PAPER • OPEN ACCESS

Coastal land-use mapping along Johore Straits using Sentinel 1-SAR data with maximizing parameterization of machine learning classification

To cite this article: Shin Yi Chen *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1064** 012022

Coastal land-use mapping along Johore Straits using Sentinel 1-SAR data with maximizing parameterization of machine learning classification

Shin Yi Chen[1], Mazlan Hashim[1,2] and Nurul Hazrina Idris[2,3]
[1]Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia.
[2]Geoscience and Digital Earth Centre, Research Institute for Sustainability and Environment, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia.
[3]Tropical Resource Mapping Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia

Corresponding author: nurulhazrina@utm.my

Abstract. In tropical regions, cloud cover accounts for a major obstacle to detect the coastal land-use while adopting remote sensing technology. The advent of the latest Sentinel-1 C-band synthetic aperture radar (SAR) satellite provides advantages to collect data in all-weathered conditions. In this study, Sentinel-1 images are processed using Lee filter and GLCM-Mean texture analysis in order to enhance the classification results. Several sets of parameters have been tested and these resulted in the optimum overall accuracy by Neural Network with 79.00% in 2015 and 68.29% in 2019. In contrast, Support Vector Machine classifiers obtained overall accuracies of 77.44% and 71.26% in 2015 and 2019 respectively. The results were accessed and it is found that Support Vector Machine outperforms the Neural Network classifier in discriminating data with high heterogeneity properties. Besides, Support Vector Machine has more consistent results in parameter testing compared to Neural Network.

1. Introduction

Coastal land-use issues have recently raised greater awareness in countries all over the world. The coastal region includes the shorelands, coastal uplands, and coastal-influence lands [1]. The Johore Straits is an important developing strait for Malaysia, with significant changes in land use as a result of human activities such as the building of ports, cities, commercial centres, industries, agriculture fields, artificial islands, and so on. Imbalanced coastal land use project management and planning may lead to environmental challenges such as coastal erosion, water pollution, and deterioration of ecosystem services. As a result, mapping and recognising the pattern of land-use changes is crucial.

Remote sensing has been the primary approach for studying coastal land-use changes because of its low cost and multi-temporal variability. However, land-use classification using remote sensing techniques faces many challenges due to the large variation of inherent spatial and spectral information recorded within the data [2]. Sentinel-1 Synthetic Aperture Radar (SAR) data acquired in 2015 and 2019 are utilized in this research, in which the 2019 image is considered as data with high heterogeneity properties due to its complex land-use situation.

The aim of this study is to produce the coastal land-use maps along the Johor Straits area in 2015 and 2019. The objectives are: (1) to classify coastal land-use using Sentinel-1 SAR data with Support Vector

Machine (SVM) and Artificial Neural Network (ANN) classifiers; (2) to compare classification results of coastal land use derived from SVM and ANN classifiers; and (3) to optimize the parameter setting for SVM and ANN classification using Sentinel 1 SAR data.

The Sentinel-1 SAR data has been evaluated and reported as sound for coastal application and land use analysis [3]. The SAR data is acquired with an active sensor, which is an all-weather data acquisition system, capable of penetrating clouds and rain. Hence, it offers the best data source for the persistently high cloud cover in tropical regions such as Malaysia, particularly in the coastal areas. In Malaysia, there are several coastal-related studies utilizing SAR sensed data [4,5,6,7]. These studies have proved the efficiency of SAR data in tropical coast regions such as to map the coastal vegetation [5], to extract shoreline [4], modelling of coastline erosion [6], and to detect coastal spills and pollution [7]. However, there are studies reporting the shortcomings of using only SAR data in classifying land use features without the use of additional optical data [8,9]. The combination of optical and SAR imagery in mapping applications could complement and provide more detailed output. However, the ability of unfused SAR remote sensing in coastal land mapping remains unexplored. It is therefore the scope of this study to evaluate SAR data in coastal land-use classification and later use analysis for the land use changes.

