicate data deletion to ensure data integrity.

The preprocessed data was split into training and testing subsets using various validation splits, including 90-10, 80-20, 70-30, 60-40, and 50-50. This approach ensured a robust evaluation of the model's performance across different proportions of training and testing data. For the supervised classification task, a Multilayer Perceptron Deep Neural Network (MLP DNN) was chosen, facilitated by the Keras library. The MLP model architecture consisted of one input layer, three hidden layers, and one output layer. The input layer comprised 12 input vectors representing the 13 feature attributes obtained post-preprocessing. The hidden layers were structured with a cascading architecture, with the first layer containing 256 neurons, the second layer 128 neurons, and the third layer 64

neurons. The output layer consisted of a single neuron with Sigmoid activation for binary classification.

Activation functions played a crucial role in determining neuron activation based on weighted sums and biases. The training model employed a combined activation function approach, using Rectified Linear Unit (ReLU) activation for the hidden layers and Sigmoid activation for the output layer. For training loss calculation, the "binary_crossentropy" function was utilized to measure the disparity between predicted and actual class labels.

Hyperparameter tuning was employed to optimize model performance, experimenting with optimizers such as "Adam" and "Stochastic Gradient Descent (SGD)" to minimize training losses and optimize dataset weights. Additionally, techniques like dropout regularization and batch normalization were incorporated to prevent overfitting and enhance model generalization. Dropout regularization was applied to the hidden layers, randomly dropping a fraction of neurons during training to prevent overfitting. Batch normalization was employed to normalize the inputs of each layer, stabilizing the learning process and speeding up convergence. Each model was configured with 12 input vectors, a batch size of 5, and trained over 1000 epochs.

Performance metrics such as accuracy, precision, recall, and F1-score were evaluated to assess the classification performance across different validation splits. Confusion matrices were generated to visualize the classification results and identify any misclassifications. The model's robustness and generalizability were further tested using cross-validation, ensuring the model's performance was consistent across different subsets of the data.

The optimized model was validated on a separate validation set to ensure its ability to accurately identify Mahogany tree species, providing insights into its practical applicability and performance in real-world scenarios. The results were documented, highlighting the effectiveness of the preprocessing steps, the chosen architecture, and the optimization strategies employed.

In conclusion, this methodology provided a structured approach to developing an accurate and reliable model for Mahogany tree species identification. The use of GEE for data collection and preprocessing, combined with the application of advanced machine learning techniques using the MLP DNN model, ensured a robust and effective solution for this classification task. The comprehensive evaluation and optimization strategies employed ensured the model's applicability and reliability in real-world scenarios.

