Applied Mathematics and Computational Intelligence
Volume 13, No. 3, 2024 [186-219]

Batik Pattern Classification Using Machine Learning Approaches

Ida I'aza Toyib Hosen[1] and Saadi Ahmad Kamaruddin[2]

[1,2]UUM College of Arts and Sciences, School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Corresponding author : idaiaza98@gmail.com

Received: 22 April 2024
Revised: 28 July 2024
Accepted: 24 September 2024

ABSTRACT

Batik is one of Malaysia traditional textile art form that is well known. It can be difficult to accurately classify batik due to batik's wide range of motifs, colors, and complex patterns. Moreover, lack of thorough comparison research analyzing the outcomes of several machine learning algorithms restricted the understanding of the most appropriate and efficient methods for batik pattern classification. Most of previous research employed Artificial Neural Network (ANN), k-Nearest Neighbors (k-NN) and Decision Tree methods to classify batik patterns. By using these 3 methods, this study aims to classify Kelantan batik designs according to flora, fauna and geometry motifs. 133 images of the three categories were collected from social media and underwent pre-processing techniques. Image augmentation was done to enhance the diversity and quality of available training data for machine learning models. Digital transformation on the images based on colour features was developed to classify the three types of batik motifs. Image embedding was employed which returned a vector representation of each image in a data table in Orange software. The data divided into training and testing dataset, with a ratio of 8:2. The performance of each machine learning techniques were compared. The findings showed that ANN produced an excellent performance as indicated by classification accuracy (CA) of 87.9% for original images and 82.2% using Luminance Gradient image transformation technique. ANN also returns the highest Area under ROC Curve (AUC) score indicating it is the best in distinguishing the images motif. Results from confusion matrix showed that ANN have the least misclassification for batik classification. Based on the result, it signifies that ANN performs the best classification among the three approaches.

Keywords: Batik, Image classification, Artificial neural network, Image transformation, Classification accuracy

1 INTRODUCTION

Batik is a traditional textile art form characterized by the application of wax and dye onto fabric to create intricate and colorful designs. The term batik comes from the Old Javanese word "amba" and "tik", which means writing or painting dots. The charm of Nusantara batik lies not only in its beauty but also in its deep cultural significance. The detailed designs, created carefully by talented craftsmen, hold numerous stories, beliefs, and historical accounts specific to each area. They represent a blend of art, history, and identity, showcasing the richness of each region's heritage.

Centuries of cultural exchange have molded the development of Nusantara batik, integrating diverse influences from local customs, religious beliefs, and global trade networks. The vibrant mix of patterns, ranging from floral to geometric designs, demonstrates the blend of cultural elements interwoven into every fabric design. While the artistry of batik has thrived, the identification and classification of its diverse patterns have posed significant challenges. Traditional methods, dependent on human expertise and visual examination, encounter limitations in scalability, consistency, and accuracy when dealing with the extensive range of batik designs. In Malaysia, batik industry is dominated by two states, which are Kelantan and Terengganu. Amidst the digital age, machine learning, a branch of artificial intelligence (AI), presents a transformative opportunity to close the gap between traditional craftsmanship and modern technology. Through the utilization of machine learning algorithms, the classification of complex batik patterns will be more accurate and efficient. The goal of this study is to explore machine learning-based pattern recognition to classify diverse batik patterns, which address the complexities inherent in their motifs, colors, and designs.

Batik is highly valued for its elaborate patterns and cultural importance, but due to the wide range of motifs, colors, and complex patterns woven into its fabric, it can be difficult to accurately classify. This is supported by Minarno, Soesanti and Nugroho [1] that state amidst its diversity, the study of batik encounters various challenges, particularly in identifying and classifying for specific patterns. The wide and complex repertory of batik designs is too much for traditional procedures that depends on human ability and manual inspection [1]. To close the gap between traditional technique and modern innovations, a technical approach is required. Moreover, there is a significant research gap in the field since few thorough research has been done on the efficacy of different machine learning algorithms for classifying the batik motifs. The understanding of the most appropriate and efficient methods for batik pattern classification is restricted by the lack of thorough comparison research analyzing the outcomes of several machine learning algorithms. Furthermore, there is a considerable practical gap due to the batik images low resolution and quality. The efficiency of machine learning models is directly impacted by the dataset's lack of images quality. Consequently, these models' practical value may be restricted due to their poor effectiveness in real-world applications such as product verification or cultural preservation campaigns. Therefore, the mentioned shortcomings emphasize the urgent necessity of conducting a thorough study and developing techniques to fill these gaps.

The objectives of this study are (i) To identify the different motifs that can be used to classify the batik patterns, (ii) To develop classification models based on different motifs in (i), and (iii) To determine the best machine learning approach in classifying batik pattern based on (ii). This research focuses on the implementation and evaluation of machine learning methodology for the classification of Nusantara batik patterns, specifically Kelantan batik. The scope includes the acquisition and pre-processing of a comprehensive dataset of Kelantan batik patterns, encompassing a wide range of motifs, colors, and design variations. However, this study does not explore the physical production or historical analysis of batik but instead focuses explicitly on the application of machine learning techniques for pattern classification.

The findings of this study will redound to the benefit of economic, fashion industry and others. Firstly, since batik is frequently a source of revenue for artists and communities, it has economic relevance. As the study supports design innovation in industries by giving access to potential tools for designers that can create new patterns based on traditional iconic motifs, this is an acknowledgement of batik motif through providing a market and long-term sustainability to local economy and also encourage the revisit from wearied audience towards nurturing our rich traditions. Next, batik motifs are frequently used in modern clothing and design. Understanding the motifs of batik is crucial for designers who attempt to bridge the gap between traditional craftsmanship and modern design by taking inspiration from traditional patterns. Furthermore, tourists find Nusantara batik motifs appealing, and they can improve their trip experience. Gaining an appreciation and awareness of these patterns promotes a greater understanding of the region's many cultures by facilitating cultural exchange and interaction. Other than that, this study will also contribute to the benefit of education and research. Academic study on batik motifs can be conducted in numerous fields, including sociology, anthropology, art history, and cultural studies. Understanding such concepts is essential for educational reasons, as it contributes to the spread of information regarding the cultural legacy of the area. Besides, the application of machine learning algorithms to classify batik patterns is a significant advancement in the use of technology in the field of traditional art. The study advances our knowledge of machine learning methods while also applying them to cultural art. This study is an example of how tradition and modernity may coexist peacefully since it establishes a connection between traditional craft and modern technology. By showcasing how flexible ancient artistic techniques like batik are in modern technological environments, it promotes an appreciation for cultural variety and interdisciplinary cooperation.

2 LITERATURE REVIEW

2.1 History of Batik

Steelyana [2] in her study states that samples of dye resistance patterns on cloth can be traced back 1,500 years ago to Egypt and the Middle East, despite the fact scholars cannot agree about the exact origins of batik. Historical samples have also been discovered in West Africa, Turkey, India, China, and Japan. Steelyana [2] also mentioned that even though people in those nations used the dye-resisting decoration technique, none of them advanced batik to its present-day form as the highly developed intricate batik found on the island of Java in Indonesia.

Mohamad Akhir et al. [3] mentioned that the history of batik is recorded in the seventeenth-century Malay Annals, when Sultan Mahmud gave Laksamana Hang Nadim the command to go to India in order to get 140 pieces of sarasah cloth (batik), each of which featured 40 various types of flowers. Due to the challenge of meeting the sultan's standards, he made up on his own. Unfortunately, he only brought four pieces of batik to the sultan when his ship sank.

The history of batik industry in Malaysia is formed by the arrival of a group of Japanese immigrants who landed on the Malay Peninsula and established modest home production facilities in the early 20th century. They live on the East Coast, close to Kuala Terengganu and Kota Bharu [4]. Hence, it is believed that Malaysian batik originated in the states of Kelantan and Terengganu and dominated by these two states in Malaysia.

2.1.1 Nusantara Batik

Evers [5] found that the term "Nusantara" originates from two Sanskrit words: "nusa" signifying 'island' and "antara" conveying 'in between' or 'including'. It entered ancient Javanese texts, where "nūsāntara" was interpreted as 'other islands'. The term batik comes from the Old Javanese word “amba” and “tik”, which means writing or painting dots. However, there are also those who say that the batik is derived from the word “mbatik” which means making a point [6]. In the simplest sense, Akhir et al. [3] defined batik as an artistic work of decoration that is influenced by the batik artist. Batik is defined as "a process of dyeing fabric by using a resist technique, covering areas of cloth with a dye-resistant substance to prevent them from absorbing colors." This is the deep definition of the batik. Hence, Nusantara batik is an ancient textile art form, holds deep cultural and historical significance in the regions of Indonesia, Malaysia, and surrounding areas.

2.1.2 Batik Development

Over centuries, batik has evolved, reflecting the diverse cultural influences from trade, religious practices, and indigenous traditions across the archipelago. Batik patterns were not merely decorative, but also dissolves into moral principles, customs, taboos and symbols [7]. The evolution of Nusantara batik patterns exhibits an exquisite fusion of artistic elements, incorporating motifs inspired by nature, mythology, religion, and daily life. Distinct regional styles emerged, each with its unique characteristics and symbolism. For instance, the motif found in Gresik batik differ markedly from those in Lamongan batik, showcasing the richness and diversity of this art form [7]. The colors, designs, and techniques employed in batik production vary significantly, reflecting the cultural nuances of different communities.

In recent decades, efforts to preserve and promote batik in Nusantara have gained momentum, leading to its recognition as a United Nations Educational, Scientific and Cultural Organization (UNESCO) Intangible Cultural Heritage of Humanity [8]. Contemporary batik artisans have been continuously exploring various creative processes since the establishment of batik's early origins [9]. In Malaysia, batik is not only being used for cultural icon, but for formal and non-formal wear, contemporary fashion and soft furnishings.

2.1.3 Kelantan Batik Motifs

Known as a "cultural pot of Malay culture," Kelantan state is in Northeast region in Malaysia. In addition to being distinctive, Kelantan's tangible and intangible legacy is renowned for its realistic, traditional Malay conception of art. Many aspects of Malaysian culture are said to have originated in Kelantan; even if they were subsequently identified as belonging to the Malay heritage, Kelantan art is exceptional and cannot be found anywhere else in Malaysia [10]. According to Jahangiri et al. [4], the two east coast states of Malaysia, Terengganu and Kelantan, are the most well-known for producing batik and are considered the beginnings of the batik industry. Subsequently, industry expands to other regions, especially the Peninsular Malaysia.

