JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Ahmad Radzali, N. S., Abu Bakar, A., & Zamahsasri, A. I. (2023). Machine learning models for behavioral diversity of asian elephants prediction using satellite collar data. *Journal of Information and Communication Technology, 22*(3), 363-398. https://doi.org/10.32890/jict2023.22.3.3

Machine Learning Models for Behavioural Diversity of Asian Elephants Prediction Using Satellite Collar Data

*[1]Nurul Su'aidah Ahmad Radzali, [2]Azuraliza Abu Bakar & [3]Amri Izaffi Zamahsasri
[1]Department of Information System and Communication, Politeknik Sultan Idris Shah, Selangor, Malaysia
[2]Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[3]Supritendent's Office, Kelantan National Park, Kelantan, Malaysia

*[1]nurulsuaidah@psis.edu.my
[2]azuraliza@ukm.edu.my
[3]amriizaffi@gmail.com
*Corresponding author

Received: 22/9/2022 Revised: 20/12/2022 Accepted: 6/4/2023 Published: 24/7/2023

ABSTRACT

Analysis of animal movement data using statistical applications and machine learning has developed rapidly in line with the development and use of various tracking devices. Location and movement data at temporal and spatial scales are collected using the Global Positioning System (GPS) to estimate the location of animals. In contrast, installing a satellite collar can ensure continuous monitoring, as the received data will be sent directly to the electronic mailbox. Nevertheless, identifying an exact pattern of elephant activity from satellite collar data is still challenging. This study aimed to propose a machine learning model to predict the behavioural diversity of Asian elephants. The study involved four main phases, including two levels of model development, to produce initial and primary classification models. The phases were data collection and preparation, data labelling and initial classification model development, all data classification, and primary classification model development. The elephant behaviour data were collected from the satellite collars attached to five elephants, three males and two females, in forest reserves from 2018 to 2020 by the Department of Wildlife and National Parks, Malaysia. The study's outcome was a novel classification model that can predict the behaviour of the Asian elephant movement. The findings showed that the XGBoost method could produce the predictive model to classify Asian elephants' behaviour with 100 percent accuracy. This study revealed the capability of machine learning to identify behaviour classes and decision-making in setting initiatives to preserve this species in the future.

Keywords: Machine learning, XGBoost algorithm, Satellite collar data, Behaviour classification.

INTRODUCTION

Wild animals live in the wild and do not interact with humans, and consist of thousands of mammals, birds, reptiles, fish, amphibians, and insects (Chandrakar, 2018). Wildlife refers to an animal that is not tame and has its habitat, such as forests, mountains, deserts, and even the ocean. It has a vital role in maintaining the balance of the environment because it can provide stability to different natural processes. The revolution of animal movement analysis is in line with improved positional accuracies and temporal frequency detection devices, such as ARGOS, Radio Frequency Identification (RFID), IDentification (Tag), Geotag, and Global Navigation2 Satellite System (GNSS) (Cooke et al., 2004). Cargnelutti et al. (2007) and Mattisson et al. (2010) used mobile collars to test the performance of the Global Positioning System (GPS) in monitoring wildlife, while Tomkiewicz et al. (2010) used GPS to research animal behaviour and ecology. There is much importance to be gained from forecasting the behaviour and movement of wildlife. According to Wijeyakulasuriya et al. (2020), realistic forecasts can help form conservation strategies to combat the decline in biodiversity and understand the spread of infectious diseases through animal populations.

In Peninsular Malaysia, wildlife management by the government is carried out entirely by the Department of Wildlife and National Parks (DWNP) under the Ministry of Natural Resources, Environment, and Climate Change (NRECC). The Wildlife Conservation Division is responsible for conducting in-situ monitoring and management programmes of wildlife species and their habitats. In addition, this division collects all information and photographs available on wildlife and other related species to provide a complete database and Geographic Information System (GIS) (PERHILITAN, 2021). The Asian elephant, scientifically known as *Elephas maximus*, is the largest mammal on land. They are protected under the Wildlife Conservation Act 2010, Act No. 716, where it is an offence to shoot, kill, possess, or possess part of an elephant's limbs (PERHILITAN, 2021). DWNP actively carries out monitoring of Asian elephants around Peninsular Malaysia. Apart from live tracking by visiting the field areas, more monitoring is done continuously by installing satellite collars on some selected Asian elephants.

The satellite collar used on the Asian elephant species is manufactured by Africa Wildlife Tracking (AWT). This model is a 10D-cell GPS collar with the ability to record the location of the collar or animal over a long period. The model used by DWNP is the Iridium Satellite (IR-SAT), which uploads data via the Iridium satellite system. Each collar model is equipped with a different battery size that can change the amount of data recorded in a single spread. The larger the battery, the heavier the collar is (Manual, 2013). Data recorded by the satellite will be sent directly to the electronic mailbox every few hours. Spatial data, or geospatial data, is a group of information about physical objects represented by numerical values in a geographic coordinate system. In general, spatial data represents an object's location, size, and shape on the earth's surface, such as animals, buildings, lakes, rivers, hills, mountains, or plains (Fontecchio, 2013). Using data collected via satellite collars, researchers can study animal movement patterns and estimate animal habitat size over time. They can also recognise animal behaviour, find the most preferred habitat type, and even identify if other animals inhabit the same perimeter.

Installing a satellite collar can ensure continuous monitoring, as the received data will be sent directly to the electronic mailbox. Nevertheless, the pattern of elephant activity from satellite collar data is challenging to identify. Determining the causes that drive the ecological process is difficult because the acquisition of information on elephant ecology is minimal. The low accuracy of predicting Asian elephant behaviour provides new ideas for predicting using machine learning. The application of machine learning on satellite collar data in identifying patterns of activity and movement of animals is increasingly employed. Wijeyakulasuriya et al. (2020) used Random Forest, Neural Networks, and Recurrent Neural Networks to predict animal movements. Valletta et al. (2017) noted that machine learning techniques could extract knowledge from complex behaviour data using the Random Forest algorithm. Rew (2019) proposed a Long-Billed Curlew Bird movement prediction using a Recurrent Neural Network algorithm. Despite the widespread use of machine learning methods, acquiring labelled elephant behaviour data for classification models is challenging. In Malaysia, in line with the increasing number of Asian elephant movement data and other animals, machine learning methods are also very suitable for determining labels from satellite collar data for predictive classification purposes.

This paper presents a machine learning approach to predict the behavioural diversity of Asian elephants. The Asian elephant data involved five male and female Asian elephants in different forest areas around Peninsular Malaysia. These five elephants are one over four of the number of elephants fitted with satellite collars. The estimated elephant population in Peninsular Malaysia is around 1,000 to 1,100 individuals. Before the collar is installed, it will be turned on for a while to ensure it functions correctly. The recorded data started around 2018 until 2020 with 15,389 data. The authors conducted two levels of predictive modelling: the initial and primary classification model development. The initial level included developing a descriptive semi-supervised model to determine the behaviour data labels from a sample of Asian elephant data.

This study devised an evaluation scheme to validate the labelled data to develop the initial classification model. The initial model then classified the remaining collected Asian elephant data. A novel predictive model was created from 11,629 classified data at the primary model development level. The contributions of the study are a novel methodology for preparing the elephant behaviour data label, two-level predictive model development for labelling and classifying Asian elephant behaviour from collar satellite data, and a novel machine learning predictive model for classifying Asian elephant behaviour.

RELATED WORKS

Data mining is one of the various fields that make up data science. Other areas include machine learning, statistics, mathematics, visualisation, data engineering, and domain knowledge and expertise (Hamdan et al., 2018). Data mining is a logical process of obtaining useful information or patterns from large amounts of data to make specific decisions involving three main steps: exploration, pattern identification, and application (Dey et al., 2017).

Worldwide, the prediction models of animal behaviour change in response to global change are through new investigations and understandings of spatiotemporal patterns of animal movement that change according to the environment (Somayeh et al., 2012). One of the most important and pressing global conservation challenges addresses the loss and fragmentation of wildlife habitats, especially for large-bodied animals and the vast terrestrial mega-fauna ecosystems (Torre et al., 2019). The Federal Government of Malaysia developed the Forest Backbone Master Plan for Ecological Linkages (CFS) in 2010 to protect biological diversity and ecosystem services by securing landscape connectivity between major forest blocks of Peninsular Malaysia (Torre et al., 2019). In their study, Torre et al. (2019) presented an evaluation of the effectiveness of the CFS master plan to promote functional relationships of the Asian elephant species, in particular, to identify the most critical forest areas for maintaining elephant species relationships in Peninsular Malaysia.

