Pertanika J. Sci. & Technol. 32 (6): 2699 - 2722 (2024)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Machine Learning and Remote Sensing Applications for Assessing Land Use and Land Cover Changes for Under-monitored Basin

Muhammad Amiruddin Zulkifli[1], Jacqueline Isabella Anak Gisen[1]*, Syarifuddin Misbari[1], Shairul Rohaziawati Samat[1] and Qian Yu[2,3]

[1]*Faculty of Civil Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya PersiaranTun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia*

[2]*China Institute of Water Resources and Hydropower Research, State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, 100038 Beijing, China*

[3]*Ministry of Water Resources of the People's Republic of China, Research Center on Flood and Drought Disaster Reduction, 100038 Beijing, China*

ABSTRACT

Urban sprawling caused by industrial and economic growth has significantly affected land use and land cover (LULC). Using satellite imagery for real-time examination in Kuantan has become exceedingly expensive due to the scarcity and obsolescence of real-time LULC data. With the advent of remote sensing and geographical information systems, LULC change assessment is feasible. A quantitative assessment of image classification schemes (supervised classification using maximum likelihood and deep learning classification using random forest) was examined using 2022 Sentinel-2 satellite imagery to measure its performance. Kappa coefficient and overall accuracy were used to determine the classification accuracy. Then, 32 years of LULC changes in Kuantan were investigated using Landsat 5 TM, Landsat 8 OLI, and Sentinel-2 based on the best classifier. Random forest classification outperformed maximum likelihood classification with an overall accuracy of 85% compared to 92.8%. The findings also revealed that urbanisation is the main factor contributing to land changes in Kuantan, with a 32% increase in the build-up region and 32% in forest degradation. Despite the subtle and extremely dynamic connection between ecosystems, resources, and settlement, these LULC changes can be depicted using satellite imagery. With the precision of using a suitable classification

ARTICLE INFO

Article history:
Received: 20 December 2023
Accepted: 29 April 2024
Published: 25 October 2024

DOI: https://doi.org/10.47836/pjst.32.6.15

E-mail addresses:
amiruddinzulkifli7@gmail.com (Muhammad Amiruddin Zulkifli)
isabella@umpsa.edu.my (Jacqueline Isabella Anak Gisen)
syarifuddinm@umpsa.edu.my (Syarifuddin Misbari)
shairul@umpsa.edu.my (Shairul Rohaziawati Samat)
yuqian@iwhr.com (Qian Yu)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

scheme based on comprehensive, accurate and precise LULC maps can be generated, capturing the essence of spatial dynamics, especially in under-monitored basins. This study provides an overview of the current situation of LULC changes in Kuantan, along with the driving factors that can help the authorities promote sustainable development goals.

Keywords: Geographical information system (GIS), image classification, LULC changes, maximum likelihood, random forest

INTRODUCTION

Rapid spawning of population growth significantly affected the LULC rates, especially with the economic and industrial revolution over the past few years (Talukdar et al., 2020). LULC can act as an efficient method to measure land transformation for land use management. With temporal information used in land use maps, a better understanding of dynamic LULC changes to serve different purposes such as urban and town planning, ecosystem and environmental assessment, hazard monitoring and management, natural resources exploration and management, and soil erosion and desertification detection is significant, especially for water resource management. Considering global dynamics and the response to environmental and socioeconomic factors, land use and land cover change is an essential subject.

Conventionally, Land Use and Land Cover (LULC) changes are determined through field surveys with a manual classification that requires more time and energy. It consequently causes data redundancy due to constant changes in land use in a short period. Besides, LULC is important in flood risk assessment, as well as hydrology, meteorology, and geomorphology. Different LULC profiles provide varying surface water retention values, affecting surface runoff. The scarcity of updated LULC data has contributed to increased flooding incidents (Zaidi et al., 2014). High costs, substantial time investment, and considerable human resources often characterise traditional methods of studying LULC. It underscores the growing demand for more convenient, user-friendly, and innovative approaches for identifying and analysing LULC, particularly in regions such as Kuantan. While some governmental agencies and local authorities provide LULC data, accessing this information can often prove to be a significant challenge. It tends to be expensive and difficult to obtain, frequently requiring extensive paperwork and time to access. Another major issue is the lack of updated information. The data these entities provide is often outdated, failing to reflect the current state of land use and land cover in the region. This situation further amplifies the need to develop and implement more efficient and accessible LULC study methods.

In situations where updated and accurate LULC data is lacking, access to remote sensing technology becomes crucial as it can provide this data. Advancement of data acquisition through online databases and advanced equipment based on remote sensing technology such as hyperspectral satellites, airborne, light detection and ranging (LIDAR), and unmanned aerial vehicle (UAV) has helped in providing a relatively good and accurate data (Gaur & Singh, 2023). These high-accuracy and low-error data tend to be expensive and difficult to obtain. In contrast, lower-quality data is freely accessible and requires less computation time. The advent of open-source satellite imagery, like Sentinel-2 and Landsat, has expanded its technology to meet the needs of various fields. The capacity of satellites to identify surface features and create classifications using different algorithmic approaches makes them ideal for LULC studies. Furthermore, the geographic information systems (GIS) technology has improved time and energy efficiency in surveying LULC analysis. Data acquisition is a critical component as it acts as a foundation for successful analysis and decision-making. These upgrades in remote sensing and GIS technology integration also enormously contribute when handling inaccessible and unavailable information in poorly gauged areas, such as disaster assessment, environmental monitoring, and urban management (Maryantika & Lin, 2017). LULC has also significantly helped provide mitigation solutions, policy-making decisions, and sustainable development (Isola et al., 2023; Yulianto et al., 2020; Shahbudin et al., 2009).

Land cover classification through remote sensing imagery analysis is associated with modification and spectral distinction based on each pixel to define the category. The techniques include supervised and unsupervised classifications with different computational algorithms (classifiers). However, the classification of satellite imagery for LULC changes studies can contain several obstacles. Despite the diverse acquisition periods, obtaining similar multi-temporal and spatial satellite image data sets involves extensive work. Other disturbances, such as atmospheric and radiometric interference, may necessitate extra care in image processing (Vicenteserrano et al., 2008). In this regard, advancement in the quality and quantity of algorithms and mathematical equations for image analysis and manipulation has enhanced the capability of extracting information for many applications (Abdullah et al., 2019). Artificial intelligence (AI) and machine learning have been used to integrate remote sensing and geospatial analysis and interpretation of Earth data in recent years (Lary et al., 2016). Among the best algorithms used include random forest (RF) (Abdullah et al., 2019; Balha et al., 2021), support vector machine (SVM) (Balha et al., 2021), k-Nearest Neighbour Network (KNN) (Gondwe et al., 2021; Ngondo et al., 2021), Artificial Neural Network (ANN) (Saini & Rawat, 2023) and Dynamic Time Warping (DTW) (Viana et al., 2019). Given that all the approaches pose different outcomes and accuracy levels, the result highly depends on the training input and regression output. High training input and high regression value generate better classification. Additionally, quantitative measures of LULC mapping based on satellite observations through a machine learning algorithm have been proven efficient (Talkudar et al., 2020). It is important to note that each of these algorithms has unique strengths and application areas. Choosing the right algorithm is crucial to the success of any analytical project.

LULC generation can vary significantly based on software, data availability, study area, and project prerequisites. A pixel-based classification model can improve the classification of satellite imagery based on the training samples. As for ANN, the classification requires an enormous amount of quality data set and poses several problems, including potential overfitting without proper regularisation, computationally intensive and resource-consuming, and high sensitivity to input data quality and pre-processing. However, supervised classification through maximum likelihood (ML) is also considered an excellent parameter for obtaining good results (Nath et al., 2018; Geidam et al., 2020). The probability distribution based on the statistical model of normal distribution has made an ML classifier among the established classification functions (Shahbudin et al., 2009; Seyam et al., 2023). This likelihood is calculated using the highest probability of a certain occurrence within the data. It is worth noting that this approach is grounded in the assumption of a normal distribution within each class. ML can also provide a consistent result that will eventually converge to the true value of the parameter (Gaur & Singh, 2023). Meanwhile, the classifier based on RF imputation compromises the decision tree from backscatter training variables to compute the land cover class of interest. It entails creating several decision trees from a provided training dataset and then producing output classes for each tree. Due to its robustness and efficiency, it has been deemed one of the best and most commonly used methods for classifying satellite imagery. This spatial analysis can provide significantly good results, especially in tropical and subtropical sites (Aja et al., 2022; Abdullah et al., 2019). RF technique serves a relatively good accuracy value with acceptable results, outperforming other machine learning SVM, ANN, and KNN (Saini & Rawat, 2023). The RF algorithm is also considered a stable and consistent overall accuracy compared to ANN and SVM classifiers.