In the beginning, the production of the motifs was more focused on flora motifs than fauna and geometric motifs. In the early stages, batik making in Kelantan was strongly influenced especially in terms of motifs used by North Javanese batik from Indonesia. However, as time went on only a few motifs from Java were shown. Motif designs that depict elements of animals and mythical creatures will be eliminated. This is due to the Malay community at that time preferring to use motifs suitable for Muslim Malays as well as Islamic practices. In Malaysia, Islam is the main religion and Malays are not given permission by the Muslim community to wear motifs that have elements of wild animals or mythical creatures. Most batik motif designs in Kelantan are based on nature such as plants or flowers [11]. Nawi [11] in his study stated that Kelantan batik motif elements are divided into two, namely flora motifs and geometric motifs. The flora motifs include rice flower, star anise, hibiscus and others while geometric motifs include bamboo shoot.

2.2 Machine Learning in Pattern Classification

Before the existence of machine learning, the classification of Batik patterns has heavily relied upon human expertise and manual examination by artisans and experts in the field. Since batik has a diverse range of motifs, colors and designs, the subjective approach of manual examination poses difficulties in terms of accuracy, scalability, and consistency. Over time, machine learning techniques have transformed the field of pattern recognition by providing advanced methods for classifying complex patterns across several domains. There are several machine learning methods used in previous study for various industries. The study of technical solutions for pattern classification has been prompted by the shortcomings of manual classification techniques.

2.2.1 k-Nearest Neighbors (k-NN)

Early techniques such as the k-NN is a non-parametric classification method, which is simple but effective in many cases [12]. There are some shortcomings with k-NN technique, in which this technique is low deficiency and dependency on k, hence Guo et al. [12] in their research presented a novel solution for dealing with the problems by using Clementine software. Guo et al. [12] chose a subset of representatives from the training dataset along with supplementary information to effectively represent the entire dataset. To remove the reliance on k without requiring user's intervention, the optimum but distinct k determined by the dataset itself was used when choosing each representative. Guo et al. [12] concluded that the experimental results conducted on six public datasets demonstrate that the k-NN model is a highly competitive method for classification. Its average classification accuracy across these six datasets is comparable to that of C5.0 and k-NN. Furthermore, the k-NN model notably reduces the number of data tuples in the final classification model, achieving an average reduction rate of 90.41%.

Research of batik classification but focusing on Batik Lampung motifs has been done by using k-NN method [13]. This study focused on the ability of a system to classify Lampung Batik motif using K-Nearest Neighbor method. Total data collected are 25 images for each motif. The known motifs of Batik Lampung, including Jung Agung, Siger Kembang Cengkih, Siger Ratu Agung, and Sembagi, have original image samples stored in RGB format. Initially, these images are resized to 50 x 50 pixels and then converted into grayscale. To facilitate recognition, the Gray Level Co-Occurrence Matrix (GLCM) feature is extracted. Subsequently, the k-NN algorithm is employed for classification, using various values of k (3, 5, 7, 9, 11) and orientation angles (0°, 45°, 90°, 135°) to classify these motifs. Andrian et al. [13] concluded that k-Nearest Neighbor classification method has been successfully implemented in the process of pattern recognition of Lampung's Batik, with the highest accuracy of 97.96% obtained on testing at the orientation angle of 135° at k = 7 and lowest level accuracy of 80.49% in testing at the orientation angle of 45° with a value of k = 9.

2.2.2 Decision Tree

With the advent of more advanced algorithms, many classification algorithms have been proposed in the machine learning and data science literature [14]. For instance, neural networks and decision tree has significantly enhanced the accuracy and scalability of pattern recognition systems. In [15], Yoo et al. conducted a study to examine the feasibility of identifying COVID-19 from CXR pictures using a deep learning-based decision-tree classifier. The proposed classifier was made up of three binary decision trees, each trained by a deep learning model using a convolution neural network based on the PyTorch framework. The CXR images were categorized as normal or abnormal using the first decision tree. While the third tree did the same for COVID-19, the second tree recognized the abnormal pictures that have symptoms of tuberculosis. They concluded that the average accuracy of the third decision tree is 95%, whereas the accuracies of the first and second choice trees are 98 and 80%, respectively. Prior to the availability of RT-PCR findings, patients may be pre-screened using the suggested deep learning-based decision-tree classifier to accelerate decision-making and perform triage [15].

2.2.3 Artificial Neural Networks (ANN)

In 2020, Kasim, Bakri & Septiarini [16] in their study utilized artificial neural networks (ANN) to distinguish batik motifs and non-batik fabric motifs. The collection of batik and non-batik images, their pre-transformation into grayscale forms, the extraction of texture features from grayscale images, and the use of artificial neural networks to recognize patterns are some of the crucial stages that must be taken. The process of acquiring photos involves gathering batik and non-batik images, from several motifs. 30% of processed data sets are used for testing, while 70% are used for training. By comparing the Levenberg-Marquardt algorithm (trainlm) and the Scaled conjugate gradient algorithm (trainscg) training methods, the artificial neural network models employed in this study utilize the Backpropagation learning process. With an accuracy value of 84.12%, the Scaled conjugate gradient algorithm (trainscg) training technique produced better accuracy results than the Levenberg-Marquardt algorithm (trainlm) approach by 86.11% [16].

2.3 Image Transformation

Image transformation refers to the process of changing the image's properties using various methods. In image processing, computer vision and machine learning applications, these transformations are frequently used to improve images, extract features, or prepare data for analysis. Typical image alterations include scaling, brightness adjustments, rotation and sharpening.

Ibrahim et al. [17] conducted an examination on the contrastive analysis of rice grain classification performance between multi-class support vector machine (SVM) and artificial neural network (ANN). The analysis was conducted on three types of rice grain images: Ponni, Basmati, and Brown rice. A digital image transformation analysis focusing on shape and color features was developed to classify these rice grain types. The performance of the proposed method was evaluated using 90 testing images for each rice variation. In this research, three different preprocessing techniques were employed, namely Level Sweep with sweep: 1, width: 110, opacity: 1, Gradient Magnitude with max: 94.73, min: 0, mean: 3.41 and Hysteresis Threshold with the parameters max: 255, min: 0, mean: 3.97, standard deviation: 31.58. The ANN demonstrated superior classification accuracy, reaching

93.34% when employing the Level Sweep image transformation technique. These results show that the morphological features of both shape and color are adequate for utilization as input variables.

In 2008, Ruppertsberg et al. [18] evaluated human sensitivity to variations in color and luminance gradients resulting from shifts in the light source position within a complex scene. They examined viewers' ability to distinguish between gradients caused by varying light source positions in Experiment 1. When the light source location deviated from the reference scene by a maximum of 4 degrees, it was discovered that observers could correctly identify a change in the gradient information. The gradient's brightness data served as the primary foundation for this sensitivity (Experiments 2 and 3). When attempting to distinguish between gradients, some observers rely on the spatial distribution of chromaticity and brightness values within them (Experiment 4). The idea that gradients hold information that could help with the recovery of 3D shape and scene configuration features is supported by the high sensitivity to gradient changes.

Mudrova & Procházka [19] used two different ways that Principal Component Analysis (PCA) is used in image processing. In the first application, the three colour components of the image are reduced to one that contains the majority of the information. The second application of PCA determines the chosen object orientation by utilizing the attributes of eigenvectors. There are several techniques for detecting prior objects. The quality of picture segmentation also affects the outcomes of the subsequent PCA-based object orientation evaluation procedure.

2.4 Summary

The cultural importance and artistic value embedded in Batik patterns necessitate specialized approaches for classification. Previous research shows that various machine learning approaches have been done and most of the methods gave a satisfying result in classification. From the literature review, three methods will be considered to undergo the analysis, which are k-NN, decision tree and ANN because most of previous research used these three methods to classify batik patterns. While previous research establishes a basis for automated Batik pattern classification, it has difficulties in managing the great variety and complexities present in Batik patterns. Objective of this research is to contribute to the convergence of technology and traditional artistry in a culturally mindful and influential manner. Besides, very few of the research focuses on classification of Kelantan batik patterns while the batik industry dominated by this state. Hence, this study will be concentrated on classifying Kelantan batik patterns. Moreover, image transformation will be applied to the Kelantan batik images with three techniques, which are Level Sweep, Luminance Gradient and Principal Component Analysis.

3 METHODOLOGY

The understanding and acceptance of information from past studies related to batik and application of machine learning in pattern classification have been seen in the literature review. In order to achieve the objectives, related methodology research techniques were implemented. In this section, explanations on the related research method were discussed.

3.1 Data Collection

This study focuses on solving specific practical problems by applying existing knowledge or methodologies. Aside from that, it focuses on applying theories, principles, and methodologies to real-world situations or issues to develop practical solutions. Hence, this study is applied research study.

In this study, secondary data were used to do the analysis. The data consists of high-resolution images of Kelantan batik motifs. The Kelantan batik pattern images were taken from social media sites, specifically Facebook. Permission asked from the pages or group admins to use the batik images in their sites for research purpose.