Spatial data mining and geographical knowledge discovery have emerged as active research areas with a focus on developing theories, methodologies, and practices for extracting useful information and knowledge from large and complex databases (Sharma et al., 2017; Li & Wang, 2016). According to Sarkar et al. (2015), a large-scale spatial data mining approach can be used to extract repetitive spatiotemporal patterns and identify animal habitat choices. Although machine learning is still not widely used for modelling animal behaviour data, Valletta et al. (2017) stated that machine learning has an appropriate methodology and can analyse complex behaviour data. Devices mounted on animals and sophisticated statistical and machine learning techniques in analysing animal movement data support the development of movement ecology studies. GPS transmitters are used to estimate the animal location and tri-axial acceleration to obtain location and movement data at spatial and temporal scales for machine learning modelling and advanced statistical methods to identify the ecological and physiological mechanisms behind animal behaviour and movement (Wang, 2019). Nowadays, accelerometers embedded in smartphones are widely used to track human activity, as mentioned by Mohamed et al. (2018). The acceleration data are classified into multiple classes to determine the type of activities, such as walking, running, and doing sports. Rast et al. (2020) tracked animal behaviours using acceleration data and machine learning on several species in captive conditions to study the ecology of animals in natural habitats. Besides studies on the behaviour of animals living on land, machine learning models are also used for life living in the air and water, such as diving seabirds (Browning et al., 2018), sharks (Brewster et al., 2018), Japanese medaka fish (Ochiai et al., 2013), gull birds (Wijeyakulasuriya et al., 2020), and even penguins (Carroll et al., 2014).

Studies on animal activity and behaviour using machine learning have been conducted, including Wijeyakulasuriya et al. (2020), which found that predicting the state of movement behaviour and velocity of animals is more accurate with machine learning compared to parametric movement models. Meanwhile, Rast et al. (2020) concluded that wild animal behaviour from acceleration data through training data captured animals' behaviours without observing them in real life. The result showed accurate behaviour classification for the Artificial Neural Network machine learning model. For example, Mathis and Mathis (2020) used deep learning approaches to measure animal behaviour, and Rui and Yu (2020) detected animal characteristics based on Cascade Convolutionary Neural Networks. Peng et al. (2019) identified representations of movement patterns for different groups of animals using a Recurrent Neural Network. Rew (2019) proposed an animal movement prediction scheme using Recurrent Neural Network architecture forecasting to predict animal movement based on studies on Long-Billed Birds. Wang (2019) reviewed general principles and applications of Hidden Markov models, Random Forests, and Support Vector Machines in inferring an animal's behaviour from movement data.

Chen et al. (2019) described the latest approach to deep learning to automatically identify and isolate species-specific activities from still images and video data. Yadav and Bist (2019) classified changes in the spawning activity of Drosophila (fruit flies) via a three-hour video using a Deep Residual Network. Teimouri et al. (2018) divided homogeneous trajectories in terms of movement parameters into segments and grouped similar segments according to movement parameters that had been estimated using Decision Tree Hierarchical Clustering. Browning et al. (2018) trained a deep learning model to predict the diving activities of three seabird species, namely European shags, Guillemots, and Razorbills. Brewster et al. (2018) evaluated the performance of several machine learning classifications to find out five behaviours performed by a young Lemon Shark equipped with an accelerometer in Bimini, Bahamas. Finally, Valletta et al. (2017) introduced unusual or unfamiliar animal behaviours through machine learning, promising that this technique can analyse complex behaviour data. These related works show that machine learning and deep learning models perform better in predicting and classifying animal behaviours than traditional parametric models, such as linear regression, Poisson distributions, and exponential distributions, which can have certain restrictive assumptions leading to low prediction accuracy.

Semi-supervised machine learning studies have been conducted utilising the extensive spatiotemporal structure in behavioural videos (Wu et al., 2020). Spatial statistics are subjected to physical constraints and temporal smoothing from frame to frame. The results showed a partial model. The resulting sequence verified both labelled and unlabelled frames to achieve more accurate and robust tracking that required users to label fewer training data. Besides, trained classifiers can accurately classify unlabelled data. Camargo et al. (2020) proposed a more effective and efficient active semi-supervised learning framework, including new active learning methods. They found that active semi-supervised learning approaches by labelling small samples with specialist help showed better conversion (competitive computational accuracy and timing) than active supervised learning. Han et al. (2016) produced an efficient combination of confidence-based active learning and self-training to minimise the need for humans for sound classification training model annotations. The experimental results indicated that this approach required smaller labelled data to achieve similar performance in both scenarios compared to passive learning, active learning, and self-training.

Tanha et al. (2012) developed a semi-supervised multi-class learning algorithm that used primary classification combined with a similarity function applied to find the classification that maximised the margin and consistency of all data. Bishop et al. (2012) investigated standard decoder behaviour without retraining daily, i.e., neural signal behaviour on both datasets over time. The study found that signal characteristic changes were minimised mainly in a few days. Zhu et al. (2009) explored the relationship between machine learning and human learning in semi-supervised classification. Based on examples of past studies, semi-supervised learning can be used to predict animal behaviour in different domains, such as identifying cases of extortion in banking activities, classifying website content, analysing images, classifying text documents, and in speech recognition.

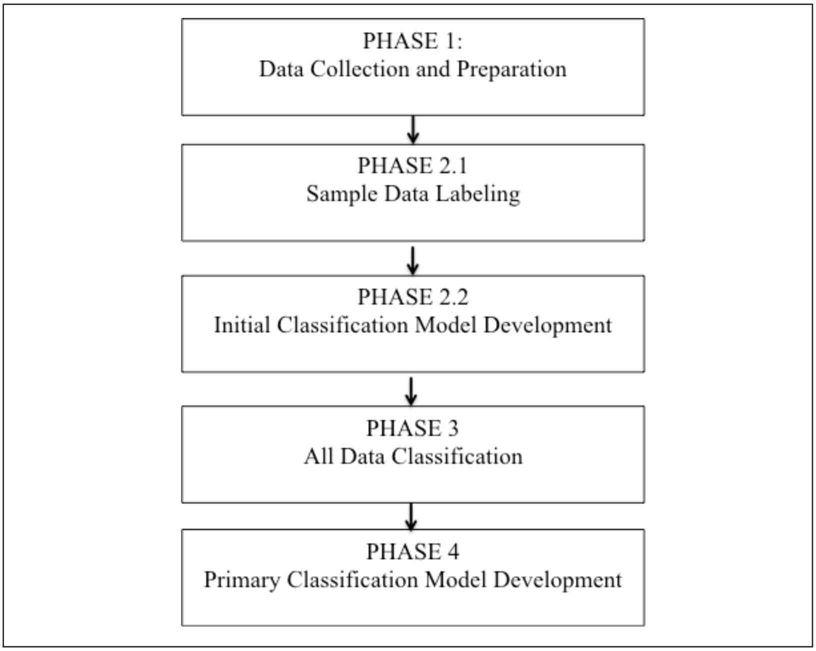
This study will adapt semi-supervised machine learning models where a small portion of the unlabelled dataset will be labelled manually with the help of a domain expert. The classified data will be trained by using the machine learning models proposed by previous works. After that, only a few models with the best accuracy and result will be chosen to classify the satellite collar data of Asian elephants in Malaysia and subsequently record them for study reporting. The most important part is that the result of classification will be validated whether it is correctly classified. The classification of Asian elephants' behaviour using satellite collar data for machine learning models has never been done before, and these models can be used to classify other new Asian elephant satellite collar data.

MATERIALS AND METHODS

This study was carried out in four main phases: data collection and preparation, data labelling and initial classification model development, all data classification, and primary classification model development. In Phase 1, the data collection and preparation phase, the elephant behaviour data were collected from the satellite collars attached to five male and female Asian elephants in forest reserves in Kelantan, Terengganu, Pahang, Perak, and Johor from around 2018 to 2020 by the Department of Wildlife and National Parks (DWNP), Malaysia. Phase 2 involved data labelling and initial classification model development. In this phase, the study employed the semi-supervised machine learning method to label and validate a sample of 100 Asian elephant behaviour data into four classes: walking, eating, bathing, and resting behaviours. These four behaviours are basic daily activities of Asian elephants based on the information gained from a domain expert at DWNP, Malaysia. Then, an initial classification model was developed using machine learning algorithms, namely XGBoost, Gradient Boost, Random Forest, and AdaBoost. The model with the highest classification accuracy was then used to classify the rest of the dataset in Phase 3. Finally, descriptive evaluation and analysis were performed to obtain a complete classified dataset. The study developed the primary classification model using the same algorithms in Phase 2 in the final phase. The purpose is to obtain the best classification model, also known as a classifier, that can be used to detect future elephant activity and to acquire the novel Asian elephant behaviour predictive model. Figure 1 depicts the methodology of the study.

