Application of Sentinel-2 Satellite Data to Map Forest Cover in Southeast Sri Lanka through the Random Forest Classifier

Thakshila Gunawansa
School of Surveying and Built Environment,
University of Southern Queensland,
West Street, Toowoomba, 4350 QLD, Australia
Department of Engineering Technology,
Faculty of Technological Studies,
Uva Wellassa University
Passara Road, Badulla, Sri Lanka.
u1129332@umail.usq.edu.au, thakshila@uwu.ac.lk

Kithsiri Perera
School of Surveying and Built Environment,
University of Southern Queensland,
West Street, Toowoomba, 4350 QLD, Australia
Kithsiri.Perera@usq.edu.au

Armando Apan
School of Surveying and Built Environment,
University of Southern Queensland,
West Street, Toowoomba, 4350 QLD, Australia
Armando.Apan@usq.edu.au

Nandita Hettiarachchi
Department of Mechanical and Manufacturing Engineering,
Faculty of Engineering,
University of Ruhuna,
Hapugala, Galle, Sri Lanka
nandita@mme.ruh.ac.lk

ABSTRACT

Sentinel-2 satellite data has been used for forest cover monitoring for almost five years. Mapping with Sentinel data will be a cost-effective solution for Sri Lanka, where the lack of updated land cover maps with high spatial resolution is a significant challenge in the land resource management of the country. A study area of about 5,000 km^2 located in southeast Sri Lanka was selected for this study. Agricultural lands, forests including Yala national park, and villages with perennial crops make up the region. A Level-2A Sentinel-2 image with less than 10 percent cloud cover was used in the European Space Agency's (ESA) SNAP software version 8.0.0 for image processing and the forest cover of the study area was mapped through the Random Forest classifier (RFC). Normalized Difference Vegetation Index (NDVI) is also calculated as a Sentinel product to support RFC output. For RFC, ground truth data were collected through the reference of Google Earth high-resolution data. The classification accuracy was assessed using the Google Earth image as the reference dataset. Furthermore, RFC results were compared with NVDI greenness values. The classification accuracy was calculated using a confusion matrix (error matrix) through randomly selected 100 sample points. The overall accuracy of the land cover map was 85 percent, with a 96 percent accuracy for forest cover identification. The study found RFC as an effective method to isolate forest cover in Sri Lanka.

KEYWORDS: *Sentinel-2, Random Forest Classifier, Land cover classification, Land cover mapping, Normalized Difference Vegetation Index.*

1 INTRODUCTION

Humans rely on forests for necessities such as livelihoods, shelter, food, and energy. Beyond these, the forest ecosystem is essential for biogeochemical cycle regulation, carbon sequestration, climate, and support services needed to maintain waste management and detoxification (Aznar-Sánchez et al., 2018; Sheeren et al., 2016). The global land cover has been rapidly changing due to urbanization, agricultural expansion, fire, fuelwood gathering, invasive species, and climate changes (Murayama & Ranagalage, 2020). Adverse impacts include biodiversity loss, decreasing land productivity, and changing climate conditions. Due to these changes impacting human life, effective monitoring mechanisms ensure that natural resources are used correctly. As a result, accurate and up-to-date forest cover mapping has become a valuable resource for agriculture monitoring, land policy development, land cover assessment, forest monitoring and management, scientific research, urban planning, and conservation (Grabska et al., 2019; Thanh Noi & Kappas, 2017). Several studies have been conducted globally on deforestation, degradation, and interventions (Ranagalage et al., 2020).

South Asia is an economically developing region with a rapidly growing population. Sri Lanka is a 65,525 km^2 tropical island in the Indian Ocean, located between 5°55'–9°51' N latitude and 79°52'–81°51' E longitude. Sri Lanka has a unique and diversified ecosystem due to its central hilly topography and extensive waterways. As a result, the geographical patterns of wind, rainfall, temperature, relative humidity, and other climatic components are influenced by topographical variances of diverse regions, such as the central highlands and lowland plains extending to the coastline zone. The country's annual rainfall ranges from less than 1000 mm per year in the southeast and northwest to more than 2500 mm per year in the central mountains.

Population growth has increased in Sri Lanka's agriculture sector activities, particularly in rural areas. The rural population has increased dramatically from 14.1 million in 1990 to 17.8 million in 2020 with the advancement of the country's free healthcare system (Rural population Sri Lanka, 2021). Population and agriculture growth have put tremendous pressure on land utilization (Perera & Tateishi, 2012). Due to the open economic policy implemented in the late 1970s, socio-economic and political changes occurred, and many development projects began. These projects included the development of the Mahaweli river basin, transportation and highway, and rural urbanization (Rathnayake et al., 2020). From 1980 to 2009, Sri Lanka has suffered a civil war, particularly in the North and the East (Höglund & Svensson, 2009). Unfortunately, land cover mapping records for these periods are limited.