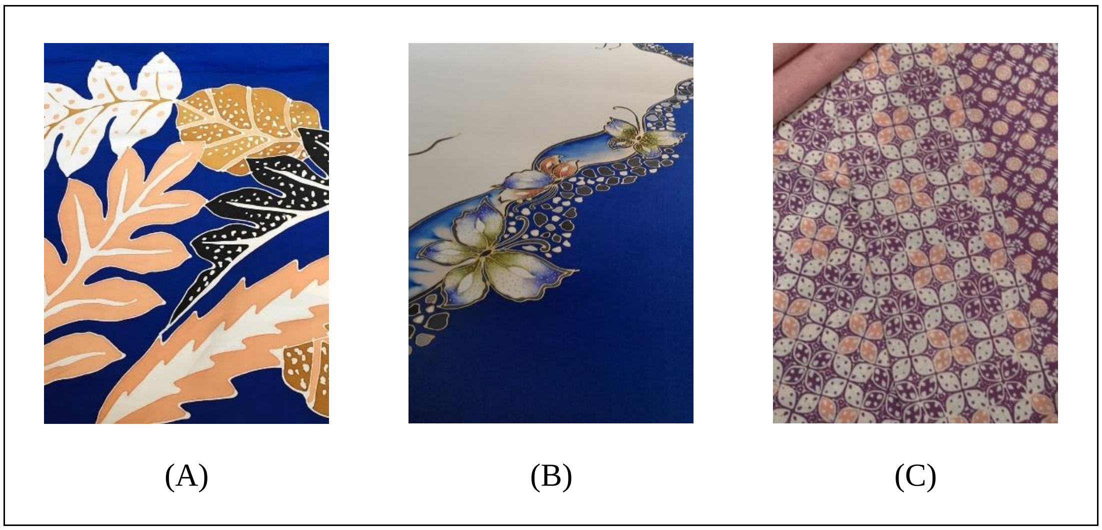


Figure 1 : Type of Batik Images (A) Flora, (B) Fauna and (C) Geometry

Examples of the Kelantan batik images taken from social media can be viewed in Figure 1. A total of 133 Kelantan batik images have been used for analysis in this study, where 60 for flora motif, 15 for fauna motif and another 58 is geometry motif.

3.2 Research Flowchart

This study consists of three main parts which are literature review, data pre-processing and machine learning (ML) model development. The first part provides information on the history of batik, type of machine learning in pattern classification and image transformation.

The second part explains the data pre-processing method, which are divided into three: image augmentation, image transformation and image embedding. The pre-processed data is then split into 8:2 ratio that aims to train the model to improve the pattern classification level.

In the last part of the study, the ML model is developed. The ML model includes ANN, k-NN and decision tree. The ML models then undergo hyperparameter tuning that involves the adjustment of a model's parameters to attain optimal performance. Then, test and compare the models by evaluation technique. From the result, the best model is selected.

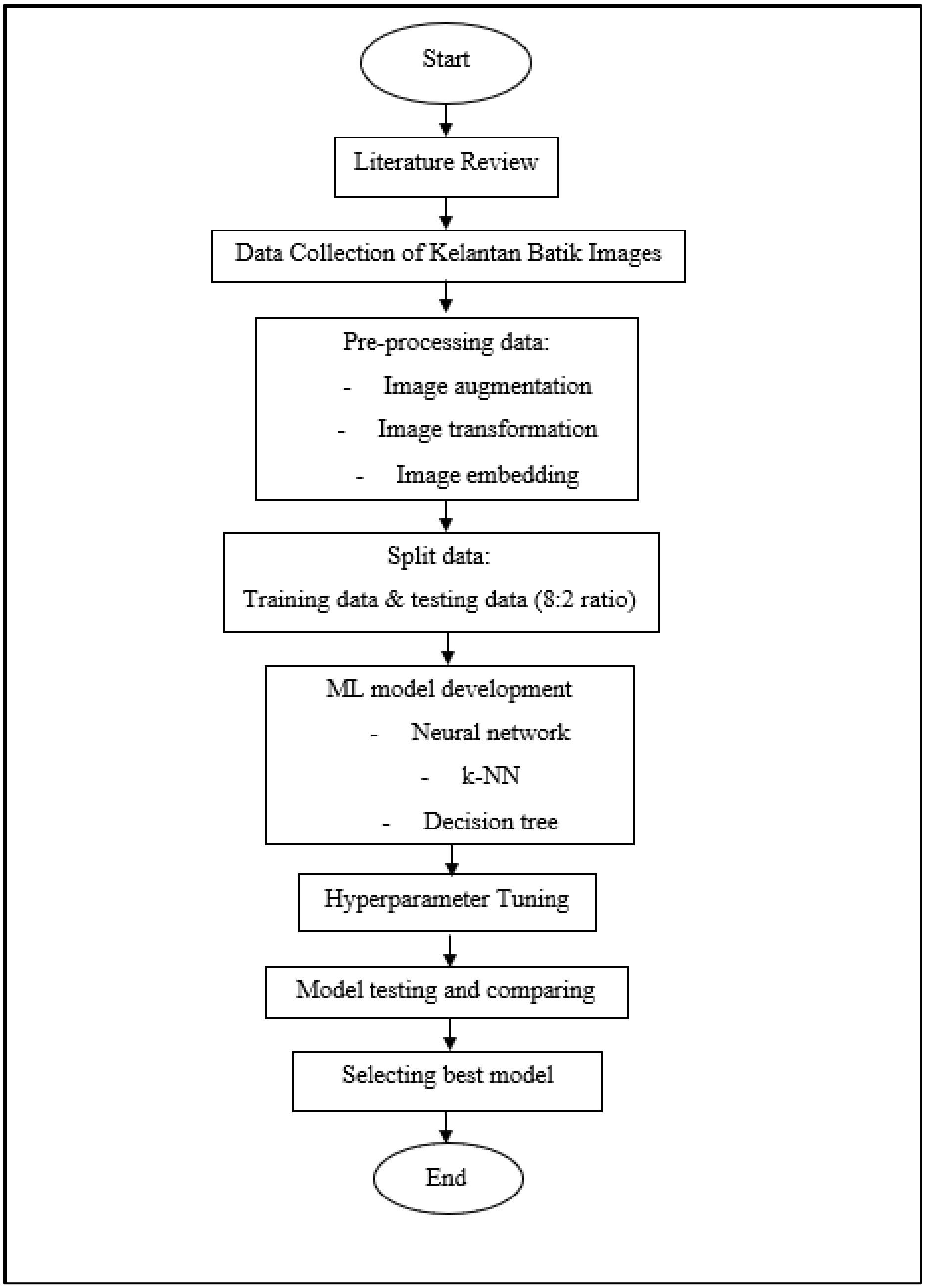


Figure 2 : Flowchart of Batik pattern classification

3.3 Machine Learning Approaches

Few methods were used in this study, which are k-NN, decision tree and ANN. The k-NN algorithm is a generalized approach for nearest neighbor rules in machine learning. Its inductive inference involves assigning the class label of the k-sample, which is most similar to the one being tested. Unlike the nearest neighbor, the k-NN algorithm expands beyond a single nearest neighbor to consider 'k' neighbors during the decision-making phase. This extension allows the k-NN algorithm to access and use more data. Unlike other classification algorithms with separate training phases, it omits the learning processing step [20].

The second method, which is decision tree is one of the most popular approaches for representing classifiers in data classification. Decision tree classifiers have been suggested for usage in many different disciplines, including medical illness analysis, text categorization, user smartphone classification, images, and many more. Decision tree is a sequential model that effectively and cohesively combines a number of fundamental tests in which a numerical feature is compared to a threshold value in every test. Compared to the numerical weights in the neural network of connections between nodes, the conceptual principles are far simpler to create. Decision tree is used mostly for grouping reasons. Additionally, decision tree is a classification model that is frequently used in data mining. The nodes and branches are composed of each tree. Every subset specifies a value that may be accepted by the node, and each node represents features in a category that needs to be categorized [21].

The third method is artificial neural network (ANN), which is considered a very important part in the new industry of artificial intelligence. An artificial neuron is a computational model inspired by natural neurons [22]. A neural network is a computational model comprised of several interconnected nodes, sometimes known as neurons. A distinct output function known as the activation function is represented by each node. The artificial neural network's memory is represented by the connection between every two nodes, which serves as a weight for the signal that passes through the connection. The network's weight value, incentive function, and network connectivity will all affect the network's output [23]. Wu & Feng [23] states that neuron processing unit in an artificial neural network can represent a variety of objects, including features, letters, concepts, or a meaningful abstraction pattern. The three types of processing units in a network are an input unit, an output unit, and a hidden unit. The core of an artificial neural network is the network transformation and dynamic behavior of a parallel distributed information processing functions, and in different degrees and levels of imitation of people's information processing of the brain and nervous system. Artificial neural networks are non-programme, adaptive, brainstyle information processors.

For this study, Orange software was used to analyze the data in order to achieve the objectives of the study. Orange is a great software to illustrate the network of data analysis and it consists of variety of tools that can extract features from data and do the hyperparameter tuning.

3.4 Method of Data Analysis

In this study, objective 1 was achieved by going through the literature review. Then, the data collection process was the start of achieving objective 2. After collecting the secondary data from various sources, the data underwent pre-processing, which is a crucial initial step that encompasses various procedures aimed at cleansing, normalizing, and improving the dataset containing the batik patterns. This process ensures that the data is correctly formatted and optimized, preparing it for effective training of machine learning models. Data pre-processing consists of image augmentation, image transformation techniques and image embedding. Throughout the pre-processing phase, ethical considerations hold utmost importance to ensure the respectful and responsible utilization of cultural artifacts.

Within this study, image augmentation refers to the deliberate manipulation and expansion of the dataset containing batik pattern images to enhance the diversity and quality of available training data for machine learning models. Image augmentation techniques involve applying various transformations to the batik images, such as rotation, cropping and brightness adjustment to create modified versions while retaining the essential visual characteristics of the patterns. The rotation and cropping of images were performed by using Photos application. In this study, the ratio for the images width to length is 3:4. Three different techniques of pre-processing were employed in https://29a.ch/photo-forensics/#pca as follows:

i) Level Sweep - Sweep: 1, Width: 110, Opacity: 1.

ii) Luminance Gradient - Normalize, Intensity: 1, Opacity: 0.5.

iii) Principal Component Analysis (PCA) - Input: Color, Mode: Projection, Component: 1, Enhancement: Equalize Histogram, Opacity: 1.

These techniques were implemented to check the best pre-processing technique for the batik pattern dataset. Figure 3(A) illustrates the original image while Figure 3(B-D) are the sample outcome of the image transformation for Level Sweep, Luminance Gradient and PCA.