Mapping LULC regularly with remote sensing data can provide insight into the environmental impact of human activities, especially regarding forest disturbances. Forests are among the primary land use categories that play an enormous role, especially in providing various resources for humankind. It helps to maintain the hydrological cycle and atmospheric temperature, reduce natural disaster impacts, and prevent global warming (Ngondo et al., 2021). Terrestrial forests near urbanised areas play significant roles in maintaining carbon flux. Over time, the depletion of total forest cover deteriorates the ecosystem as it causes loss of biodiversity and clean water, the emergence of zoonotic diseases and health issues, and environmental degradation such as floods, soil erosion, and heat island effects. Most of the forest in low-lying regions, like Kuantan, contains a collection of trees and shrubs that can withstand extreme conditions like high salinity, high temperature, and less humidity. In Malaysia, a total of 612580.11 ha of land use and land cover (LULC) area is defined as mangrove forest, with about 17% covering Peninsular Malaysia (Lokman, 2004). Due to rapid population and economic growth, climate change, and global warming, there have been significant shifts in forest distribution. This necessitates improved analysis of LULC changes, especially in cities near high biodiversity and environmental ecosystems.

There is a need for convenient and user-friendly methods to identify LULC changes in Kuantan that enable the efficient identification of forest distribution while minimising time, human resources, and cost. These LULC changes can be identified through satellite image classification, which describes the land uses. This study analysed LULC changes in Kuantan over 32 years (1994 to 2022) using Landsat 5 TM, Landsat 8 OLI, and Sentinel-2 satellite imagery. These varied satellite imageries were used following the need for more unavailability of data in a satellite. Variation in spatial resolution based on satellite imagery is less significant to affect the quality of analysis on primary output (Fisher et al., 2017). Most LULC classifications based on satellite imagery for change detection studies used more than two image satellites with more than 25 years of timeframe. Higher spectral imagery has many elements to be considered because of limitations in terms of cost, processing resources, and data accessibility. Moreover, a high spectral resolution image tends to sacrifice temporal resolution, which implies the frequency of available data in a single pass. The viability of a lower spectral with suitable pixel resolution imagery of Landsat and Sentinel imagery that are openly accessible and cost-efficient has contributed massive amounts of interest to the LULC study. Landsat 5 TM and Landsat 8 OLI have occupied the whole territory of the study area, so it is possible to conduct the study in the study area. However, for Sentinel-2 imagery, the study area of Kuantan required two different distinctive images despite the higher spatial resolution.

The objectives of this study are (1) to perform LULC classification based on the ML classifier and RF classifier and (2) to analyse the LULC changes in the city of Kuantan between 1994 and 2022 using the best classifier between ML and RF algorithm. A reliable and accurate classification based on remote sensing is expected to improve the LULC databases in Kuantan. The need for LULC as a baseline study for managing urban sprawling conservatively and comprehensively without damaging the natural ecosystem and habitat is highly significant, especially in cities in low-lying regions near the river mouth. These regions are exposed to various threats, from sea level rise, climate changes, global warming, and flooding, because of their highly sensitive ecology, which leads to societal vulnerability. Therefore, the study of LULC changes can effectively help in planning, managing, and monitoring the development of cities to ensure sustainable urban development is achieved.

METHODOLOGY

Study Area

This study focused on the LULC patterns in Kuantan (Figure 1), located on the latitude of 3° 48’ 27.72” N and longitude of 103° 19’ 33.60” E, east coast of Peninsular Malaysia. Kuantan is the capital city of Pahang and was originally dominated by natural forests along the Kuantan River. Over time, the increasing demand for various industries has significantly changed the LULC. Following its location at the lower part of the Kuantan River, Kuantan is characterised by a high surface temperature with a low range for minimum and maximum temperature. It is also considered one of the largest cities on the East Coast of Peninsular Malaysia. Kuantan's rapid urban development is mainly prompted by the government's support through economic corridors, which has significantly influenced the LULC by expanding transportation facilities and physical infrastructure. As a result, highly rapid development has influenced the distribution of the environmental ecosystem.

Kuantan also has a unique community rich with values, culture, history, and economic power that contributes to a prosperous quality of life for its population. The local population was recorded at 548,014 people and is expected to increase significantly with 2.1% annual growth. Two of the major economic activities in Kuantan are tourism and industries. The emergence of the petrochemical, timber, and fishing industries has had a significant influence on the city's population growth. Furthermore, the Special Economic Zone was introduced to catalyse the fast-tracking economic development on the East Coast, subsequently leading to a significant increase in employment opportunities. Consequently, there is a possibility of LULC due to the need to explore forests for civilisation.

The local weather in Kuantan has been almost uniform over the years. The area is exposed to the Northeast monsoon that annually happens from early November to March and eventually affects the wind flow patterns. During this period, the region is

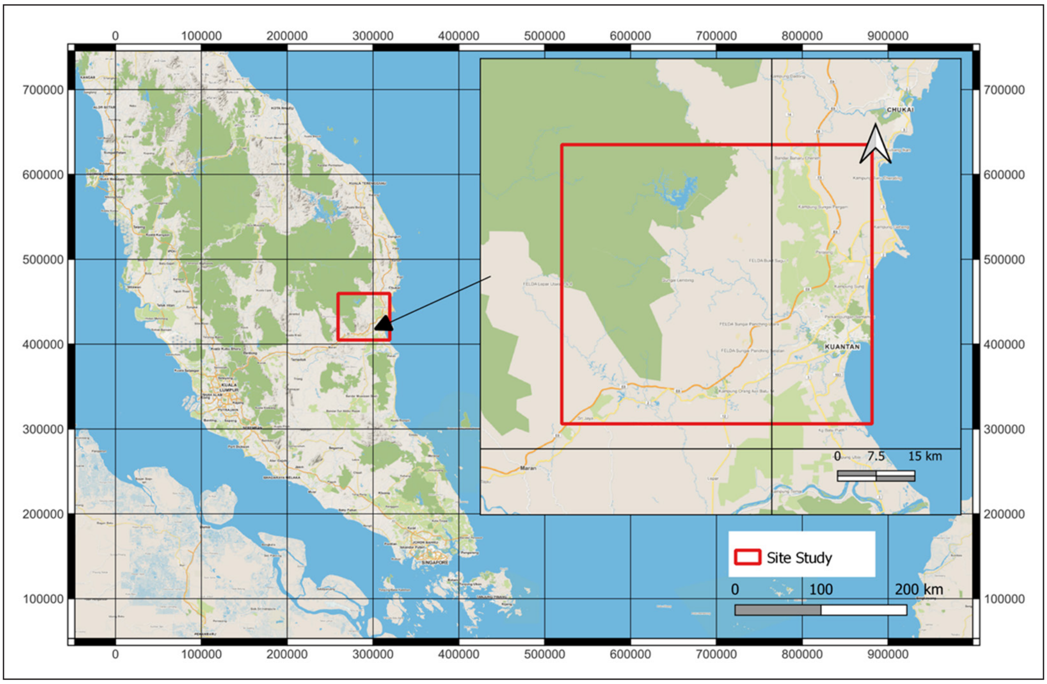

Figure 1. The study area of LULC Classification in Kuantan

susceptible to flooding because the rainfall usually reaches the maximum from November to January, while the driest is usually between June and July. According to the Malaysian Meteorological Department, the average monthly rainfall in Kuantan is between 3000 mm and 3500 mm, which indicates high humidity. As the Kuantan is relatively located near the equator, the region receives an average of 6 hours of sunshine daily. Moreover, the study area also has a uniform temperature throughout the year with relatively small changes in annual variation.

Geographically, Kuantan is among the low-lying cities in Malaysia that are exposed to flooding threats. Rapid development because of increased population growth and extensive industrial development has contributed to increased flooding incidence in Kuantan in recent years. Significant changes in land cover over the past few years have promoted the demand for flood hazard assessment. Furthermore, Kuantan is exposed to the threat of flooding due to its proximity to the estuary and low-lying coastal region, especially during heavy rainfall in the monsoon season. Historically, flooding has happened annually but hardly hit in 1926, 1967, 1999, 2001, 2007, 2011, and 2013, where it caused most areas in Kuantan City to be submerged, resulting in significant damage and loss of life (Zaidi et al., 2014).

Data Acquisition and Preparation

Figure 2 illustrates the processes involved in the implementation of this study. Five satellite imagery data were obtained from the United States Geological Survey (USGS) website (Table 1). Between 1994 and 2002, Landsat 5 TM was the preferred imaging satellite; however, it was replaced by Landsat 8 OLI satellite imagery with a resolution of 30 metres in 2013 and 2017, followed by 10-metre resolution Sentinel-2 satellite imagery in 2022. This selection of satellite images was based on the quality of the images, the availability of satellite images with less cloud coverage, and the absence of line stripping or shot noise. All three satellites provide identical optical images with non-false colour composites. However, Sentinel and Landsat images have slightly different spectral band indications.