Forest loss in Sri Lanka has escalated in the last few decades. It shows that 23,217 km^2 of forests in 1976 decreased to 21,936 km^2 in 2014, a net loss of 1281 km^2 (5.5%) in four decades (Nisansala et al., 2020; Sudhakar Reddy et al., 2017). However, to understand the patterns of forest cover changes and their drivers, Sri Lanka needs comprehensive studies that use robust analytical techniques and available data (Rathnayake et al., 2020). The demand for reliable information on the county's forest cover has grown in the past decades. Furthermore, the available studies are highly localized ,obsolete and restricted to bitemporal comparisons (Rathnayake et al., 2020).

Remote sensing is the technology of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from a satellite or an aircraft). Unique cameras collect remotely sensed images, which help researchers "sense" things about the Earth. Furthermore, remote sensing is an effective method for collecting spatially explicit information on the Earth's surface, such as forests (Belgiu & Drăgu, 2016).

The Sentinel-2 multispectral products produced by the ESA and the European Union (EU) as a part of the Copernicus Programme have contributed to effectively monitoring the Earth's surface. Sentinel-2 satellites contain multispectral scanners onboard and are the second constellation of the ESA Sentinel missions. Sentinel-2's main objective is to deliver high-resolution satellite data for spatial planning, agricultural and environmental monitoring, water monitoring, forest and vegetation monitoring, land carbon monitoring, natural resource monitoring, and global crop monitoring (SUHET, 2015). In addition, many research studies on land cover/use classification have used Sentinel-2 pictures since the launch of the Sentinel-2 multispectral detectors in 2015. In addition, Sentinel-2 can map and monitor forest regions and measure biophysical structures of vegetation (Askar et al., 2018).

Sentinel-2 is regularly used to map forests in many research studies. Table 1 presents the Sentinel-2 product types. Sentinel-2 satellite data shows ample potential for improving forest classification at medium to large scales because of the availability of high spatial resolution.

Table 1. Sentinel-2 product types

Name	Description
Level-0	Not released to users, compressed raw image data in Instrument Source Packet (ISP) format
Level-1A	Not released to users, obtained by decompressing the Level-0 raw image data. A geometric model is developed, allowing any pixel in the image to be located
Level-1B	Not released to users, provides radiometrically corrected imagery with top-of-atmosphere radiance values, and the product includes the refined geometry used to produce the user accessed Level-1C products
Level-1C	Available to users, top-of-atmospheric reflectance are the most commonly used products in land cover/use mapping, systematic generation, and online distribution
Level-2A	Available to users, provides bottom-of-atmospheric reflectance, is the most commonly used product in land cover/use mapping, systematic generation, and online distribution and generation on the user side (using Sentinel-2 Toolbox)

Sentinel-2 sensed data used to determine the status of Sri Lanka's southeast forest cover and forecast future changes. Also, land cover mapping with Sentinel data will be a cost-effective solution for Sri Lanka, where the lack of updated land cover maps is a significant challenge in land resource management.

RFC is a powerful machine learning classifier that produces multiple decision trees using a randomly selected subset of training samples and variables. It provides an algorithm for estimating missing values and flexibility in classification and regression problems (Belgiu & Drăgu, 2016; Rodriguez-Galiano et al., 2012). Higher classification accuracy and the ability to measure variable importance in land-cover mapping are two advantages of RFC (Jin et al., 2018).

The present study prepared a forest cover map of a selected area in the southeast Sri Lanka using Sentinel-2 satellite data through the RFC to investigate the accuracy of the forest map.

2 DATA AND METHODS

2.1 Study Area

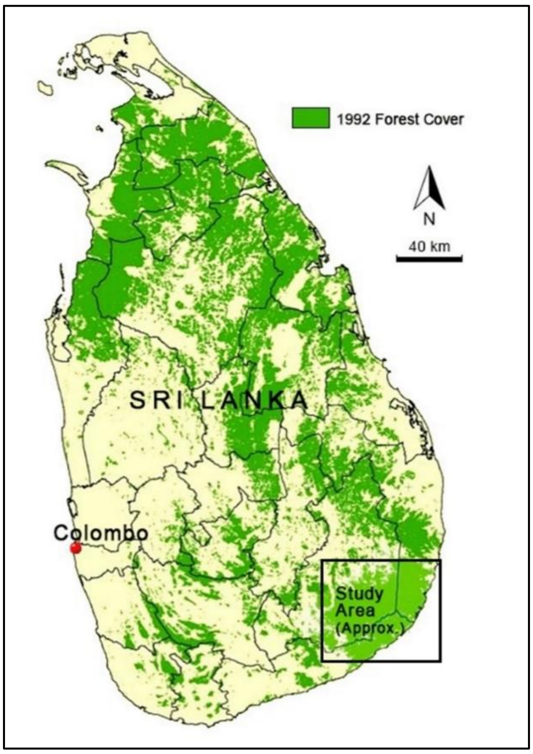


Figure 1. The study area (base map source: www.fao.org)(Sathurusinghe A, 2017)

