UTILIZATION OF GOOGLE EARTH ENGINE (GEE) AND DEEP LEARNING ALGORITHMS FOR OIL PALM MAPPING OVER PENINSULAR MALAYSIA

WAI CHEE SAN

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

UTILIZATION OF GOOGLE EARTH ENGINE (GEE) AND DEEP LEARNING ALGORITHMS FOR OIL PALM MAPPING OVER PENINSULAR MALAYSIA

WAI CHEE SAN

SUBMITTED TO FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

CHAPTER 1: INTRODUCTION

1.1 Background of Study

Malaysia is a Southeast Asian country where its territory covers Peninsular Malaysia and two states on the Borneo Island, Sabah and Sarawak. Due to its location nearby the north of equator, Malaysia enjoys hot and humid weather with generous rainfall throughout the year. This offers an ideal cultivation environment for major crops such as rubber, cocoa and oil palm (Shaharum et al., 2020; Uning et al., 2020). Out of those plantation crops, oil palm has been playing an essential part in sustaining the economic growth of the country. It has been known as the major contributor of industrial plantation crops to the gross domestic product (GDP) of the country. Meanwhile in 2019, oil palm has contributed 37.7% to the value added of agriculture sector (DOSM, 2020).

Despite its important contribution to economy, unstructured oil palm plantations can lead to large-scale deforestation and loss of biodiversity. Overloading biomass waste is also another threat to the environment brought by oil palm plantations (Shaharum et al., 2020; Mohd Najib et al., 2020). Proper governance and control on the oil palm cultivation activities are required to preserve environmental sustainability. The total planted area of oil palms in Malaysia in 2020 has reached about 5.87 million hectares (MPOB, 2020). The management and planning of oil palm plantations have been challenging because a huge area of plantations is involved. Furthermore, conventional survey methods make data acquisition costly, labour-intensive and time-consuming. The more sophisticated, time-saving and economical tool, remote sensing, has evolved to address the limitations of traditional techniques. Through remote sensing, people in the oil palm industry can manage plantations efficiently and investigate the corresponding impacts to the environment (Chong et al., 2017).

Previously, most of the researchers have done their mapping studies using remote sensing data through personal computers and they were limited to small areas. Therefore, Google Earth Engine (GEE), an application working on cloud-computing concept developed by Google, was discovered and used to process large-sized data without any cost and consuming storage on the personal computers. The latest mapping study over the planted oil palm in Peninsular Malaysia using GEE was done by Shaharum et al. (2020). The authors focused to map oil palm planted areas using several machine learning algorithms, followed by comparing how they performed in fulfilling the task. Inspired by this research, remote sensing data will be further explored using GEE platform together with deep learning algorithms to identify the oil palm planted area over Peninsular Malaysia in this study.

1.2 Statement of the Problem

The latest research studied the use of machine learning models on the Landsat satellite imagery data provided by GEE in classifying the land used for planting oil palm trees over Peninsular Malaysia, yet there has been no research done to fulfill the similar task but employing deep learning algorithms on the remote sensing data collected from GEE. As deep learning algorithms are well-known for classification tasks, in this study, they will be integrated with GEE to look into their performance in mapping the oil palm plantation over Peninsular Malaysia and the results obtained will be compared with the best performed machine learning algorithm in previous study, which is RF.

1.3 Research Objectives and Questions

The aims of this study are (i) to classify oil palm planted area and produce a map covering the identified plantation areas over Peninsular Malaysia using remote sensing imagery on GEE and deep learning algorithms, (ii) to assess and compare the performance of deep learning algorithms in mapping oil palm planted areas over Peninsular Malaysia and (iii) to perform comparison of performance of deep learning algorithms with machine learning algorithm in mapping land coverage by oil palm over Peninsular Malaysia.

Besides, this study investigates the following research questions:

i. What hyperparameters of deep learning models can be optimized to perform oil palm land cover classification with satisfactory performance?
ii. Which algorithm among the chosen deep learning algorithms has the best performance in identifying the oil palm planted areas across Peninsular Malaysia?
iii. How well do deep learning algorithms perform compared to machine learning algorithm in performing the oil palm land cover classification task over Peninsular Malaysia?

1.4 Scope of Study

This study is going to integrate deep learning algorithms with remote sensing data available on GEE to classify oil palm plantations over Peninsular Malaysia. The remote sensing data between years 2020 and 2021 will be created using GEE with the aid of high-quality images available on Google Earth and land area use statistical records from the Malaysia's Department of Agriculture (DOA). The deep learning models to be involved in this study are ResNet50, InceptionResNetV2 and Xception whereas machine learning model is RF. Next, collected remote sensing data will be divided into training and testing components, followed by training both deep learning and machine learning models where their hyperparameters will be further optimized to improve their performance. Eventually, the classification results will then be compared with Malaysia Palm Oil Board (MPOB) inventory.

1.5 Significance of Study

As deep learning models are to be applied on the remote sensing data in GEE, it can be known that which algorithm will perform the best among the chosen models. Additionally, hyperparameters which can be optimized to further improve the model performance will be studied in this research project. Through this work, it can also see how well the deep learning algorithms will perform compared to conventional machine learning algorithm. The ultimate outcome of this research is to classify the land covered by oil palm trees with satisfactory performance on the land of Peninsular Malaysia working with deep learning algorithms which in future can be used as a monitoring tool in oil palm industry.

There will total 5 chapters included in this report in which Chapter 2 will be reviewing the relevant literatures, Chapter 3 will be explaining the methodology involved, Chapter 4 will be showing and discussing the results obtained whereas Chapter 5 includes conclusion and possible extensions of this study or future work.

CHAPTER 2: LITERATURE REVIEW

2.1 Study Area

The study area will be covering Peninsular Malaysia as this work aims to map the land used for oil palm plantations. Peninsular Malaysia (Figure 2.1) is also known as West Malaysia which belongs to a part of Malaysia together with Sabah and Sarawak separated by South China Sea. The location of Peninsular Malaysia is close to the Equator, between 1° and 7° in the northern latitude and longitudes of 100° and 104° E longitude. The area of Peninsular Malaysia is about 132, 090 km^2 which is composed of 11 states as shown in Figure 2.2. It experiences hot and humid rainforest tropical climate with northeast (from October to March) and southwest monsoons (from May to September) throughout the year. The annual average recorded precipitation ranges between 1500 – 5000 mm. With high temperatures and humidity, the weather of Peninsular Malaysia is generally warm where the mean temperature ranges between 23°C and 35°C (Shevade and Loboda, 2019, Hazir et al., 2020). These climatological factors have contributed to make Peninsular Malaysia as suitable cultivation site for oil palm (Mohd Najib et al., 2020).

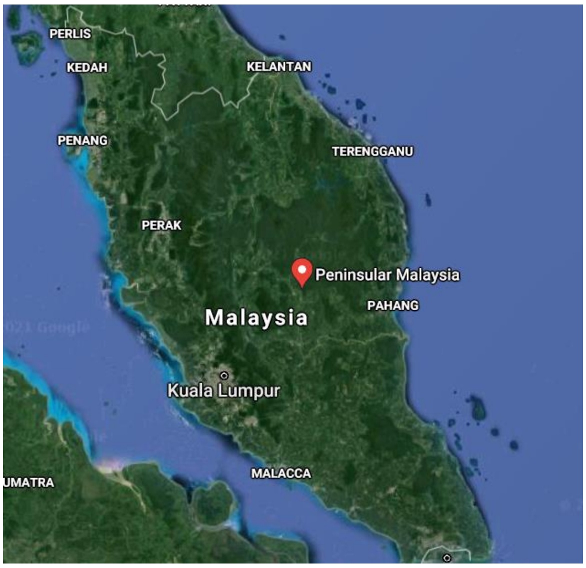


Figure 2.1: The map of Peninsular Malaysia (Photo sourced from https://www.google.com.my/maps)

Figure 2.2: Peninsular Malaysia map (Photo sourced from https://www.semanticscholar.org/)

2.2 Oil Palm

Oil palm is a remarkable crop around the world due to its highest capability to yield oil out of its fruits compared to other oil-bearing crops. As an example, one tonne of palm oil can be produced from only 0.26 hectares of oil palm land whereas 2.22, 2.00 and 1.52 hectares are required for soybean, sunflower and rapeseed to produce the equal amount of oil (Ratnam, 2020). In 2020, palm oil still outperformed other vegetable oils as previous years which accounted for 75.45 million MT or 36.1% from worldwide vegetable oil production (Statista, 2021). Oil palm does not only produce oil, but also other non-food products such as furniture, plywood and building materials (Chong et al., 2017).