Figure 1

Methodology of the Study

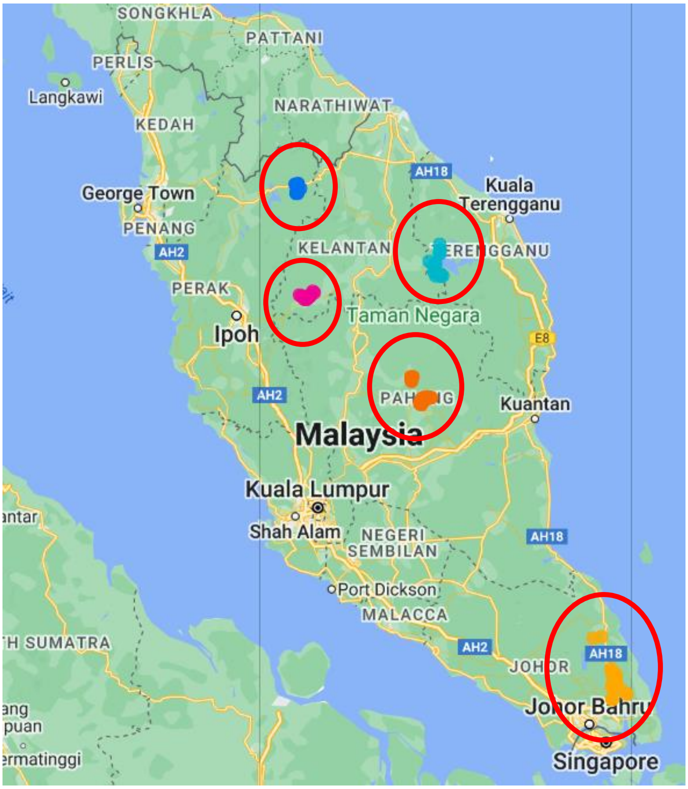


Phase 1: Data Collection and Preparation

This study utilised an original dataset from five satellite collars attached to Asian elephants. The collected data involved five male and female Asian elephants in different forest areas around Peninsular Malaysia. These five elephants were 25 percent of the total number fitted with satellite collars. The estimated elephant population in Peninsular Malaysia is around 1,000 to 1,100 individuals. Before the collar is installed, it will be turned on for a while to ensure it functions correctly. The data were recorded around 2018 until 2020 with 15,389 data. Figure 2 shows the location distribution of the five Asian elephants in forest reserves in Peninsular Malaysia.

Figure 2

Location Distribution of Five Asian Elephants in Forest Reserves in Peninsular Malaysia



This labelling process was carried out by mapping the location of coordinates on a topographic map using GIS applications and generating plot point graphs for tri-axis acceleration data (x, y, z) and guided by movement speed data. The mean for the acceleration data point is illustrated in Figure 3. Table 1 shows the descriptions for the raw datasets obtained.

Figure 3

Accelerometer Orientation in Satellite Collar

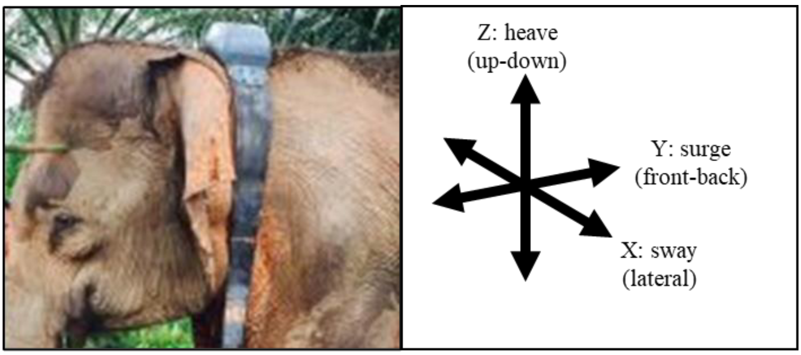


Table 1

Satellite Collar Dataset Description

Attribute (Data Type)	Data Description
Index (Varchar)	Data numbering
Station (Varchar)	Satellite station name
Tag (Varchar)	Satellite tag name/id
Type (Varchar)	Satellite type
GPS Lock Time-Stamp (UTC) (Datetime)	Universal standard date and time
Location (Varchar)	Collar location (latitude/longitude)
DOP (Integer)	Location data accuracy measure (metres)
Speed (Integer)	Animal movement speed (km/h)
Movement (Varchar)	Movement (yes/no)
Alarms (Varchar)	GPS-based alarm (yes/no)
Temperature (Integer)	Ambient temperature (°C)

(continued)

Attribute (Data Type)	Data Description
Accel (Roll, Pitch) [X,Y, Z] (Varchar)	The rate of acceleration on the x, y, and z axes (mm) and the degree of rotation of the animal's movement (°) on the x and y axes
Elevation (Integer)	Elevation (metre)
Coverage (Varchar)	Coverage (yes/no)

Five datasets from five Asian elephants of different locations totalling 15,389 data records with 15 attributes were merged into one extensive dataset. This dataset then went through several data preparation phases to ensure that clean and complete data were provided. The study handled incomplete data, outliers, and inconsistent data. Changing the position of the collar attached to the elephant's neck will cause the signal from the transmitter to not be received by the satellite. This act will lead to some records in the data, such as location coordinates and accelerometer data, not being accepted. The data should be removed because inserting a new value into the data is unreasonable as it will change the meaning of the data.

Outliers may occur due to an error while data are being recorded. There were 11 records of elephant movement speeds that exceeded the maximum limit of 55 to 127 km/h. Asian elephants' maximum movement speed is up to 30 to 35 km/h, depending on the size of the elephant. All 11 movement speed outliers were replaced with a mode value of 0 km/h. Some attributes with inconsistent data, such as single value and original data after transformation, were removed.

As a result of the above process, 11,729 records were prepared. Attributes can be categorised into three categories: attributes that are deleted as many as four attributes, attributes that are retained as many as 12 attributes, and new attributes created to replace the attributes that are deleted as many as six attributes. After the data cleaning and integration process, the study removed four attributes: Station, Type, GPS Lock Time-Stamp (UTC), and Temperature.

The study selected 17 attributes for machine learning algorithms' data transformation and feature selection. In contrast, the attribute 'Distance' was used in the descriptive analysis. The data transformation process is critical to obtain high-accuracy forecasts. This study converted several attributes to binomial and nominal representation forms, as shown in Table 2.

Table 2

Satellite Collar Dataset Description

Attribute	Representation
Tag	2 = tag satelit 2102, 3 = tag satelit 2103, 4 = tag satelit 2104, 5 = tag satelit 2105, 6 = tag satelit 2106
Month (created)	January–December
Year (created)	1 = 2018, 2 = 2019, 3 = 2020
Dayparting (created)	1 = morning, 2 = afternoon, 3 = evening, 4 = night
Location (created)	1 = Johor, 3 = Kelantan, 6 = Pahang 8 = Perak, 11 = Terengganu
Speed	0 = No speed, 1 = Speed 1 km/h, 2 = Speed 2 km/h, 3 = Speed 3–32 km/h
Movement	0 = No movement, 1 = Yes movement
Alarms	0 = No alarm, 1 = Yes alarm
Weather (created)	1 = Warm, 2 = Hot, 3 = Very hot, 4 = Extremely hot
Elevation	0 = No elevation, 1 = Yes elevation
Coverage	0 = No coverage, 1 = Yes coverage
Behaviour	1 = Eat, 2 = Bath, 3 = Rest, 4 = Walk

Attributes in a dataset have different levels of importance and capabilities in determining the success of a classification model. Feature selection aims to address high-dimensional data by discarding irrelevant and redundant data, thus improving learning accuracy, reducing computation time, and creating a better understanding of data and learning models (Cai et al., 2018). The selection of features to be eliminated considers the views of domain experts and technical analysis. Two methods of technical analysis for feature selection are Gain Ratio and Information Gain. Table 3 lists the position of the attributes in descending order after going through the feature selection process using both techniques.

Table 3

Ranking of Attributes from Information Gain and Gain Ratio

	Position	Gain Ratio	Position	Information Gain
1	Speed	0.66	Speed	0.93
2	Roll	0.41	Month	0.31

(continued)

	Position	Gain Ratio	Position	Information Gain
3	Pitch	0.36	Dayparting	0.16
4	Z	0.36	Weather	0.15
5	X	0.30	Y	0.10
6	Movement	0.28	Z	0.10
7	Y	0.22	X	0.08
8	Weather	0.13	Roll	0.08
9	Month	0.09	Tag	0.08
10	Dayparting	0.08	Location	0.08
11	Elevation	0.06	Pitch	0.06
12	Tag	0.03	Movement	0.02
13	Location	0.03	DOP	0.02
14	DOP	0.02	Elevation	0.02
15	Year	0.01	Year	0.01
16	Coverage	0.01	Coverage	0.01
17	Alarm	0.01	Alarm	0.01

Phase 2: Sample Data Labelling and Initial Classification Model Development

The study manually labelled a sample of 100 data prepared for initial classification model development in this phase. The purpose was to obtain a reliable initial classifier to classify and label the rest of the dataset. Two steps were involved: sample data labelling and initial classification model development. The output for the sample data labelling was the labelled data appropriate for modelling. At the same time, the developed initial model would be used to classify and label the rest of the dataset.

Descriptive analysis

Eleven thousand seven hundred and twenty-nine rows of clean datasets were analysed to obtain meaningful information and description of the data. Walking distance was an attribute created that could be obtained from the distance of one location point to another point. Observation made for walking distance based on dayparting could give additional knowledge about Asian elephants' most active walking time of the day. The time was divided into four intervals: morning, afternoon, evening, and night. The farthest walking distance recorded during the evening, which started at 1800 hours and was followed by night at 2400 hours, morning, and lastly, afternoon, matched the explanation made by domain experts. According to the experts, elephants' most active walking time was at night or in a dark environment, as depicted in Figure 4.