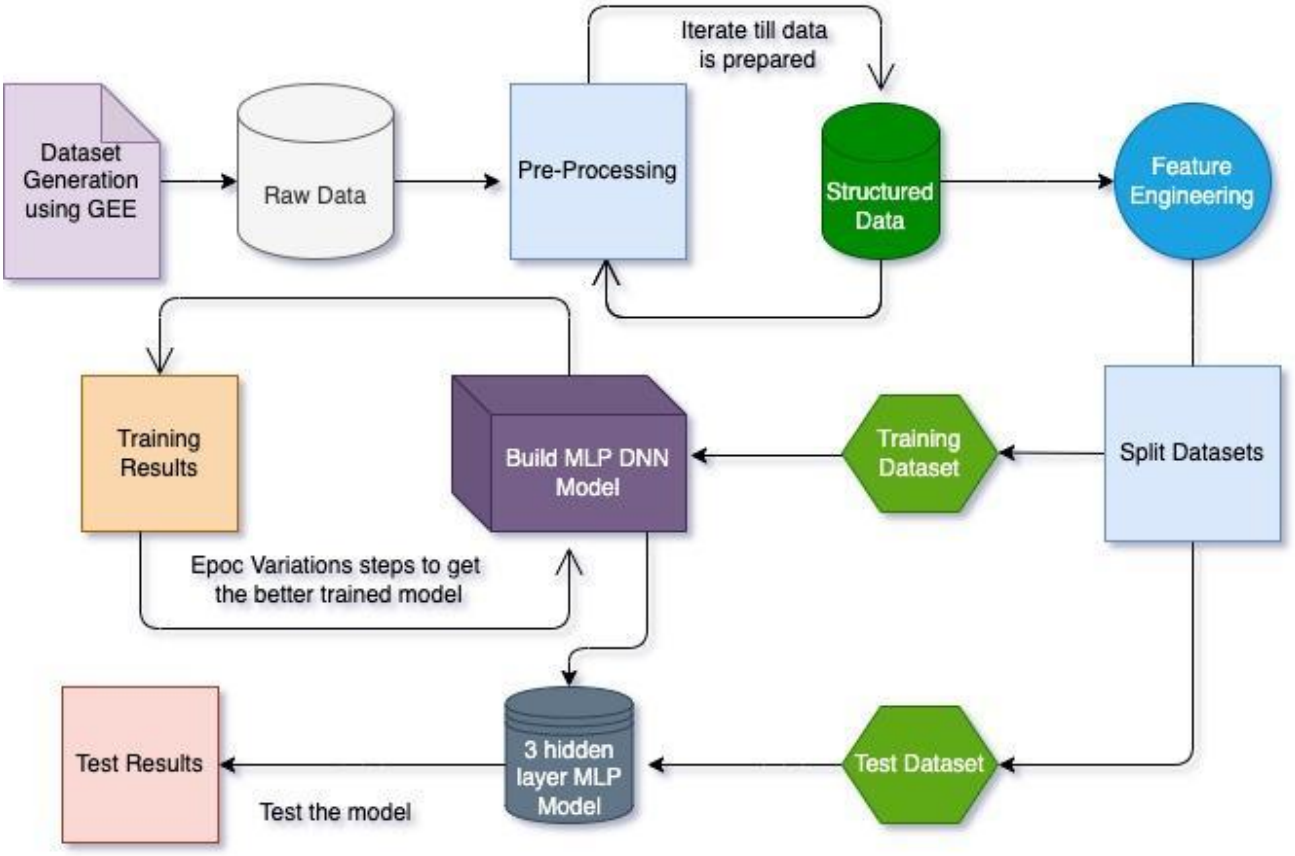

Fig 5: Process flow for Supervised Classification using Multilayer Deep Neural Network

5. EXPERIMENTAL RESULTS

In this section, the outcomes of the experiment were obtained using the dataset and deep learning model. All experiments were conducted on Google Colab utilizing Google Cloud resources, which provided 12.72 GB RAM and 107.77 GB storage capacity and facilitated the Keras library. The performance is assessed across different categories in terms of accuracy, recall, and precision. The results are categorized into two sections

5.1 SR satellite spectral band wavelength dataset results

For Mahogany tree species identification, MLP DNN is utilized. Initially, the MLP DNN model is trained using the SR satellite spectral band wavelength dataset, comprising 2339 data points. This dataset is split into training and testing sets for model validation: (90, 10), (80, 20), (70, 30), (60, 40), and (50, 50). Each split allocates a specific percentage of the dataset for training and testing, resulting in varying numbers of data points for each subset. Additionally, two spectral band wavelength datasets for SR Satellite are generated, covering periods from May 1st, 2019 to July 30th, 2019, and May 1st, 2020 to July 30th, 2020. The subsequent analysis is based on these datasets, considering the distribution of data points across the splits.

5.1.1 SR satellite spectral band wavelength dataset of 2019 results

Initially, the MLP DNN model used the SR satellite spectral band wavelength dataset from May 1st, 2019 to July 30th, 2019, employing various splitting methods outlined earlier. For training loss calculation, binary cross-entropy was utilized, and two optimizers were experimented with: Adam and SGD. Below, the results of the 2019 spectral band wavelength dataset were discussed, and the performance achieved with the Adam and SGD optimizers was analyzed.

Table 2: SR satellite spectral bands (2019) results using Adam optimizer

Split Method	Training Accuracy	Testing Accuracy	Precis-ion	Recall	F1 – Score
(90, 10)	99.71	**96.58**	**98.36**	96.26	97.3
(80, 20)	98.5	93.8	97.34	97.86	**97.6**
(70, 30)	**99.76**	96.3	97.45	95.71	96.58
(60, 40)	99.64	95.3	92.4	**99.05**	95.61
(50, 50)	99.4	95.47	97.27	97.37	97.32

Table 3: SR satellite spectral bands (2019) results using SGD optimizer

Split Method	Training Accuracy	Testing Accuracy	Precis-ion	Recall	F1 – Score
(90, 10)	**96.96**	92.74	92.89	**97.86**	95.31
(80, 20)	91.56	89.10	**97.74**	92.51	95.05
(70, 30)	95.85	92.31	88.21	94.82	91.39
(60, 40)	94.23	91.24	96.39	94.43	**95.40**
(50, 50)	94.10	**93.08**	95.38	94.97	95.18