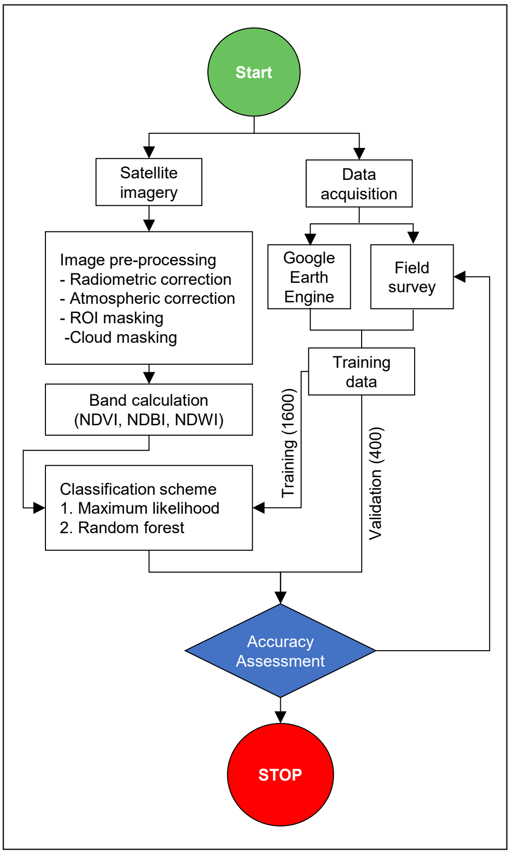

Figure 2. Flow chart of the study for LULC classification in Kuantan

Table 1
Satellite imagery used in the study for LULC classification

Data Type and Date	Description	Source
Landsat 5 TM (1994, 2002)	The data was atmospherically corrected, and four visible and near-infrared bands were obtained: two short-wave infrared bands and a thermal infrared band. The spatial resolution of 30-m, single scene data, Sun-synchronous	Google Earth Engine database—USGS
Landsat 8 OLI (2013, 2017)	The data was atmospherically corrected, and five visible and near-infrared bands, two short-wave infrared bands, and two thermal infrared bands. The spatial resolution of 30-m, single scene data, Sun-synchronous	Google Earth Engine database—USGS
Sentinel-2 (2022)	The data was atmospherically and geographically corrected, and three visible bands, five near-infrared bands, four short-wave infrared bands, and an ultra-blue band were obtained. The spatial resolution of 10-m, double scene data, Sun-synchronous	Google Earth Engine database—USGS

A relatively high sensor of Sentinel-2 provides a good spectral indication for detailed vegetation analysis compared to Landsat 8.

Once the satellite images were obtained, radiometric correction techniques were applied. For Landsat 5 TM imagery, the sensors convert the data quantisation scale at an 8-bit digital number (DN), equal to 256 greyness level. Then, the DN values were converted to radiance values using bias and gain values. Then, the radiance values were converted to a Top-of-Atmosphere (ToA) reflectance unit. Meanwhile, for Landsat 8 OLI, the data was rescaled to 16-bit DN values, equal to 65536 greyness level, which also required conversion from the DN value to the radiance value. Based on the radiance values, the data were converted to a ToA reflectance unit. As for Sentinel-2 imagery, the DN values were converted directly to the ToA reflectance unit. After radiometric correction, atmospheric correction through Dark Object Subtraction 1 (DOS1) was applied using the Sentinel Application Platform (SNAP) engine. Noise from atmospheric and spectral interference on satellite images was minimised through this process. Then, the satellite images underwent cloud removal through the threshold value of the reflectance range on the Near-Infrared band (NIR). Then, the masking process focuses on the region of interest (ROI) area after the image stacking from different bands is performed.

While Landsat imagery required a single satellite image, Sentinel-2 required two-scene images to cover the study area. Therefore, image mosaicking was required for the prepossessing analysis. The characteristics of each pixel for image classification were identified based on the colour, shape, size, pattern, location, and DN associated with each object. The classification was performed with the help of spectral indexing on the satellite imagery to allow each class to be prominent compared to other classes. Subsequently, it will reduce the redundancy of pixel classification. The index-based technique was also utilised by manipulating the spectral properties of satellite imagery through the derivative equations of Normalised Difference Built-up Index (NDBI) (Equation 1), Normalised Difference Vegetation Index (NDVI) (Equation 2), and Normalised Difference Water Index (NDWI) (Equation 3) for better visualisation.

$$\mathrm{NDBI} = \frac{\mathrm{SWIR} - \mathrm{NIR}}{\mathrm{SWIR} + \mathrm{NIR}} \qquad [1]$$

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \qquad [2]$$

$$\mathrm{NDWI} = \frac{\mathrm{GREEN} - \mathrm{RED}}{\mathrm{GREEN} + \mathrm{RED}} \qquad [3]$$

Training Data Based on Field Data Survey

The satellite image was selected to complement the site investigation, where a set of training areas with a training-to-accuracy sample ratio of 4:1 was generated for each class. It adhered to the recommendation by Aja et al. (2022) to avoid any systematic error caused by redundant pixels for training and validation. A total of 2000 randomly selected training sample points were used, with 1600 points for training and 400 points for validation. These training data, consisting of positional coordinates and LULC information, were collected through a mixture of field measurements and Google Earth Engine. It is to ensure that the training data are correctly assigned as they reflect the characteristics of each classified object. Upon completing the image classification, 400 points were used for accuracy analysis. Some classes of LULC tend to have similar spectral properties and less heterogeneity, making LULC classification difficult to process. Therefore, several LULC classes were combined to represent a major class, as presented in Table 2.

Table 2
Description of LULC types for the study

ID Class	LULC Types	Description
1	Urbanisation/Build-up area	Consists of all build-up regions, residential, industrial, commercial regions, villages, transportation, infrastructure, and buildings.
2	Vegetation	Consists of mixed forest, dense forest, shrub and mangrove forest, and agriculture plants.
3	Water Body	Consists of water bodies, including rivers, oceans, lakes, reservoirs, ponds, and others.
4	Barren Land	Consists of less to no plantation with only soil exposed at the Earth's surface.

Satellite Image Classifier Test

LULC Classification

For classifier testing, the satellite data of the year 2022 from Sentinel-2 was acquired from the USGS website and utilised for LULC mapping by employing consistent and stable classification algorithms, which are RF and ML techniques. Equation 4 presents the mathematical representation of the RF algorithm, showcasing its function in reducing errors and improving efficiency. Meanwhile, the ML model is mathematically represented as L(θ|x), a function of the parameters (θ) and the observed data (x), as shown in Equations 5 to 7. Here, θ denotes the model parameters that influence the final result. The observed data, labelled x, refers to the collected and analysed real data. For optimal classification, the parameters of θ are characterised by the mean (μ) and standard deviation (σ). The mean is the central value of a data set, providing an average point, while the standard deviation measures the data's dispersion or variation.

$$\mathrm{RF} = 1 - \sum_{i=1}^{n} (P_i)^2 \qquad [4]$$

$$L(\mu, \sigma | x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \qquad [5]$$

$$\mu_{ML} = \frac{1}{n} \sum_{i}^{n} x_i \qquad [6]$$

$$\sigma_{ML} = \sqrt{\frac{1}{n} \sum_{i}^{n} (x_i - \mu)^2} \qquad [7]$$

Additionally, the QGIS applications and SNAP engine were applied. The RF algorithm was run using 100 trees based on 4 input features (Table 2) from 1600 points training dataset. As for ML, a classification based on normal estimation with 1600 points of training dataset was applied.

Accuracy Analysis

The overall accuracy and Kappa coefficient were calculated using Equations 8 and 9. The training dataset (400 points) was undertaken in a specific area based on a random location. The Kappa coefficient (K) (Equation 7) and overall accuracy (OA) (Equation 8) analyses based on the confusion matrix were applied to conduct an accuracy analysis in determining the classification's validity and reliability, where P_0 indicates the relative probability of observed agreement while P_e indicates the hypothetical probability of chance of agreement.

$$K = \frac{p_0 - p_e}{1 - p_e} \qquad [8]$$

$$\mathrm{OA} = \frac{\text{Number of correct prediction}}{\text{Number of total prediction}} \times 100 \qquad [9]$$

Time Series Comparison

Once the classification testing was done for the LULC map in 2022, LULC maps for 1994, 2002, 2013, and 2017 were produced using the best image classifier. The highest overall accuracy and kappa coefficient were the definitive indicators for selecting the best classifier to determine the LULC changes. A total of 5 LULC maps were constructed using computational analysis, Google Earth Engine, and field surveys in 1994, 2002, 2013, 2017, and 2022. The distribution area of each LULC class for 1994, 2002, 2013, 2017, and 2022 was also calculated. This inconsistent interval of satellite images from 1994 to 2022 is due to the unavailability of good-quality satellites caused by high cloud cover and spectral noise in the study area. Sentinel and Landsat satellites provide the same optical sensor for non-false colour composite imagery, allowing for comparative analysis. A comparative analysis examined the effects of differing spatial and spectral resolutions. Sentinel-2 appeared to offer superior technical satellite properties compared to the Landsat satellite. However, due to the higher quality of Landsat 9's 16-bit radiometric resolution, compared to Sentinel-2's 12-bit radiometric resolution, the accuracy of classification results was nearly identical. Hence, the same technique (normalised treatment) was used to compensate for these different resolutions. This distinction is necessary to identify the differential impacts on the results from the same approach. Moreover, resampling of the training for different approaches is not required in this context, as the technique acts equivalently to the same training sample.