2.2.1 Botanical Description

Oil palm has male and female inflorescences occurring separately on the same tree and this characteristic makes it a monoecious crop. The average heights of oil palm trees are between 15 and 30 m, besides having lifespans of about 300 years. An oil palm tree is made up of a long vascular stem ending in a crown of pinnate-shaped leaves.

The growth of leaves and flowers will start from the single bud in the crown base. Regularly, there will be two leaves produced by the oil palm tree in a month. On the other hand, the formation of flower buds happens in axils of the oil palm leaves where flowers of both sexes are produced at different times in separate inflorescences. The flower primordium has both male and female parts. The male inflorescence has a long stalk with a length of about 40 cm, with finger-shaped spikelets while each of them has about 600-1500 yellow flowers. One inflorescence is able to produce 10-30 g of pollen. Female inflorescence has spikelets in shorter lengths despite its similar structure as male does. There will be 5-30 flowers on each of those spikelets being receptive for 2-3 days. Fruits will be produced after fertilization and they take half a year to become mature (Nair, 2021).

The oil palm fruit is the important source of one of the main oil palm products, which is palm oil. A palm tree can have bunches varying in weight between 10 and 40 kg with thousands of fruitlets per bunch depending on the age, genetic traits and environment (Ithnin and Kushairi, 2020). The shape of the oil palm fruitlet is usually spherical. At the early stage, it is in deep violet to black colour and will turn to reddish orange at the ripening stage. The whole structure of an individual fruit is shown in Figure 2.3. An oil palm fruit generally has a weight between 6 and 20 g. It can be divided into several layers: outer skin (exocarp), fleshy mesocarp and endocarp or shell. The kernel or seed is encased in the shell. Mesocarp and kernel are where the crude palm oil and kernel palm oil extracted, respectively (Poku, 2002).

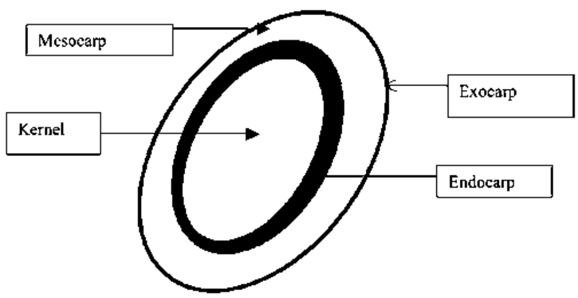


Figure 2.3: Structure of the palm fruit (Photo sourced from http://www.fao.org)

2.2.2 Historical Background of Oil Palm Industry in Malaysia

The dominant oil palm species cultivated across Malaysia is African oil palm or can be referred by its species name, *Elaeis guineensis* Jacquin. As indicated by its name, it originates from West Africa. Malaysia, back then known as Malaya, had its first oil palm tree brought by British as an ornamental plant in 1870. Oil palm trees started to be planted for commercial use in 1917 at Tennamaram Estate located in Batang Berjuntai, Selangor. This was marked as a significant starting point of large-scale oil palm plantations in Malaysia (Teoh, 2002).

Under the agricultural diversification programme initiated by government in early 1960s, oil palm plantation has become active markedly and this was when oil palm found a firm foothold in Malaysia. The aim of this programme was to expand the country's economic dependence on other commodities besides tin and rubber. The government continued the effort to launch land settlement schemes to aggressively pursue oil palm cultivation among smallholders and as such, Malaysia was ahead of Nigeria as dominant palm oil manufacturer and exporter in the world later in the 1960s and the industry has not looked back since then. The status of oil palm has transformed from a minor crop in 1917 to the backbone of the country's income and this demonstrates its vital role in uplifting the nation's livelihood over the last century (Kushairi et al., 2017).

2.2.3 Latest Performance of Oil Palm Industry in Malaysia

A summary of how Malaysian oil palm industry performed in 2020 will be explained in this section. 2020 was a challenging year to most of the industries due to COVID-19 pandemic outbreak globally and oil palm industry was also among the affected ones. There was a temporary suspension in export demand and prices in the first half of the year. However, it only resumed towards the second half when global economic sectors reopened and Malaysian government lifted movement control order (MCO). A higher export revenue of RM73.25 billion was recorded due to higher mean crude palm oil (CPO) price and lower closing stock. However, declines were observed in other key indicators such as CPO production, fresh fruit bunch (FFB) yield, national oil extraction rate (OER), palm oil exports and palm oil stocks.

CPO production in 2020 decreased by 3.6% to 19.14 million tonnes compared to that in 2019. Lower amount of FFB processed and OER performance achieved were the contributing factors to this decline. Besides, average FFB yield experienced a marginal decrease of 2.7% to 16.73 tonnes per hectare compared to the record of 17.19 tonnes per hectare in 2019. The lower quality of FFB processed by mills also lead to a decline of 1.4% in the national OER performance in 2020.

Another industry indicator, total export amount of palm oil and its related products has dropped 4.1% to 26.73 million tonnes in 2020 relative to the amount of 27.88 million tonnes in 2019's record. In 2020, India still remained as the largest palm oil export market for Malaysia, followed by China, the European Union (EU), Pakistan, the Philippines, Turkey and USA. Approximately 59.0% of total palm oil was exported from Malaysia to these seven markets (MPOB, 2020).

2.2.4 Oil Palm Planted Area in Malaysia

The total land area used for planting oil palms in Malaysia in year 2020 was 5.87 million hectares, which showed a slight decline of 0.6% in comparison with 5.90 million hectares recorded in 2019. Sarawak remained its top position among all states as it had the largest oil palm planting area of 1.58 million hectares followed by its neighbouring state, Sabah with 1.54 million hectares of land planted with oil palm trees. On the other hand, the sum of land areas cultivated with oil palms in Peninsular Malaysia recorded in 2020 was 2.74 million hectares or 46.7%. It could be observed that East Malaysia has contributed higher oil palm yields than Peninsular Malaysia.

Last year, 89.2% of the entire oil palm planted land areas in Malaysia was covered by matured oil palm trees. 90.3% of the land area used for oil palm plantation in Sarawak was having matured palm trees or represented in numbers, 1.43 million hectares. The size of land covered with matured palm trees in Sabah was approximately 1.34 million hectares which made up 87.1% of the total oil palm plantations in the state. On the other side, statistical record showed that matured oil palm trees had accounted for 89.6% of oil palm plantation areas in Peninsular Malaysia. The detailed statistics of distribution of oil palm plantation areas by state in Malaysia as of December 2020 is shown in Table 2.1.

Table 2.1: Distribution of oil palm plantation areas by state in Malaysia as of December 2020 (Source: MPOB, 2020)

State	Matured		Immatured		Total	
	Hectares	%	Hectares	%	Hectares	%
Johor	688,291	92.9	52,537	7.1	740,828	12.6
Kedah	80,210	89.3	9,572	10.7	89,782	1.5
Kelantan	131,768	78.6	35,831	21.4	167,599	2.9
Melaka	51,672	91.7	4,689	8.3	56,361	1.0
Negeri Sembilan	173,490	91.1	16,972	8.9	190,462	3.2
Pahang	702,163	89.8	80,084	10.2	782,247	13.3
Perak	352,557	90.0	39,211	10.0	391,768	6.7
Perlis	689	99.3	5	0.7	694	0.0
Pulau Pinang	12,540	97.7	289	2.3	12,829	0.2
Selangor	113,911	90.0	12,614	10.0	126,525	2.2
Terengganu	148,244	83.0	30,384	17.0	178,628	3.0

Peninsular Malaysia	2,455,535	89.7	282,188	10.3	2,737,723	46.7
Sabah	1,344,608	87.1	198,446	12.9	1,543,054	26.3
Sarawak	1,431,600	90.3	152,920	9.7	1,584,520	27.0
Malaysia	5,231,743	89.2	633,554	10.8	5,865,297	100.0

2.3 Google Earth Engine (GEE)

GEE (website: https://earthengine.google.com) was introduced and launched by Google in 2010. It is a cloud computing platform which houses multi-petabyte satellite imagery data catalogue while these images are ingested daily and then made ready for global-scale data mining and analysis It is a multi-disciplinary tool applied in various studies such as forest planning, crop yield estimation, environment assessment, disease risk mapping and urban monitoring. There are two methods for users to access GEE: one is by Application Programming Interface (API) while another is online interactive development environment (IDE). Both of these methods allow users to prototype and visualize their results systematically (Gorelick et al., 2017, Shaharum et al., 2020).

