Mapping & Classifying Paddy Fields Applying Machine Learning Algorithms with Multi-temporal Sentinel-1A in Ampara district

W. M. R. K Wanninayaka[1], R. M. K. T Rathnayaka[2], and E. P. N Udayakumara[3]

[1] *Computing & Infromation Systems, Sabaragamuwa university of Sri Lanka Belihuloya, Sri Lanka*

[2] *Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka Belihuloya, Sri Lanka*

[3] *Department of Natural Resources, Sabaragamuwa University of Sri Lanka Belihuloya, Sri Lanka*

corresponding_author@kdu.ac.lk

Abstract: In Sri Lanka, Seasonal paddy field area mapping is still doing based on the traditional methods with poor technologies. Therefore this research focuses on the machine approach of mapping paddy fields area accurately on remote sensing data taken from the satellite. Multi-temporal Sentinel-1A Synthetic Aperture Radar(SAR) data was used to map the spatial distribution of the secretary's divisions paddy area in the Ampara district during the period from April 2019 to September 2019. The classifying algorithms were mainly used under the multi-temporal spectral filter classification with 11 dual-polarization(VH/VV) SAR using SNAP, QGIS, ENVI tools. The Time series model was used for each VH and VV bands separately. According to minimum and maximum value of both VH and VV bands, paddy field area was classified using deference of min and max value respectively The overall precision of paddy fields is shown to be 0.92 Also use random forest classification method to processed images with ENVI and It shows 0.86 accuracy rate. Each divisional secretary area showed accurate paddy classification according to non-remote sensing data provided by the district agriculture office of Ampara. This method can easily be used to classify paddy cultivation areas than its traditional methods. Also, it is low cost and very fast method. As further development, Rice prediction model is proposed using the same classified area with vegetation indexes of Sentinel 2 imagery.

Keywords: Rice Yield, Sentinel-1A, Nearest Neighbour, SAR, VV, VH, Time Series.

Introduction

Agriculture is regarded as one of the most important factors of human existence. When it comes to Sri Lanka, Rice is a main agricultural product and widely consumed product. Paddy crops are cultivated in wet-land in every district. There are two main seasons to cultivate paddy in Sri Lanka, which are Yala and Maha. Two seasonal monsoons are greatly affecting the existence of paddies. According to Paddy statistics of 2019 of Yala season which was published by the department of census and statistics, (*Paddy statistics 2018/2019*, 2019) total harvested paddy extent in Sri Lanka was 855,000 acres. Also, paddy production was 1,519,00MT and was 4,896Kg per net hectare. In present, the Agriculture section uses various kinds of technologies. But Sri Lanka still not using its latest remote sensing mechanisms for agriculture field. So that surveys take too many times to provide results about agricultural facts. Almost all document related to surveys was published after the crop cutting. It is not useful for irrigation section and other crop prediction models. Because important data are very helpful for getting the actual situation of agriculture and to take rapid decisions. Lack of real-time information about the agriculture field affects the country's economy. Most of the Researchers used NDVI and its related vegetation indexes based classification for mapping paddy fields. (Filgueiras *et al.*, 2019) Most of the time, its accuracy was very low when considering satellite data imagery which not be in particular harvested time. Also, the accuracy of MODIS satellite data-based researchers is low. (Dammalage and Shanmugam, 2018) Because the resolution is only 200m. thus Sentinel 1 based SAR time series model is more suitable for that purpose. Considering about Yala season in Sri Lanka, There is 8 district mainly participate in the paddy production in Sri Lanka. Among them is Ampara, which is the second-highest paddy production district. It is provided considerable paddy yield compares with other districts. It has 138,515 acres of gross extent sown area, 136,196 acres of gross extent harvested area, and 130,163 acres of net extended harvest. Most of the paddy yields were cultivated from Mahaveli project's water and tanks water in Yala season.

This research is focused on a mapping paddy field in an accurate method. It's more helpful for decision-makers to get real-real time solutions for the affecting problem of paddy. Also Its helpful to forecast actual production of rice. Remote sensing was chosen for that purpose. Paddy life cycle is too small because it is very easy to identify deference in a shorter time duration. It's life cycle changes within 3 months or 4 months. In that case, a suitable way to measure paddy area is to use time serious algorithms collecting much-related satellite imagery. Sentinel-1A is the optimal solution for measuring this kind of changes using many of imagery. Considering the difference between those imageries, it can be a mapped paddy area.