RESULTS AND DISCUSSION

Training Classifier

LULC Classification

The supervised classification of the ML method was used to classify the LULC map (Figure 3). By using threshold probabilities to look at the binary profiles of individual pixels, it was possible to classify the data using the spectral band's normal distribution and determine the statistics and odds for each class. Five classes were developed: water, urban area, vegetation, bare land, and an unnamed class. Due to the unknown parameters in the defective proportion, the normal distribution adopted in maximum livelihood, this unidentified class was established as a defective item. The LULC map based on the RF technique (Figure 4) was divided into four primary categories: water, urban area, vegetation, and barren terrain. This classification was done using a decision tree that included multiple regression and classification trees from a random subset of samples. Subsequently, the data from the ML tends to avoid misinterpretation compared to RF classification.

Accuracy Assessment

A confusion matrix based on pixels was used to provide a statistically accurate assessment of both classifications. The accuracy achieved a 0.05 confidence interval for the statistically significant correlation of the two LULC classifications using ML

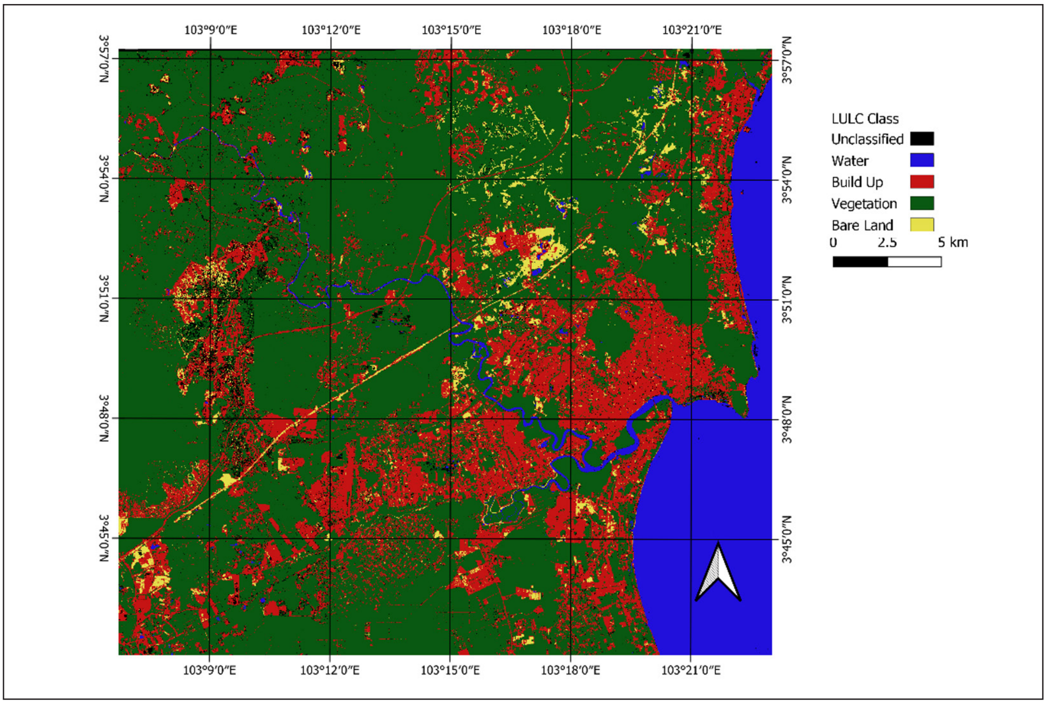

Figure 3. LULC map using ML algorithm in 2022

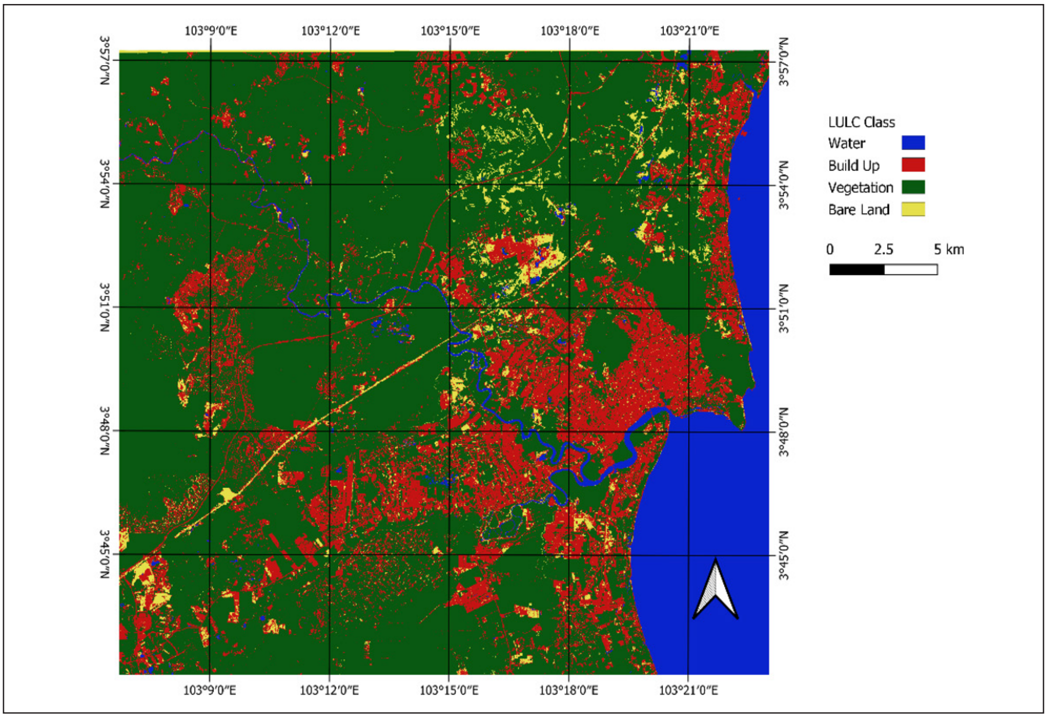

Figure 4. LULC map using RF algorithm in 2022

and RF, which were significantly good at 92.8% and 85%, respectively (Tables 3 & 4). The large number of estimates in the training dataset caused the large differences in the classifier results between RF and ML classifications. RF classifiers can handle large-scale heterogeneous distributions and complex data with less data bias. A study by Talukdar et al. (2020) proved that LULC classification poses a significant result using an RF classifier. Lary et al. (2016) also found that RF classifiers can reduce classification bias compared to other classifiers. However, RF classification has certain problems with the initial variable biases and overfitting of the regression model. Misclassification and overlaps significantly influenced the accuracy of the classification (Saini & Rawat, 2023). The bias of the majority RF classification emerges from uneven classes in the training dataset. It can also be deduced that the RF classifier produced less overall accuracy compared to the ML classifier because of the quality of the training dataset. The training dataset for the classification was monitored in a highly dense area with various classifications summarised into four main classes. These available datasets might be insufficient for training the RF algorithm, limiting the variety of predictions.

The effectiveness of the RF classifier in handling highly heterogeneous features can be proven with a well-trained sample. Abdullah et al. (2019) showed that training dataset preparation is highly important in generating a good image classification. The authors also emphasised that the training dataset for the RF model is relatively sensitive to changes that can cause a biased result. Talukdar et al. (2020) further emphasised that the RF classifier can pose a good result when dealing with heterogeneous data containing various types of features. As for ML, the classification of the imaging satellite was relatively good since the data was derived from the normal distribution of the dataset, making the classifier suitable for a relatively immense dataset. Lower variance estimation affected the classifier's ability to withstand a broad sample size, as a high sample size can produce less biased outcomes. Therefore, the data was consistent and almost identical to its actual value.