GEE acts as data warehouse storing over 40 years of geospatial datasets collected from different types of satellite sensors. The examples of imagery datasets available are Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Land Observing Satellite (ALOS), Landsat and Sentinel. The Earth Engine’s data archive also contains climate weather and geophysical related datasets for geospatial enthusiasts or researchers to perform analysis in whichever manner they want (Kumar and Mutanga, 2018).

GEE platform greatly reduces computational time by fully utilizing Google’s cloud infrastructure and allowing parallelized geospatial data processing. Users can also send requests to GEE servers by making API calls within their development environment using JavaScript or Python languages. The client libraries for calling the GEE API are hosted on GitHub. Moreover, GEE has an exclusive Git repository for users to store, share and conduct script versioning of their codes which further encourages the collaborations among researchers.

Another highlighted feature of GEE is its code editor. Users can write, develop and run their scripts using client libraries in popular programming languages which are Python and JavaScript. There are many packages available, for instance image collection and processing, machine learning, reducer and specialized algorithms to simplify the script writing process and at the same time users can obtain their desired outputs (Tamiminia et al., 2020). In fact, GEE has become a valuable tool for researchers to harness it for studies in various fields especially those related to remote sensing and environment, without suffering from the need of computers with high-end processors and large storage capacity when dealing with very large geospatial datasets. The higher the resolution of the remote sensing images, the larger the size it will have.

2.4 Deep Learning

In this study, deep learning algorithms are going to be applied on the remote sensing data obtained on GEE platform. Deep learning is a sub-branch under machine learning where it progressively learns and extracts meaningful traits from raw input data layer by layer. The term ‘deep’ in ‘deep learning’ actually does not imply the approach will achieve a deep understanding but it refers to the idea of deep learning models treating data as a layered structure and the number of layers represents their depths (Chollet, 2017).

A neural network that is composed of more than three layers, including the inputs and output, is automatically included as a deep learning algorithm. It transforms input data to outputs while learning meaningful representations automatically (Litjens et al., 2017). Hidden layers are located between input and output layers. Their job is to receive information from raw input and process them. For a neural network to be considered as ‘deep’ neural network, it needs to have multiple hidden layers (Ma et al., 2019).

Deep learning has gained popularity in the domains of computer vision and remote sensing in the last few years. Deep learning models are preferred over shallow machine learning models as machine learning models tend to capture only first two layers of data features. As mentioned, they are able to learn features progressively from data, have better discriminative ability relative to handcrafted features, deliver outputs with higher accuracies, outperforming conventional machine learning and image processing techniques (Cué La Rosa et al., 2019). Previous remote sensing research have employed different types of deep learning models like Long Short Term Memory (LSTM) networks, Convolutional Neural Network (CNN) and autoencoders (AE). CNN will be the focus of this work and its concepts will be further discussed in the next section.

2.5 Convolutional Neural Networks (CNN)

CNN is a well-known algorithm under deep learning with complex network structure which can perform convolution operations. It has become a preferred tool of computer vision researchers after Krizhevsky et al. implemented its architecture and won a competition known as ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012. The structure of CNN is composed of hierarchical structures with stacked arrangement: input layer, convolutional layer, pooling layer, fully connected layer and output layer (Figure 2.4). The convolutional layer and the pooling layer alternate few times and no full connection is required when the neurons of these two layers are connected to each other. The working principle of CNN is as follows: the input image will be convolved with a number of K kernels $W = \{W_1, W_2, \ldots, W_K\}$ and added biases $\gamma = \{b_1, b_2, \ldots, b_K\}$ in each layer, followed by producing a new feature map X_k in each round. These extracted features will then undergo an elementwise nonlinear transformation $\sigma(.)$. Each convolutional layer l: $X_k^l = \sigma(X_k^{l-1} * X^{l-1} + b_k^{l-1})$ will go through the exact same steps. Usually at the final part of the CNN architectures, fully connected layers will be added accordingly if they exist (Litjens et al., 2017, Ma et al., 2019). There are some popular

CNN architectures used for image classification tasks such as VGG networks, MobileNetV2 and DenseNet. The CNN architectures chosen to be used in this study are ResNet50, InceptionResNetV2 and Xception. This is based on the research paper prepared by Mahdianpari et al. (2018) where these models were the top three convolutional networks with accuracies between 93-97% when classifying complex remote sensing scenes on the wetlands in Canada.

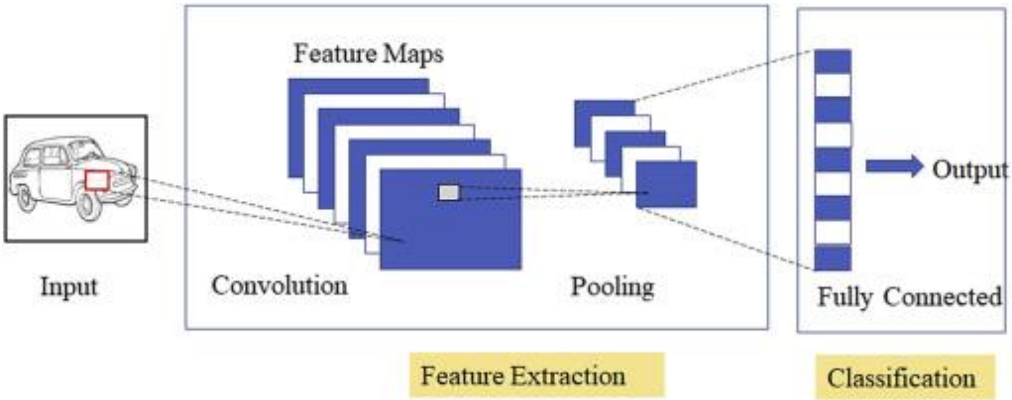


Figure 2.4: CNN architecture (Photo sourced from: Nisha & Meeral, 2021)

2.5.1 ResNet50

ResNet or its full name, Residual Network, was the champion of ILSVRC in classifying, detecting and localizing images, as well as the winner of Common Objects in Context (COCO) classification competition in 2015. It was invented after He et al. (2016) discovered there is a maximum threshold with the conventional CNN models. In this research, the authors plotted the training and test error of CNN with 20 layers and another with 56 layers. The results showed that models going into 'deep' did not necessarily perform better than the 'shallow' one. The training error of 56-layer CNN was higher than the 20-layer network. This failure was named as vanishing gradient problem. ResNet, with a deep residual learning block, was introduced to tackle this problem (Mahdianpari et al., 2018). There are different variations of ResNet, including ResNet18, ResNet34, ResNet50 and so on. The numbers on the names denote the number of layers in the model. ResNet is similar to VGG network, another well- known CNN architecture, but its depth is about eight times greater compared to that of VGG (Maeda-Gutiérrez et al., 2020). ResNet50 is chosen to be used in this study. Generally, it contains 49 convolutional layers, 2 pooling layers, a fully connected layer and softmax in its architecture (Kim et al., 2020). The general compressed view of ResNet model to be used in this work is shown in Figure 2.5.

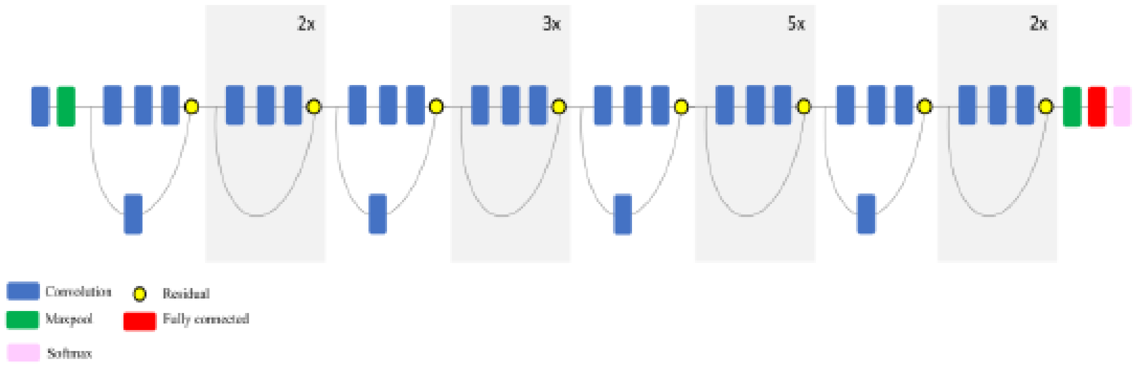


Figure 2.5: Schematic diagram of ResNet50 in compressed view (Photo sourced from: Mahdianpari et al., 2018)

2.5.2 InceptionResNetV2

InceptionResNetV2 is the second CNN architecture chosen for this study. It is the product of integrating Inception structure and ResNet (Szegedy et al., 2016). Both Inception and ResNet have been commonly used in recent years to perform image recognition tasks due to its outstanding performance without the need of high-end computer resources. There are multiple sized convolutional filters in the Inception-ResNet block combining with residual connections. The employment of ResNet helps to address the degradation issue caused by deep network architectures besides reducing the training time at a great extent (He et al., 2016). In InceptionResNet, batch-normalization concept is applied only upon the traditional layers of the architecture, rather than above the residual summations (Mahdianpari et al., 2018). The compressed architecture of InceptionResNetV2 model is shown in Figure 2.6.