Figure 1 shows the study area (approximately 5,000 km^2) in the southeast of Sri Lanka. The region mainly accounts for Yala national park, agricultural lands, forests, and villages with perennial plants. Yala national park covers a vast proportion of the selected area.

2.2 Sentinel-2 Data Collection and Pre-processing

Sentinel-2 has 13 spectral bands, three spatial resolution levels of 10 m, 20 m, and 60 m (ESA. Sentinel-2 Missions-Sentinel Online; ESA: Paris, France, 2014), a 290 km swath, a radiometric resolution of 12 bits, and five days of revisit times two satellites (SUHET, 2015). The remainder of the study was focused on 10m bands. Table 2 illustrates the characteristics of ESA Sentinel-2 satellite images.

Table 2. Characteristics of ESA Sentinel-2 satellite images

Spatial resolution (m)	Bands	Wavelength (nm)
10	Band 2 – Blue Band 3 – Green Band 4 – Red Band 8 – NIR	490 560 665 842
20	Band 5 – Red Edge Band 6 – Red Edge Band 7 – Red Edge Band 8A – Narrow NIR Band 11 – SWIR Band 12 – SWIR	705 740 783 865 1610 2190
60	Band 1 – Coastal Aerosol Band 9 – Water Vapour Band 10 – Cirrus	443 940 1375

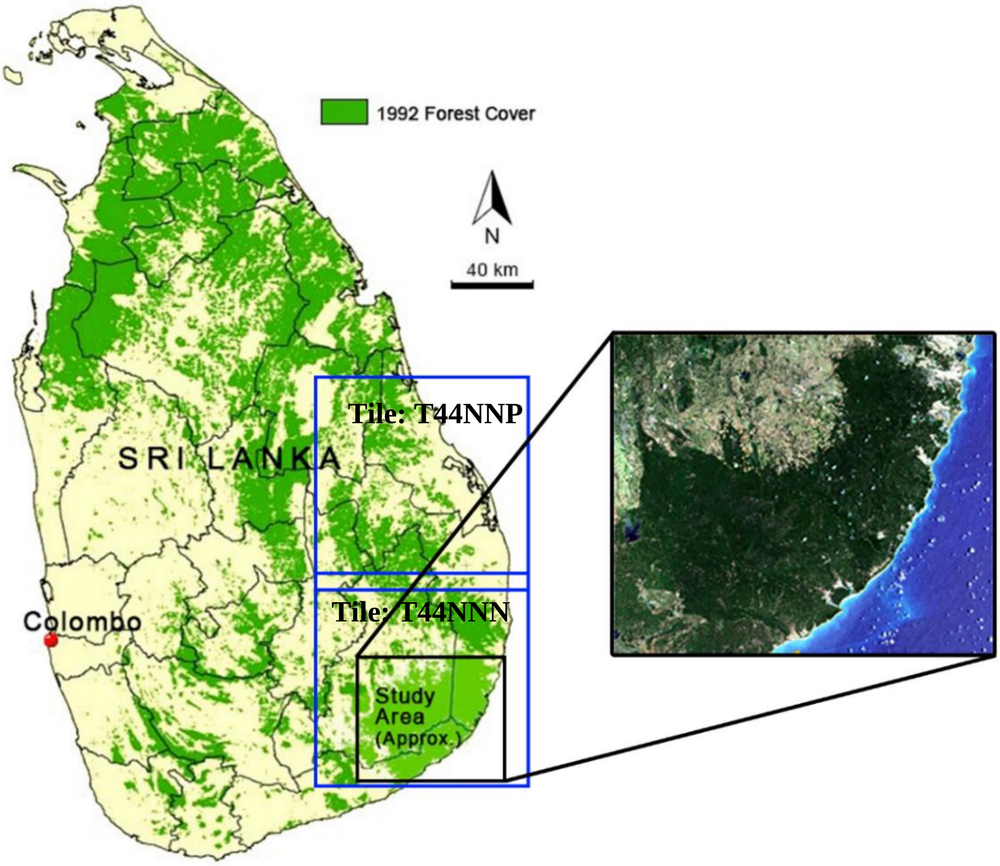


Figure 2. Sentinel-2 imagery of study area

In this study, a Level-2A Sentinel-2 image (20200402T080805) from image tile T44NNN was downloaded to retrieve the bottom of atmosphere reflectance in Sen2Cor's cartographic geometry. Copernicus Open Access Hub is a map browser that allows users to download data from all operational Sentinel missions using the ESA's graphical user interface. Less than 10 percent cloud cover over the entire area and full tile coverage was considered for this study. Figure 2 shows the Sentinel-2 imagery of the study area.