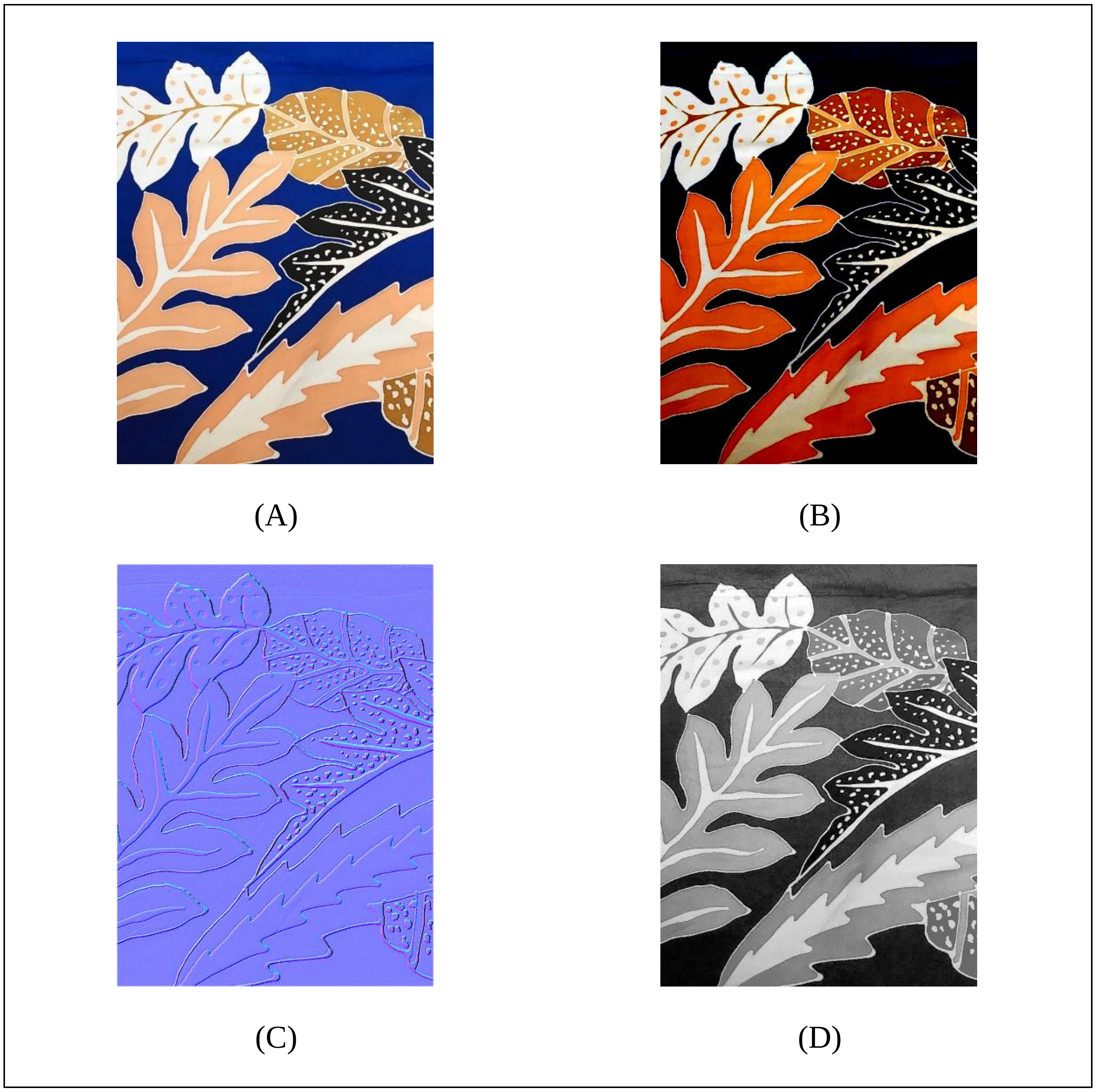


Figure 3 : (A) Original Image of Kelantan Batik Pattern. Outcomes of Images Transformation: (B) Level Sweep, (C) Luminance Gradient and (D) Principal Component Analysis

Next, the data were imported into Orange and image embedding was implemented on the data. Image embedding is a reduced-dimensional representation of the image. It is a detailed vector representation of the image that may be applied to a variety of tasks, including classification [24]. This will make the processing of the semantic and visual characteristics of visual data by machine learning models easier. From image embedding, we were able to get 2048 additional features that describes the content of the images.

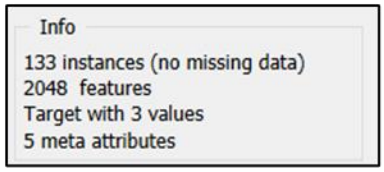


Figure 4 : Info of Data Table Widget in Orange

The data then were split into training and test data with a ratio of 8:2 [25] following the image embedding process. The purpose of splitting the secondary data is to train the model to improve the pattern classification level. The splitting of data was done in Orange after importing the data by setting 80% for fixed proportion of data in Data Sampler widget as shown in Figure 5. This returns the desired percentage from the entire data.

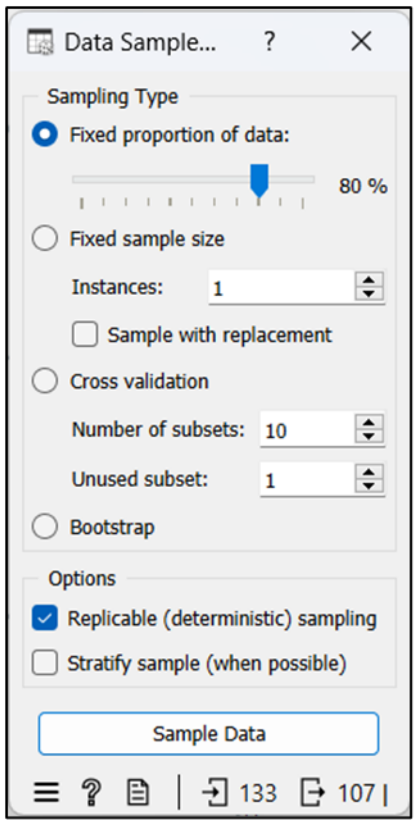


Figure 5 : Data Sampler Widget in Orange

Then, machine learning models were created and underwent hyperparameter tuning. Hyperparameter tuning involves the adjustment of a model's parameters to attain optimal performance. This process is crucial for enhancing the accuracy and efficiency of any machine learning models. In this study, parameters such as number of neurons and number of maximum iterations will be adjusted. Next, test the machine learning models and finally compare the three models by evaluation techniques such as the Area under ROC Curve (AUC), Classification Accuracy

(CA), Recall and F1 score. AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1).

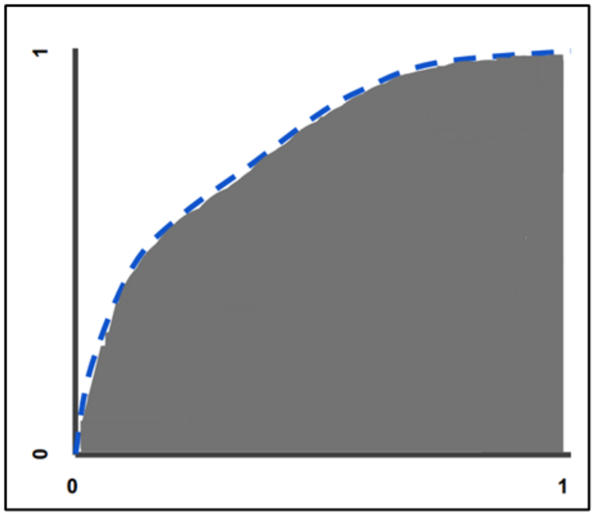


Figure 6 : Area Under ROC Curve

The formula for the CA is shown below:

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

where

$TP = True\ Positive$
$TN = True\ Negative$
$FP = False\ Positive$
$FN = False\ Negative$

Then, confusion matrix has also been generated to allow more detailed analysis, which visualizes the performance of the algorithms. Finally, comparison from the results for each algorithm was made to find out which performs best.

4 DATA ANALYSIS AND RESULT

The most crucial part of doing research is data analysis. In this section, the research findings will be discussed. This section will provide more details on the comparison between the performance of the original images and the images transformed. In addition, the performance for each machine learning algorithm can be seen within this section. The intended and targeted outcomes have been achieved via a series of trials.

4.1 Different Motifs to Classify Batik Pattern

Nawi [11] mentioned about 3 types of batik motifs, which are flora, fauna and geometry. These 3 types of motifs can be found during the data collection. The type of motifs was used as an indicator to decipher the batik images according to the pattern. Hence, flora, fauna and geometry are the different elements that can be used to classify the batik images in Orange.

4.2 Performance of Machine Learning Algorithms

There are 133 of original data collected that have undergone the pre-processing with 3 techniques, Level Sweep, Luminance Gradient and Principal Component Analysis, which make a total of 532 data. However, each type of image was tested separately to check which type of image performs highest accuracy.

Machine learning algorithms used are ANN, k-NN and decision tree. The evaluation of the 3 methods can be checked by classifier performance metrics which include AUC and CA.

AUC, which stands for area under ROC curve, provides an overall performance metric across all potential classification criteria. The likelihood that an index test will correctly classify a randomly chosen subject from a sample is shown by the AUC value [26]. The higher the AUC, the better the model is at distinguishing the classification.

Classification accuracy (CA) tells us the percentage of correct prediction by a model. CA is a commonly utilized and uncomplicated measure to convey theDD quality of the complete categorization is the percentage of cases accurately assigned [27]. This can be computed using:

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

where

$TP = True\ Positive$
$TN = True\ Negative$
$FP = False\ Positive$
$FN = False\ Negative$

Recall measures the classifier's accuracy in properly identifying positive classes by labelling them as such when they are in fact positive. Furthermore, recall falls short of providing all the details regarding how accurately the positive instances were classified. The probability of properly guessing a positive case is expressed as recall [27]. Recall can be calculated from:

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

Meanwhile, precision is a different measure to recall that concentrates on the positive cases [27]. Precision is defined as the ratio of true positives (TP) to the total number of positives that a model predicts [28]. Precision is computed as follows:

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

Foody [27] mentioned that F1 is another commonly used statistic that basically combines the data from two of the fundamental metrics. To be more precise, the F1 metric is the harmonic mean of precision and recall. When neither precision nor recall are zero, the magnitude of F1 is 0. Otherwise, it is 1.0 when both precision and recall show a perfect classification. Below is the formula for F1:

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} = \frac{2TP}{2TP + FN + FP}$$

The correlation between the genuine classes and the predicted labels is measured by the Matthews correlation coefficient (MCC) [29]. Chicco et al. [29] stated that Matthews correlation coefficient produce a high score only when the classifier accurately predicts the majority of positive and negative data class, as well as when the majority of its positive and negative predictions are correct. The formula for MCC is:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

Table 1 shows the evaluation results for all types of images. The table gives sight on the AUC, CA, F1, Prec, Recall and MCC.