Table 3
Area-based error matrix between LULC-based ML by 400 points

		Training Data Set					Total
	Classes	**Unclassified**	**Water**	**Urban Area**	**Vegetation**	**Barren Land**	
Maximum likelihood	Water	0	28	0	0	0	28
	Urban Area	5	1	161	4	0	171
	Vegetation	3	0	6	163	0	172
	Barren Land	4	0	4	2	19	29
Total		12	29	171	169	19	1
Overall Accuracy (Confident interval of 95%)							92.8
Kappa Hat Classification							0.85

Table 4
Area-based error matrix between LULC-based RF by 400 points

		Training Data Set				Total
Classes		**Water**	**Urban Area**	**Vegetation**	**Barren Land**	
Random Forest	Water	25	2	1	0	28
	Urban Area	1	145	24	1	171
	Vegetation	0	13	150	9	172
	Barren Land	0	6	3	20	29
Total		26	166	178	30	400
Overall Accuracy (Confident interval of 95%)						85
Kappa Hat Classification						0.75

LULC Changes

LULC Changes Analysis

The classifier testing results indicate that classification using ML yields higher overall accuracy compared to RF. Therefore, the LULC changes for the years 1994, 2001, 2013, and 2017 were performed using the ML classifier. The results of the LULC classification for 1994, 2001, 2013, 2017, and 2022 are presented in Figures 6 to 10, respectively. The total coverage of LULC in all studied years is presented in Table 5. In general, LULC observation in Kuantan based on image classification from 1994 to 2022 was dominated by urbanisation and vegetation. There was a significant inverse relationship between the build-up and vegetation classes. Figure 5 shows a constant increase of build-up coverage by approximately 32% of the total coverage area changes found from 1994 to 2022. As for vegetation, the coverage dropped from 1440 ha of the total area to approximately 1171 ha before it bounced back with a slight increment of 30 ha in 2017. Meanwhile, 2022 recorded

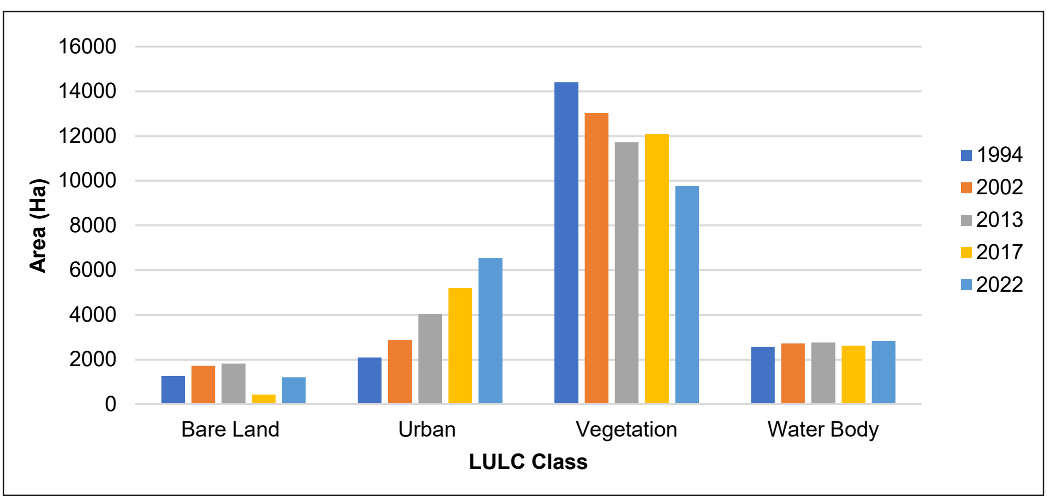

Figure 5. LULC changes using the ML classifier in Kuantan by area coverage from 1994 to 2022