The data were pre-processed using the ESA SNAP version 8.0.0 software developed by ESA. The downloaded image opened in ESA SNAP with profile as Sentinel-2 Multispectral Instrument (MSI) natural colours, responsible for standard RGB composition of the image bands by assigning corresponding spectral bands to the red, blue, and green composition were explored further.

Sentinel-2 photos have varied pixel sizes (10, 20, and 60 meters) based on the spectral band. Resampling is to adjust the pixel size of an image to the required size. Thus, it enables uniform presentation and processing of data with a varying spatial resolution. In addition, most of the functions in the ESA SNAP software require the exact spatial resolution of every band in the source image. The ESA SNAP software was resampled at 10m, and a B2 reference band from the source product was used.

Sentinel data cover large areas, which negatively affects the processing time. As a result, the ESA SNAP program allows users to limit the data's geographic range or select the required bands for processing. The subset image will only show the selected bands that have been chosen. As a result, the area of interest and size of the image was limited to forest boundaries. To generate a subset, one enters the subset range in the geocoordinate section and selects the required bands for the band subset. For this study, B2, B3, B4, and B8 were selected.

2.3 Data Processing

Data processing was based on the training and validation of polygons of Sentinel-2 imagery. Vector data is essential to identify training sites for the RFC. Therefore, training data sites had to be defined and located on the map for the classification. Table 3 summarizes the land cover classification schemes used in this study.

Table 3. Land cover classification schemes

Classification Scheme	Description
Crop Land	The land planted for crops, including cultivated land
Forest	The land grow trees, bushes, covered by natural, newly forested, or planted forests
Residential Land	Land used for townships, and rural settlements
Sea	Sea and lagoon
Inland water	Reservoirs, ponds, and tanks
Clouds	Clouds
Other	Coastal areas and bear land

2.3.1 Random Forest Classifier

Classification of the individual trees is the output in RFC. Therefore, the number of training samples and the number of trees to generate are the required input parameters for RFC. The number of training samples and the number of trees used in this study to generate RFC were fixed at 5000 and 1000, respectively.

2.3.2 Normalized Difference Vegetation Index

According to a variety of sources, NDVI is the most widely utilized index for vegetation monitoring. The computation equation was derived from the ESA SNAP algorithm requirements. Red corresponds to the B4 band, while near-infrared (NIR) corresponds to the B8 band.

$$NDVI = \frac{NIR-Red}{NIR+Red} = \frac{B8-B4}{B8+B4} \tag{1}$$

2.4 Accuracy Assessment

For the study, a measure of validation of each classified map was provided by constructing a confusion matrix between the training areas and the RFC maps. A confusion matrix (or error matrix) is usually used as the quantitative method to calculate image classification accuracy. One hundred ground truth data (pixels) was used and manually digitized an image recorded to build the confusion matrix. Google Earth imagery is an open source and offers a clear view of land cover and land use features with details and can be best utilized for field verifications. In addition, Google Earth also provides ground information such as sharp longitude/latitude readings and elevation data. In this study, Google Earth image was used as the source of ground truth data collection. Figure 3 shows the flow chart of the methodology.