Table 1 : Evaluation Results for All Model

Image Type	**Model**	**AUC**	**CA**	**F1**	**Prec**	**Recall**	**MCC**
Original	ANN	0.951	0.879	0.877	0.877	0.879	0.793
	kNN	0.876	0.701	0.676	0.816	0.701	0.541
	Tree	0.756	0.738	0.723	0.711	0.738	0.547
Level Sweep	ANN	0.944	0.822	0.817	0.822	0.822	0.700
	kNN	0.868	0.682	0.647	0.775	0.682	0.503
	Tree	0.743	0.71	0.709	0.714	0.710	0.508
Luminance Gradient	ANN	0.946	0.822	0.820	0.823	0.822	0.698
	kNN	0.891	0.804	0.801	0.819	0.804	0.674
	Tree	0.750	0.720	0.715	0.715	0.720	0.518
Principal Component Analysis	ANN	0.930	0.804	0.791	0.785	0.804	0.661
	kNN	0.842	0.701	0.677	0.743	0.701	0.512
	Tree	0.755	0.692	0.690	0.690	0.692	0.480

From Table 1, AUC of ANN is the highest for all type of images with values higher than 0.93, followed by k-NN and decision tree. This shows that ANN is the best model in distinguishing classification out of the 3 models used in the analysis. The CA for ANN is also the highest with values ranging from 0.80 to 0.88. Therefore, we can say that neural network gives the best accuracy for batik pattern images classification.

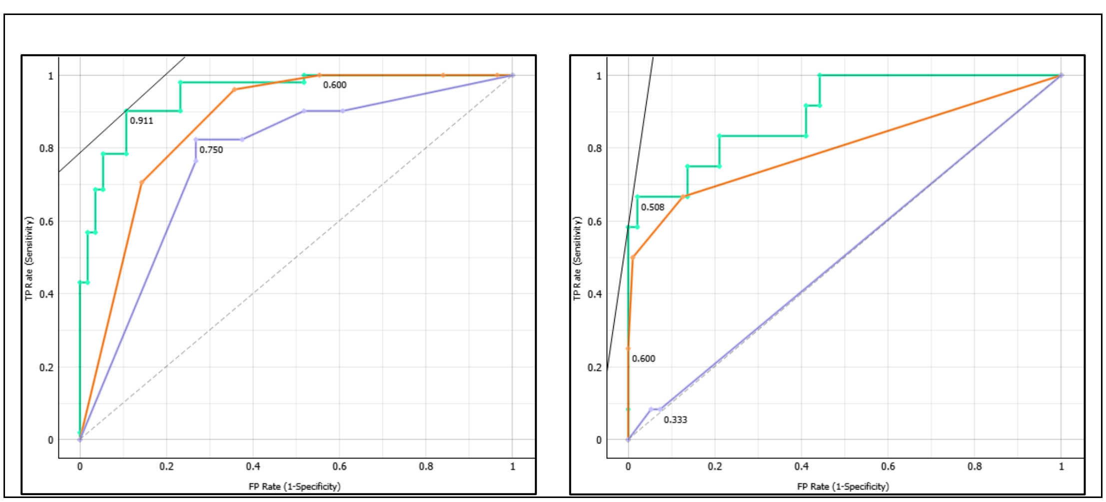
ANN also returns the highest precision values. From luminance gradient images, the highest precision value is 0.823, which means that when it predicts an image is a flora motif, it is correct around 82.3% of the time. From the precision results, the decision tree returns the lowest value ranging from 0.69 – 0.715 across all images type. Referring to level sweep images, it has the lowest recall value by k-NN method, followed by decision tree and neural network. In this case, recall tells us that 68.2% of the images are correctly identified as the correct motifs by k-NN classifier.

The high score of F1 demonstrates a well-balanced performance of the model. The highest F1 scores for original, level sweep, luminance gradient and PCA are 0.877, 0.822, 0.822 and 0.804 respectively. A high score of F1 can concurrently accomplish high recall and high precision. Except for luminance gradient images, the lowest F1 scores come from k-NN model, while lowest F1 score of luminance gradient is from decision tree.

Across the MCC values, the values mostly range around 0.48 – 0.55 which is quite an average value. This means that most of the model accurately predicts most of the positive and negative data class, as well as its positive and negative predictions are correct around 48% to 55% only. Original images data maintain a higher MCC of 79.3% by ANN. Overall, we can say that ANN gives the best performance for batik pattern images classification. This is consistent with results from [30] that concluded ANN classifier performs better than decision tree and [31] that concluded ANN performance is better than k-NN in classification.

The overall result from Table 1 reflects that original photo is the best images type to be used for classification. However, the other 3 pre-processing techniques could not be underestimated as the results are not much different from the original images. Out of the 3 pre-processing techniques, luminance gradient is the best pre-processing technique for batik dataset classification.

Following the evaluation results in Table 1, the ROC curve is observed. ROC curve or receiver operating characteristic curve is a graphical representation used to illustrate the diagnostic performance of binary classifiers. The closer the ROC curve is to the upper left corner of the graph, the higher the accuracy of the test. The upper left corner of the ROC curve represents perfect classification, where sensitivity is 1 (true positive rate) and the false positive rate is 0 (specificity is 1). Consequently, the ideal ROC curve indeed has an Area Under the Curve (AUC) value of 1.0 [32]. Since the original images show the highest AUC for classification, it will be used to observe the analysis of ROC curve for each category.



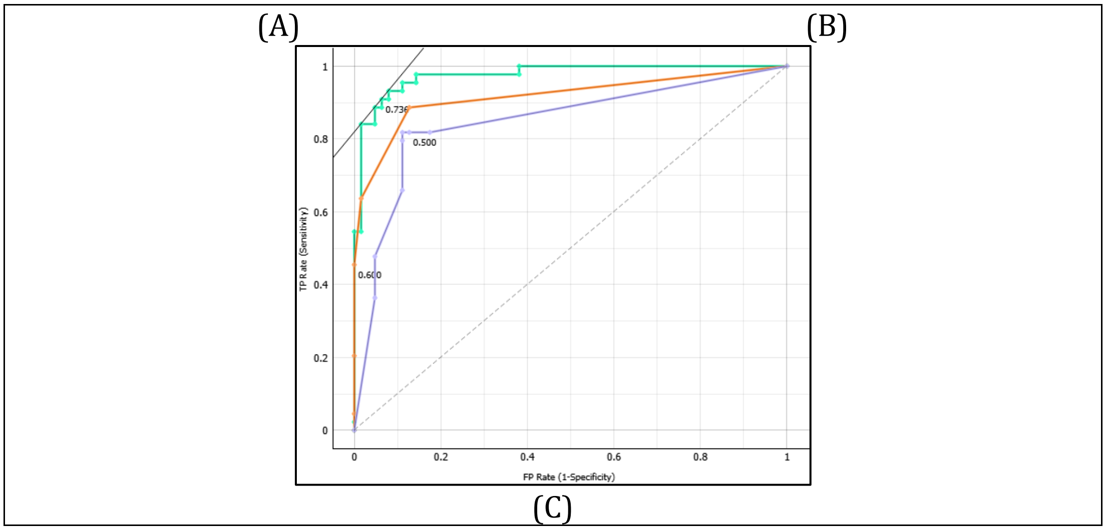


Figure 7 : ROC Curve of Original Image for Category (A) Flora, (B) Fauna and (C) Geometry

Figure 7 shows the ROC curve of all categories. The green line in the figure indicates ANN, orange for k-NN and purple line is decision tree method. From the figure, the ROC curve for ANN is the closest to the top left corner for each category, followed by k-NN and decision tree. This proves that ANN has the best performance in batik image classification. The ROC curve in geometry motif is the closest to the top left corner, which indicates all models performed the best in classifying geometry images. The ROC curve for fauna category is further from the top left corner which suggests all models have the least accuracy for fauna category.

Batik pattern classification is better performed by ANN than k-NN and decision tree for a number of significant reasons. ANNs are very good at learning robust features from raw image data simply because they are capable of extracting complex ones automatically. As for the batik patterns, they are characterized with detail-oriented and multi-layered designs, which can only be recognized if ANN is used to extract complex features from them. In contrast to this fact, decision tree and k-NN take advantage of pre-defined criteria in splitting data, which may not be effective in capturing the fine details as well as subtleties that exist in batik designs. Data with non-linear relationships are captured well by ANN because of their multilayered structure and nonlinear activation functions. On the other hand, batik patterns often have intricate non-linear attributes which can be effectively modeled by ANN unlike k-NN and decision tree that face a challenge of such complexities. ANNs are thus very amenable to capturing the distinctiveness of batik patterns. They can be customized through different architectures and optimization strategies. This elasticity is important because it can handle varied image data and complicated pattern recognition tasks that cannot be achieved by k-NN and decision tree. ANNs also have higher ability to withstand noise and image quality variations which are usually experienced in practical scenarios where batik images are concerned.

4.3 Misclassification of Data

Misclassification of data can be detected from confusion matrix. For an n-class classification problem, the confusion matrix is a square matrix with n rows and n columns. Each row represents the actual samples (instances) belonging to a particular class, which serve as inputs to the classifier. Similarly, each column represents the predicted samples, which are the outputs of the classifier, categorized into the respective classes [33]. Table 2 displays the scheme of confusion matrix for binary classification.