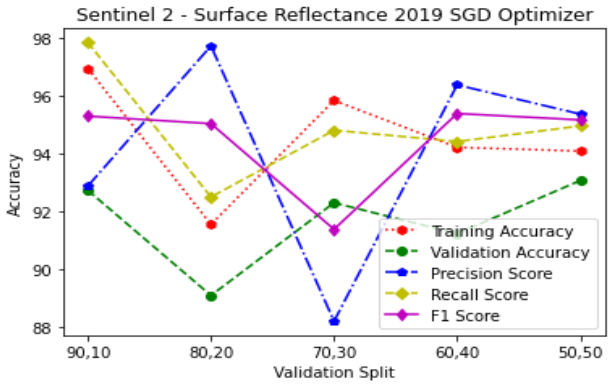

Fig 6: SR satellite spectral band wavelength dataset of 2019 results using SGD optimizer

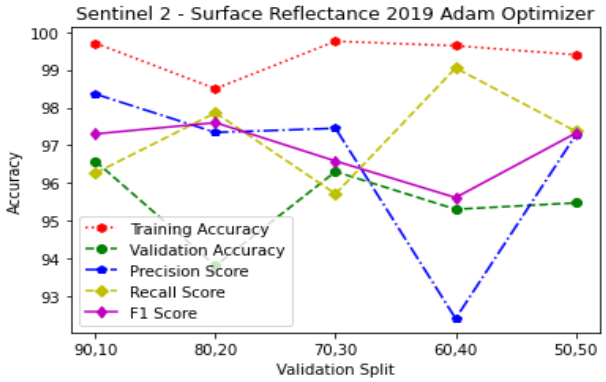

Fig 7: SR satellite spectral band wavelength dataset of 2019 results using Adam optimizer

5.1.2 SR satellite spectral band wavelength dataset of 2020 results

For the analysis of the SR satellite spectral band wavelength dataset from May 1st, 2020, to July 30th, 2020, the MLP DNN model was trained using different splitting ratios. Binary cross-entropy was used for training loss calculation, and two optimizers were experimented with: Adam and SGD. Below are the results of the 2020 spectral band wavelength dataset, focusing on the performance achieved with the Adam and SGD optimizers.

Table 4: SR satellite spectral bands (2020) results with Adam optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall Score	F1 - Score
(90, 10)	99.19	94.44	94.27	93.93	96.04
(80, 20)	98.61	**94.66**	**97.3**	96.26	**96.77**
(70, 30)	99.57	93.73	93.88	**98.57**	96.17
(60, 40)	**99.79**	94.02	96.24	94.02	95.12
(50, 50)	94.7	90.26	93.07	98.47	95.69

Table 5: SR satellite spectral bands (2020) results with SGD optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall Score	F1 - Score
(90, 10)	**93.35**	89.74	93.12	94.12	93.62
(80, 20)	92.73	89.74	**94.65**	94.65	**94.65**
(70, 30)	91.39	**89.74**	94.27	93.93	94.1
(60, 40)	93.44	89.32	89.94	**98.37**	93.96
(50, 50)	92.13	89.91	94.23	94.64	94.43

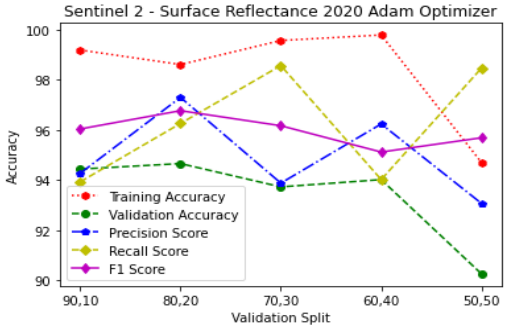

Fig 8: 2020 SR spectral band wavelength results using Adam optimizer

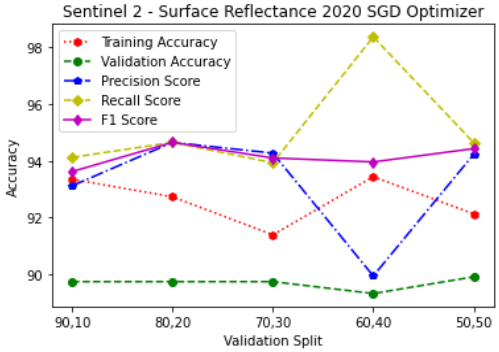

Fig 9: SR satellite spectral band wavelength dataset of 2020 results using SGD optimizer