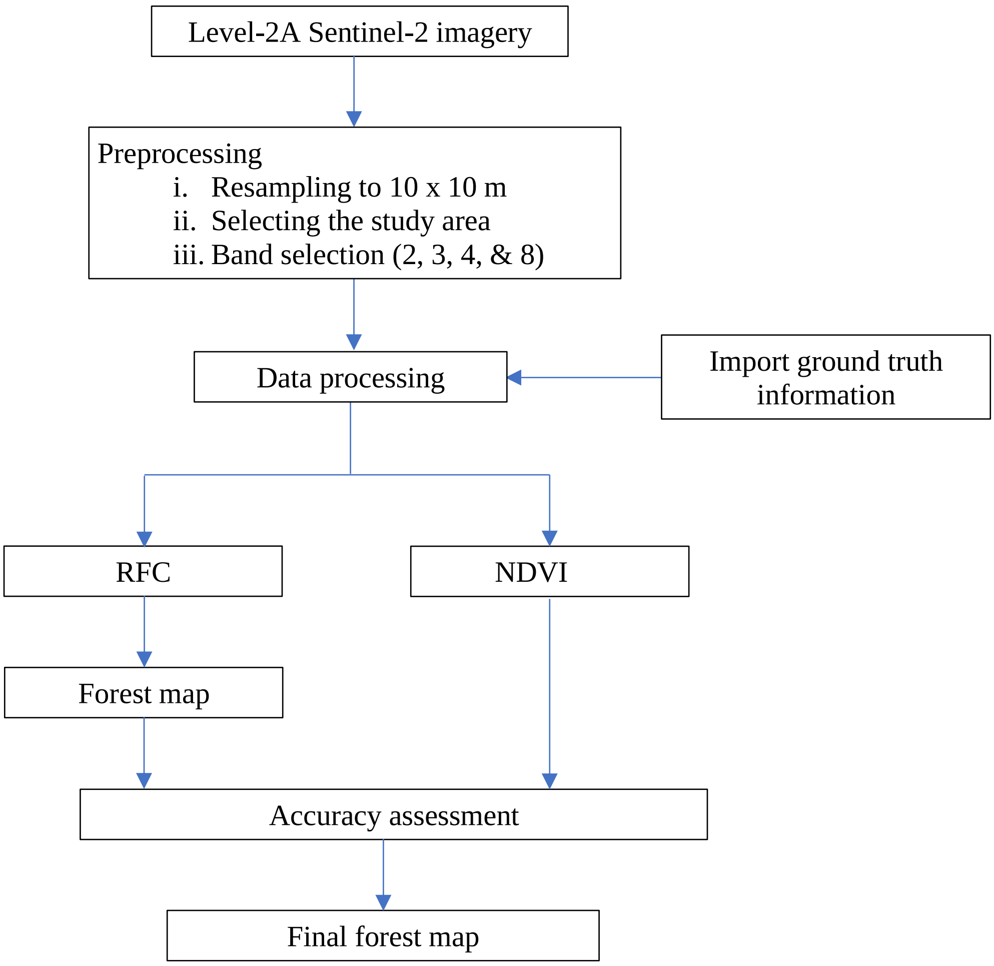


Figure 3. Flowchart of the methodology applied in the study

3 RESULTS

Accuracy assessment of the random forest classified image was verified by comparing the randomly developed dataset with 100 ground truth points. Figure 4 presents the forest cover in the southeast Sri Lanka derived using the RFC and Table 4 illustrates the confusion matrix of the RFC with Google Earth images (figure 5). This visual interpretation was coupled with the expert and prior knowledge of the study area's features. The accuracy assessment was essential for determining the quality of the classified cultivated area derived through the RFC in remote sensing.

For each classification scheme, the overall accuracy (OA) was calculated as the classifier performance estimators. The OA was defined as,

$$OA = \frac{\sum Correct\ predictions}{Total\ number\ of\ predictions} \tag{2}$$