Table 2 : Confusion Matrix for Binary Classification

		Predicted	
		Positive	Negative
Actual	Positive	TP	FP
	Negative	FN	TN

Table 3 : Confusion Matrix of Original Images by ANN

		Predicted			
		Flora	Fauna	Geometry	∑
Actual	Flora	8	4	0	12
	Fauna	1	46	4	51
	Geometry	1	3	40	44
	∑	10	53	44	107

Table 4 : Confusion Matrix of Original Images by k-NN

		Predicted			
		Flora	Fauna	Geometry	∑
Actual	Flora	4	8	0	12
	Fauna	0	51	0	51
	Geometry	0	24	20	44
	∑	4	83	20	107

Table 5 : Confusion Matrix of Original Images by Decision Tree

		Predicted			
		Flora	Fauna	Geometry	∑
Actual	Flora	1	8	3	12
	Fauna	5	42	4	51
	Geometry	1	7	36	44
	∑	7	57	43	107

Table 3, Table 4 and Table 5 display the confusion matrix for ANN, k-NN and decision tree methods respectively. From Table 3, it can be observed that not many of the data is misclassified. There are 4 fauna data that is misclassified for flora and 4 flora data that misclassified as geometry, which is the highest value of misclassification for this method.

From Table 4, by k-NN method, even though there's not much misclassification, but a lot of geometry data misclassified as flora, with value of 24. The confusion matrix of decision tree in Table 5 justifies that 8 of fauna data are misclassified as flora and 3 misclassified as geometry. 7 images are misclassified as flora while it belongs to geometry category. Overall, ANN shows the least misclassification in the batik images classification. This is in line with conclusions from [34] and [35] where they stated that ANN performs the best accuracy and least misclassification against the other classification methods.

5 CONCLUSION

As a conclusion, the objectives have been magnificently achieved. The three different motifs used to classify batik patterns are flora, fauna and geometry. The application of ANN, k-NN and decision tree for batik classification to a variety of testing images has also been successful. With the high classification accuracies of more than 0.80 and AUC values of higher than 0.93, ANN is proven to be the best approach for the batik patterns classification on different types of images. ANN could be used as a tool to achieve more precise assessment of the batik patterns classification in sorting and cataloging batik designs in future, making it easier for researchers to study this traditional art form.

The study is of importance in so far as it can lead to the drastic changes in batik industry. The use of ANN makes it possible to authenticate batik products, thus serving as trusted means for distinguishing real from counterfeit batiks. This verification is significant because it helps safeguard batik industry's reputation and guarantee fair reward and recognition system for craftsmen involved in batik production. Furthermore, high precision rates of ANN when identifying patterns on a given sample have opened up possibilities for innovations within this sphere. Such models can be used by designers when composing new designs that are associated with traditional motifs; thus, they will potentially attract many customers and stimulate the growth of market demand. More importantly, even though there may be variations in images' quality, these networks are robust enough to serve practical purposes in cases where the input data differs significantly or even contains noise.

There are few recommendations that emerge after conducting this study. First, it is recommended to try using other machine learning model to classify the batik images pattern. This could involve experimenting with deep learning architectures, such as convolutional neural networks (CNNs) as suggested by [25] and [36] that outperform traditional methods due to their ability to learn and select features automatically and efficiently.

Moreover, it is suggested to increase the scope of study, not only in one state so that enough dataset can be collected, namely, at least 100 images per category. Sufficient datasets are required in order for the machine learning to learn and improve classification accuracy. Furthermore, future research may explore additional categories or refine the classification process further to enhance the accuracy of the batik patterns classification. In a nutshell, by embracing these recommendations, we can advance the batik classification and unlock new opportunities for cultural appreciation in the digital age.

5.1 Limitations of the Study

The study on classifying Kelantan batik motifs using machine learning approaches focusing on flora, fauna and geometry patterns has some limitations. First, the availability of a limited dataset of Kelantan batik patterns may restrict the diversity of the patterns captured. Since Kelantan batik focuses on flora motifs and not many of the motifs were based on living creatures, hence only few of fauna motifs are found. This may cause biases in the classification model, wherein the model may perform well on the majority class while exhibiting poorer performance on minority class.

Next, variability in image quality and resolution within the dataset poses challenges in standardizing preprocessing techniques, which may affect the performance of the model. Models trained on datasets with significant variability in image quality may exhibit bias towards certain quality levels.

Moreover, classifying batik motifs into categories involves subjectivity. Different individuals may interpret differently whether a motif is flora, fauna or geometry. This limitation may affect the consistency of classification results.

In summary, while this study aims to classify the batik pattern through machine learning approaches, it is essential to address the limitations inherent in the dataset and external factors that may impact the models' performance. By acknowledging these limitations, we will be able to achieve continuous improvement for future research.

ACKNOWLEDGEMENT

Numerous people have been contributing to the success of this project. There is an abundance of appreciation for what they've done.

First and foremost, I would like to thank Allah for letting me through all difficulties during the stages of this project. I would like to acknowledge and give my warmest thanks to my supervisor, Dr. Saadi Bin Ahmad Kamaruddin for his guidance and advice that carried me through all the stages of writing this project. Without his support, this work will not be possible.

Not forgetting School of Quantitative Sciences, Universiti Utara Malaysia for giving me an opportunity to do this project. Big appreciation to Facebook pages' admins that gave me permission to use their batik images for the purpose of this project. I would also like to give thanks to my family as a whole for their continuous support and understanding when undertaking my project. Special thanks to my Amirah Hanum Binti Amil Amran for her ideas and moral support.

REFERENCES

[1] A. E. Minarno, I. Soesanti, and H. A. Nugroho, "Batik Nitik 960 Dataset for classification, retrieval, and generator," *Data*, vol. 8, no. 4, p. 63, Mar. 2023, doi: 10.3390/data8040063.

[2] E. Steelyana, "Batik, a beautiful cultural heritage that preserve culture and support economic development in Indonesia," *Binus Business Review*, vol. 3, no. 1, p. 116, May 2012. [Online].

Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2193017

[3] N. H. Mohamad Akhir, N. W. Ismail, R. Said, and S. R. Kaliappan, "Traditional Craftsmanship: The origin, culture, and challenges of batik industry in Malaysia," in *Islamic Perspectives Relating to Business, Arts, Culture and Communication: Proceedings of the 1st ICIBACC 2014*, 2015, pp. 229–237.

[4] P. Jahangiri, S. M. Kahn, S. A. S. Abubakar, and M. Hedayat, "A BRIEF REVIEW OF MALAYSIAN BATIK DEVELOPMENTS FROM 1950s TO 1980s," *ATRAT: Jurnal Seni Rupa*, vol. 1, no. 3, Sep. 2013, doi: 10.26742/atrat.v1i3.395.

[5] H. D. Evers, "Nusantara: History of a Concept," *Journal of the Malaysian Branch of the Royal Asiatic Society*, vol. 89, no. 1, pp. 3–14, 2016, doi: 10.1353/ras.2016.0004.

[6] D. R. Rini, I. K. Hidayat, J. Samodra, and S. Sillah, "MOTIF BATIK MAPPING IN EAST JAVA AS LOCAL CULTURAL HERITAGE," *International Journal of Social Science Research*, vol. 1, no. 1, pp. 1–7, Oct. 2019, [Online]. Available: https://myjms.mohe.gov.my/index.php/ijssr/article/view/7038

[7] I. Ismurdiyahwati and A. Nasor, "THROUGH THE CONCEPT OF VISUAL LANGUAGE, BATIK MOTIFS IN THE NORTHERN REGION OF EAST JAVA, HAVE a COMMON MEANING," *Ismurdiyahwati | Prosiding ISBI Bandung*, 2021, vol. 1, no. 1.

[8] S. I. Syed Shaharuddin *et al.*, "A review on the Malaysian and Indonesian batik production, challenges, and innovations in the 21st century," *SAGE Open*, vol. 11, no. 3, p. 215824402110401, Jul. 2021, doi: 10.1177/21582440211040128.

[9] S. T. F. Poon, "The Journey to Revival: thriving revolutionary batik design and its potential in contemporary lifestyle and fashion," *International Journal of History and Cultural Studies (IJHCS)*, vol. 3, no. 1, pp. 48–59, 2017, doi: 10.20431/2454-7654.0301006.

[10] Ab. Shuaib and F. Olalere, "The Kelantan Traditional Arts as Indicators for Sustainability: An Introduction to its Genius LocI," *Journal of Social Sciences*, vol. 2, no. 2, pp. 87–100, Apr. 2013, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2246406

[11] M. J. Nawi, "Ekplorasi reka bentuk motif kain batik Kelantan sebagai subjek kajian dalam penghasilan karya Catan," Final Year Project thesis, Universiti Malaysia Kelantan (UMK), 2021.

[12] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[13] R. Andrian, M. A. Naufal, B. Hermanto, A. Junaidi, and F. R. Lumbanraja, "K-Nearest Neighbor (K-NN) classification for recognition of the Batik Lampung motifs," *Journal of Physics Conference Series*, vol. 1338, no. 1, p. 012061, Oct. 2019, doi: 10.1088/1742-6596/1338/1/012061.

[14] I. H. Sarker, "Machine learning: algorithms, Real-World applications and research directions,"

SN Computer Science, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[15] S. H. Yoo *et al.*, "Deep Learning-Based Decision-Tree classifier for COVID-19 diagnosis from chest x-ray imaging," *Frontiers in Medicine*, vol. 7, p. 427, Jul. 2020, doi: 10.3389/fmed.2020.00427.

[16] A. Kasim, M. Bakri, and A. Septiarini, "The Artificial neural networks (ANN) for batik detection based on textural features," in *Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia*, Jan. 2020.

[17] S. Ibrahim, S. Ahmad Kamaruddin, A. Zabidi, and N. A. Md. Ghani, "Contrastive analysis of rice grain classification techniques: Multi-class support vector machine vs artificial neural network," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, p. 616, doi: 10.11591/ijai.v9.i4.pp616-622.

[18] A. I. Ruppertsberg, M. Bloj, and A. Hurlbert, "Sensitivity to luminance and chromaticity gradients in a complex scene," *Journal of Vision*, vol. 8, no. 9, p. 3, Jul. 2008, doi: 10.1167/8.9.3.

[19] M. Mudrova and A. Procházka, "Principal component analysis in image processing," *Proceedings of the MATLAB Technical Computing Conference*, 2005. [Online]. Available: http://dsp.vscht.cz/konference_matlab/matlab05/prispevky/mudrova/mudrova.pdf

[20] W. Xing and Y. Bei, "Medical Health big data classification based on KNN Classification Algorithm," *IEEE Journals & Magazine | IEEE Xplore*, vol. 8, pp. 2169–3536, Nov. 2019, doi: 10.1109/ACCESS.2019.2955754.