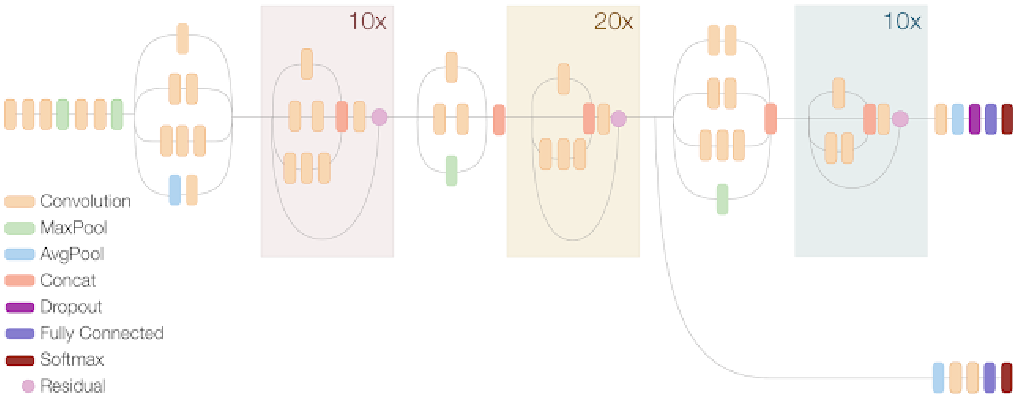


Figure 2.6: Compressed architecture of InceptionResNetV2 model (Photo sourced from: https://medium.com)

2.5.3 Xception

Xception has an architecture inspired by another model developed by Google and called Inception. It is also known as the 'extreme' version of Inception. It differs from Inception by replacing the Inception module with depthwise separable convolutional layers. In short, Xception is built by stacking the aforementioned convolutional layers together with residual connections (Chollet, 2017). There are two main convolutional layers in the configuration of Xception: spatial convolution and depthwise convolution. Spatial convolution has 3x3 convolution for each channel of input data whereas depthwise convolution carries out its task by having 1x1 convolutions to map the concatenated channels to a new channel space (Mahdianpari et al., 2018). Figure 2.7 illustrates the compressed architecture of Xception.

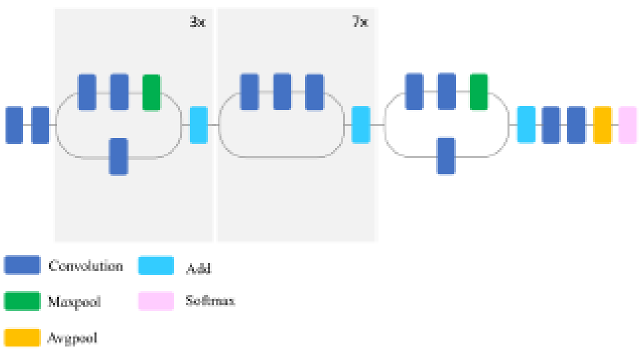


Figure 2.7: Compressed architecture of Xception model (Photo sourced from: Mahdianpari et al., 2018)

2.6 Transfer Learning

To explore the CNN architectures discussed in the previous sections, transfer learning needs to be performed as training deep networks from scratch requires fitting millions of model weights and exceptional long training time. In transfer learning, a model will be firstly trained and developed for a task and reused as a beginning point for another task in the related domain. This means what a model has learned in one setting will be applied to improve optimization in another setting. By having a pretrained model in the related task, the time spent to build a model for a new task can be reduced notably. Besides, it helps to prevent overfitting and randomization effects on the weights and better training performance. In this work, a standard dataset much larger in size will be selected to initialize the models with latent image features and followed by training them specifically for oil palm mapping over Peninsular Malaysia (Hussain et al., 2018).

2.7 Machine Learning

The general definition of term ‘machine learning’ is to give computers the ability to imitate human actions by continuously learning from surroundings. A machine will learn from what it has experienced when conducting a task and improve itself from time to time. Machine learning is a domain which needs the intersection of methods, algorithms and data analysis skills. The advanced data analytics questions which can be solved by machine learning algorithms are classification, regression or forecasting, reinforcement learning, anomaly detection and clustering. In the opposite side of deep learning, machine learning algorithms are considered as shallow learning because they used to learn only one or two layers of data’s representations (El Naqa and Murphy, 2015).

Machine learning algorithms can be further categorised as supervised, unsupervised, semi-supervised and reinforcement models. The first category, supervised learning algorithm, needs observations and their respective labels to work. The models will then learn from the labelled dataset to perform classification or prediction tasks. The examples of algorithms that belong to this category are Naïve Bayes, SVM and RF. On the other hand, unsupervised learning does not require labelled observations. The algorithms in this category will study and infer a function to reveal the patterns hidden among the unlabelled data. They can be further grouped into clustering and association models. One of the well-known algorithms of this family is K-means clustering. Semi-supervised learning models use both labelled and unlabelled data for training purpose whereas reinforcement learning methods uses observations collected by interacting with surroundings to decide whether maximizing the rewards or reducing the penalties.

As stated in the previous chapter, one of the study objectives is to investigate the differences in performance of deep learning algorithms with machine learning algorithm. According to the study done by Shaharum et al. (2020), RF had the best performance in extracting the oil palm information among the models used and hence in this work, RF will be utilized for comparison with deep learning algorithms.

2.7.1 Random Forest (RF)

RF is a tree-based algorithm (Figure 2.8) which involves building a number of decision trees, combining their outputs through a process called ensemble learning and obtain the average to improve the learning capability of the model. It is a model trained by bagging or bootstrap aggregation algorithm which was introduced by Breiman (1996) to decrease the model complexity that overfits the training data (Levantesi and Piscopo, 2020). Through this way, RF mitigates the drawback of decision tree which tends to have data overfitting issue due to high variance when it goes deeper and it has better generalization capabilities compared to decision tree (Sarker & Salah., 2019). Both regression and classification tasks can be fulfilled by using RF. It still manages to maintain its performance at satisfactory level with non-linear datasets or datasets with missing values besides having lower risk of overfitting.

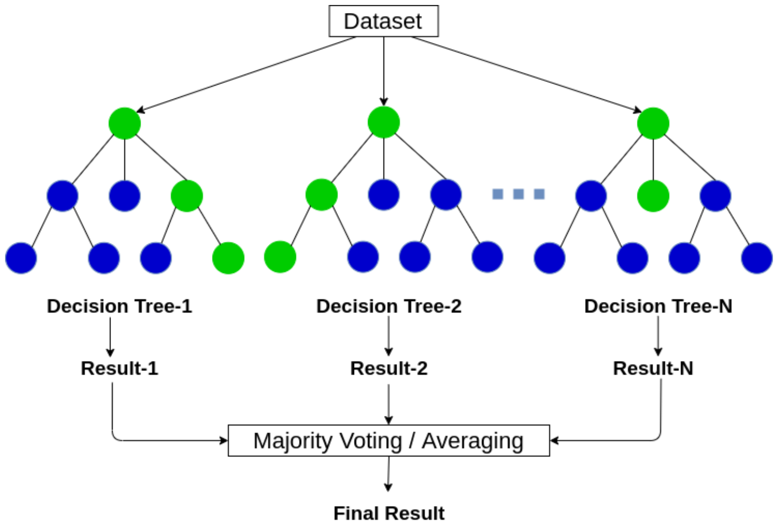


Figure 2.8: The skeleton of RF algorithm (Photo sourced from: https://www.analyticsvidhya.com/)

2.8 Related Work

Oil palm plantations have been expanding rapidly in many countries over the last few decades to continuously supply one of the dominant edible oils, palm oil. However, this expansion rate has become a concern for observers because it brings unwanted impacts to the environment such as global warming and habitat destruction due to deforestation activities. Proper monitoring of oil palm plantations is of the utmost importance to make them environmental-friendly and sustainable and because it always involves a large area coverage, people start exploring remote sensing to fulfil their tasks. Remote sensing data is able to provide critical geographical information over large spatial areas for better monitoring of land use or some other related purposes like routing the palm oil trucks. The applications of remote sensing to keep track of the extent of oil palm cultivations are getting common in the industry.