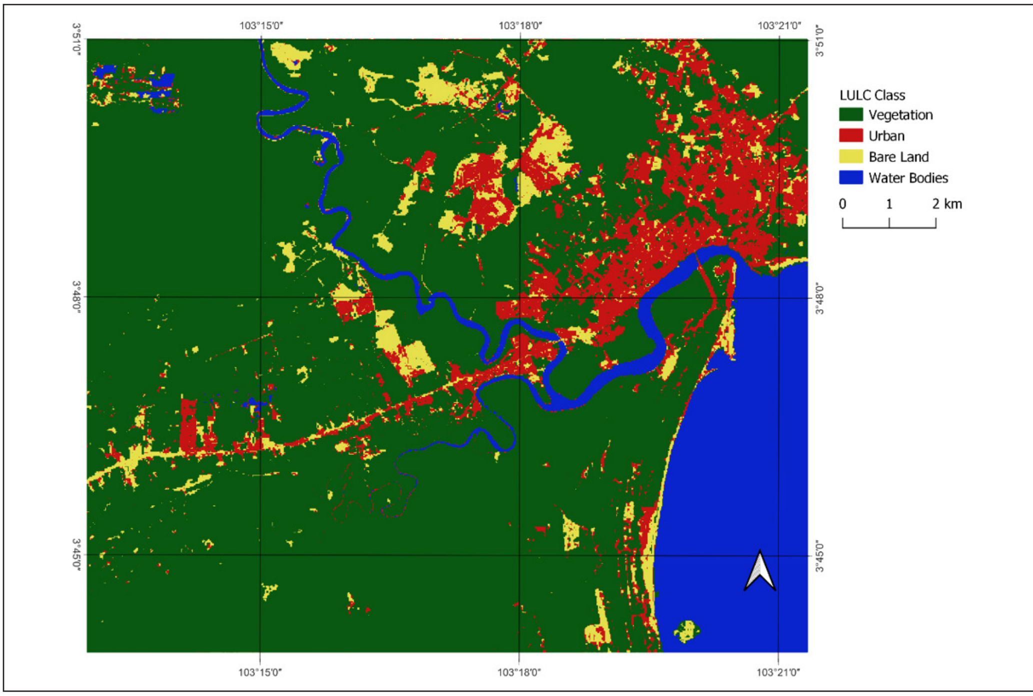

Figure 6. LULC coverage map for 1994 in Kuantan using ML classifier

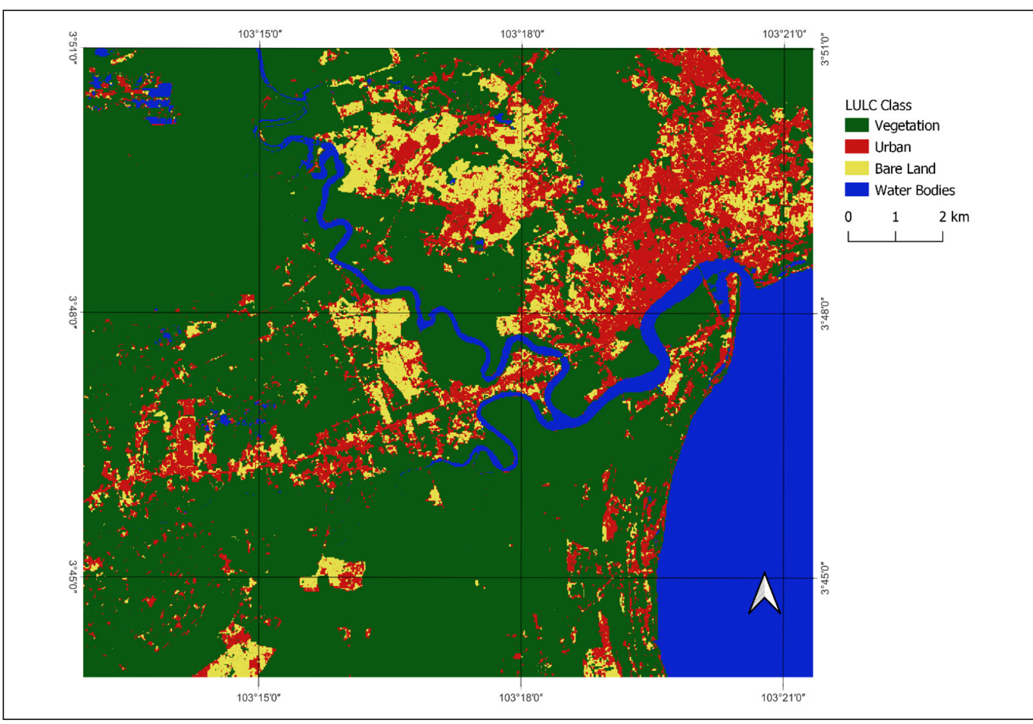

Figure 7. LULC coverage map for 2001 in Kuantan using ML classifier

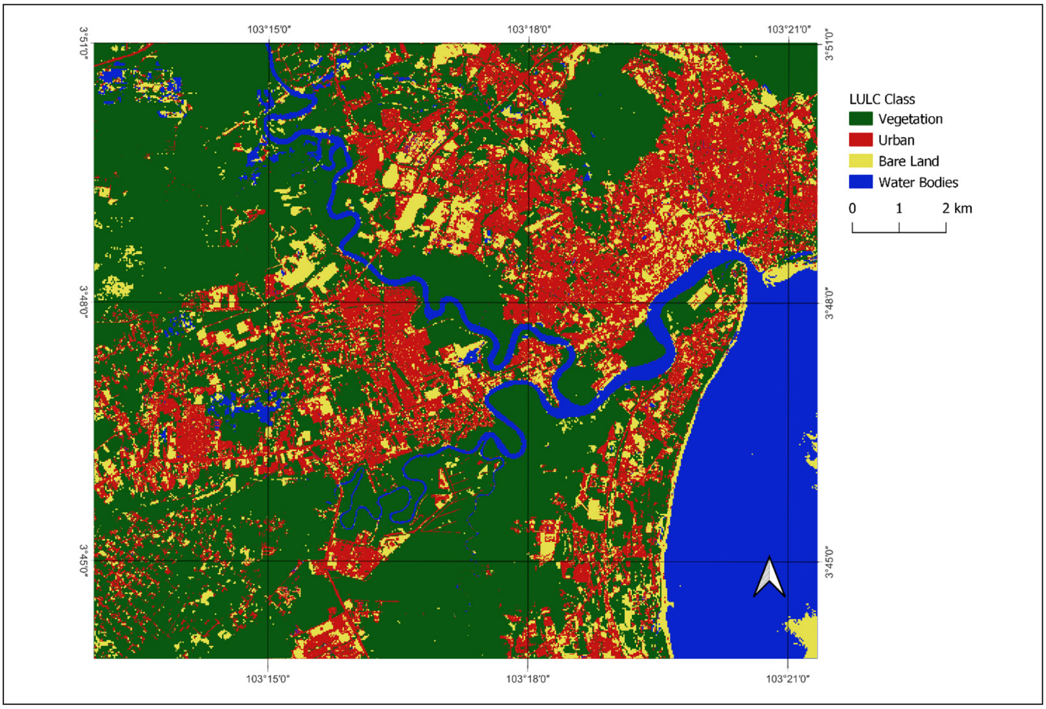

Figure 8. LULC coverage map for 2013 in Kuantan using ML classifier

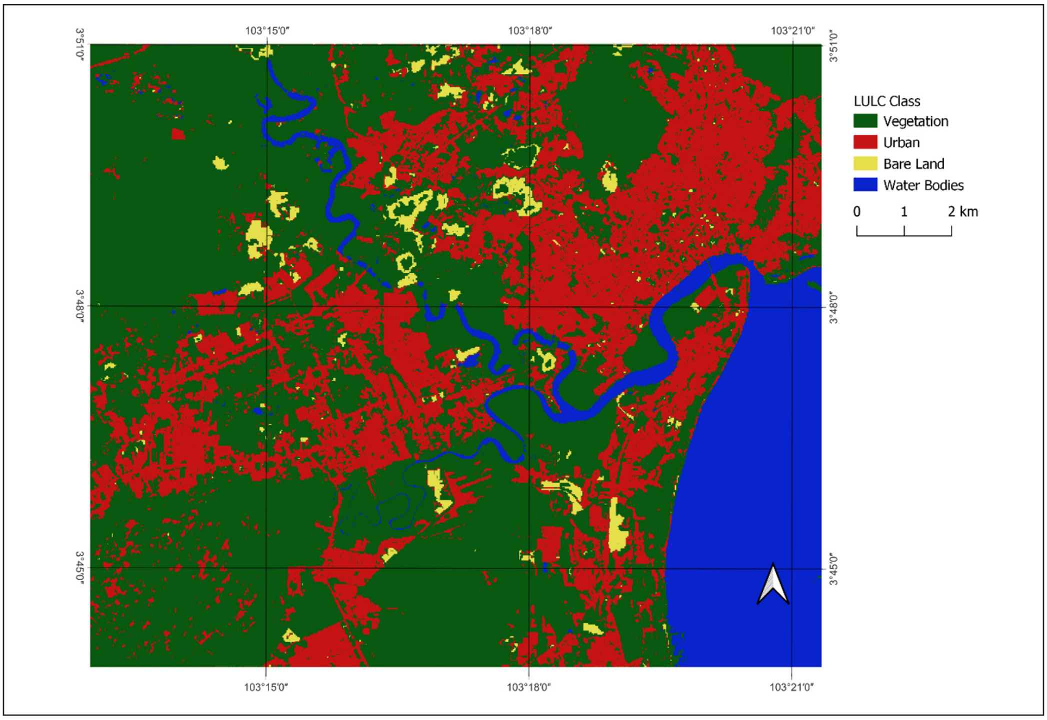

Figure 9. LULC coverage map for 2017 in Kuantan using ML classifier

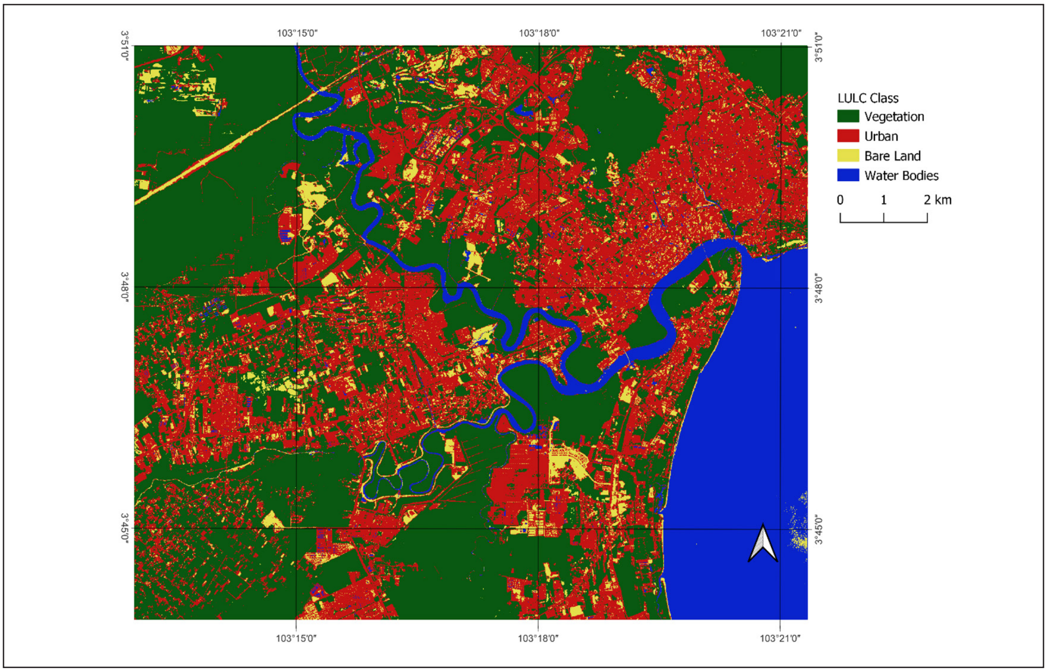

Figure 10. LULC coverage map for 2022 in Kuantan using ML classifier

the lowest coverage with less than 1000 ha. Overall, the vegetation coverage depicted 32% of the total coverage in the study area. Bare land faced a slight fluctuation over the past 28 years. There were also less significant changes in water outlet and storage in Kuantan. Furthermore, LULC in Kuantan has been experiencing major urban development due to the increasing demand for various needs. The increment of build-up coverage from 1994 to 2022 has been affecting the vegetation distribution as well as the bare land. The increase in socioeconomic growth and development has also attracted more migration into the city. Therefore, the build-up coverage expanded at a higher rate starting from 2013 compared to the early mid-90s.

According to the Kuantan Municipal Council, the city of Kuantan has been experiencing an annual population growth of 2.68%, causing changes in land use and extending the urban limit. Such a change has significantly increased the flooding frequency following the fluctuation of the runoff-discharge relationship. It is important to note that in 2013, the land area of Kuantan experienced the highest coverage during the past three decades when a massive flooding event occurred at the year's end. Zaidi et al. (2014) propounded that prolonged rainfalls and land use changes in 2013 had exposed low-lying regions, including Kuantan, to the risk of massive flooding. A constant change in build-up coverage and slight increments of bare land may bring about significant effects on river capacity, especially in Kuantan, as both the area coverage of bare land and built-up increased from 1995 to 2013. It has been supported by a study by Konrad (2014) that validated the effect of

Table 5
LULC changes for the years 1994, 2001, 2013, 2017, and 2022

LULC Classification	Yearly LULC Coverage (Ha)				
	1994	2001	2013	2017	2022
Bare Land	1267.0165	1715.6422	1815.7364	434.6605	1205.9481
Build-up	2098.2353	2859.1095	4071.0507	5193.9036	6546.9908
Vegetation	14409.1381	13044.0127	11719.9452	12088.9712	9779.5833
Water Body	2564.4092	2720.0977	2762.1378	2621.3023	2824.0054

urbanisation and massive land openings for development activities in flood-prone regions, increasing the flooding risk. Besides, changes from dense vegetation cover to agricultural and residential areas affect the hydrological ecosystem, including soil degradation imbalance and streamflow (Tewabe & Fentahun, 2020). It has been proven that, in late 2013, massive flooding hit the city of Kuantan, affecting local communities and the environment. Besides that, this flooding event significantly deteriorated buildings, properties, utility structures, the transportation system, agricultural crops, and vegetation (Rahman, 2014; Romali et al., 2019).