Methodology and Experimental design

A. Study Area

This study area location in Sri Lanka. The district of Ampara falls in the eastern province in the southeast of Sri Lanka. It occupies an area of 4415 square kilometres (1,705 sq mi). (Department of Census & Statistics, 2015) It is bordered by northern districts of Batticaloa and Polonnaruwa, eastern Indian Ocean, south Hambantota district, south-east Badulla and north-western districts of Matale and west-south Monaragala. Paddy production in Ampara District was accounted for 19% of paddy production of the country. (*Paddy statistics 2018/2019*, 2019) It is the second-largest paddy production in a particular season. The winters of Ampara are short, dry, muddy, and mostly gloomy and humid during the year. Ampara's summits are hot and rainy. Temperature common during the year ranges between 24 °C and 34 °C and rarely reaches 22 °C or 36 °C. (*Average Weather in Ampara, Sri Lanka, Year Round - Weather Spark*, no date) The site of the study covers mainly an agricultural zone in which 12.7% of the research area is occupied by paddy. Furthermore, there are also built-up areas and other minority crops. In this analysis, the classification is considered only for the principal land area. The remaining groups are classified as "other." The location of the study area is shown in Figure 1.

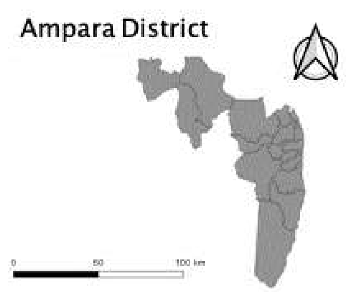


Figure 1. Safe Area

B. Datasets used in the study

The simulation model for process-oriented crop growth uses input variables for crop, weather and soil parameters. Data were

accessed and incorporated into the model from various sources and platforms, such as satellite, weather observations and in situ crop data. This section describes the data sources and pre-processing used in the approach suggested.

Satellite Data: In the Copernicus Programme, the ESA(European Space Agency) launched the Sentinel-1 satellite. The Sentinel-1 mission consists of two satellites that operate on C-band Synthetic Aperture RADAR (SAR), day and night, which allow them to obtain imagery regardless of any weather aberration. The instrument Sentinel-1A SAR works at 5.405 GHz (C-band of approx. 5.6 cm RADAR wavelength), includes 12 day revisits with VH and VV polarizations, and spatial resolution of 5 m by 20 m, respectively, in the range and azimuth directions. The picture size is equal to five looks and 10 meters. We accessed Google Earth Engine(GEE) Sentinel-1A backscatter images. The GEE set comprises the scenes of the S1 Ground Range, calibrated, and updated with the use of the Sentinel-1 Toolbox. For thermal noise reduction, radiometry calibration and terrain correction, each scene is pre-processed with a toolbox from the Sentinel-1. By log-scaling, the final terrain corrected values are converted into decibels (dB). We used VV and VH polarization in this analysis. The pictures S-1A and S-1B from 1 April onwards. 2019 to 30 September. 2019 to 30 September. Downloaded satellite imagery of Sentinel 1A shows below(Table 1)

Ground Truth and other Data: This research was mainly focused on paddy data of Ampara district. It was collected paddy field area and paddy harvest extent related to each district divisional area of Ampara. Paddy statistic information was got from an annual report provided by the department of census and statistics in Sri Lanka. (*Paddy statistics 2018/2019*, 2019) CTDroid Sri Lanka application collected GPS co-ordinates

Table 3. Downloaded sentinel 1A imagery details

DATE	Days different	DOY(2019)	Track	Orbit
2019.05.26	-	146	27	27924
2019.06.07	12	158	27	27924
2019.06.19	12	170	27	27924
2019.07.01	12	182	27	27924
2019.07.13	12	194	27	27924
2019.07.25	12	206	27	27924
2019.08.06	12	218	27	27924
2019.08.18	12	230	27	27924
2019.08.30	12	242	27	27924

for the paddy fields and forest on mobile phones and collected about 60 paddy samples. (Dammalage and Shanmugam, 2018) For the classification, the ground truth data were used and Google Earth application established paddy area and sample collections (approx. 35 samples) of the remaining areas.