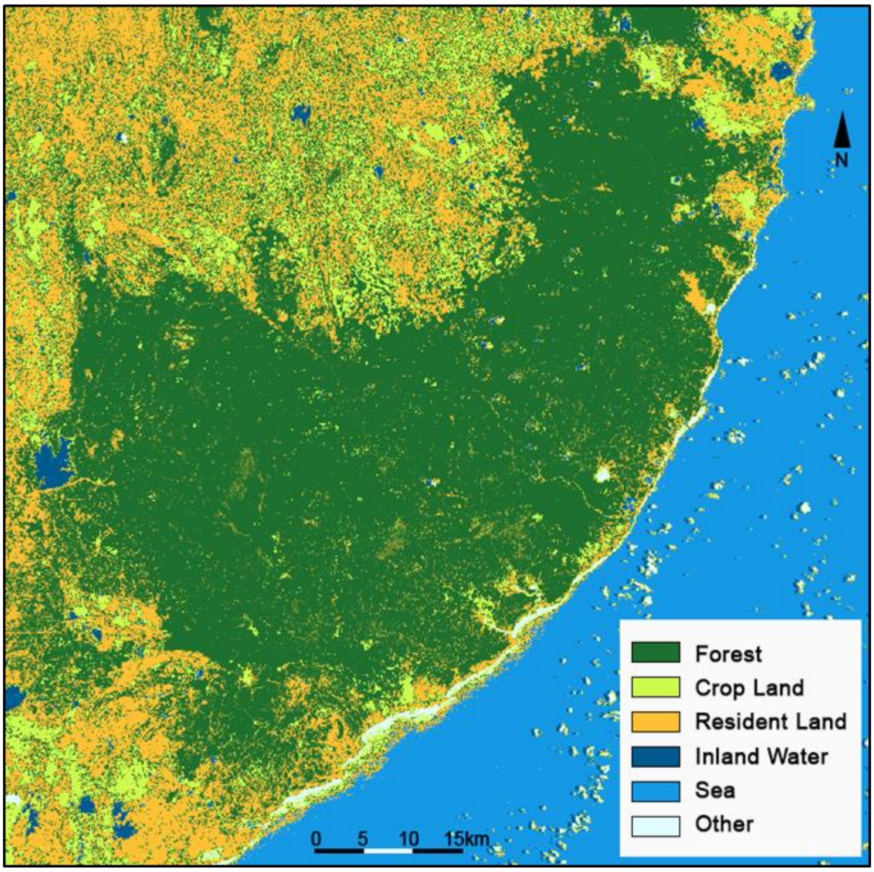


Figure 4. Forest cover in southeast Sri Lanka derived using the RFC

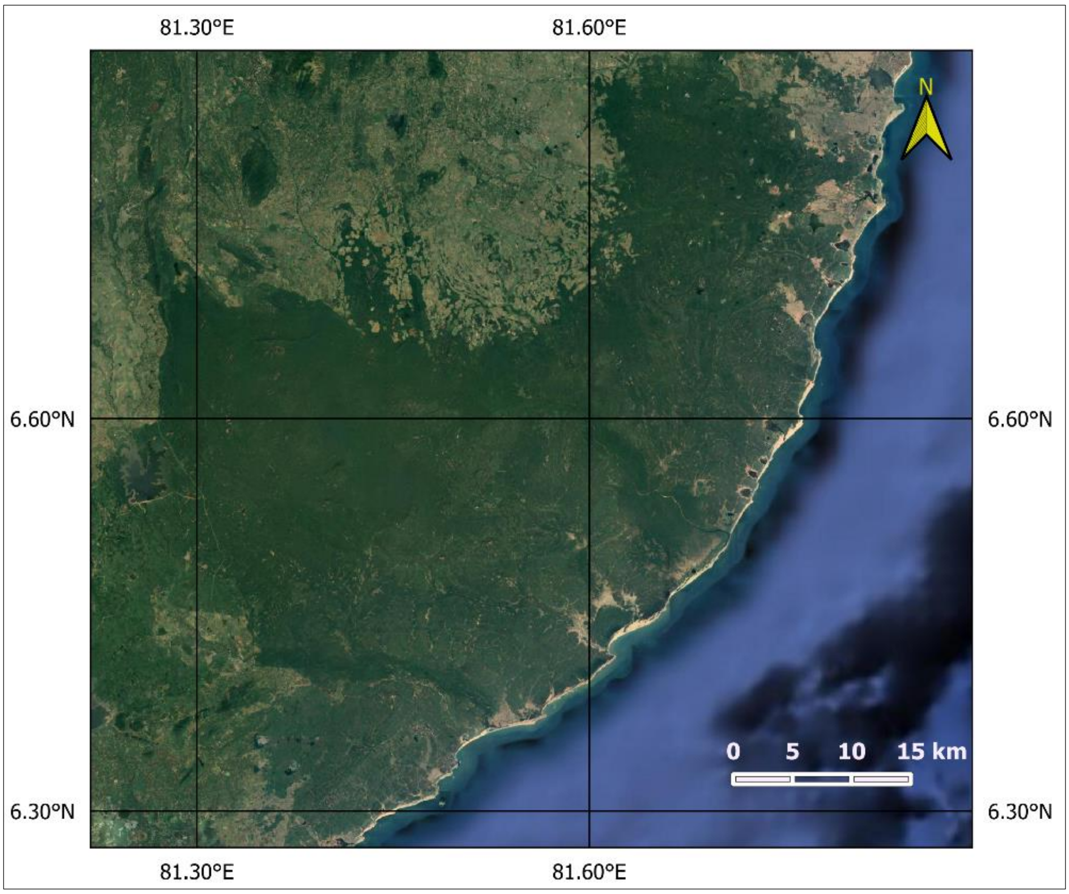
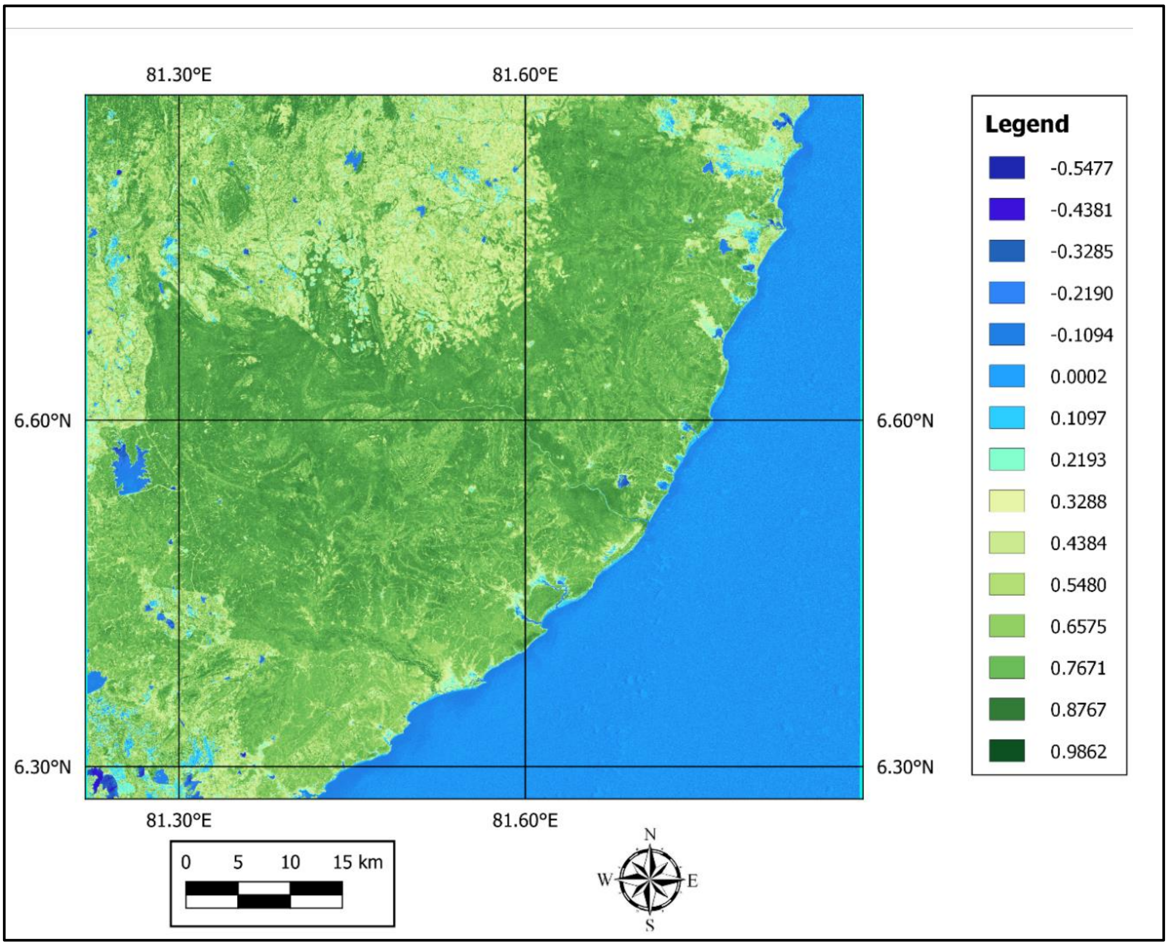