5.2 TOA satellite spectral band wavelength dataset results

In the Mahogany tree species identification analysis, an MLP DNN was trained with the SR satellite spectral band wavelength dataset consisting of 2339 data points. This dataset was split into training and testing subsets using various ratios: (90,10), (80,20), (70,30), (60,40), and (50,50). Additionally, two distinct spectral band wavelength datasets were generated for TOA Satellite: one spanning from May 1st, 2019, to July 30th, 2019, and the other covering May 1st, 2020, to July 30th, 2020. The following sections will discuss the results of the analysis of these datasets and their effectiveness in Mahogany tree species identification.

5.2.1 TOA satellite spectral band wavelength Dataset of 2019 results

Initially, the MLP DNN model was trained using the TOA satellite spectral band wavelength dataset from May 1st, 2019, to July 30th, 2019, employing various splitting methods as discussed earlier. For training loss calculation, binary cross-entropy was utilized, and two optimizers were experimented with: Adam and SGD. Below are the results of the 2019 spectral band wavelength dataset, analyzing the performance achieved with the Adam and SGD optimizers.

Table 6: TOA satellite spectral band wavelength dataset of 2019 results using Adam optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall	F1 – Score
(90, 10)	**96.91**	**94.02**	94.24	**96.26**	**95.24**
(80, 20)	94.55	90.17	94.15	94.65	94.4
(70, 30)	93.22	90.74	**95.19**	91.79	93.45
(60, 40)	89.88	89.42	92.38	92.26	92.32
(50, 50)	87.08	86.07	92.64	88.18	90.36

Table 7: TOA satellite spectral band wavelength dataset of 2019 results using SGD optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall	F1 – Score
(90, 10)	**89.79**	**88.89**	92.61	87.17	89.81
(80, 20)	85.46	81.41	89.23	93.05	91.1
(70, 30)	87.42	87.46	90.64	91.61	91.12
(60, 40)	86.74	84.83	**93.1**	89.81	**91.42**
(50, 50)	85.12	85.81	87.19	**95.3**	91.06

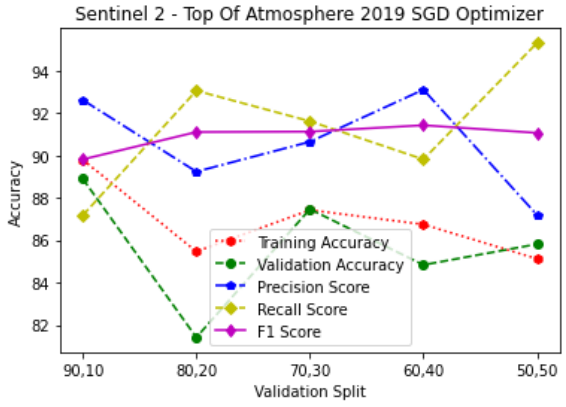

Fig 10: TOA satellite spectral band wavelength dataset of 2019 results using SGD optimizer

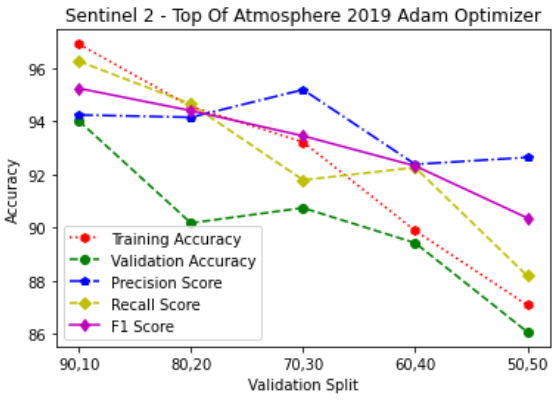

Fig 11: TOA satellite spectral band wavelength dataset of 2019 results using Adam optimizer

5.2.2 TOA satellite spectral band wavelength dataset of 2020 results

In this study, the MLP DNN model is trained using the TOA satellite spectral band wavelength dataset covering the period from May 1st, 2020, to July 30th, 2020. Two optimization algorithms are experimented with: Adam and SGD. In the subsequent sections, the results obtained from the spectral band wavelength dataset of 2020 are analyzed and displayed, focusing on the performance achieved with the Adam and SGD optimizers.