As previously mentioned, Ampara district was selected for this research as an area for mapping paddy fields. It is the second-largest district in Sri Lanka and produces a higher amount of paddy yield. It has 19 Divisional Secretary's divisions. So this research focuses on mapping paddy areas for every division and validates accuracy with every district. Also, use google earth to identify sample paddy fields and validate those field classifications. Paddy area related pixels have a very high variety of value than other every object. Because paddy has changed with more variety in their low life cycle duration. Minimum paddy variety value can be selected using non-remote sensing data collected from the district agriculture office in Ampara. For mapping paddy fields, I used multi-temporal Sentinel-1A Synthetic Aperture Radar(SAR) with dual polarization(VH/VV). The whole data set is between April 2019 to September 2019. So Remote sensing data includes 11 imagery which includes these ranges. Considering Sentinel 1 Imagery, There was no need for cloud removal because sIt is not affected with cloud cover. But it needs to deferent preprocessing mechanism according to the selected task.

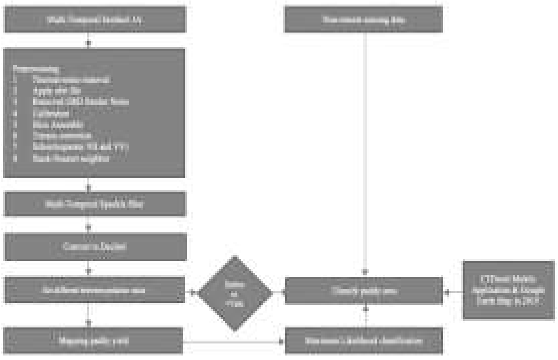


Figure 2. Flowchart of mapping and classification model

C. Sentinel 1 Pre-processing

The preprocessing chain for the correction of the SAR products was needed once the images had been downloaded. Because of speckle noise, SAR images must be corrected radiometric, terrain corrected and filtered. An additional step to quantify a relationship between the spread radiation of the VH and VV bands has also been added to this pre-processing chains. The entire process is carried out with the Sentinel1 products' intensity bands. The Sentinel Application Platform (SNAP) program that is used for this process is supplied without charge by the ESA. Below figure() displayed preprocessing steps for sentinel 1.

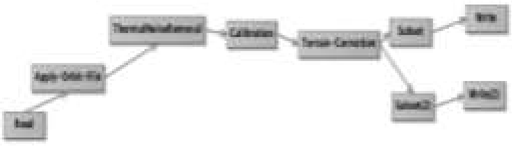


Figure 3. Sentinel 1 Preprocessing Steps

Subsets: Sentinel 1 Preprocessing imageries are subset as VV and VH separately in this process. Finally, it has 22 imagery with dual bands. Those two types of imagery band was stack separately using NEAREST NEIGHBOUR resampling type.

Multi-Temporal Speckle filter: For this study, the Lee sigma filter was used. The preference is based on previous studies showing that the Lee Sigma algorithm has a better performance than other filters for speckles. (Rahman and Thakur, 2018) Most of the researchers were using Multi-Temporal Speckle filtering before the terrain correction. But Ampara district is a flat area. so it is not affected to accuracy.

Conversion to dB scale: Due to the high dynamic range of SAR imagery such as the large range of value, Both stacks of satellite imageries are converted to dB scale. The decibel transformation is used to improve visualization and data analysis.

D. Paddy Area Mapping

Paddy mapping is a critical step in this research. (Mohite *et al.*, 2019) Considering about paddy, It specializes than other crops. Paddy only has a small life cycle such as 2 or 3 months. During this time, It has a large variety of colour and other facts. Sentinel 1 imagery backscatter can easily identify those values with more variation. Helping this scenario it provides more accurate paddy classification than NDVI paddy classification. Thing is NDVI classification is difficult to classify paddy area only. In that case, it must use Random forest or related classification for mapping. Mainly used those two stacks which are VV and VH time-series imagery sets. Separately calculate VV max and VV min values and VH min and VH max values. It can compare the different value of VV and VH and find extract paddy area. When considering the differences between those who value in time series can find real paddy area in Ampara district. The false colour composition was helpful to visualize a more colourful image to identifying paddy field in Ampara district.

E. Paddy Area Classification

Paddy classification of Ampara district mainly uses two types in this research.

Classifying using Divisional Secretary Division: Mapping images accuracy check considering whole divisional district area. There is 19 divisional district area which are Ampara, Alayadivembu, Padiyathalawa, Nintavur, Sainthamaruthu, Pottuvil, Lahugala, Karaitivu, Navithanveli, Mahaoya, Tirukkovil, Sammanthurai, Uhana, Dehiattakandiya, Damana, Kalmunai, Irakkamam, Akkaraipattu, Addalaichenai. Paddy area with manual way considering all divisional district crop yield value.