[21] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[22] C. Gershenson, "Artificial neural networks for beginners," *arXiv.org*, Aug. 20, 2003. https://arxiv.org/abs/cs/0308031

[23] Y.-C. Wu and J.-W. Feng, "Development and application of artificial neural network," *Wireless Personal Communications*, vol. 102, no. 2, pp. 1645–1656, Dec. 2017, doi: 10.1007/s11277-017-5224-x.

[24] R. G. Tiwari, A. Misra, and N. Ujjwal, "Image embedding and classification using Pre-Trained Deep Learning architectures," in *2022 8th International Conference on Signal Processing and Communication (ICSC)*, pp. 125–130, Dec. 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10009560/

[25] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1430–1437, Aug. 2020, doi: 10.11591/eei.v9i4.2385.

[26] Ş. K. Çorbacıoğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic

accuracy Studies: A guide to interpreting the area under the curve value," *Turkish Journal of Emergency Medicine*, vol. 23, no. 4, p. 195, 2023, doi: 10.4103/tjem.tjem_182_23.

[27] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS ONE*, vol. 18, no. 10, p. e0291908, Oct. 2023, doi: 10.1371/journal.pone.0291908.

[28] T. Zeugmann *et al.*, "Precision and recall," Encyclopedia of Machine Learning, 2011.

[29] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Mining*, vol. 14, no. 1, pp. 1–22, Feb. 2021, doi: 10.1186/s13040-021-00244-z.

[30] A. Beucher, A. B. Møller, and M. H. Greve, "Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark," *Geoderma*, vol. 352, pp. 351–359, Oct. 2019, doi: 10.1016/j.geoderma.2017.11.004.

[31] E. K. Kim *et al.*, "COMPARISON OF NEURAL NETWORK AND K-NN CLASSIFICATION METHODS IN VOWEL AND PATELLAR SUBLUXATION IMAGE RECOGNITIONS," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 07, no. 04, pp. 775–782, Aug. 1993, doi: 10.1142/s0218001493000388.

[32] F. S. Nahm, "Receiver operating characteristic curve: overview and practical use for clinicians," *Korean Journal of Anesthesiology*, vol. 75, no. 1, pp. 25–36, Feb. 2022, doi: 10.4097/kja.21209.

[33] M. F. Amin, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial," *Journal of Engineering Research - Egypt/Journal of Engineering Research*, vol. 6, no. 5, p. 0, 2022, doi: 10.21608/erjeng.2022.274526.

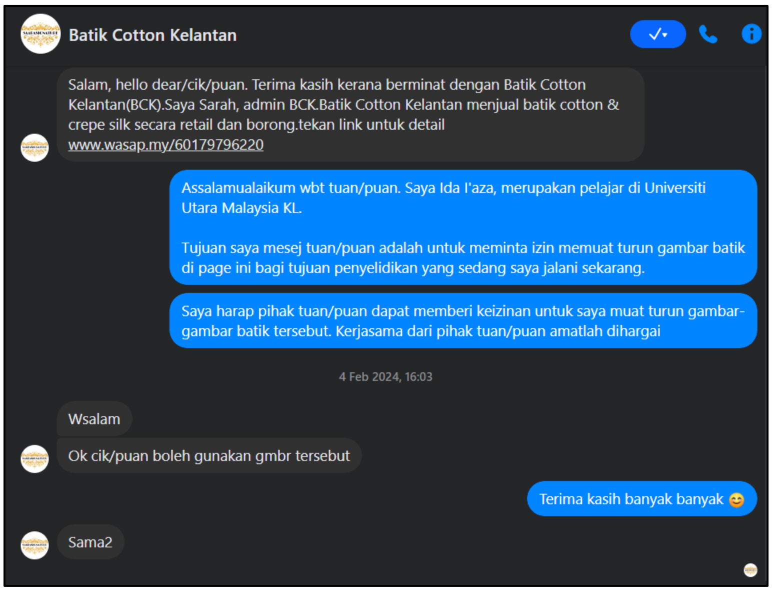
[34] J. Chen *et al.*, "A comparison of four data mining models: Bayes, neural Network, SVM and Decision trees in identifying syndromes in coronary heart disease," *Advances in Neural Networks–ISNN 2007: 4th International Symposium on Neural Networks, ISNN 2007, Nanjing, China, June 3-7, 2007, Proceedings*, 2007, pp. 1274–1279.

[35] G. B. Mahanya and S. Nithyaselvakumari, "Analysis and comparison of ventricular cardiac arrhythmia classification using sodium channel parameters with ANN and KNN classifier," *CARDIOMETRY*, no. 25, pp. 911–918, Feb. 2023, doi: 10.18137/cardiometry.2022.25.911918.

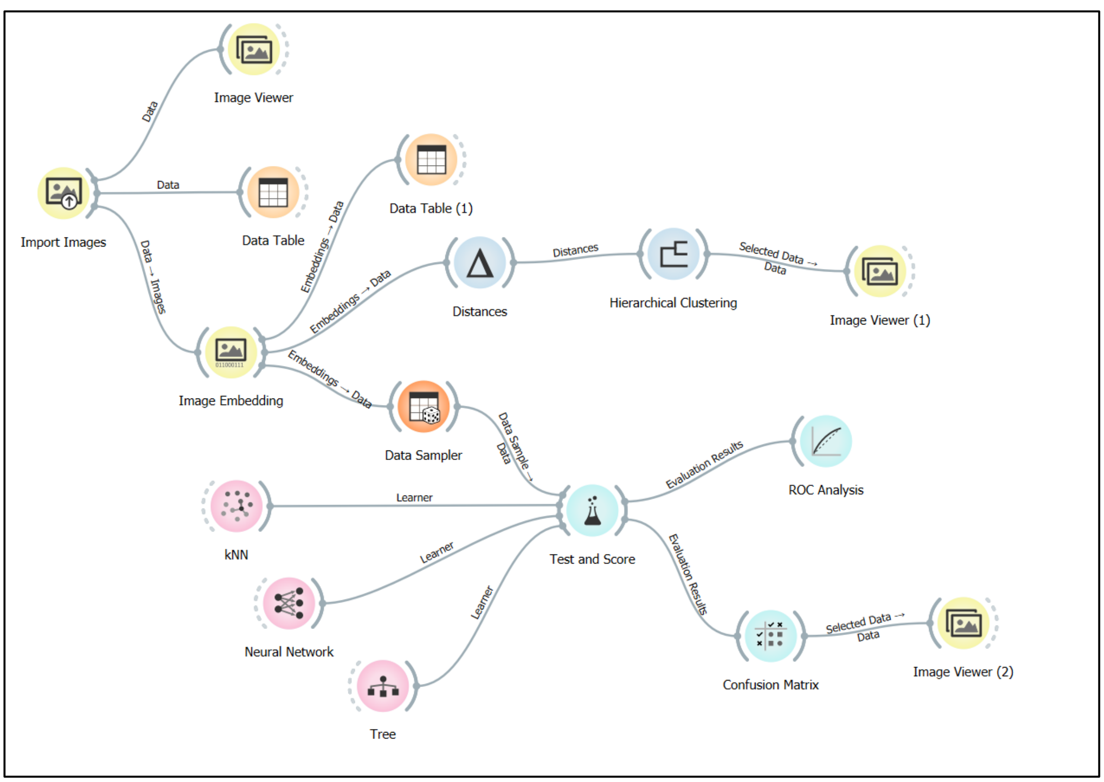
[36] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0276-2.

APPENDIX

Permission from Facebook admin

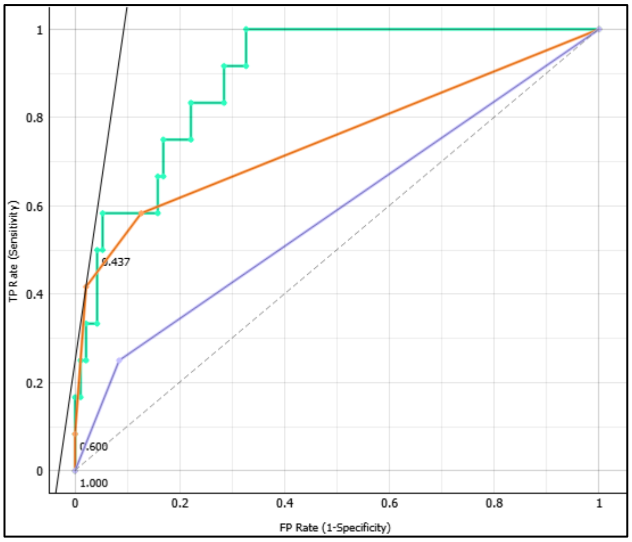


Network Arrangements in Orange

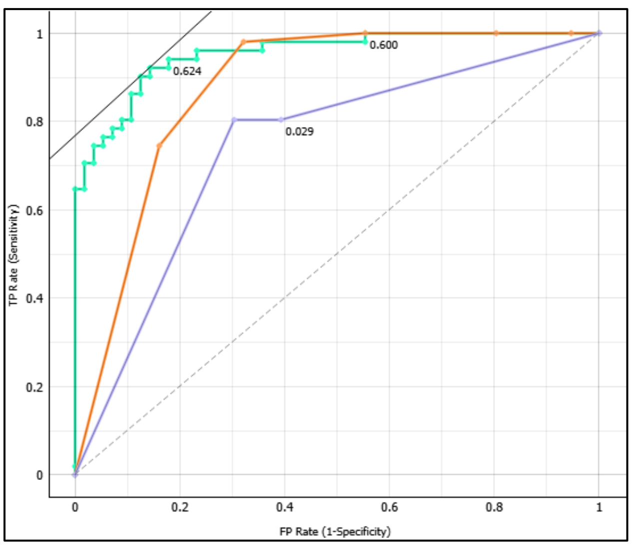


Results for Level Sweep:

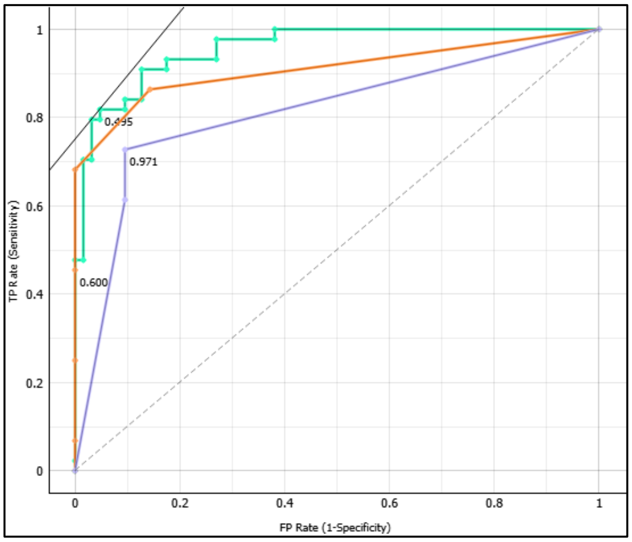
ROC Curve Plot for Fauna



ROC Curve Plot for Flora



ROC Curve Plot for Geometry



Confusion Matrix for Neural Network

Actual \ Predicted	Fauna	Flora	Geometry	∑
Fauna	5	6	1	12
Flora	1	48	2	51
Geometry	3	6	35	44
∑	9	60	38	107

Confusion Matrix for k-NN

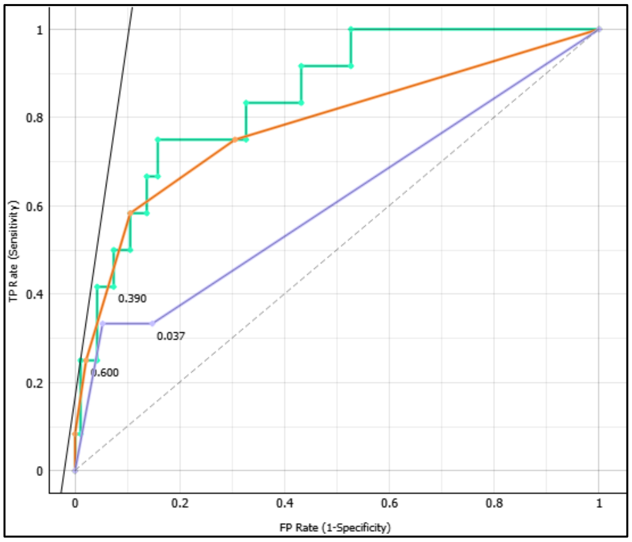
		Predicted			
		Fauna	Flora	Geometry	∑
Actual	Fauna	2	10	0	12
	Flora	0	51	0	51
	Geometry	1	23	20	44
	∑	3	84	20	107

Confusion Matrix for Decision Tree

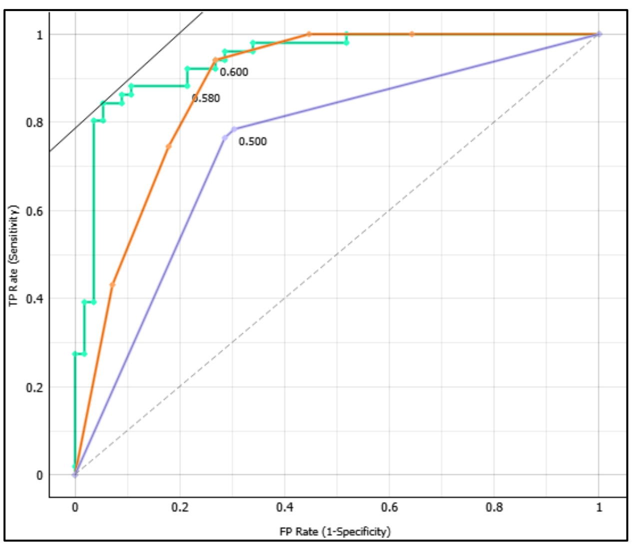
		Predicted			
		Fauna	Flora	Geometry	∑
Actual	Fauna	3	8	1	12
	Flora	5	41	5	51
	Geometry	3	9	32	44
	∑	11	58	38	107

Results for Luminance Gradient

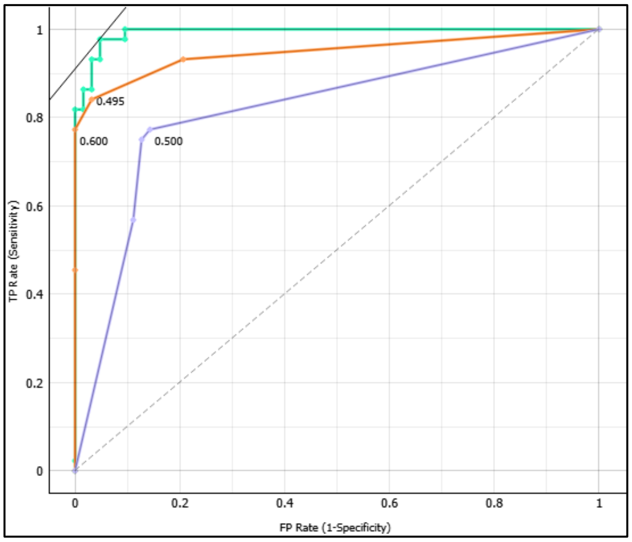
ROC Curve Plot for Fauna



ROC Curve Plot for Flora



ROC Curve Plot for Geometry



Confusion Matrix for Neural Network

Actual \ Predicted	Fauna	Flora	Geometry	∑
Fauna	5	7	0	12
Flora	4	45	2	51
Geometry	1	5	38	44
∑	10	57	40	107

Confusion Matrix for k-NN

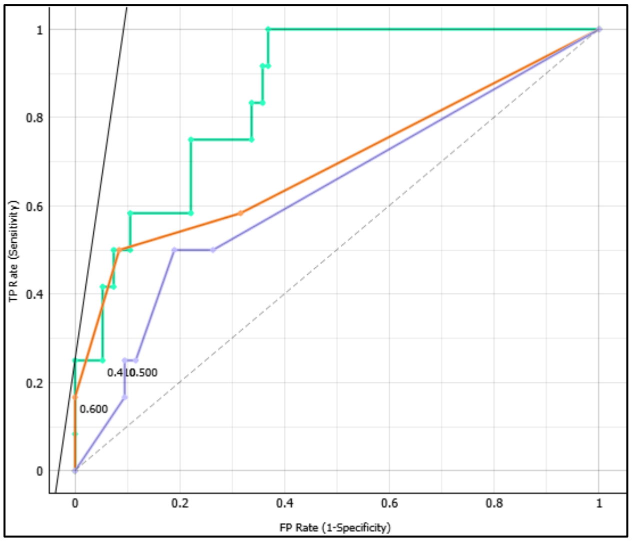
		Predicted			
		Fauna	Flora	Geometry	∑
Actual	Fauna	4	8	0	12
	Flora	3	48	0	51
	Geometry	3	7	34	44
	∑	10	63	34	107

Confusion Matrix for Decision Tree

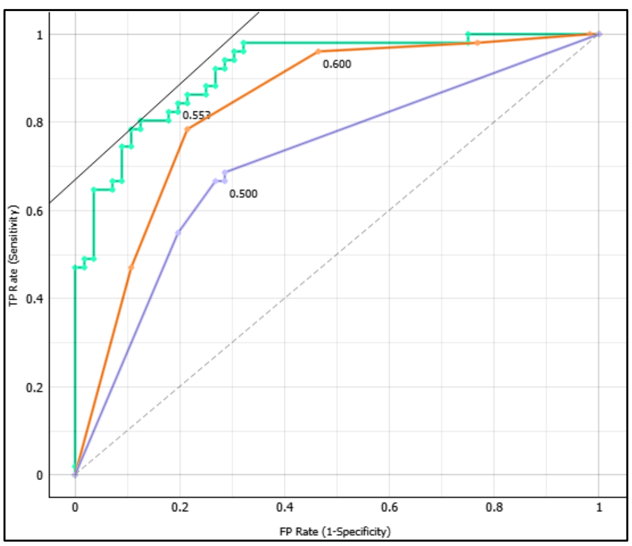
		Predicted			
		Fauna	Flora	Geometry	∑
Actual	Fauna	4	8	0	12
	Flora	3	40	8	51
	Geometry	2	9	33	44
	∑	9	57	41	107

Results for Principal Component Analysis

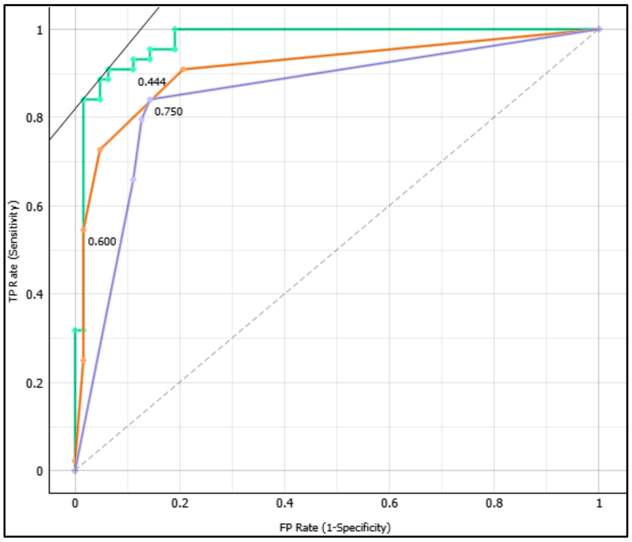
ROC Curve Plot for Fauna



ROC Curve Plot for Flora



ROC Curve Plot for Geometry



Confusion Matrix for Neural Network

Actual \ Predicted	Fauna	Flora	Geometry	∑
Fauna	3	9	0	12
Flora	2	43	6	51
Geometry	2	2	40	44
∑	7	54	46	107

Confusion Matrix for k-NN

Predicted

Actual	Fauna	Flora	Geometry	∑
Fauna	2	10	0	12
Flora	1	49	1	51
Geometry	2	18	24	44
∑	5	77	25	107

Confusion Matrix for Decision Tree

Predicted

Actual	Fauna	Flora	Geometry	∑
Fauna	3	9	0	12
Flora	8	34	9	51
Geometry	1	6	37	44
∑	12	49	46	107