Figure 4

The Difference between Asian Elephants' Walking Distance by Dayparting

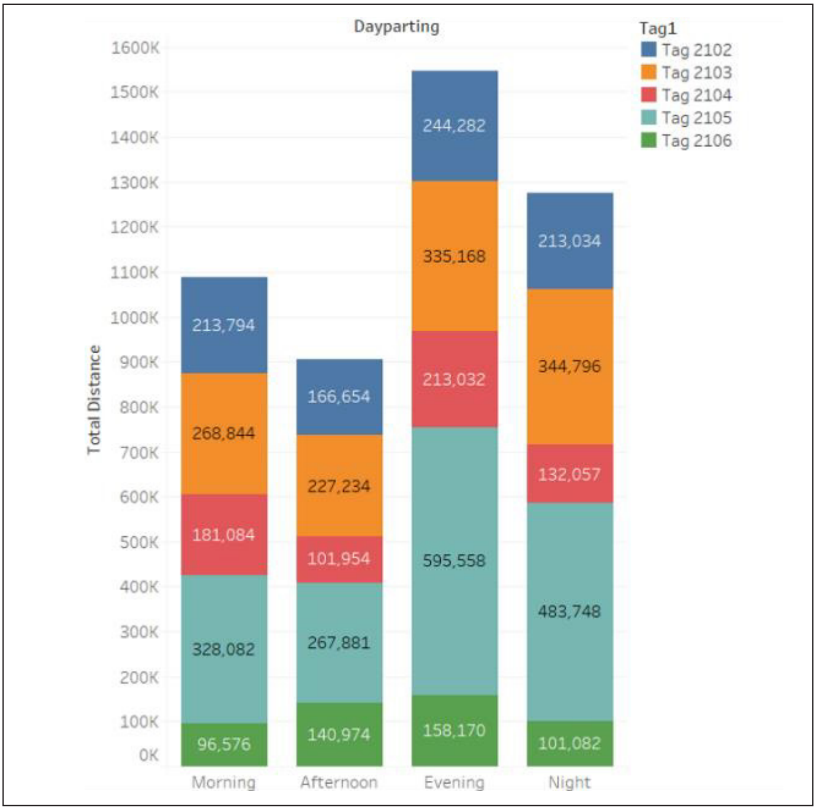
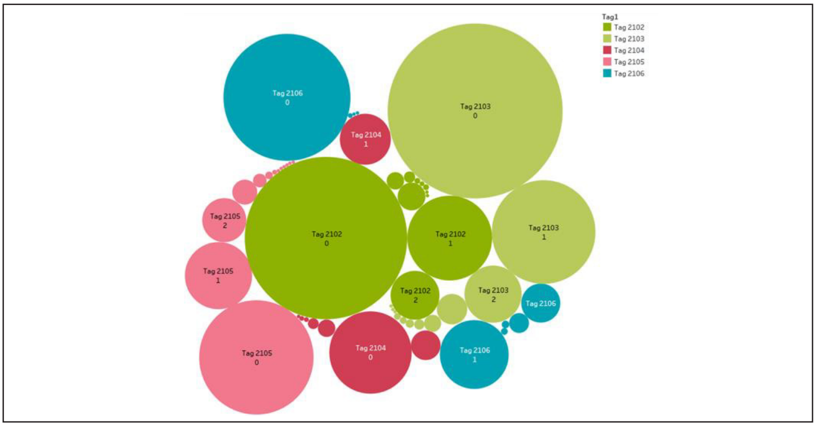


Figure 5 shows the bubble chart for the speed of Asian elephants while walking at km/h. 0 km/h represented activities other than walking: bathing, eating, and resting. Most often, the walking speed was 1 km/h, followed by 2 km/h and 3 km/h. This observation supported the statement of domain experts that elephants usually walk at a low speed, which is 1 or 2 km/h, and will gradually increase the speed under certain circumstances, such as threatened with danger. The finding indicated that the area was convenient and had deficient elements of provocation for the Asian elephant habitat.

Figure 5

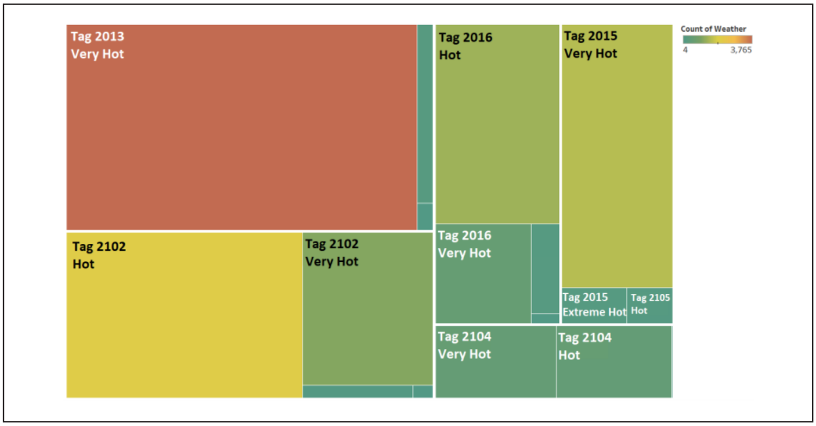
Asian Elephants' Walking Speed



Malaysia has hot and humid weather all year round. From the treemap chart in Figure 6, the weather distribution was mainly very hot and hot in all five forests reserve. Elephants usually perform activities in open places, such as rivers, lakes, swamps, and fields exposed to direct sunlight compared to canopy forest areas, which explains the weather conditions. Nevertheless, the experts did not deny that such weather conditions are likely to indicate a slight ecological decline in the Asian elephant habitat due to uncontrolled activities by parties who are not responsible.

Figure 6

Weather Distribution in Asian Elephant Habitat



Phase 2.1: Sample Data Labelling

The classification model development required data with a class label. Since the obtained satellite collar data had no class labels, a dependent variable that acted as a multi-class label attribute was created. The 'Behaviour' label attribute had four class values: 'walk', 'eat', 'rest', and 'bath'. A total of 100 data records were randomly selected, i.e., approximately 20 records from each elephant for data labelling with the help of domain experts.

The labelling process was performed based on graphs of accelerometer acceleration data on the x, y, and z axes, a topographic map of the current location of the Asian elephant, and movement speed data. For example, 'bath' and 'eat' behaviours could be distinguished by the location of the elephant's position, such as being in a river, lake, farm, or farm area. If the elephant's behaviour was 'rest', the acceleration graph readings on the x and y axes were close to 0 and vice versa.

Using the ArcGIS application, the study identified the elephant locations by mapping coordinate data on a topographic map. If the location of the Asian elephant at that time were in a river or lake, with a movement speed of 0 km/h, the elephant would be labelled as 'bath' behaviour. Suppose the speed data were other than 0 km/h; it meant that the elephant's behaviour was 'walk'. The 'walk' behaviour was the easiest to identify by simply referring to the 'Speed' data. At the same time, the 'bath' behaviour could be referred to the topographic map, while 'eat' and 'rest' were distinguished through acceleration data graphs and topographic maps as described. In the end, 100 labelled training datasets and 11,629 unlabelled datasets made a total of 11,729 datasets ready for modelling.

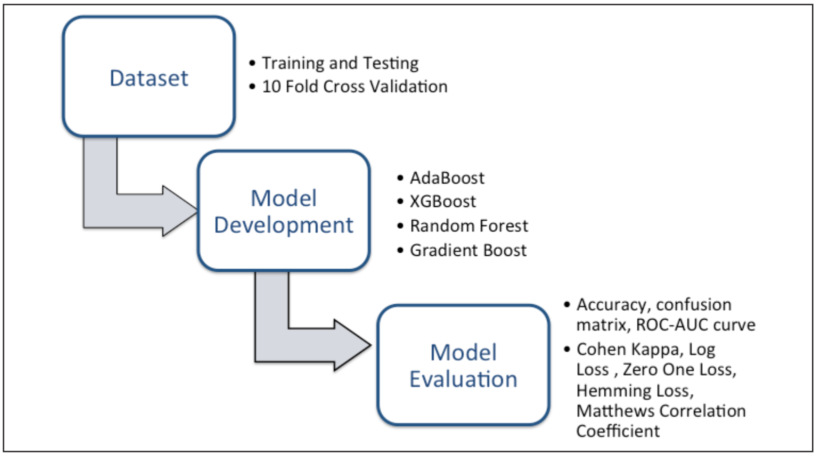
Phase 2.2: Initial Classification Model Development

Many past studies on wildlife behaviour using machine learning forecasting models have been conducted.

The current study employed four machine learning classification algorithms to model the labelled dataset, namely, XGBoost, Random Forest, AdaBoost, and Gradient Boost. XGBoost, or Extreme Gradient Boost, implements a Gradient Boosting Decision Tree Algorithm, allowing the machine learning model to undergo computational errors and then train the model to predict errors before adding the ensemble model. Random Forest develops a forest from a group of decision trees trained with bagging techniques to improve overall results that are more precise and stable. AdaBoost is one of the most popular ensemble techniques, which uses many classification models to obtain ensemble classification models for the best prediction. The Gradient Boost algorithm consists of three elements: loss function, weak learner, and additive model, and four improvements: shrinkage, penalised learning, random sampling, and tree constraints, making it an excellent prediction model. Figure 7 depicts the classification model development scheme used in this study.