There has been a growing interest among researchers to study remote sensing for oil palm-related usage as well after it has started to be used in oil palm monitoring since 1990s. Figure 2.9 implies that the employment of satellite sensor data in oil palm-associated problems has shown an upward trend over the last decade. The possible usage of remote sensing imagery data in oil palm plantation activities have been explored continuously. The examples of applications of remote sensing are classifying crop land use, detecting the presence of trees, tree counting and assessing tree health. Among those applications, land cover application and tree detection are related to the focus of this work, which is to classify and map the land area covered by oil palm over Peninsular Malaysia. Following this direction, publications are found and reviewed. Table 2.3 presents an overview of past publications which studied the potential usage of remote sensing techniques for mapping oil palm land coverage and detecting trees.

From the reviewed literatures, it can be observed that it is getting popular among researchers to employ open-source data collected using different satellite sensors such as Sentinel, QuickBird, WorldView and Landsat images. Some of the cited works show that researchers did not only use a single sensor type data but also tried the combination of data from different sensors. Besides, researchers have adopted a variety of methods or algorithms to conduct oil palm mapping or tree detection. There are several research gaps identified when reviewing those research papers which employed deep learning models. Firstly, the study location did not cover the whole area of Peninsular Malaysia. There is also no effort has been made yet to assess and compare how do multiple deep learning models fulfil the mapping task in one study. Thus, in this work, it is going to integrate the chosen deep learning algorithms with data extracted from GEE to map the land used for oil palm plantation over Peninsular Malaysia. Additionally, comparison among those deep learning models with selected machine learning classifier, RF will also be conducted.

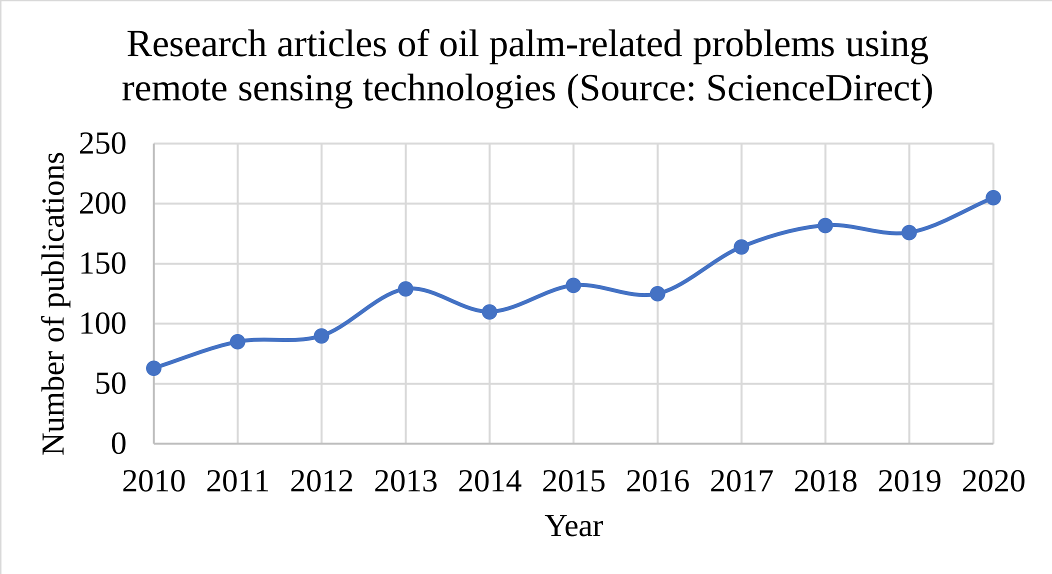


Figure 2.9: Research trend related to the adoption of remote sensing technologies for oil palm-related problems over the last 10 years

Table 2.3: An overview of past publications for oil palm mapping and tree detection using remote sensing technologies

Paper	**Research Focus**	**Dataset/Satellite**	**Method/Approach/Algorithm**	**Accuracy obtained**	**Remarks**
Shafri et al. (2011)	Perform oil palm feature extraction and counting the trees using high-quality aerial images	AISA airborne imaging spectrometer sensor images	Various image analysis techniques such as texture analysis and segmentation process	95%	• Aerial imagery data with high spatial resolution will be a useful tool to supervise oil palm plantation activities provided with suitable assessment skills
Nooni et al. (2014)	Map oil palm plantations with SVM because it always has better classification results compared to other algorithms	Landsat ETM+	SVM and maximum likelihood classifier (MLC)	SVM 78%, MLC 72%	• Both chosen classifiers did not meet the prerequisite accuracy of 85%.

Srestasathiern and Rakwatin (2014)	Detect the presence of oil palm trees using high quality multi-spectral satellite imagery	QuickBird images	Rank transformation and non-maximal suppression	90%	• The proposed method produced higher tree detection rate when compared to manual labelling.
Li et al. (2015)	Map the land planted with oil palm trees in Cameroon using machine learning models	ALOS-PALSAR 50-m orthorectified mosaic images	SVM, decision tree and K-Means	SVM 92%, DT 93% and K-Means 88%	• DT was the best performing model for large-scale mapping in the aspects of speed and performance.
Cheng et al. (2016)	Discover satellite flown PALSAR instrument in identifying and mapping oil palm plantations in cloudy regions	30-m resolution Landsat TM image and 20-m resolution ALOS PALSAR image	SVM and Mahalanobis Distance (MD)	**SVM** Landsat 90%, PALSAR 75%, Landsat + PALSAR 95% **MD** Landsat 84%, PALSAR 89%, Landsat + PALSAR 94%	• The combined satellite data was the best to be used for oil palm mapping.
Lee et al. (2016)	Employ GEE to classify industrial oil palm plantations in Tripa, Aceh, Indonesia	Landsat 8 top-of-atmosphere reflectance images	Classification and Regression Trees (CART), RF and Minimum Distance (MD)	Top 3: CART (all bands) 94%, RFT (all bands) 91% and CART (RGB) 85%	• The classification accuracies and Kappa coefficients of CART and RF were better than MD.

Li et al. (2017)	Detect the presence of oil palm trees and count them with newly proposed deep learning architecture and high-resolution remote sensing images for Malaysia	QuickBird image	LeNet convolutional neural network	96-99%	• The proposed framework could detect nearly all the oil palm trees and had higher accuracy compared to manual counting and other tree detection methods.
Li et al. (2017)	Propose a deep CNN framework for oil palm tree detection covering large land areas using high-resolution remote sensing images in Malaysia	QuickBird image	AlexNet-based deep convolutional neural network	92-97%	• Further optimization or combination of the proposed deep CNN with other algorithms are needed to improve overall accuracy.
A. Baklanov et al. (2018)	Suggest a new approach to monitor the development of oil palm plantations based on fully CNNs to solve semantic segmentation problem	Landsat 8 cloud-free composite produced in GEE	Slightly modified version of FCN-8s fully convolutional network, FCN-8s-EL	95-99%	• Authors would like to improve accuracy of the model by including all spectral bands in Landsat imagery. • The same methods could be applied to other regions and diverse types of industrial plantations.
Cheng et al. (2018)	Produce and publish oil	ALOS-2 PALSAR-2 product	Supervised classification method only	95-99%	• There was still largely unconfirmed oil

	palm maps for the plantation area is more than 10,000 hectares	with 25 m resoltuion and Fine Mode Dual Polarization images	was mentioned.		palm cover in Africa and Indonesia.
De Alban et al. (2018)	Combine Landsat and L-band Synthetic Aperture Radar (SAR) data to estimate used land area	Landsat data, L-band SAR data and combination of both satellite data	RF	Landsat data 91-92%, L-band SAR data 56-71%, combined data 93-94%	• Combined satellite data produced the highest classification accuracies compared to single sensor data.
Shaharum et al. (2018)	Utilize SVM on open source (Python) and commercial (ENVI) platforms to perform oil palm mapping	Landsat 8 images	SVM	Open source 91%, commercial 98%	• Python scikit-learn module could produce convincing results.
Li et al. (2019)	Perform large-scale detection using a two-stage convolutional neural network	QuickBird satellite image	Two-stage AlexNet model	91-96%	• The proposed method achieved much higher average F1-score compared with existing methods and less confusions with other vegetation and buildings.
Miettinen et al. (2019)	Discover the potential usage of manual and automated mapping	CIFOR visual mapping and CRISP automated mapping results	Compare those visual and automated mapping results	-	• Thorough monitoring of oil palm plantation activities can be achieved by using manual