The machine learning classification technique used the Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers to generate the coastal land-use map. These classifiers have been widely reported to be used for the classification of complex image scenes with large variations in irregular land use classes and sizes, typically found in coastal land use features [10, 11]. Both of these classifiers can synthesise regression or classifying functions based on discrete or continuous data sets, are insensitive to noise or over-training, and can handle imbalanced data sets well [12]. Therefore, these machine learning approaches using SVM and ANN, which are relatively robust for accurate and reliable classification, were adopted in the study.

2. Materials and Methods

2.1. Study Area

In this research, the study area covers the coastal area along the Johore Straits (see Figure 1). It is centrally located at the southern part of Johor state at a latitude and longitude of 1°16’19.18" N and 103° 34’00.35" E. The landward corridor limit of the study area is set to be approximately 5 km from the existing shoreline, whereas seaward corridor limits are at the national demarcation boundary line between the Malaysian coastline and Singapore. The study area is relatively flat near the coastline and slightly undulated along the inland boundary. The entire region received an annual rainfall of 2338 mm per year. The total area of the study area is 379383.872 km2.

2.2. Materials This study analyses coastal land-use trends using Sentinel-1 SAR data collected on 2015-11-30 and 2019-10-28. These source data were processed in accordance with the processes outlined in section 2.3. During image classification, the Johor Land Use Map and topographic maps from government agencies are utilised to validate and plot ROI samples. Tables 1 and 2 include information on the data utilised in the study.

Figure 1. Coastal region along Johore Straits (source: Google Earth, url:https://earth.google.com/web/@1.34801803,104.43350010,438861.89573173a,0d,35y,-0.0097h,0.0000t,0.0000r?utm_source=earth7&utm_campaign=vine&hl=en)

Table 1. Specification of Sentinel-1 SAR data

	Details
Sensor-Name	Sentinel-1
Sensor Type	C-Band Radar
Acquisition Date	2015-11-30 and 2019-10-28
Polarization	Dual-mode: VV+VH
Data ID	S1A_IW_GRDH_1SDV_20151130T224731_20151130T224800_008840_00CA00_C769
	S1A_IW_GRDH_1SDV_20191028T224758_20191028T224823_029665_036109_BF21
Spatial Resolution	10m
Product Type	Level 1 Ground Range Detected (GRD)
Mode	Interferometric Wide (IW)
Revisit Period	Every 12 days

Table 2. Source of ancillary data used

Type of data	Source
Johor Land Use Map (2010)	Department of Agriculture, Peninsular Malaysia
Topographic Map of Johor (Sheet number: BG13, BG31, BG32, BG41)	Department of Survey and Mapping, Malaysia

2.3. Satellite Image Processing The processing of SAR data in this study is critical for producing reliable results. The Lee filter with a kernel window size of 5 pixels by 5 pixels was used to decrease the speckle noise. Lee filter with kernel window size of 5 x 5 is selected as it is widely used and has been proved efficient for urban mapping due to its ability in preserving the pixel values in heterogeneous areas [13]. The corrected images are then utilised to create a Grey Level Covariance Matrix (GLCM) mean layer in order to improve SAR imaging visualisation. Window size of 9 by 9 pixels is applied in GLCM mean texture as it is able to present urban regions as a class and has better delineation between swamp forest and oil palm classes [14]. Figure 2 depicts the many stages of SAR image processing.