Figure 7

Classification Model Development Process



Initial Model Evaluation

The model development employed the 10-fold cross-validation setting to obtain various training and testing samples randomly to observe the capability of the most stable classifier. The machine learning algorithms learn the training data to produce the classifier. The accuracy of a classifier is evaluated based on the percentage of test data that are correctly classified, as shown in Equation 1. The highest accuracy classifier was chosen to classify the rest of the 11,629 unlabelled data.

$$Accuracy = \frac{TP + TN}{N} \tag{1}$$

where,

TP: The number of actual class behaviours that are accurately classified

TN: The number of other class behaviours that are accurately classified

N: The number of records in the test data

Phase 3: Data Classification

The classification model obtained in Phase 2 was used to classify the 11,629 rows of data. The correctness of the data class was evaluated using a descriptive analysis method based on several rules to determine whether the label for each behaviour class was correct. There were four behaviour classes: walking, resting, bathing, and eating. Walking behaviour was the easiest to determine through the speed attribute (Speed), whereby if the speed was other than 0 km/h, the elephant showed walking behaviour. The steps are illustrated in Algorithm 1.

Algorithm 1: Steps to Identify Elephant's Walking Behaviour

Step 1: Scan the 'walking' class data
Step 2: Scan the 'speed' attribute
    If 'speed' ≠ 0, the class is correctly labelled
    If 'speed' = 0, the class is incorrectly labelled

The evaluation for the resting class was based on the values from the x, y, and z (accelerometer) acceleration data plotted on the line graph to analyse its shape and the elephant location data referenced on the topological map. The rule for determining the resting class was that the values of the x and y axes on the acceleration data should be small, indicating that the elephant's body was at rest. Based on the topographic map, the location of the elephant resting should be on land instead of rivers, swamps, or lakes. The steps for determining the assessment of the resting behaviour class are specified in Algorithm 2.

Algorithm 2: Steps to Identify Elephant's Resting Behaviour

Step 1: Scan the 'rest' class data
Step 2: Break down the 'rest' data by tag/location
Step 3: Scan the 'speed' attribute
    If 'speed' = 0, the class is labelled correctly
    If 'speed' ≠ 0, the class is not labelled correctly

Step 4: Plot the values of the acceleration data x, y, z to the line graph
If the value of 'x & y' is small, the class is correctly labelled
If the value of 'x & y' is large, the class is incorrectly labelled
Step 5: Plot the latitude and longitude on a topographic map
If the location is onshore, the class is correctly labelled
If the location is not on land, the class is incorrectly labelled

The bathing class was assessed through observations on topological maps. Attributes for latitude and longitude were plotted on a topological map to see the location of elephants classified as bathing. Elephants in the water, namely rivers, swamps, and lakes, showed bathing behaviour as long as the movement speed was 0 km/h. The generated rule algorithm for assessing the bathing behaviour class is in Algorithm 3.

Algorithm 3: Steps to Identify Elephant's Bathing Behaviour

Step 1: Scan the 'bathing' class data
Step 2: Break down the 'bathing' data by tag/location
Step 3: Scan the 'speed' attribute
If 'speed' = 0, the class is correctly labelled
If 'speed' ≠ 0, the class is incorrectly labelled
Step 4: Plot the latitude and longitude on a topographic map
If the location is on water, the class is correctly labelled
If the location is not on the water, the class is incorrectly labelled

Locations that focused on eating activities were on lands with food sources, such as plant areas, agricultural areas, and riverbanks. The eating class should also record movement speed readings at 0 km/h. Assessment for the eating class was performed using Algorithm 4.

Algorithm 4: Steps to Identify Elephant's Eating Behaviour

Step 1: Scan the 'eating' class data
Step 2: Break down 'eat' data by tag/location
Step 3: Scan the 'speed' attribute
If 'speed' = 0, the class is correctly labelled
If 'speed' ≠ 0, the class is incorrectly labelled
Step 4: Plot the latitude and longitude on a topographic map
If the location is on land and plants, the class is labelled correctly
If the location is not on land, the class is not labelled correctly.

Phase 4: Primary Classification Model Development

The machine learning algorithms in Phase 2 were employed to obtain the primary classification model from the classified dataset in Phase 3. In this phase, the model development followed the same steps as in Figure 7 in the previous phase. The algorithms were developed using Python programming in Google Colab. Nevertheless, the model evaluation considered several comprehensive metrics, such as classification accuracy Equation 2, confusion matrix, Area Under the Curve of the Receiver Operating Characteristic (ROC-AUC), and aggregate matrices, namely Cohen Kappa, Log Loss and Zero One Loss, Hemming Loss, and Matthews Correlation Coefficient. The subsequent measurement was the recall measure Equation 3 to determine the classifier model's ability to find all the correct records for each class. Details of the evaluation are presented in the next section.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

where,

TP (True Positive): The number of behaviours correctly classified

FP (False Positive): The number of behaviours incorrectly classified

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{3}$$

where,

TP (True Positive): The number of behaviours correctly classified

FN (False Negative): The number of behaviours incorrectly classified

RESULTS AND DISCUSSION

This section presents the experimental result of Phase 2 to Phase 4 of the methodology. It evaluates the performances of the initial classification model, data classification, and primary classification model.

Initial Classification Model Performance

Classification accuracy is the primary measure of model performance in machine learning. This study recorded the classification accuracy of 10-fold cross-validation models for the XGBoost, Random Forest, AdaBoost, and Gradient Boost classification models in Table 4. The highest and average accuracy of F1–F10 models showed that XGBoost outperformed other methods with the highest and average values of accuracy at 90 percent and 69 percent, respectively. Nevertheless, different ways that gave competitive results in individual models, such as Random Forest in F4, could also be observed. Therefore, the XGBoost model was used to classify the remaining unlabelled data in this study.

Table 4

Performance of Initial Classification Model (in Accuracy)

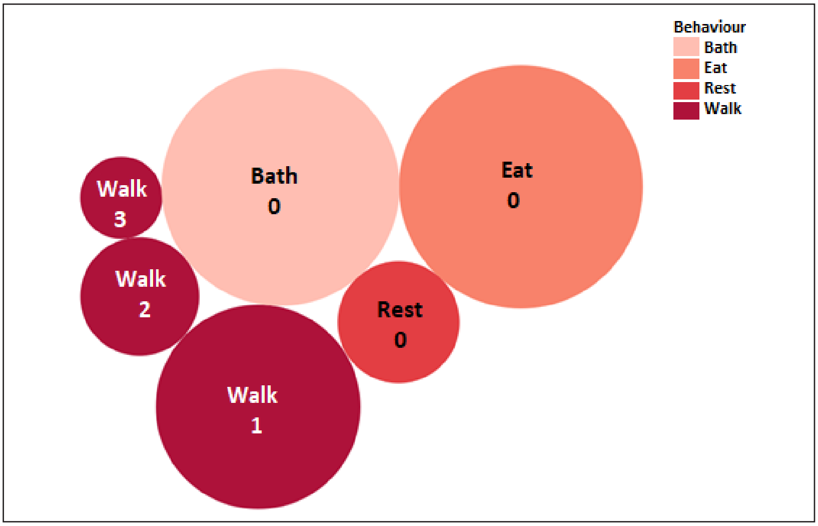
Model Accuracy (%)	Highest Accuracy	Average Accuracy
XGBoost	90	69
Random Forest	90	66
AdaBoost	80	63
Gradient Boost	90	62

Data Classification Performance

The XGBoost model was the best classifier to classify the remaining 11,629 unlabelled data. The model had classified 3,772 data as walking, 3,568 data as eating, 3,393 as bathing, and 896 as resting. This study analysed the walking activity in conjunction with the attribute of movement speed (Speed). Data that recorded non-zero speed values other than 0 km/h were classified as walking. Figure 8 depicts the classification of the walking class in a bubble chart. Based on the chart, the model correctly classified 3,772 data as walking, as these data had a non-zero speed. The data with a speed of 1 to 32 km/h were correctly classified as walking.

Figure 8

Bubble Chart for Walking Behaviour Class



The XGBoost model classification of resting activities from unlabelled datasets was 896. The resting class was assessed through speed, accelerometer acceleration data, and topographic maps. The screening was done on the speed data to ensure the resting class recorded a 0 km/h. The values of x and y were minimal for the resting activities, indicating that the elephant's body barely moved. Table 5 shows that 847 records were correctly classified, while 49 records were incorrectly classified.

The rules for analysing the bathing class classification were based on the 'Speed' attribute, and the topographic map was plotted using latitude and longitude location data. The 'Speed' attribute was filtered to ensure only a 0 km/h speed data record. Then, location point mapping on a topographic map was performed to observe Asian elephants that were not in wet areas, such as rivers, swamps, or lakes. At this stage, many observations were done together with domain experts to obtain more accurate information and views. Five sets of topographic maps for five elephants showed that 198 labels were incorrectly classified as bathing class activities, while 3,195 data were correctly classified. Table 5 shows the correct and incorrect numbers for the bathing behaviour class.

The eating class was assessed by filtering speed data and referring to topographic maps. The movement speed for eating activity was 0 km/h. Careful observation of topographic map diagrams was needed as elephants fed in forests, agricultural areas, rivers, and swamps. Table 5 shows that the correct classification of eating classes was 3,484, while 84 were misclassified.