Climate change and global warming also contribute significantly to the hydrological system of the watershed through extreme evapotranspiration and imbalanced water components, thus increasing the rate of overflow events (Neidhardt & Shao, 2023; Johnson et al., 2022). According to Amini et al. (2022), the fluctuation of vegetation in the past few decades was caused by extreme drought. However, a study by Husain et al. (2023) suggested that massive population per area and vegetation cover can influence the surface temperature, which can cause global warming and heat islands. Furthermore, there is a substantial correlation between water abundance obtained from the LULC map and surface temperature, which affects glacier reduction and sea level rise (Samra, 2021). This indicates that LULC changes play an enormous part in contributing to climate change.

The special Economic Zone proposed by the Malaysian government in Kuantan City has a significant effect on the LULC, where high rates of anthropogenic activities contribute to the LULC changes. Extensive bare land exploration and reduction of forest cover for urbanisation and industrial expansion can modify the surface runoff and infiltration rates, percolation, and lateral flow, exposing the city to the threat of flash floods. As the data from 2022 indicates, the built-up area in Kuantan has seen substantial growth, expanding to three times its size compared to 1994. This considerable expansion underscores the rapid urbanisation and development this area has experienced over the span of nearly three decades. Extensive industrial and economic growth movement in the main city has highly contributed to the formation of LULC (Amini et al., 2022). A study by Saddique et al. (2020) suggested that LULC changes can massively influence the water balance components in a river basin, such as surface runoff, base flow, water yield, and evapotranspiration. The influence of agriculture and the main forest may also contribute to the increase in LULC change rates. The phenomenon also happened in Tanzania, where agriculture is the main source of income for the communities living there (Ngondo et al., 2021). The coastal forest has indicated a significant decline as urbanisation rates increased (Maryantika & Lin, 2017). These changes have a substantial effect on urban growth, especially in Kuantan.

LULC Accuracy Analysis

The Kappa coefficient and confusion matrix were generated based on the classification of LULC for the years 1994, 2002, 2013, 2017, and 2022 (Figure 11). The accuracy assessment seemed reliable and acceptable for all classifications across the years. Nevertheless, the image classification 1994 recorded a notably low accuracy, which may have resulted from the misclassification and high noise in the image satellite. The overall accuracy of image classification in 2013 and 2022 was slightly less significant compared to those in 2017 due to the differences in spatial resolution and image quality. Despite the high resolution of Sentinel-2, the classification could not provide distinctive features for recognising real classes because of the insufficient training sample. Since all images used the same training dataset, the high-resolution imagery would be affected the most as the image size ratio did not influence the classification algorithm. It suggests that although Sentinel-2 data provides images with higher spatial resolution, the Landsat data may be a better source to investigate LULC changes following its capability to monitor Earth for more than five decades.

Based on the results, it is evident that image classification classifiers can distinguish more relevant features and provide an accurate prediction. Different spatial quality and quantity may cause poor overall accuracy. The image with a larger pixel size can serve a huge number of features due to the increased level of detail. It may lead to better discriminative power and improved accuracy of the classification. Fisher et al. (2017)

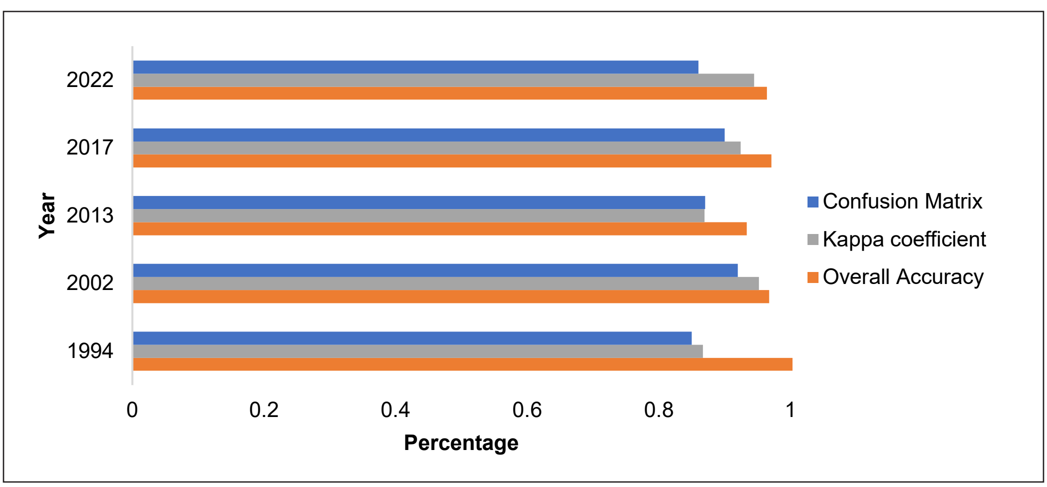

Figure 11. Graph of accuracy analysis for LULC classification in 1994, 2002, 2013, 2017 and 2022 based on ML

indicated that spatial resolution has a significant impact on the outcome of the classification, with higher resolution providing higher overall accuracy. However, this result has caused a significant increase in processing time and cost. Aja et al. (2022) highlighted that image classification using Sentinel satellite imagery produced less significant results compared to Landsat satellite imagery. It is consistent with the results of this study, where sentinel imagery, which provided higher spatial resolution imagery, required a detailed assessment compared to less spatial resolution. Moreover, handling a high-quality image requires significantly higher computational resources and processing time than a low-quality image. The amount of training data and inference speed of image classification also varied with the image pixel size. Therefore, it is recommended to consider the trade-off between computational complexity and accuracy, depending on the application requirement.

LIMITATION AND IMPLICATION

Comprehensive and cost-effective monitoring technology for LULC studies associated with forest distribution holds imperative importance. Introducing remote sensing satellite technology and GIS offers significant advantages for a highly dense city such as Kuantan, which has high land use heterogeneity. The capability of the Sentinel-2 satellites to identify surface features and generate classifications based on different algorithmic approaches stands as an acceptable method for LULC studies. While machine learning techniques have shown promising results in identifying LULC changes, some limitations include insufficient training pixels. Obtaining the training datasets in highly concentrated areas, especially in residential areas, is relatively difficult as the course satellite image with a low level of detail covers a huge area of a single training dataset. Eventually, misclassification of certain pixels tends to occur, especially during the execution of RF algorithms that are highly sensitive to input data.

Integrating a diverse data source from different satellite imageries for LULC changes makes it possible for recent data with the new remote sensing technology but not for historical data, especially satellite images from the 1990s to 2000s in Kuantan. The outdated and incomplete LULC maps are constrained to the LULC analysis, especially for the image availability before the 2000s. Recent remote sensing and GIS technology to obtain LULC data has benefited the stakeholders in terms of time, cost, and effort to study the LULC changes. Complimentary field surveys, online database monitoring or real-time supporting data can increase the study's reliability and accuracy. As the study progresses, LULC will be crucial in flood risk management as it helps identify susceptible flooding areas, especially in under-monitor regions like Kuantan. Furthermore, extensive changes in LULC coverage, especially in natural vegetation and wetlands, can alter the environmental ecosystem. Considering the LULC function, comprehensive green development practices can be implemented in developing policies and designing management strategies.

CONCLUSION AND RECOMMENDATION

LULC maps have a wide application, including natural resource management, baseline mapping for GIS input, and legal boundaries for tax and property evaluation. Generating LULC maps is impossible without the help of other geospatial datasets. This study applied remote sensing and GIS technology to analyse the significant LULC changes underlined by the continuous development in Kuantan. RF and ML classifiers generated an acceptable LULC classification based on the satellite imagery. ML posed a higher overall accuracy of 92% compared to RF, with 85% overall accuracy. RF tends to be overfitted and requires a huge training dataset to compute the image classification. However, it is most suitable for the high heterogeneity of LULC with detailed features compared to the ML classifier. It is recommended that both RF and ML classifiers be integrated with other classifiers to increase their validity and reliability. Given that both classifiers have significantly good results, an integrated, comprehensive classification scheme that compromises supervised classification and machine learning can be achieved in the near future.