	approaches to monitor oil palm activities in Borneo Island				and automated mapping approaches.
Abdani & Zulkifley (2019)	Propose a modified DenseNet with spatial pyramid pooling (SPP) for tree detection of different land sizes without tree age as a limiting factor	Planet Satellite imagery captured from PlanetScope	DenseNet with SPP	96-99%	• The accuracy could be improved by approximately 2% with the support of SPP. • The DenseNet with the shallowest structure (121 layers) performed the best. • Semantic segmentation module can be included in the proposed method to predict plantation size in future.
Descals et al. (2019)	Assess viability to map land used for planting oil palm trees by distributions of typology and age	Sentinel-1 and Sentinel-2 images	RF	90-94%	• Oil palm industry related parties can map large-scale matured oil palm areas without the employment of advanced models but only Sentinel-1 features.
Freudenberg et al. (2019)	Propose a U-Net type neural network to achieve better accuracy or speed	Digital Globe's WorldView-2 satellite images	U-Net	89-92%	• The model proposed could perform comparably with the existing methods. It is also proved to be reliable under

	compared to simple neural networks used in the existing studies				hard conditions and can easily be transferred to other regions.
Mubin et al. (2019)	Combine deep learning approach and geographical information system (GIS) to build a system to detect young and mature oil palm trees	WorldView-3 image	LeNet	93-95%	• The classifier showed satisfactory results on unseen datasets. It is able to make detection correctly on oil palm from background and also differentiate oil palm from other plants.
Oliphant et al. (2019)	Map crop-covered land in Southeast and Northeast Asian countries using multi-dimensional 30-m Landsat time-series data	Landsat 8 and 7, 30-m, time series data	RF	83-96% for each of the 7 refined agro-ecological zones (RAEZs)	• High correlation value was obtained between calculated cropland areas with United Nations Food and Agriculture Organization (FAO) reports.
Puttinaovarat & Horkaew (2019)	Propose an automated method to map oil palm land coverage from remote sensing imagery	THEOS satellite images	GoogLeNet CNN used to determine the presence of oil palm plantation, delineation was done using SVM	92%	• Gabor response managed to describe representations better compared to GLCM. • Combined Gabor feature and SVM method was proved to perform better

					than machine learning models.
Shaharum et al. (2019)	To perform land cover classification on cloud computing platform provided by Remote Ecosystem Monitoring Assessment Pipeline (REMAP) using a built-in RF classifier	Optical sensing data collected from Landsat Enhanced Thematic Mapper and Operational Land Mapper	RF	80%	• It proved that REMAP is able to be used for large-scale mapping and change detection in oil palm distributions although the platform only offers to analyse Landsat data with RF classifier.
Bonet et al. (2020)	Detect oil palm units in multispectral photographs taken with unmanned aerial vehicles using the system constructed based on deep learning algorithms and transfer learning approach	Near InfraRed (NIR) images	VGG-16 network and SVM	98-99%	• The proposed framework is highly suggested to be applied in the precision agriculture domain.
Ferreira et al. (2020)	Propose a novel approach to map Amazonian palm species at the level of	UAV images	ResNet-18 incorporated in the DeepLabv3+ architecture	97-99%	• The proposed method improved the average producer’s accuracy by about 4.7%.

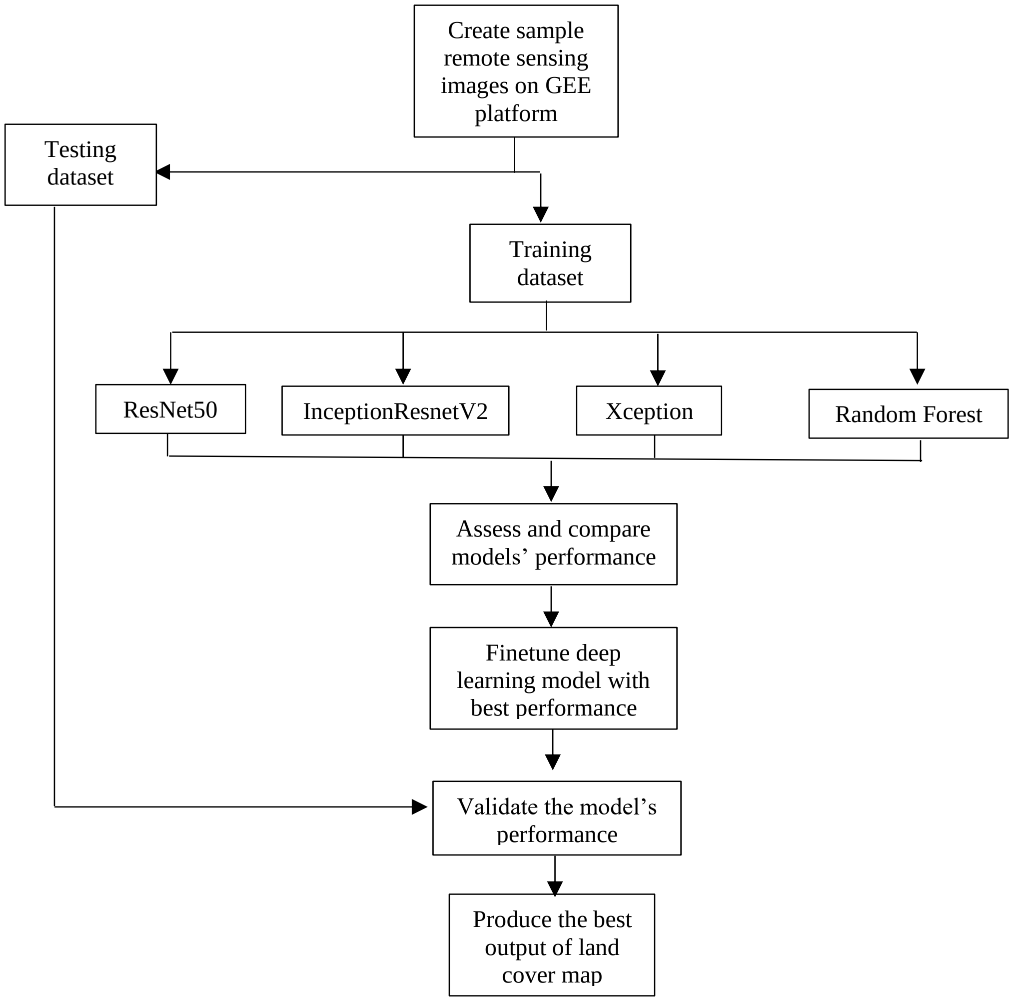
	individual tree crown (ITC) with RGB images				• The proposed method identified correctly 34.7% more ITCs compared to CSS.
Shaharum et al. (2020)	Map oil palm planted area using remote sensing data and GEE by utilizing non-parametric machine learning algorithms	30 m Landsat 8 images	CART, SVM and RF	CART 80%, RF 87% and SVM 93%	• SVM performed the best in classifying 7 classes of objects. • RF performed better than SVM when extracting oil palm information. • McNemar's test showed that performance of SVM and RF are comparable. • GEE is shown to be efficient to perform crop distribution mapping over a large area.
Sarzynski et al. (2020)	Evaluate the use of semi-automated method to classify land used for oil palm plantation in Sumatra, Indonesia	Landsat 8, SAR and combined SAR and Landsat 8 images	RF	Landsat map 78%, SAR map 74% and Combined Landsat + SAR map 84%	• Combined map produced similar amount of oil palm land coverage as shown in official government statistics. • Combining radar and optical data would perform better than using only radar or optical datasets separately in classifying oil palm planted areas.

Descals et al. (2021)	Present the latest global oil palm plantation map by typology and evaluate the ability of deep learning in classifying complex scenarios in the remote sensing domain	Sentinel-1 and Sentinel-2	Convolutional neural network, DeepLabv3+ model	99%	• The estimated area was smaller than what has reported by FAO as this research managed to identify oil-palm stands with closed canopies. • The model, actually developed in 2019, can be regularly rerun and used to monitor crop expansion in monocultural settings.