Supervised Classification: Mapping image classifying with ENVI tool with Maximum Likelihood classification. Polygon based and pixel-based classification using this with validating data respectively sample paddy area and CTDroid based geo-locations. Also, use Google earth and LULC map related to 2014 paddy area field. The best classification with great precision was chosen. The percentage and hectares areas were measured by each divisional district area, the precision measurements were evaluated and if accuracy levels were determined.

F. Accuracy assessment

Both classification methods use different methods for accuracy assessment.

Classifying using DSD(Divisional Secretary Division): The consistency of rice areas mapped with the S1A data was also checked in the district by simple regression techniques at the level of the division. A district's total rice area was determined by multiplying the number of crops per year with the harvested area of rice cultivation systems. This method based on 2019 paddy production of Ampara district related to all divisional district area which is provided from the Ampara district agricultural office. It used a trial and test method. Individually check all accuracy level of each area. Considering the suitable value of most accuracy level.

Supervised Classification: Maximum likelihood classification which is the type of Supervised classification assumes that the statistics are usually distributed for each class in each band and calculates the likelihood of the pixel belonging to a particular grade. All pixels are labelled unless choosing a likelihood threshold. [15] The class with the highest probability (i.e., the Maximum likelihood) is allocated to each pixel. The pixel remains unclassified when the overall likelihood is less than one threshold that defines. ENVI implements Maximum likelihood classification by computing for each pixel of the image the following discriminant features [16], [17] This method based on sample paddy area and geo-location-based paddy and non-paddy point to assess the accuracy of paddy, water and non-paddy based LULC classification.

Results

A. Interpretation of paddy rice backscatter temporal profile

The average time back-scatter profiles for selected rice fields for both VH and VV polarization are shown in Fig. 4. Results showed that the VV polarization was higher than VH for the selected rice fields as observed in previous studies. (Mansaray et al., 2017) In the backdrop of paddy rice fields, it revealed a persistent variance in the temporal profile of the VH polarisation. The accuracy of the backscatter was a measure of the number of cycles. For the VV polarization that reached a peak during the polarization process, it was not the same. This steady increase in VH polarization channel's temporal backscatter profile may be attributable to the signals being less influenced by changes in surface area, which makes it a more accurate channel for characterizing rice growth conditions. On the other hand, as a consequence of increased attenuation of the vertical cylinder (i.e. stems and leaves) with this channel, the sensitivity of VV polarizations to watertight surfaces is easily detected. Regardless of these distinct patterns of rice crop growth conditions, both give variations in backscatter time, which can make an enormous contribution to paddy rice field mapping, with VH and VV polarization in Figure 4 and 5. Although this research has not followed a method, the advantage of using both polarization channels is exemplified in the classification process. The contribution of multi-temporal VH and VV polarization to the process of classification as one method was evaluated in this review.

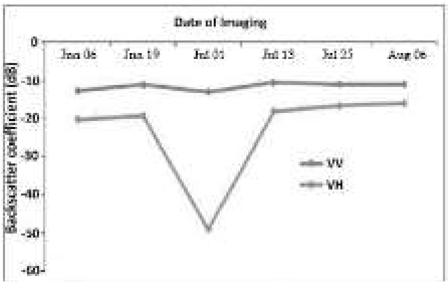


Figure 4. Temporal backscatter profiles of paddy rice fields for both polarizations (VH and VV) of Sentinel-1A imagery

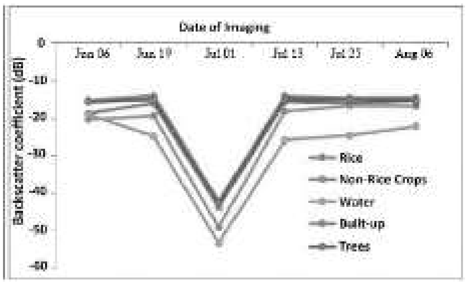


Figure 5.Temporal backscatter profiles of rice and other objects

B. Paddy Area Mapping

Paddy field mapping is very critical things using remote sensing. Using whole 11 satellite imagery, getting actual deferent among those value, because of highest changing value items are the actual paddy area. it can be easily identified using false-colour composition value, In Fig. 6, It shows the paddy area related to yellow colour.