As for LULC changes, high-intensity development in Kuantan affected the LULC pattern from 1994 to 2022. Primarily, build-up coverage significantly increased while vegetation coverage degraded. Sustainable landscape and town planning management is necessary because population growth and economic demand are the main factors influencing LULC changes. These elements are important for smart urban city planning and are aligned with the Sustainable Development Goal (SDG)-11. Therefore, LULC identification based on the surface spectrum of satellite images utilising remote sensing data is a proper technique to investigate land use status as part of the effort to provide ecosystem balance. Yet, there have been recommendations for LULC to use satellite imaging to improve coverage and spatial data observation by satellite imagery, particularly in Kuantan. It includes practising deep learning and artificial intelligence (AI) interfaces with advanced imputation for image analysis. In conclusion, accessible and transparent information on LULC changes can be introduced as a tool to support the establishment of an advanced technological revolution towards the construction of smart cities.

ACKNOWLEDGEMENTS

This work received funding from the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2021/WAB02/UMP/02/2 (RDU210120) and Universiti Malaysia Pahang Al-Sultan Abdullah under the International Grant (Grant Number: UIC221507). The research was also supported by the National Natural Science Foundation of China (Grant Number: 51909273) and the Program for the Introduction of High-End Foreign Experts (Grant Number: G2021058002L).

REFERENCES

Abdullah, A. Y., Masrur, A., Adnan, M. S., Baky, M. A., Hassan, Q. K., & Dewan, A. (2019). Spatio-temporal patterns of land use/land cover change in the heterogeneous coastal region of Bangladesh between 1990 and 2017. *Remote Sensing*, *11*(7), Article 790. https://doi.org/10.3390/rs11070790

Aja, D., Miyittah, M. K., & Angnuureng, D. B. (2022). Quantifying mangrove extent using a combination of optical and radar images in a wetland complex, Western Region, Ghana. *Sustainability*, *14*(24), Article 16687. https://doi.org/10.3390/su142416687

Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban land use and land cover change analysis using random forest classification of landsat time series. *Remote Sensing*, *14*(11), Article 2654. https://doi.org/10.3390/rs14112654

Balha, A., Mallick, J., Pandey, S., Gupta, S., & Singh, C. K. (2021). A comparative analysis of different pixel and object-based classification algorithms using multi-source high spatial resolution satellite data for LULC Mapping. *Earth Science Informatics*, *14*(4), 2231–2247. https://doi.org/10.1007/s12145-021-00685-4

Fisher, J. R., Acosta, E. A., Dennedy-Frank, P. J., Kroeger, T., & Boucher, T. M. (2017). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sensing in Ecology and Conservation*, *4*(2), 137–149. https://doi.org/10.1002/rse2.61

Gaur, S., & Singh, R. (2023). A comprehensive review on land use/land cover (LULC) change modeling for urban development: Current status and future prospects. *Sustainability*, *15*(2), Article 903. https://doi.org/10.3390/su15020903

Geidam, K. K., Adnan, N. A., & Alhaji Umar, B. (2020). Analysis of land use land cover changes using remote sensing data and Geographical Information Systems (GIS) at an urban set up of Damaturu, Nigeria. *Journal of Science and Technology*, *12*(2), 24-37. https://doi.org/10.30880/jst.2020.12.02.003

Gondwe, J. F., Lin, S., & Munthali, R. M. (2021). Analysis of land use and land cover changes in urban areas using remote sensing: Case of Blantyre City. *Discrete Dynamics in Nature and Society*, *2021*, 1–17. https://doi.org/10.1155/2021/8011565

Husain, M. A., Kumar, P., & Gonencgil, B. (2023). Assessment of spatio-temporal land use/cover change and its effect on land surface temperature in lahaul and Spiti, India. *Land*, *12*(7), Article 1294. https://doi.org/10.3390/land12071294

Isola, F., Lai, S., Leone, F., & Zoppi, C. (2023). Land take and landslide hazard: Spatial assessment and policy implications from a study concerning Sardinia. *Land*, *12*(2), Article 359. https://doi.org/10.3390/land12020359

Johnson, T., Butcher, J., Santell, S., Schwartz, S., Julius, S., & LeDuc, S. (2022). A review of climate change effects on practices for mitigating water quality impacts. *Journal of Water and Climate Change*, *13*(4), 1684–1705. https://doi.org/10.2166/wcc.2022.363

Konrad, C. P. (2014). *Effects of Urban Development on Floods [Fact Sheet]*. U. S. Geological Survey. https://doi.org/10.3133/fs07603

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Lokman, M. (2004). *Exploring the interface: The enigmatic mangroves*. Kolej Universiti Sains dan Teknologi Malaysia.

Maryantika, N., & Lin, C. (2017). Exploring changes of land use and mangrove distribution in the economic area of Sidoarjo District, East Java using multi-temporal landsat images. *Information Processing in Agriculture*, *4*(4), 321–332. https://doi.org/10.1016/j.inpa.2017.06.003

Nath, B., Niu, Z., & Singh, R. (2018). Land use and land cover changes, and environment and risk evaluation of Dujiangyan City (SW China) using remote sensing and GIS techniques. *Sustainability*, *10*(12), Article 4631. https://doi.org/10.3390/su10124631

Neidhardt, H., & Shao, W. (2023). Impact of climate change-induced warming on groundwater temperatures and quality. *Applied Water Science*, *13*(12), Article 235. https://doi.org/10.1007/s13201-023-02039-5

Ngondo, J., Mango, J., Liu, R., Nobert, J., Dubi, A., & Cheng, H. (2021). Land-use and land-cover (LULC) change detection and the implications for coastal water resource management in the Wami–Ruvu Basin, Tanzania. *Sustainability*, *13*(8), Article 4092. https://doi.org/10.3390/su13084092

Rahman, H. A. (2014). An overview of environmental disaster in Malaysia and preparedness strategies. *Iranian Journal of Public Health*, *43*(3), 17–24.

Romali, N. S., Yusop, Z., Zaki, N. I. M., Sulaiman, M. K., Ghani, N. A. A. A., & Sulaiman, S. (2019). Flood damage function model for residential area in Kuantan: A preliminary study. *International Journal of Integrated Engineering*, *11*(2). https://doi.org/10.30880/ijie.2019.11.02.022

Saddique, N., Mahmood, T., & Bernhofer, C. (2020). Quantifying the impacts of land use/land cover change on the water balance in the afforested river basin, Pakistan. *Environmental Earth Sciences*, *79*(19), Article 448. https://doi.org/10.1007/s12665-020-09206-w

Saini, R., & Rawat, S. (2023). Land use land cover classification in remote sensing using Machine Learning Techniques. In *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)* (pp. 99-104). IEEE Publishing. https://doi.org/10.1109/ihcsp56702.2023.10127126

Samra, R. M. A. (2021). Dynamics of human-induced lakes and their impact on land surface temperature in Toshka Depression, Western Desert, Egypt. *Environmental Science and Pollution Research*, *29*(14), 20892–20905. https://doi.org/10.1007/s11356-021-17347-z

Seyam, M. M., Haque, M. R., & Rahman, M. M. (2023). Identifying the land use land cover (LULC) changes using remote sensing and GIS approach: A case study at Bhaluka in Mymensingh, Bangladesh. *Case Studies in Chemical and Environmental Engineering*, *7*, Article 100293. https://doi.org/10.1016/j.cscee.2022.100293

Shahbudin, S., Zuhairi, A., Kamaruzzaman, Y., & Jalal, K. C. A. (2009, October 20-22). Impact of coastal development on mangrove distribution in Kuantan, Pahang. In *International Workshop on Integrated Coastal Zone Management* (pp. 1-11). Izmir, Turkey.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y. A., & Rahman, A. (2020). Land-use land-cover classification by machine Learning classifiers for satellite observations - A review. *Remote Sensing*, *12*(7), Article 1135. https://doi.org/10.3390/rs12071135

Tewabe, D., & Fentahun, T. (2020). Assessing land use and land cover change detection using remote sensing in the lake tana basin, Northwest Ethiopia. *Cogent Environmental Science*, *6*(1), Article 1778998. https://doi.org/10.1080/23311843.2020.1778998

Viana, C. M., Girão, I., & Rocha, J. (2019). Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sensing*, *11*(9), Article 1104. https://doi.org/10.3390/rs11091104

Vicenteserrano, S., Perezcabello, F., & Lasanta, T. (2008). Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of landsat images. *Remote Sensing of Environment*, *112*(10), 3916–3934. https://doi.org/10.1016/j.rse.2008.06.011

Yulianto, F., Suwarsono, Nugroho, U. C., Nugroho, N. P., Sunarmodo, W., & Khomarudin, M. R. (2020). Spatial-temporal dynamics land use/land cover change and flood hazard mapping in the upstream citarum watershed, West Java, Indonesia. *Quaestiones Geographicae, 39*(1), 125–146. https://doi.org/10.2478/quageo-2020-0010

Zaidi, S., M., Akbari, A., & Ishak, W., M., (2014) A critical review of flood history in Kuantan River basin: Challenges and potential solutions. *International Journal of Civil Engineering and Geo-Environment, 5*, 1-5.