CHAPTER 3: METHODOLOGY

3.1 Proposed Framework

This study aims to perform image classification on the remote sensing images available on the Google Earth Engine (GEE) platform and map the land planted with oil palm trees over Peninsular Malaysia. The pipeline of this work includes creating sample remote sensing images covering Peninsular Malaysia on GEE platform. It will be followed by splitting them into training and testing datasets. Next, train the chosen deep learning (ResNet50, InceptionResnetV2 and Xception) and machine learning (Random Forest) models and finally, deep learning model with the highest accuracy and consistency will be selected for hyperparameter optimization to have the best output of land cover map. Figure 3.1 illustrates the proposed framework of this work in detail.



3.2 Data Collection and Sampling

Landsat 8 imagery data with 30 m spatial resolution for the study location, Peninsular Malaysia, provided by the United States Geological Survey (USGS), are available in the data archive of GEE. Landsat 8 data is known to be sensitive towards clouds as it is obtained via passive remote sensing. The presence of clouds can downgrade the data quality as the information beneath the clouds might be unclassified. To overcome this issue, GEE allows users to adjust the cloud cover composition and perform fast image patching to obtain the information that might be missed due to the clouds .

To perform data collection for this work, Landsat 8 data of Peninsular Malaysia from year 2020 to 2021 will first be patched together. As adopted by Shaharum et al. (2020), Bands 1 – 7 of Landsat 8 images will be used to generate additional data. The information of each band is shown in Table 3.1. The additional data (Table 3.2) is produced by several equations and they will be stacked together as layers with the aforementioned 7 bands to perform image classification task. With the assistance of these additional layers, certain information can be extracted in a more efficient way. For example, one of the layers named Normalized Difference Vegetation Index (NDVI) extracts the information of live green vegetations using reflected wavelengths of visible and near-infrared lights. Chlorophyll contents in the green vegetations actively absorb red light and reflects near infrared (NIR). Meanwhile, NDVI is sensitive to detect their presence as it is calculated using the ratio of red and NIR reflectances (Shaharum et al., 2020).

After performing image patching to obtain the remote sensing images, it will be followed by sample creation. Samples will be obtained by creating points on all states in Peninsular Malaysia through random sampling. High-resolution Google Earth images can be a supplementary tool to create samples. The samples obtained will then be divided into train and test datasets to be used for training models and validating performances of the models respectively.

Table 3.1 Details of the Landsat 8 bands (Source: https://www.usgs.gov)

Name	Description	Resolution (m)	Wavelength (μm)
Band 1	Coastal aerosol	30	0.43 – 0.45
Band 2	Blue	30	0.45 – 0.51
Band 3	Green	30	0.53 – 0.59
Band 4	Red	30	0.64 – 0.67
Band 5	Near Infrared (NIR)	30	0.85 – 0.88
Band 6	Short wave Infrared 1	30	1.57 – 1.65
Band 7	Short wave Infrared 2	30	2.11 – 2.29

Table 3.2 Details of the additional layers used for classification (Source: Shaharum et al., 2020)

Name	Equation
NDVI	$\frac{NIR - Red}{NIR + Red}$
Normalized Difference Water Index (NDWI)	$\frac{Green - NIR}{Green + NIR}$
Blue Red	Blue-Red
Blue Green	Blue-Green

3.3 Model Development

Models to be explored in this project are ResNet50, InceptionResNetV2 and Xception. They will be first developed using transfer learning. At the same time, to achieve the objective of comparing the performance of deep learning models against machine learning model, random forest will also be trained using the same train dataset. These models will be assessed using common evaluation metrics such as models' accuracy, precision, recall and F1-scores. Next, after comparison, the best performing deep learning model will be undergoing hyperparameter tuning using transfer learning to produce the best version of land cover map.

3.4 Hardware

The hardware specifications of the computer used for model training in this study are as below:

Processor: Intel(R) Core(TM) i7-10510U CPU 1.80 GHz, 4-Core

Memory: 8 GB 2667 MHz DDR4

Graphics: Intel UHD Graphics 1 GB

3.5 Evaluation Metrics

The classification performances of the proposed algorithms will be evaluated using evaluation metrics as mentioned in the previous section. The formulae of these metrics are derived from the calculations of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Table 2.2 summarized how they are defined from the aspects of actual values and predicted values.

Table 2.2: Summary of True Positive, True Negative, False Positive and False Negative

	Actual value	**Predicted value by the model**
True Positive (TP)	True	True
True Negative (TN)	False	False
False Positive (FP)	False	True
False Negative (FN)	True	Negative

Accuracy is one of the most used metrics to evaluate a model's performance. It divides the sum of all correct predictions by summing up the total amount of predictions made by a model. It is represented by the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

On the other hand, precision is calculated by calculating the proportion of true positives by the sum of positive observations that the model has predicted. It measures the correctness or the predictive power of the model. Its formula is given as follows:

$$Precision = \frac{TP}{TP + FP}$$

Another evaluation metric to assess the model performance is recall, which can also be referred by the term 'sensitivity'. It quantifies the proportion of number of true positives over the sum of observations that were actually positive. It is calculated by using the formula below:

$$Recall = \frac{TP}{TP + FN}$$

After getting the values of precision and recall, another measure, F1-score can also be obtained. It is actually the weighted average calculated using the values of precision and recall. Since both values of false positives and false negatives are taken into consideration, F1-score is usually a better metric compared to accuracy. Its formula is shown as below:

$$F1 - score = 2\, x \frac{Precision\ x\ Recall}{Precision + Recall}$$

Other than accuracy, precision, recall and F1-score, there is another tool known as confusion matrix and it is used to summarize the performance of the model constructed. It gives a holistic view of how well the models are performing by comparing the actual target values with those predicted by the models so that the error made by the classifiers can be known (Maeda-Gutierrez et al., 2020).

3.6 Hyperparameter Tuning

In the workflow of developing either machine or deep learning model, hyperparameter tuning or optimization is usually the next step after model evaluation process has done. Hyperparameters are the configurations of a model with values that are set before training of the model starts. Different deep or machine learning models will have their respective built-in hyperparameters. Having the optimized model hyperparameters is vital as it directly controls the training behaviour of algorithms and impacts on the performance of the models under learning. Thus, hyperparameter tuning is required to make sure the models are working at the best performance level. As such, in this study, hyperparameters of the chosen models will be explored accordingly so that adjustments can be performed to ensure the models are learning at their best.

3.7 Script

The source codes written for this research work will be available in OilPalmLandCoverMap.

REFERENCES

Abdani, S. R., & Zulkifley, M. A. (2019, December). Densenet with spatial pyramid pooling for industrial oil palm plantation detection. In *2019 International Conference on Mechatronics, Robotics and Systems Engineering (MoRSE)* (pp. 134-138). IEEE.

Baklanov, A., Khachay, M., & Pasynkov, M. (2018, June). Fully convolutional neural networks for mapping oil palm plantations in Kalimantan. In *International Conference on Learning and Intelligent Optimization* (pp. 427-432). Springer, Cham.

Bonet, I., Caraffini, F., Pena, A., Puerta, A., & Gongora, M. (2020, July). Oil Palm Detection via Deep Transfer Learning. In *2020 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1-8). IEEE.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Cheng, Y., Yu, L., Cracknell, A. P., & Gong, P. (2016). Oil palm mapping using Landsat and PALSAR: A case study in Malaysia. *International journal of remote sensing*, *37*(22), 5431-5442.

Cheng, Y., Yu, L., Xu, Y., Liu, X., Lu, H., Cracknell, A. P., ... & Gong, P. (2018). Towards global oil palm plantation mapping using remote-sensing data. *International Journal of Remote Sensing*, *39*(18), 5891-5906.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 1251-1258).

Chong, K. L., Kanniah, K. D., Pohl, C., & Tan, K. P. (2017). A review of remote sensing applications for oil palm studies. *Geo-spatial Information Science*, *20*(2), 184-200.

Cué La Rosa, L. E., Queiroz Feitosa, R., Nigri Happ, P., Del'Arco Sanches, I., & Ostwald Pedro da Costa, G. A. (2019). Combining deep learning and prior knowledge for crop mapping in tropical regions from multitemporal SAR image sequences. *Remote Sensing*, *11*(17), 2029.

De Alban, J. D. T., Connette, G. M., Oswald, P., & Webb, E. L. (2018). Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing*, *10*(2), 306.