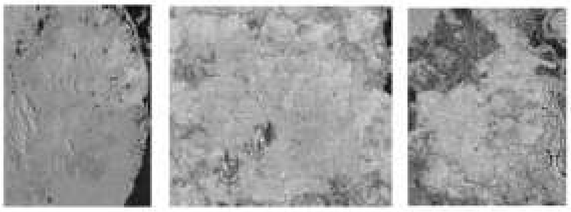


Figure 6. False-colour compositions images in sample paddy fields and mapping paddy area in yellow colour

Using this it can easily identify a particular area. but it can't be classifying accurate value, so getting separately different of dual polarize value can be classifying accurate crop field. VH and VV band can be calculated as VVmin, VVmax, VHmin, VHmax. Deferent of VV and VH value(VVmax-VVmin and VHmax-VHmin) mapping above selected area. Paddy area classifies as considering particular value related to deferent of VV and VH band value. Figure 7 shows that considering the table.

Divisional Secretary area		5.5	5	4.8	4.9	4.95
Ampara District	Actual	55,117				
	Mapping	40281.91	49054.95	55647.34	52272.92	53338.7
	%	26.91563	10.99850	-0.962213	5.160085	3.13781
Alayadivembu	Actual	3303.6				
	Mapping	3174.512	3509.479	3736.231	3534.231	3639.088
	%	3.907495	-6.231959	-13.09574	-6.981202	-10.15122
Pottuvil	Actual	1357.5				
	Mapping	960.854	1133.083	1273.049	1199.342	1235.61
	%	29.21886	16.53164	6.221069	11.65068	8.979006
Navithanveli	Actual	2487.1				
	Mapping	1985.686	2389.188	2683.052	2533.91	2607.569
	%	20.1626	3.936794	-7.878734	-1.882112	-4.843754
Damana	Actual	5246.6				
	Mapping	3562.452	4356.619	4963.704	4651.417	4805.31
	%	32.0998	16.963	5.391987	11.34437	8.410971
Irakkamam	Actual	1594				
	Mapping	1440.227	1660.2	1821.715	1740.769	1781.531
	%	9.646989	-4.153074	-14.28576	-9.207591	-11.76481
Addalaichenai	Actual	2453				
	Mapping	1554.523	1878.782	2108.887	1992.433	2050.881
	%	36.62768	23.40881	14.02825	18.75528	16.39295

Figure 7. VH and VV deferent band value related unique value to classifying real paddy area

The research was choosing deferent value related to paddy fields' accuracy. The accurate band value for deferent of VH band is 4.8. Total district mapping accuracy is 92% & and another individual Divisional area was changing with low deferent. Ampara district actual paddy area is 55,117ha, so predicted area is 55, 647.34ha. Figure 3 shows a sample of two paddy field classification. For identifying the paddy field, use google earth imagery and mapping it with false-colour composition. Finally using paddy classification model, Classifying only paddy area in Ampara district and also divisional secretary area. Figure 8 show deferent between google earth sample paddy area, mapping paddy area(with false-colour compositions) and classifying paddy area. This classifying paddy area can be used to predict paddy production using sentinel 2 vegetation index values. SAR images are very sensitive to identifying surface objects. VV and VH dual-polarization value range included separately with deferent values to identifying that. Time series of Sentinel 1 imageries are more accurate than single imagery for this kind of task. Because changing band value related to the crop age is possible in Sentinel 1. Higher and lower range band value of time series shows the actual different within sowing and harvesting related to paddy area. A value which is above 4.8 shows higher accuracy related to providing sample paddy area details.

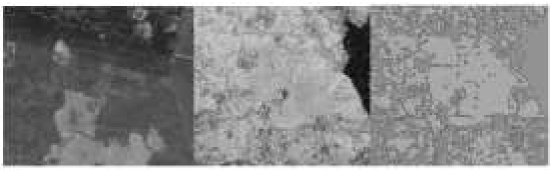


Figure 8. Google Earth imagery, False colour composition value, Classification paddy with value 4.8 show respectively

C. Supervise classification

In Supervised Classification, it used Maximum likelihood classification. Using this classification it used mainly three classes. There are Paddy, NonPaddy, Water. Classification can be displayed as below.

Figure 9. Maximum likelihood classification for Sentinel 1

Classification is done with ENVI software and accuracy assessment also done with that software. Table 2 shows the accuracy assessment report of maximum likelihood classification

Table 4. Classification details of the sample area

Class Name	References totals	Classification Totals	Number correct	Produces Accuracy	User Accuracy
Unclassified	0	0	0	0	0
Paddy	100	92	91	91%	98.9 %
Non-Paddy	100	110	82	82%	74.5 %
Water	40	38	35	87.5 %	92.1 %
Total	240	240	208		

The overall accuracy of the classification is 86%. And overall kappa statistics is 0.85.