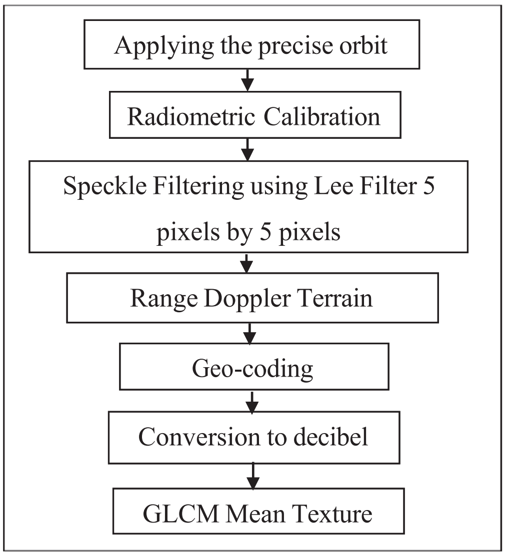
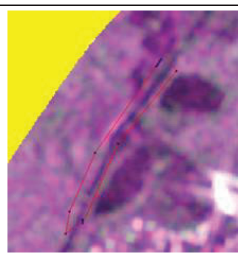


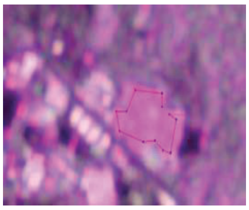
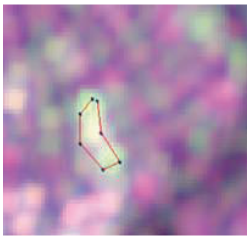
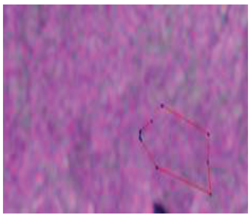
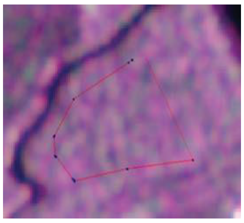
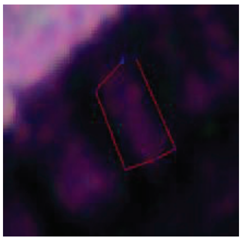
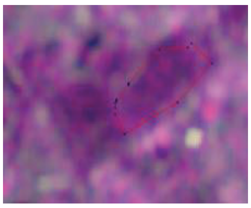
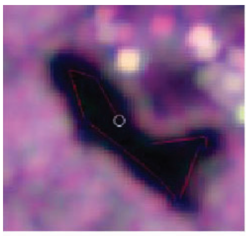
Figure 2. SAR Image Processing

2.4 Selecting Training Samples The SAR images that have been processed are stacked and loaded into ENVI 5.3. The land-use classification scheme designed by the Department of Agriculture, Peninsular Malaysia, is used as a reference to classify the land use. Table 3 shows the land use classes used in this study. To cope with back-scatter fluctuations within each class, the training samples are separated into sub-classes. A sufficient number of training samples are selected using the stratified random technique to best represent the ground area.

Table 3. Descriptions of Land Use Classes (derived from Department of Agriculture, Peninsular Malaysia)

Land Use Classes	Land Use Activities	Sub-classes	Description
1.Urban, Settlements and Associated Non-Agricultural Area	1J Highway/ Main Road	-	
	1U Urban, Residential etc.	Urban 1	

IOP Conf. Series: Earth and Environmental Science 1064 (2022) 012022 doi:10.1088/1755-1315/1064/1/012022

		Urban 2	
		Urban 3	
3. Tree, Palm and Other Permanent Crops	3O Oil Palm	-	
8. Swamps, Marshland and Wetland Forest	8F Swamp Forest	-	
9. Others (Cleared Land / Beach Sand)	9B Beach Sand	-	
	9C Cleared Land	-	
10. Water Body	-	-	

2.4. Neural Network and Support Vector Machine Classification Both classifiers are tested with several parameter sets using ENVI 5.3 to acquire the optimal results. The lists of parameter sets are as displayed in Tables 4 and 5. The Neural Network classifier is also tested with the number of hidden layers 1,2,5, and 10. They are found to have better results with hidden layers of 1.