A total of 11,298 data records labelled in the previous stage would be used in developing the classification model to predict eating, bathing, resting, and walking behaviours. The views from domain experts stated that the machine learning algorithm's performance was outstanding, simplified the task, and saved time. Figure 9 illustrates the performance of the data classification.

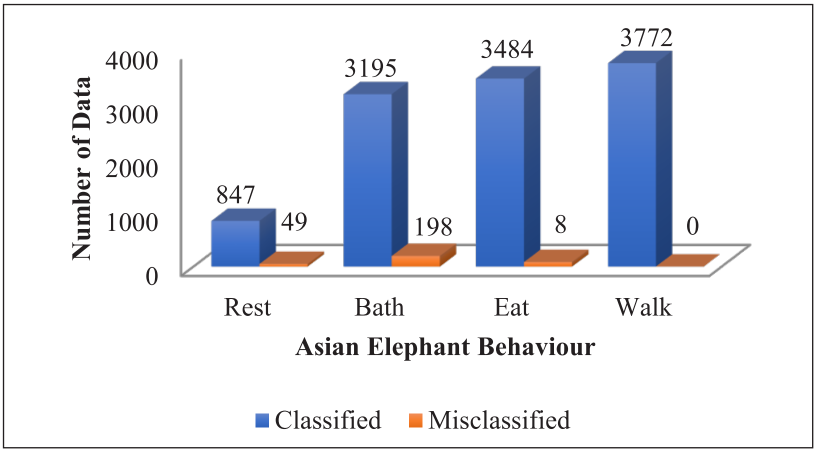
Table 5

The Result of Data Classification

Behaviour	Rest	Bath	Eat	Walk	Total Data
Classified	847	3,195	3,484	3,772	11,298
Misclassified	49	198	84	0	331
Total Sample	896	3,393	3,568	3,772	11,629
Accuracy	94.53	94.16	97.65	100	n/a

Figure 9

Unlabelled Data Classification Performance



Primary Classification Model Performance

The classification accuracy was recorded for analysis as in Table 6. The XGBoost, Gradient Boost, and Random Forest models showed outstanding predictive accuracy for walking, resting, and eating behaviours of 1.0. All three models were also very good at predicting bathing activity, with an accuracy of 0.99. The AdaBoost classification model showed difficulty and inability to predict resting behaviour with a score of 0.00, while the ‘bathing’ class prediction gained an accuracy value of 0.79. The ‘walking’ and ‘eating’ classes were predicted with an accuracy of 1.0.

Table 6

Precision Scores for Four Classification Models by Class

	XGBoost	GBoost	Random Forest	AdaBoost
Walk	1.00	1.00	1.00	1.00
Rest	1.00	1.00	1.00	0.00
Eat	1.00	1.00	1.00	1.00
Bath	0.99	0.99	0.99	0.79

Table 7 shows the recall values for the four classification models. The XGBoost model outperformed other models with 1.0 recall for walking, resting, and bathing class behaviours. The eating class also performed well, with a recall of 0.99. The Gradient Boost and Random Forest models indicated the same predictive performance across all classes, i.e., the walking, eating, and bathing classes were 1.00, and the resting class was 0.99. The AdaBoost model managed to achieve a recall of 1.00 in the walking, eating, and bathing classes but was very weak in the resting class, with a retrieval score value of 0.00.

Table 7

Recall Values for Four Classification Models by Class

	XGBoost	GBoost	Random Forest	AdaBoost
Walk	1.00	1.00	1.00	1.00
Rest	1.00	0.99	0.99	0.00
Eat	0.99	1.00	1.00	1.00
Bath	1.00	1.00	1.00	1.00

Based on Table 8, the F1 scores for each model by class were recorded. The XGBoost model ranked first with the walking and resting class behaviour scores at 1.00, while eating and bathing classes were at 0.99. Again, the Gradient Boost and Random Forest models showed equivalent performance with an F1 score of 1.00 for walking and eating classes, a score of 0.99 for the bathing class, and 0.99 for the resting class. The AdaBoost model was in the last position with the walking and eating classes at a value of 1.0, the bathing class at 0.88, and the resting class at 0.00.

Table 8

F1 Scores for Four Classification Models by Class

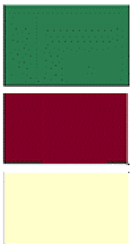
	XGBoost	GBoost	Random Forest	AdaBoost
Walk	1.00	1.00	1.00	1.00
Rest	1.00	0.99	0.99	0.00
Eat	0.99	1.00	1.00	1.00
Bath	0.99	0.99	0.99	0.88

The number of correctly predicted classes could be identified through the confusion matrix. Figure 10(a) depicts the confusion matrix for the XGBoost classification model. All behavioural activities were correctly predicted according to the actual class with a score of 100 percent. The confusion matrix for the following classification model was Gradient Boost, as shown in Figure 10(b). This model successfully predicted walking, eating, and bathing classes with 100 percent accuracy. The resting class was 99 percent successfully predicted correctly, while 1 percent was incorrectly predicted as bathing. The Random Forest model had the same results as the Gradient Boost model, i.e., it successfully predicted walking, eating, and bathing with 100 percent accuracy and 99 percent accuracy for the resting class, whereby 1 percent was classified in the bathing class, as illustrated in Figure 10(c). The model that showed poor performance from the confusion matrix assessment was the AdaBoost model. Although all three bathing, eating, and walking classes were successfully predicted at 100 percent, the resting class was incorrectly classified at 100 percent to the bathing class. Figure 10(d) displays the confusion matrix for the AdaBoost classification model.

Figure 10 (a-d)

Confusion Matrix for Classification Models

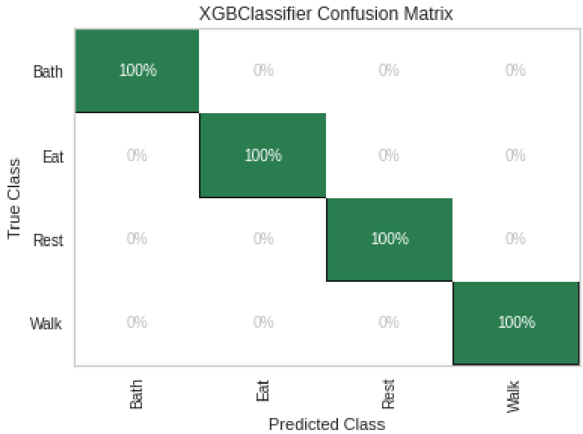
Legend:



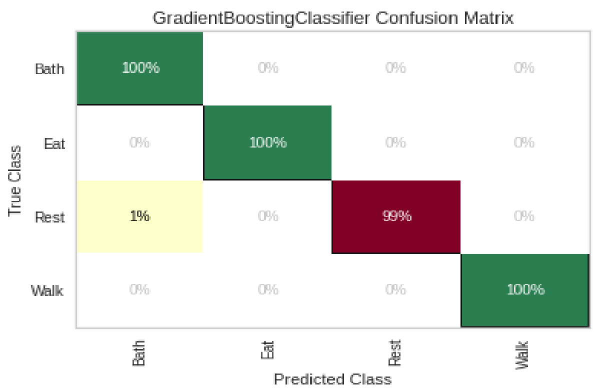
All correctly predicted

Correctly predicted

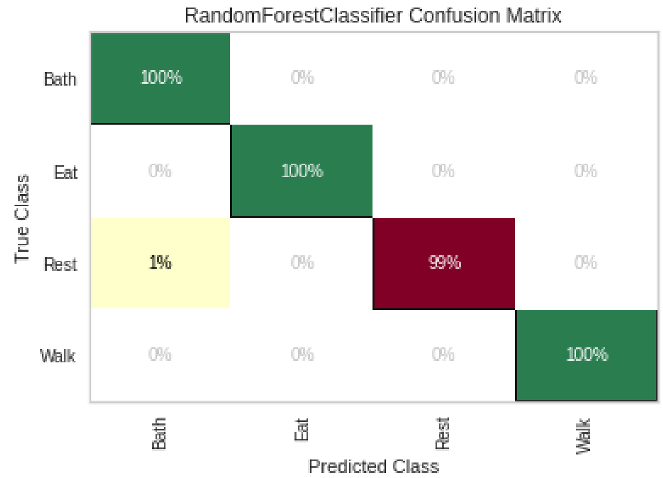
Incorrectly predicted



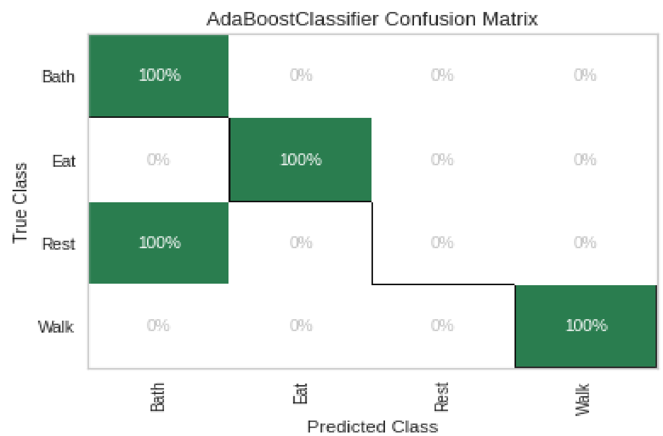
(a) XGBoost



(b) Gradient Boost



(c) Random Forest



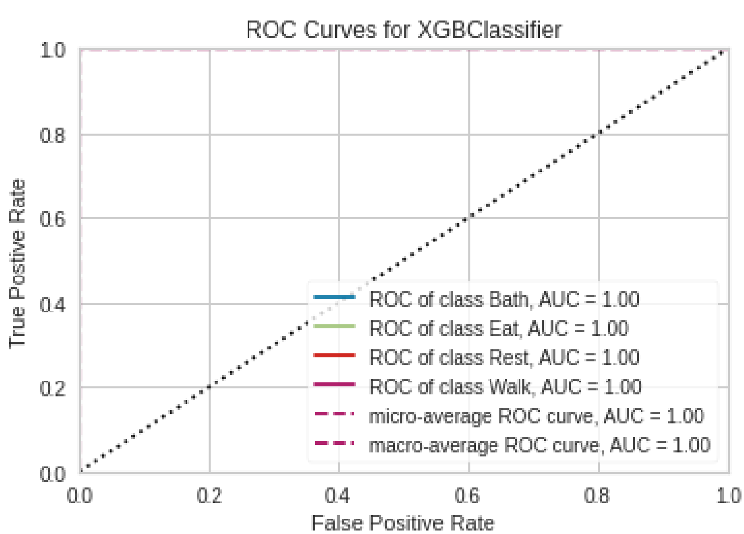
(d) AdaBoost

The ROC and AUC values for the models were recorded for evaluation and analysis. Curves for macro and micro averages were also plotted. For the predictive classification of various classes of unbalanced datasets, micro mean values are more appropriate because they take into account the frequency of each class. Figure 11(a) shows the ROC curve and the XGBoost classification model's macro and micro mean curves. XGBoost displayed outstanding performance with readings of 1.0 for the ROC curve, AUC score, and micro and macro averages.

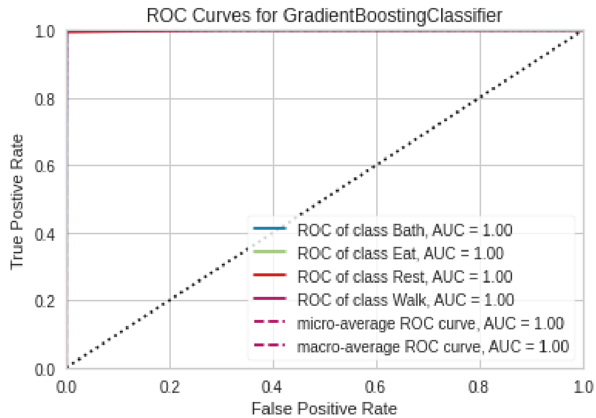
The Gradient Boost model demonstrated the exact shape of the ROC curve, AUC score, micro mean, and macro average as the XGBoost model with a reading value of 1.00. Nevertheless, a slight difference in the resting class's ROC line on closer observation indicated that the Gradient Boost model was slightly behind the XGBoost model, as shown in Figure 11(b). Based on Figure 11(c), the ROC curve, AUC score, micro mean curve, and the macro average for the Random Forest classification model were also outstanding, with all readings of 1.00. However, the detailed observation found a slight appearance of the figure's macro and micro average lines. The AdaBoost model showed lower classification performance in the ROC curve rating metric. The ROC curve for bathing and resting class behaviours still demonstrated good performance with an AUC score of 0.95 and 0.85, respectively, for both classes. The micro and macro average curves gave promising readings of 0.99 and 0.95, while the ROC curves for the eating and walking classes gave values of 1.00, as shown in Figure 11(d).

Figure 11 (a-d)

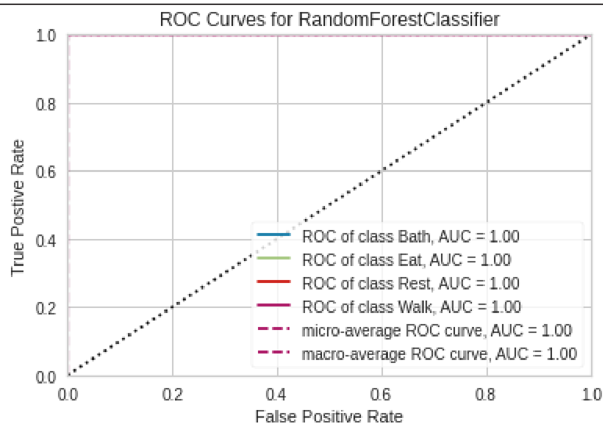
ROC Curves, AUC Scores, Micro and Macro Averages for Classification Models



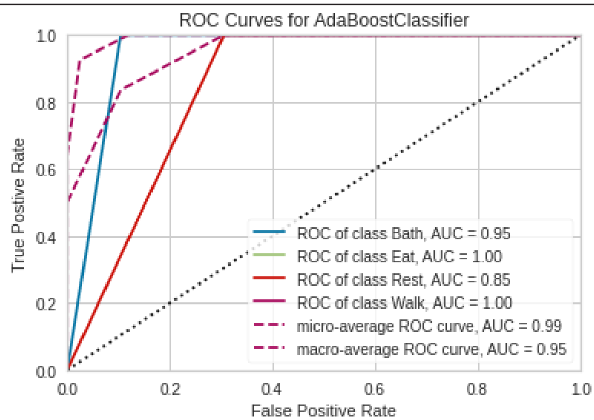
(a) XGBoost



(b) Gradient Boost



(c) Random Forest



(d) AdaBoost

The aggregate evaluation metrics were measured using four methods: Cohen's Kappa Statistics, Logarithmic Loss Function, Blank-One Loss Function, and Matthews Correlation Coefficient (MCC). Table 9 shows the performance records for each model using aggregate evaluation metrics. Through the evaluation conducted, the XGBoost, Gradient Boost, and Random Forest models achieved a very excellent Cohen's Kappa score of 0.99. In contrast, the AdaBoost model gained a score of 0.89. As for the Log Loss Function, the XGBoost model performed well, with a shallow score close to 0, which was 0.002. In second place was the Gradient Boost model with a score of 0.01, followed by Random Forest with 0.07, and the last was the AdaBoost model with a score of 0.25. The Blank-One Loss Function can measure the extent to which the output produced differs from the actual, meaning the smaller the value or, the closer it is to 0, indicating that the model's performance is good. The XGBoost, Gradient Boost, and Random Forest models had an excellent score of 0.00, while AdaBoost had 0.08. The last aggregate rating metric was the Matthews Correlation Coefficient, with scores of 0.99 for the XGBoost, Gradient Boost, and Random Forest prediction classification model, while the AdaBoost model achieved a score of 0.90.

Table 9

Aggregate Metrics for Prediction Classification Models

	XGBoost	**GBoost**	**Random Forest**	**AdaBoost**
Cohen's Kappa	0.99	0.99	0.99	0.89
Log Loss Function	0.002	0.01	0.07	0.25
Loss 0-1 Function	0.00	0.00	0.00	0.08
MCC	0.99	0.99	0.99	0.90

The descriptive and predictive approaches for the unlabelled Asian elephants' satellite collar data using semi-supervised machine learning for the behavioural prediction can save time, cost, and workforce to observe the capabilities of machine learning algorithms. This method is practical compared to the statistical predictive model and clustering algorithm. Based on the problem addressed in the study, identifying and determining activity patterns through collar data satellites using partially supervised machine learning by developing an initial classification model on 100 labelled data can be done quickly and accurately. The difficulty in identifying the causes that drive ecological processes because of limited access to information on elephant ecology can be overcome through descriptive analysis of satellite collar data. An evaluation scheme developed to assess the classification accuracy has produced a standard and adequately classified satellite collar dataset.

Based on the analysis of the speed of elephant movements, the causes of the rise and fall of ecological processes in forest reserves in Peninsular Malaysia are still in place under control. Nevertheless, the analysis of the weather distribution showing prolonged hot conditions gives little indication that there is an ecological decline stemming from human-elephant conflicts, activities in agriculture, development, logging, settlement, mining, and conversion of other land use. The capability of the classification model is promising as 97.15 percent of the total of 11,629 data were correctly labelled. The XGBoost classification model can save time, simplify predicting behavioural activity, and achieve high classification accuracy.

CONCLUSION

This study developed a predictive analytics model of wildlife behaviour using data obtained from the satellite collars attached to five Asian elephants in several forest reserves in Peninsular Malaysia. This study contributes to three main aspects: studies in machine learning, predictive analytics of wildlife behaviour, and support for big data analytics initiatives. In addition, further studies can be conducted using various algorithms, a more extensive dataset size, different evaluation metrics, and the involvement of more domain experts to achieve better study results. Experiments of the Asian elephant behavioural prediction model have created a new form of research in Malaysia. The model can be embedded into the intelligent system to detect elephant activities. The new knowledge obtained from the process of knowledge analysis can help officers in DWNP compile and formulate work strategies in the field to be more efficient and save time, cost, and energy. In addition, this study fosters an awareness of the importance of caring for ecosystems so that biological stability can be preserved against wildlife, forest areas, and local communities.