Discussion and Conclusion

Sri Lankan government and farmers are facing a lot of problems, because of the lack of yield statistics before harvest. There are no standard methods to getting statics of agriculture during the cultivation period. Real-time mapping method should be included for helping those problems. So, This research tries to focus to solve that kind of problems. When considering those non-remote sensing data which get from the District Agricultural department of Ampara is not more accurate. Because some paddy data of non-remote sensing data are not accurate because manual processing uses previous year data when they can't find real value related to the particular year. Total accuracy of this research is 92%. And Ampara district error is 0.96. It Mapped paddy field in every DS divisions and most of the paddy classifying paddy area validate with google earth imagery. Maximum Likelihood classification which is the part of Random Forest classification is used to this research as another method classifying and validating mapping image. It shows 86% accuracy for mapping area. Research can added predicted model using NDVI and other vegetation indexes with Sentinel 2 imagery as further development for predict Harvest in the whole district.

References

Average Weather in Ampara, Sri Lanka, Year Round - Weather Spark (no date). Available at: https://weatherspark.com/y/110322/Average-Weather-in-Ampara-Sri-Lanka-Year-Round (Accessed: 25 August 2020).

Dammalage, T. L. and Shanmugam, T. (2018) 'Use of Satellite Remote Sensing for Rice Yield Estimation: A Case Study of Polonnaruwa District, Sri Lanka', *Asian Journal of Advances in Agricultural Research*, 7(4), pp. 1–9. doi: 10.9734/ajaar/2018/43843.

Department of Census & Statistics, S. L. (2015) 'Area of Sri Lanka by province and district', *Statistical Abstract*, p. 1.

Filgueiras, R. *et al.* (2019) 'Crop NDVI monitoring based on sentinel 1', *Remote Sensing*, 11(12). doi: 10.3390/rs11121441.

Mansaray, L. R. *et al.* (2017) 'Mapping rice fields in urban Shanghai, southeast China, using Sentinel-1A and Landsat 8 datasets', *Remote Sensing*, 9(3). doi: 10.3390/rs9030257.

Mohite, J. *et al.* (2019) 'Spatialization of rice crop yield using Sentinel-1 SAR and oryza crop growth simulation model', *2019 8th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2019*. IEEE, pp. 1–6. doi: 10.1109/Agro-Geoinformatics.2019.8820245.

Paddy statistics 2018/2019 (2019) *Central Bank Report*.

Rahman, M. R. and Thakur, P. K. (2018) 'Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara District of Orissa State of India', *Egyptian Journal of Remote Sensing and Space Science*. National Authority for Remote Sensing and Space Sciences, 21, pp. S37–S41. doi: 10.1016/j.ejrs.2017.10.002.

J. Abbreviations and specific symbols

DOY-Day of the year

ESA-European Space Agency

EVI2-Two-band Enhanced Vegetation Index

GRD-Ground Range Detected

NDVI-Normalized Difference Vegetation Index

NIR-Near Infra-Red

RADAR-Radio Detection And Ranging

RGB-Red-Green-Blue

SAR-Synthetic Aperture RADAR

SNAP-Sentinel Application Platform

SPOT-Satellite Pour observation de la Terre

UTM-Universal Transverse Mercator

VI-Vegetation Index

WGS-World Geodetic System

ANN-Artificial Neural Network

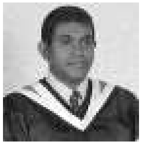
Acknowledgment

I would like to thank Dr. S.V Priyan, Head, Senior Lecturer of Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, for the deepest sentiment of thankfulness and heartfelt appreciation for his assistance, encouragement, guidance not only for the success of this report but also for his continuous concern on the research works all over the time. Furthermore, I also wish to express my gratitude to DR. Kapila Rathnayake for providing me with the technical guidance, advice, all encouragement, and endless help for the success of this research. And also the staff of the department, demonstrators need to remember thankfully

Author Biographies

Mr. W. M. R. K Wanninayaka is undergraduate of Sabaragamuwa university of Sri Lanka. He highly interested to do research in the fields of data mining and spatial sciences.

Dr. R. M. Kapila Tharanga Rathnayaka is presently working as a Senior Lecturer in Statistics, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. He highly interested to do research in the fields of Financial Mathematics, Time series Modelling and Data Mining approaches.



Dr. EPN Udayakumara is presently working as a Senior Lecturer in Faculty of Applied Sciences, Sabaragamuwa university of Sri Lanka. He highly interested to do esearch in the field of Soil Science & Hydrology, Soil Erosion & Conservation, Forestry