Table 4. Parameter Sets for Neural Network classification

Set	Training Rate	Training Momentum
1	0.2	0.9
2	0.2	0.8
3	0.3	0.8
4	0.3	0.7
5	0.2	0.7

Table 5. Parameter Sets for Support Vector Machine classification

Set	Training Rate	Training Momentum
1	0.5	100
2	0.5	90
3	0.5	80

3. Results

3.1. Neural Network Classification

For Neural Network classification, five sets of parameters are tested. Table 5 shows the outcomes that were acquired and documented. Overall, the Sentinel-1 SAR data from 2015 obtained a fair overall accuracy level, ranging from 77.75 to 79.00%. The highest overall accuracy level is demonstrated by the parameters in Set 5, which use a training rate of 0.20 and a training momentum of 0.70. Set 3 has the lowest overall accuracy of 77.75%, indicating that parameter setting is more difficult in this study. Meanwhile, the total accuracy of Neural Network classification acquired in 2019 images varied from 56.87 percent to 68.29 percent, which is lower than the findings obtained in 2015. In comparison to 2015 results, Set 2 parameters attained the greatest overall accuracy of 68.29 percent with a training rate of 0.2 and training momentum of 0.8, rather than Set 5.

Table 6. Results of Neural Network Classification (2015)

Set	Overall Accuracy	Kappa Coefficient
1	77.86%	0.7159
2	78.27%	0.7221
3	77.75%	0.7146
4	78.17%	0.7205
5	79.00%	0.7313

Table 7. Results of Neural Network Classification (2019)

Set	Overall Accuracy	Kappa Coefficient
1	67.39%	0.5928
2	68.29%	0.6022
3	63.80%	0.5509
4	56.87%	0.4663
5	65.73%	0.5739

3.2. Support Vector Machine Classification On support vector machine classification, three sets of parameters are examined. Tables 7 and 8 show the results in 2015 and 2019, respectively. Significantly, for 2015 results, all three parameter sets produced acceptable results, with overall accuracy ranging from 77.03% to 77.44% and a kappa coefficient ranging from 0.7072 to 0.7124, indicating relatively good classification performance. In this research, Set 3 parameters (gamma parameter = 0.25 and penalty parameter = 80) have the maximum overall accuracy of 77.44% and a kappa value of 0.7124. Table 8 shows that the total accuracy attained by the three sets of parameters in 2019 varied from 70.99% to 71.26%, with a kappa coefficient ranging from 0.6355 to 0.6385. Set 1 has the highest overall accuracy, at 71.26%.

Table 8. Results of Support Vector Machine Classification (2015)

Set	Overall Accuracy	Kappa Coefficient
1	77.03%	0.7072
2	77.03%	0.7073
3	77.44%	0.7124

Table 9. Results of Support Vector Machine Classification (2019)

Set	Overall Accuracy	Kappa Coefficient
1	71.26%	0.6385
2	70.99%	0.6355
3	71.12%	0.6369

3.3.. Coastal Land-Use Map The optimum results from SVM and NN classification are used to map the coastal land use.

IOP Conf. Series: Earth and Environmental Science 1064 (2022) 012022 doi:10.1088/1755-1315/1064/1/012022

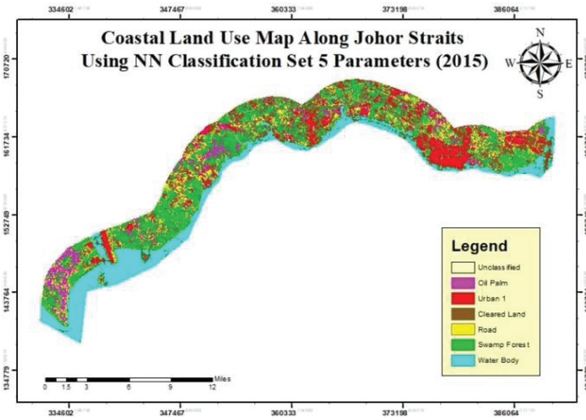


Figure 3. Coastal Land Use Map along Johor Straits using NN Classification Set 5 Parameters (2015)