Figure 5. Map of the study area (sky view source: Google map)

Table 4. Confusion matrix of the RFC with Google Earth images

	Reference Image (Google Earth)								
		Forest	**Crop land**	**Resident land**	**Sea**	**Inland water**	**Clouds**	**Other**	**Total**
Classified Image (RFC Map)	**Forest**	48	2	0	0	0	0	0	50
	Crop land	1	12	1	0	0	0	0	14
	Resident land	1	0	3	0	0	0	3	7
	Sea	0	0	0	15	0	0	0	15
	Inland water	0	1	0	0	5	0	0	6
	Clouds	0	0	0	5	0	0	0	5
	Other	0	0	0	0	1	0	2	3
	Total	50	15	4	20	6	0	5	N=100

The accuracy of the land cover map was 85 percent, with a 96 percent higher accuracy of forest cover identification. The accuracy of residential land area and cropland was 43 percent and 86 percent, respectively. During the analysis, the inland water resulted in an 83 percent accuracy. The study found RFC to be an effective method for isolating forest cover in the southeast Sri Lanka.



S

Figure 6. Level sliced NDVI result of the study area

Figure 6 shows the level sliced NDVI result of the study area. The values of more than 0.6575 of the legend indicate the forest's maximum possible areas. NDVI is used to check the accuracy of RFC and is used only as a supporting product to validate where the forest is.

4 CONCLUSIONS

This study examines the performance of RFC using Sentinel-2 satellite data to produce enhanced quality forest cover mapping for Sri Lanka. Sri Lanka's southeast land cover map was downloaded with less than 10 percent cloud cover, and a land cover classification with 7 land cover categories was performed using RFC. Furthermore, randomly selected 100 ground truth points were compared with super-resolution Google Earth images. The super-resolution Google Earth images were used as reference datasets to assess the classification accuracy of forest cover identification. The results show that RFC gives an overall classification accuracy of 85 percent in forest identification. RFC also has 83 percent and 86 percent performances corresponding for inland water and cropland, respectively. The result of RFC is compared to the NDVI to verify the accuracy of the forest cover. NDVI images produced from the same dataset have values more than 0.6575 in NDVI indicate the forest pixels with a higher probability. In the study, RFC was determined to be an efficient method for identifying forest cover in Sri Lanka.