Table 8: TOA satellite spectral band wavelength dataset of 2020 results using Adam optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall	F1 – Score
(90, 10)	**97.77**	**91.03**	**95.03**	91.98	93.48
(80, 20)	97.54	90.38	92.99	92.25	92.62
(70, 30)	89.8	84.9	91.46	**97.5**	**94.38**
(60, 40)	94.58	89.96	93.7	92.93	94.21
(50, 50)	94.78	86.32	93.94	93.22	93.57

Table 9: TOA satellite spectral band wavelength dataset of 2020 results using SGD optimizer

Split Method	Training Accuracy	Testing Accuracy	Precision	Recall	F1 – Score
(90, 10)	86.46	80.77	87.62	94.65	91.00
(80, 20)	82.26	82.05	89.62	94.65	92.07
(70, 30)	86.99	86.32	89.37	96.07	92.6
(60, 40)	85.53	86.32	87.98	**96.47**	92.03
(50, 50)	**88.96**	**88.97**	**91.19**	95.08	**93.09**

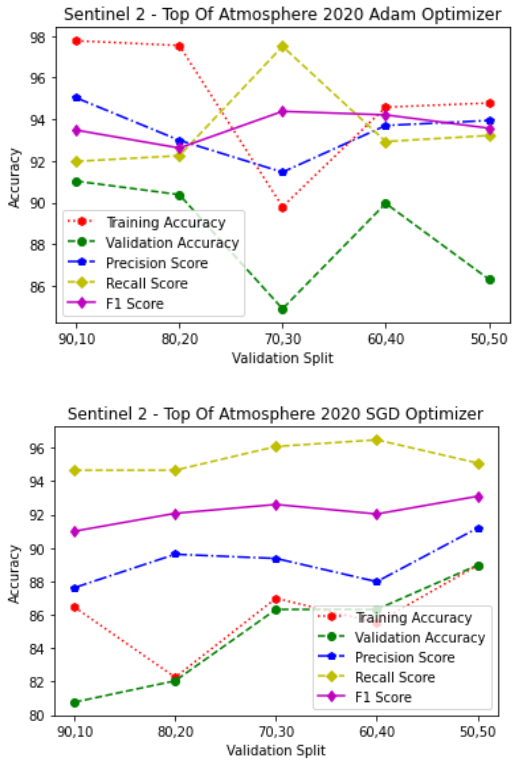

Fig 12:TOA satellite results with Adam & SGD optimizers

6. RESULT ANALYSIS AND DISCUSSION

In the SR satellite spectral band wavelength datasets from 2019 and 2020, as well as the TOA satellite spectral band wavelength datasets from the same years, distinct performance trends were observed when employing two different optimization algorithms, Adam and SGD. For both the SR and TOA datasets from 2019 and 2020, the MLP DNN models utilizing the **Adam optimizer** consistently outperformed those using the SGD optimizer. Across all validation split methods, models trained with Adam exhibited higher training accuracy, testing accuracy, precision scores, recall scores, and F1-measure scores. Additionally, the average testing accuracy was also maximized in models trained with Adam. In particular, the MLP DNN models trained with the Adam optimizer achieved remarkable performance metrics. For the SR dataset of 2019, the training accuracy reached 99.76%, while the testing accuracy peaked at 96.58%. Similarly, for the SR dataset of 2020, the training accuracy attained 99.79%, with a testing accuracy of 94.66%. The precision, recall, and F1-measure scores were also notably higher for the Adam-trained models compared to those trained with SGD.On the other hand, while the models trained with SGD optimizer still demonstrated respectable performance, they consistently lagged behind those trained with Adam. In terms of training and testing accuracy, precision, recall, and F1-measure scores, the models trained with SGD consistently scored lower across all validation split methods.

Based on a comprehensive analysis, the conclusion is that the MLP DNN model is best suited for the SR satellite spectral band wavelength datasets from 2019. These datasets consistently yielded superior performance metrics compared to the TOA datasets and those from 2020. Moreover, the Adam optimizer proved to be the optimal choice for training the model, consistently producing superior results across all datasets and split methods. **The most efficient combination is the SR satellite spectral band wavelength dataset from 2019 using the (90, 10) split method.** This combination achieves the highest testing accuracy of 96.58% and an F1-score of 97.30, making it the optimal choice for robust and reliable model performance.