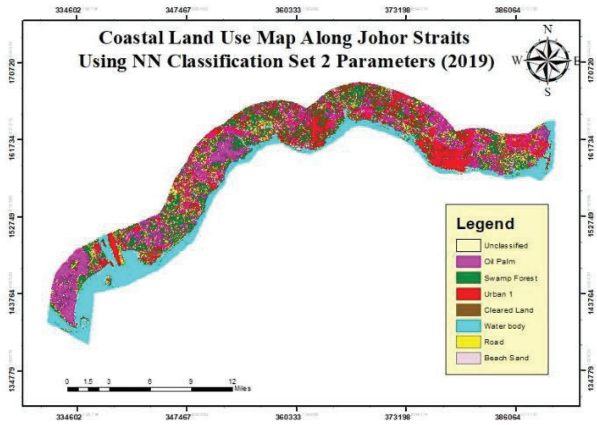


Figure 4. Coastal Land Use Map along Johor Straits using NN Classification Set 2 Parameters (2019)

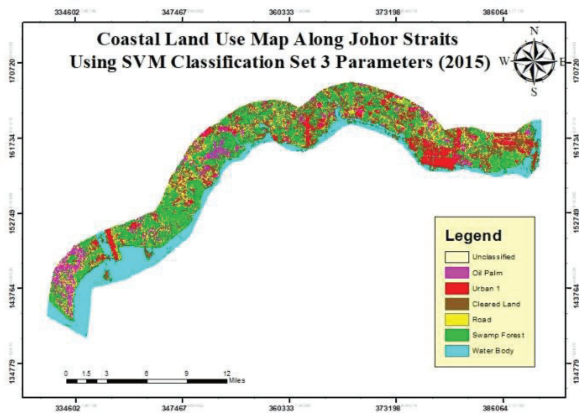


Figure 5. Coastal Land Use Map along Johor Straits using SVM Classification Set 3 Parameters (2015)

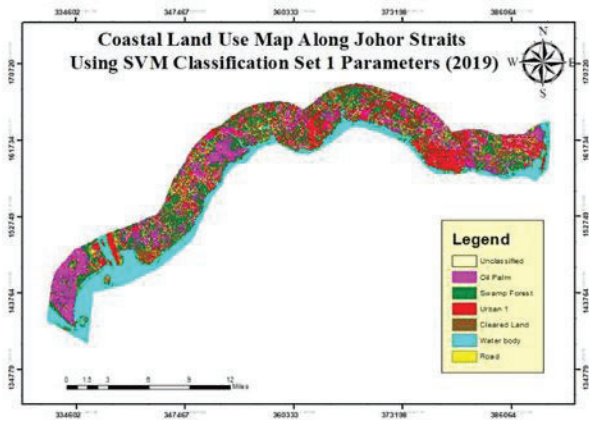


Figure 6. Coastal Land Use Map along Johor Straits using SVM Classification Set 1 Parameters (2019)

4. Discussion

Both Support Vector Machine and Neural Network classifiers perform well in distinguishing coastal land-use classifications. The optimal overall accuracies attained by Neural Network are 79.00% in 2015 and 68.29% in 2019. Meanwhile, the Support Vector Machine classifier produced the best results, with overall accuracies of 77.44% and 71.26% in 2019. Furthermore, SVM classifier outperforms Neural Network in terms of consistency, being less dependent on parameter settings, and able to handle high heterogeneity data (2019 data). The resultant overall accuracies implied the dependency of classifiers on parameter setting, with the SVM classifier having a lower difference between different parameters.