5 ACKNOWLEDGEMENTS

The first author is indebted to the University of Southern Queensland Australia for offering an international Ph.D. fees scholarship. Moreover, the first author acknowledged the institutional facilities from the University of Ruhuna and the Uva Wellassa University of Sri Lanka to carry out this study.

REFERENCES

Askar, Nuthammachot, N., Phairuang, W., Wicaksono, P., & Sayektiningsih, T. (2018). Estimating aboveground biomass on private forest using sentinel-2 imagery. *Journal of Sensors*, *2018*. https://doi.org/10.1155/2018/6745629

Assessment, M. E. (2005). Ecosystems and Human Well-Being. *United States of America: Island Press*, *5*, 563. http://www.bioquest.org/wp-content/blogs.dir/files/2009/06/ecosystems-and-health.pdf

Aznar-Sánchez, J. A., Belmonte-Ureña, L. J., López-Serrano, M. J., & Velasco-Muñoz, J. F. (2018). Forest ecosystem services: An analysis of worldwide research. *Forests*, *9*(8). https://doi.org/10.3390/f9080453

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest stand species mapping using the sentinel-2 time series. *Remote Sensing*, *11*(10), 1–24. https://doi.org/10.3390/rs11101197

Höglund, K., & Svensson, I. (2009). Mediating between tigers and lions: Norwegian peace diplomacy in Sri Lanka's civil war. *Contemporary South Asia*, *17*(2), 175–191. https://doi.org/10.1080/09584930902870792

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. *International Journal of Remote Sensing*, *39*(23), 8703–8723. https://doi.org/10.1080/01431161.2018.1490976

Murayama, Y., & Ranagalage, M. (2020). remote sensing Sentinel-2 Data for Land Cover / Use Mapping : A Review. *Remote Sensing*, *2291*(12), 14.

Nisansala, W. D. S., Abeysingha, N. S., Islam, A., & Bandara, A. M. K. R. (2020). Recent rainfall trend over Sri Lanka (1987–2017). *International Journal of Climatology*, *40*(7), 3417–3435. https://doi.org/10.1002/joc.6405

Perera, K., & Tateishi, R. (2012). Supporting elephant conservation in Sri Lanka through MODIS imagery. *Land Surface Remote Sensing*, *8524*(February), 85241N. https://doi.org/10.1117/12.979382

Ranagalage, M., Gunarathna, M. H. J. P., Surasinghe, T. D., Dissanayake, D., Simwanda, M., Murayama, Y., Morimoto, T., Phiri, D., Nyirenda, V. R., Premakantha, K. T., & Sathurusinghe, A. (2020). Multi-decadal forest-cover dynamics in the tropical realm: Past trends and policy insights for forest conservation in Dry Zone of Sri Lanka. *Forests*, *11*(8), 1–24. https://doi.org/10.3390/F11080836

Rathnayake, C. W. M., Jones, S., & Soto-Berelov, M. (2020). Mapping land cover change over a 25-year period (1993-2018) in Sri Lanka using landsat time-series. *Land*, *9*(1). https://doi.org/10.3390/land9010027

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*(1), 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rural population-Sri Lanka. (2021). The World Bank Group. https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=LK

Sathurusinghe A. (2017, May 25). *Forests and poverty alleviation in Sri Lanka*. Forests and Poverty Alleviation in Sri Lanka. https://www.fao.org/fileadmin/templates/rap/files/meetings/2017/2_Day1.pdf

Sheeren, D., Fauvel, M., Josipovíc, V., Lopes, M., Planque, C., Willm, J., & Dejoux, J. F. (2016). Tree species classification in temperate forests using Formosat-2 satellite image time series. *Remote Sensing*, *8*(9), 1–29. https://doi.org/10.3390/rs8090734

Sudhakar Reddy, C., Manaswini, G., Jha, C. S., Diwakar, P. G., & Dadhwal, V. K. (2017). Development of National Database on Long-term Deforestation in Sri Lanka. *Journal of the Indian Society of Remote Sensing*, *45*(5), 825–836. https://doi.org/10.1007/s12524-016-0636-8

SUHET. (2015). Microbial Adaptation. In *Sentinel-2 User Handbook: Vol. Rev 2* (2nd ed., Issue 1). European Space Agency. https://doi.org/10.1021/ie51400a018

Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, *18*(1). https://doi.org/10.3390/s18010018