Department of Statistics Malaysia (DOSM). (2020). *Selected Agricultural Indicators, Malaysia, 2020*. https://www.dosm.gov.my

Descals, A., Szantoi, Z., Meijaard, E., Sutikno, H., Rindanata, G., & Wich, S. (2019). Oil palm (Elaeis guineensis) mapping with details: smallholder versus industrial plantations and their extent in Riau, Sumatra. *Remote Sensing*, *11*(21), 2590.

Descals, A., Wich, S., Meijaard, E., Gaveau, D. L., Peedell, S., & Szantoi, Z. (2021). High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth System Science Data*, *13*(3), 1211-1231.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In *machine learning in radiation oncology* (pp. 3-11). Springer, Cham.

Ferreira, M. P., de Almeida, D. R. A., de Almeida Papa, D., Minervino, J. B. S., Veras, H. F. P., Formighieri, A., ... & Ferreira, E. J. L. (2020). Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. *Forest Ecology and Management*, *475*, 118397.

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., & Kleinn, C. (2019). Large scale palm tree detection in high resolution satellite images using U-Net. *Remote Sensing*, *11*(3), 312.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, *202*, 18-27.

Hazir, M. H. M., Kadir, R. A., Gloor, E., & Galbraith, D. (2020). Effect of agroclimatic variability on land suitability for cultivating rubber (Hevea brasiliensis) and growth performance assessment in the tropical rainforest climate of Peninsular Malaysia. *Climate Risk Management*, *27*, 100203.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, (pp. 1026-1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *European conference on computer vision* (pp. 630-645). Springer, Cham.

Hussain, M., Bird, J. J., & Faria, D. R. (2018, September). A study on cnn transfer learning for image classification. In *UK Workshop on computational Intelligence* (pp. 191-202). Springer, Cham.

Kim, J., Beanbonyka, R., Sung, N.-J., & Hong, M. (2020). Left or right hand classification from fingerprint images using a deep neural network. Computers, Materials & Continua, 62(3), 17–30.

Kumar, L., & Mutanga, O. (2018). Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing*, *10*(10), 1509.

Kushairi, A., Singh, R., & Ong-Abdullah, M. (2017). The oil palm industry in Malaysia: Thriving with transformative technologies. *Journal of Oil Palm Research*, *29*(4), 431-439.

Lee, J. S. H., Wich, S., Widayati, A., & Koh, L. P. (2016). Detecting industrial oil palm plantations on Landsat images with Google Earth Engine. *Remote Sensing Applications: Society and Environment*, *4*, 219-224.

Levantesi, S., & Piscopo, G. (2020). The importance of economic variables on london real estate market: a random forest approach. *Risks*, *8*(4), 112.

Li, L., Dong, J., Njeudeng Tenku, S., & Xiao, X. (2015). Mapping oil palm plantations in Cameroon using PALSAR 50-m orthorectified mosaic images. *Remote Sensing*, *7*(2), 1206-1224.

Li, W., Dong, R., Fu, H., & Yu, L. (2019). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, *11*(1), 11.

Li, W., Fu, H., & Yu, L. (2017, July). Deep convolutional neural network based large-scale oil palm tree detection for high-resolution remote sensing images. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (pp. 846-849). IEEE.

Li, W., Fu, H., Yu, L., & Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, *9*(1), 22.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, *152*, 166-177.

Maeda-Gutiérrez, V., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Guerrero Méndez, C. A., & Olvera-Olvera, C. A. (2020). Comparison of convolutional neural network architectures for classification of tomato plant diseases. *Applied Sciences*, *10*(4), 1245.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., & Zhang, Y. (2018). Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, *10*(7), 1119.

Malaysian Palm Oil Board (MPOB). (2020). *Oil palm planted area 2020*. https://bepi.mpob.gov.my/images/area/2020/Area_summary.pdf

Malaysian Palm Oil Board (MPOB). (2020*). Overview of the Malaysian oil palm industry 2020*. https://bepi.mpob.gov.my/images/overview/Overview_of_Industry_2020.pdf

Miettinen, J., Gaveau, D. L., & Liew, S. C. (2019). Comparison of visual and automated oil palm mapping in Borneo. *International Journal of Remote Sensing*, *40*(21), 8174-8185.

Mohd Najib, N. E., Kanniah, K. D., Cracknell, A. P., & Yu, L. (2020). Synergy of active and passive remote sensing data for effective mapping of oil palm plantation in Malaysia. *Forests, 11*(8), 858.

Mubin, N. A., Nadarajoo, E., Shafri, H. Z. M., & Hamedianfar, A. (2019). Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. *International Journal of Remote Sensing*, *40*(19), 7500-7515.

Nair, K. P. (2021). *Tree Crops: Harvesting Cash from the World's Important Cash Crops* (pp. 249 – 279). Springer Nature.

Nisha, S. S., & Meeral, M. N. (2021). Applications of deep learning in biomedical engineering. In *Handbook of Deep Learning in Biomedical Engineering* (pp. 245-270). Academic Press.

Nooni, I. K., Duker, A. A., Van Duren, I., Addae-Wireko, L., & Osei Jnr, E. M. (2014). Support vector machine to map oil palm in a heterogeneous environment. *International journal of remote sensing*, *35*(13), 4778-4794.

Oliphant, A. J., Thenkabail, P. S., Teluguntla, P., Xiong, J., Gumma, M. K., Congalton, R. G., & Yadav, K. (2019). Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *International Journal of Applied Earth Observation and Geoinformation*, *81*, 110-124.

Poku, K. (2002). *Small-scale palm oil processing in Africa* (Vol. 148). Food & Agriculture Org..

Puttinaovarat, S., & Horkaew, P. (2019). Deep and machine learnings of remotely sensed imagery and its multi-band visual features for detecting oil palm plantation. *Earth Science Informatics*, *12*(4), 429-446.

Ratnam, M. (2020). Ensuring the Relevance of Independent Smallholder Farmers (ISHFs) Through Sustainable Sourcing Practices: A Model to Track and Trace Within the Malaysian Palm Oil Industry. In U. Akkucuk (Eds.), *Handbook of Research on Sustainable Supply Chain Management for the Global Economy* (pp. 219-237). IGI Global.

Sarker, I. H., & Salah, K. (2019). Appspred: predicting context-aware smartphone apps using random forest learning. *Internet of Things*, *8*, 100106.

Sarzynski, T., Giam, X., Carrasco, L., & Lee, J. S. H. (2020). Combining radar and optical imagery to map oil palm plantations in Sumatra, Indonesia, using the google earth engine. *Remote Sensing*, *12*(7), 1220.

Shafri, H. Z., Hamdan, N., & Saripan, M. I. (2011). Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery. *International journal of remote sensing*, *32*(8), 2095-2115.

Shaharum, N. S. N., Shafri, H. Z. M., Ghani, W. A. W. A. K., Samsatli, S., Al-Habshi, M. M. A., & Yusuf, B. (2020). Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. *Remote Sensing Applications: Society and Environment*, *17*, 100287.

Shaharum, N. S. N., Shafri, H. Z. M., Ghani, W. A. W. A. K., Samsatli, S., Yusuf, B., Al-Habshi, M. M. A., & Prince, H. M. (2018). Image classification for mapping oil palm distribution via support vector machine using Scikit-learn module. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*(4/W9).

Shaharum, N. S. N., Shafri, H. Z. M., Ghani, W. A. W. A. K., Samsatli, S., Prince, H. M., Yusuf, B., & Hamud, A. M. (2019). Mapping the spatial distribution and changes of oil palm land cover using an open source cloud-based mapping platform. *International Journal of Remote Sensing*, *40*(19), 7459-7476.

Shahbandeh, M. (2021). *Global vegetable oil consumption, 2019/20*. Statista. Retrieved May 08, 2021, from https://www.statista.com/statistics/263937/vegetable-oils-global-consumption/.

Shevade, V. S., & Loboda, T. V. (2019). Oil palm plantations in Peninsular Malaysia: Determinants and constraints on expansion. *PloS one*, *14*(2), e0210628.

Srestasathiern, P., & Rakwatin, P. (2014). Oil palm tree detection with high resolution multi-spectral satellite imagery. *Remote Sensing*, *6*(10), 9749-9774.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*, 152-170.

Teoh, C. H. (2002). *The palm oil industry in Malaysia: from seed to frying pan*. https://wwfint.awsassets.panda.org/downloads/oilpalmchainpartaandb_esri.pdf

Uning, R., Latif, M. T., Othman, M., Juneng, L., Mohd Hanif, N., Nadzir, M. S. M., ... & Takriff, M. S. (2020). A review of southeast asian oil palm and its co2 fluxes. *Sustainability*, *12*(12), 5077.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295-316.