Both Neural Network and SVM classifiers are able to discriminate water classes from other classes. It can be observed through accuracy assessment for NN classification, the producer's accuracies acquired are 92.95% and 90.78% respectively while for user's accuracies instead, NN has achieved 100.00% for both sets of data which are very efficient and reliable. Similar to SVM classifiers, it recorded producer's accuracies of 93.59% and 95.59% while for user's accuracies marked 100.00% for both sets of data. The results are foreseeable as the wavelength of C band SAR employed by Sentinel-1 satellite is well-known for discriminating between land and sea.

Overall, both SVM and NN classifiers have equal performance in classifying all coastal land-use classes. Only slight differences can be observed by the highway/ major road class and swamp forest. According to the accuracy assessment, SVM has better discrimination ability for highway and major road class while swamp forest can be better discriminated using Neural Network classifier with resulted less omission error.

Meanwhile, it has been observed that the urban class has a significant disparity between omission and commission error. According to the statistics, the urban class has relatively low commission error, but it also has a significant degree of omission error, which was about 20-30% and up to 42-43% in 2019. This is due to SAR data's inability to recognize low density residential areas [8]. In addition, for urban classifications, both classifiers recorded effective user accuracies. The user accuracies attained by the Neural Network classifier for urban class are 95.28 % and 94.35 %. Meanwhile, the SVM classifier's results have been quite stable, with 94.39 % and 94.38 % in 2019. As a result, it is claimed that both classifiers can extract urban characteristics accurately given the training data used.

5. Conclusion

The study evaluates the classification performance of machine learning approaches with Sentinel-1 SAR data. Based on the findings, it is determined that SVM image classifier outperforms ANN in identifying land-use activities. Although the NN classifier shows generally good performance in classifying all the land-use classes using SAR data, SVM has better ability to discriminate between highly mixed classes such as major road and oil palm classes. Besides, SVM is more consistent, less susceptible on parameter setting and has better ability in handling the more complex 2019 data.

Acknowledgement

The research is supported by UTM-RUG Iconic RA research grant (Vot 09G73).

References

[1] De Andres M, Barragan J M and Scherer M 2018 *J. Land Use Policy* **71** 121-128
[2] Abdullah A Y M, Masrur A, Adnan M S G, Baky M, Al A, Hassan Q K and Dewan 2019 A *J. Remote Sensing* **11** 790
[3] Phan T N and Kappas M 2018 *J. Sensors* **18** 18
[4] Billa L and Pradhan B 2011 *J. Estuarine, Coastal and Shelf Science* **95** 395-400
[5] Marghany M, Zabidi N, Kadir W H W and Hashim M 2011 *Int. J. of Physical Sciences* **6** 6623-6629
[6] Marghany M, Zabidi N, Kadir W H W and Hashim M 2011 *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (Vancuover)* vol1(IEEE) p3354-3357
[7] Marghany M, Zabidi N, Kadir W H W and Hashim M 2011 *Int. J. of Physical Sciences* **6** 566-576
[8] Tavares P A, Beltrão N E S, Guimarães U S and Teodoro A C 2019 *J. Sensors (Basel, Switzerland)* **19** 1140
[9] Whyte A, Ferentinos K P, Petropoulos G P 2018 *J. Environmental Modelling & Software* **104** 40-54
[10] Ahmad A, Hashim U K M, Mohd O, Abdullah M M, Sakidin H, Rasib A W and Sufahani S F 2018 *Int. J. Advanced Computer Science and Application* **9** 529-537
[11] Szuster B W, Chen Q and Borger M 2011 *J. Applied Geography* **31** 525-532
[12] Osisanwo F Y, Akinsola J, Awodele O, Hinmikaiye J O, Olakanmi O and Akinjobi J 2017 *J. IJCTT* **48** 128-138
[13] Sameen M I, Nahhas F H, Buraihi F H, Pradhan B and Shariff A R 2016 *Int. J. of Remote Sensing* **37** 2358-2375
[14] Fonteh M L, Theophile F, Cornelius M L, Main R, Ramoelo A and Cho M A 2016 *J. of Geographic Information System* **Vol.08 No.04** 11