7. CONCLUSIONS

There are many practical applications for detecting and classifying specific tree species. This research evaluated sample mahogany tree species detection and classification using a combination of Google Earth Engine and multi-layer perceptron deep neural networks at the University of Chittagong, Bangladesh. The findings reveal that the 3-layer MLP model demonstrates a remarkable classification accuracy of 96.58% when applied to the Sentinel-2 Satellite Surface Reflectance dataset. These results show the effectiveness of the proposed approach in accurately identifying and classifying mahogany tree species. The simplicity and efficacy of the 3-layer MLP model make it a valuable tool for identifying mahogany trees that can be extended to other tree species. By providing geographical coordinates, users can quickly determine whether a given location hosts a mahogany tree based on its spectral band wavelength.

Ultimately, this research contributes to the broader goal of promoting sustainability and conservation efforts, particularly in monitoring mahogany tree production with zero deforestation. Moreover, the methodology holds promise for classifying various tree species beyond mahogany. By expanding the dataset to include additional known tree species,

the 3-layer MLP model can achieve even greater accuracy. This capability can be extended beyond Bangladesh, offering global applicability for tree identification. This expansion opens avenues for near real-time identification and tracking of tree species expansion within tropical regions. Through continued refinement and expansion of the dataset, this approach has the potential to make significant contributions to the field of forestry management and environmental preservation.

8. REFERENCES

[1] J. F. Mongillo and L. Zierdt-Warshaw, "Encyclopedia of environmental science," 2000.

[2] S. Bridgewater, "A natural history of Belize: Inside the Maya Forest," 2012.

[3] S. H. Brown, B. Mason, and M. Gardener, "Mahogany: Swietenia mahagoni Family: Meliaceae," *gardeningsolutions.ifas.ufl.edu*

[4] D. A. Friess, E. P. Kudavidanage, and E. L. Webb, "The digital globe is our oyster," *Front. Ecol. Environ.*, vol. 9, no. 10, pp. 542–542, Dec. 2011, doi: 10.1890/11.wb.029.

[5] D. Butler, "Virtual globes: the web-wide world," *Nature*, vol. 439, no. 7078, pp. 776–778, Feb. 2006, doi: 10.1038/439776a.

[6] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, Nov. 2013, doi: 10.1126/science.1244693.

[7] "Blog." https://ai.googleblog.com/2013/11/the-first-detailed-maps-of-global.html (accessed Jan. 22, 2021).

[8] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, Mar. 2007, doi: 10.1080/01431160600746456.

[9] S. R. Phinn, "A framework for selecting appropriate remotely sensed data dimensions for environmental monitoring and management," *Int. J. Remote Sens.*, vol. 19, no. 17, pp. 3457–3463, Nov. 1998, doi: 10.1080/014311698214136.

[10] J. R. Jensen and D. C. Cowen, "Remote sensing of urban/suburban infrastructure and socio-economic attributes," *Photogramm. Eng. Remote Sens.*, vol. 65, pp. 611–622, 1999.

[11] D. Chen and D. Stow, "The effect of training strategies on supervised classification at different spatial resolutions," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 11, pp. 1155–1162, 2002.

[12] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep Learning for Plant Identification in Natural Environment," *Comput. Intell. Neurosci.*, vol. 2017, p. 7361042, May 2017, doi: 10.1155/2017/7361042.

[13] T. Mizoguchi, A. Ishii, H. Nakamura, T. Inoue, and H. Takamatsu, "Lidar-based individual tree species classification using convolutional neural network," *Videometrics, Range Imaging, and Applications XIV*. 2017, doi: 10.1117/12.2270123.

[14] K. Zhang and B. Hu, "Individual Urban Tree Species Classification Using Very High Spatial Resolution Airborne Multi-Spectral Imagery Using Longitudinal Profiles," *Remote Sensing*, vol. 4, no. 6, pp. 1741–1757, Jun. 2012, doi: 10.3390/rs4061741.

[15] J. S. H. Lee, S. Wich, A. Widayati, and L. P. Koh, "Detecting industrial oil palm plantations on Landsat images with Google Earth Engine," *Remote Sensing Applications: Society and Environment*, vol. 4. pp. 219–224, 2016, doi: 10.1016/j.rsase.2016.11.003.

[16] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT Press Cambridge, 2016.

[17] Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., & Roger, J. C. (2018, September 1). *Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences*. Remote Sensing of Environment. https://doi.org/10.1016/j.rse.2018.04.031