ACKNOWLEDGMENT

The authors would like to thank the Department of Wildlife and National Parks for providing data, the Department of Polytechnic and Community College, and the Ministry of Higher Education. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Bishop, W., Ryu, S. I., Chestek, C., Nuyujukian, P., & Yu, B. M. (2012). Semi-supervised classification for intracortical brain-computer interfaces. https://www.ml.cmu.edu/research/dap-papers/dap_bishop.pdf.

Brewster, L. R., Dale, J. J., Guttridge, T. L., Gruber, S. H., Hansell, A. C., Elliott, M., Cowx, I. G., Whitney, N. M., & Gleiss, A. C. (2018). Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data. *Marine Biology*, *165*(4), 1–19. https://doi.org/10.1007/s00227-018-3318-y

Browning, E., Bolton, M., Owen, E., Shoji, A., Guilford, T., & Freeman, R. (2018). Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Methods in Ecology and Evolution*, *9*(3), 681–692. https://doi.org/10.1111/2041-210X.12926

Camargo, G., Bugatti, P. H., & Saito, P. T. M. (2020). Active semi-supervised learning for biological data classification. *PLoS ONE*, *15*(8 August), 1–20. https://doi.org/10.1371/journal.pone.0237428

Cargnelutti, B., & Coulon, A., Hewison, A. M., Goulard, M., & Angibault, J. M., & Morellet, N. (2007). Testing global positioning system performance for wildlife monitoring using mobile collars and known reference points. *The Journal of Wildlife Management*. https://doi.org/10.2193/2006-257

Carroll, G., Slip, D., Jonsen, I., & Harcourt, R. (2014). Supervised accelerometry analysis can identify prey capture by penguins at sea. *Journal of Experimental Biology*, *217*(24), 4295–4302. https://doi.org/10.1242/jeb.113076

Chandrakar, A. K. (2018, September). *Wildlife: An introduction* [PowerPoint slides, Department of Forestry, Wildlife & Environmental Sciences, Guru Ghasidas Vishwavidyalaya University]. https://doi.org/10.13140/RG.2.2.23194.08649

Chen, R., Little, R., Mihaylova, L., Delahay, R., & Cox, R. (2019). Wildlife surveillance using deep learning methods. *Ecology and Evolution*, *9*(17), 9453–9466. https://doi.org/10.1002/ece3.5410

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., & Butler, P. J. (2004). Biotelemetry: A mechanistic approach to ecology. *Trends in Ecology & Evolution*, *19*(6), 334–43. https://doi.org/10.1016/j.tree.2004.04.003

Dey, N., Ashour, A. S., & Nguyen, G. N. (2017). Deep learning for multimedia content analysis. *Mining Multimedia Documents*, *1*(4), 193–203. https://doi.org/10.1201/b21638

Fontecchio, M. (2013). *Spatial data*. TechTarget. https://searchsqlserver.techtarget.com/definition/spatial-data?_ga=2.262667908.945866828.1606294096-1330996966.1594392127

Hamdan, A. R., Abu Bakar, A., & Ahmad Nazri, M. Z. (2018). *Sains data - penerokaan pengetahuan dari data raya (1st ed.)*. Penerbit Universiti Kebangsaan Malaysia.

Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., & Zhu, X. (2016). Semi-supervised active learning for sound classification in hybrid learning environments. *PLoS ONE*, *11*(9), 1–19. https://doi.org/10.1371/journal.pone.0162075

Li, D., Wang, S., & Li, D. (2016). Spatial data mining: Theory and application. In *Spatial Data Mining: Theory and Application* (Issue July 2017). https://doi.org/10.1007/978-3-662-48538-5

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behaviour in neuroscience. *Current Opinion in Neurobiology*, *60*, 1–11. https://doi.org/10.1016/j.conb.2019.10.008

Mattisson, J., Andrén, H., Persson, J., & Segerström, P. (2010). Effects of species behavior on global positioning system collar fix rates. *The Journal of Wildlife Management*, *74*(3), 557–563. http://www.jstor.org/stable/27760485

Mohamed, R., Shah Zainudin, M. N., Sulaiman, M. N., Perumal, T., & Mustapha, N. (2018). Multi-label classification for physical activity recognition from various accelerometer sensor positions. *Journal of Information and Communication Technology*, *17*(2), 209–231. https://doi.org/10.32890/jict2018.17.2.8252

Ochiai, T., Suehiro, Y., Nishinari, K., Kubo, T., & Takeuchi, H. (2013). A new data-mining method to search for behavioral properties that induce alignment and their involvement in social learning in medaka fish (*oryzias latipes*). *PLoS ONE*, *8*(9), e71685. https://doi.org/10.1371/journal.pone.0071685

Peng, C., Duarte, C. M., Costa, D. P., Guinet, C., Harcourt, R. G., Hindell, M. A., McMahon, C. R., Muelbert, M., Thums, M., Wong, K.-C., & Zhang, X. (2019). Deep learning resolves representative movement patterns in a marine predator species. *Applied Sciences*, *9*(14), 2935. https://doi.org/10.3390/app9142935

PERHILITAN. (2021). *Gajah (Elephas maximus) di Semenanjung Malaysia*. Department of Wildlife and National Parks (PERHILITAN) Peninsular Malaysia. http://wildlife.gov.my/images/stories/penerbitan/kertas_maklumat/gajah.pd

Rast, W., Kimmig, S. E., Giese, L., & Berger, A. (2020). Machine learning goes wild: Using data from captive individuals to infer wildlife behaviours. *PLoS ONE*, *15*(5), 1–25. https://doi.org/10.1371/journal.pone.0227317

Rew, J. (2019). Animal movement prediction based on predictive recurrent neural network. *Sensors (Switzerland)*, *19*(20), 4411. https://doi.org/10.3390/s19204411

Rui, L., & Yu, R. (2020, May). An animal behavior state estimation method using CCNN and BN based system. In *CSSE '20: Proceedings of the 3rd International Conference on Computer Science and Software Engineering* (pp. 158–163). https://doi.org/10.1145/3403746.3403921

Sarkar, D., Chapman, C. A., Griffin, L., & Sengupta, R. (2015). Analysing animal movement characteristics from location data. *Transactions in GIS*, *19*(4), 516–534. https://doi.org/10.1111/tgis.12114

Sharma, A., Jat, H., & Gupta, R. (2017). A survey of spatial data mining approaches: Algorithms and architecture. *International Journal of Computer Technology and Electronics Communication*, April, 15–22. http://earthjournals.org/ijctecpaper/IJCTEC_102.pdf

Somayeh, D., Gil, B., & Rolf, W. (2012, September). MoveBank track annotation project: Linking animal movement data with the environment to discover the impact of environmental change in animal migration. In *Conference on Geographic Information Science 2012 (GIScience)* (pp. 35–41).

Tanha, J., Van Someren, M., De Bakker, M., Bouteny, W., Shamoun-Baranesy, J., & Afsarmanesh, H. (2012, November). Multiclass semi-supervised learning for animal behaviour recognition from accelerometer data. In *Proceedings - International Conference on Tools with Artificial Intelligence (ICTAI)* (Vol. 1, pp. 690–697). https://doi.org/10.1109/ICTAI.2012.98

Teimouri, M., Indahl, U. G., Sickel, H., & Tveite, H. (2018). Deriving animal movement behaviours using movement parameters extracted from location data. *ISPRS International Journal of Geo-Information*, *7*(2), 78. https://doi.org/10.3390/ijgi7020078

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., & Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2163–2176. https://doi.org/10.1098/RSTB.2010.0090

Torre, J. A., Lechner, A. M., Wong, E. P., Magintan, D., Saaban, S., & Campos-Arceiz, A. (2019). Using elephant movements to assess landscape connectivity under Peninsular Malaysia's central forest spine land use policy. *Conservation Science and Practice*, *1*(12), 1–14. https://doi.org/10.1111/csp2.133

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Wang, G. (2019). Machine learning for inferring animal behaviour from location and movement data. *Ecological Informatics*, *49*(December), 69–76. https://doi.org/10.1016/j.ecoinf.2018.12.002

Wijeyakulasuriya, D. A., Eisenhauer, E. W., Shaby, B. A., & Hanks, E. M. (2020). Machine learning for modelling animal movement. *PLoS ONE*, *15*(7 July), 1–30. https://doi.org/10.1371/journal.pone.0235750

Wu, A., Buchanan, E. K., Whiteway, M., Schartner, M., Meijer, G., Noel, J.-P., Rodriguez, E., Everett, C., Norovich, A., Schaffer, E., Mishra, N., Salzman, C. D., Angelaki, D., Bendesky, A., Cunningham, J., & Paninski, L. (2020). Deep graph pose: A semi-supervised deep graphical model for improved animal pose tracking. *BioRxiv*, 2020.08.20.259705. http://biorxiv.org/content/early/2020/08/22/2020.08.20.259705.abstract

Yadav, S., & Bist, A. S. (2019). Residual nets for understanding animal behaviour. *Journal of Animal Behaviour and Biometeorology*, *7*(2), 97–103. https://doi.org/10.31893/2318-1265jabb.v7n2p97-103

Zhu, J., Rogers, T., Qian, R., & Kalish, C. (2009, October). Humans perform semi-supervised learning too. In *Proceedings of the National Conference on Artificial Intelligence* (Vol. 22, Issue 1, pp. 864–869.