# Automated Labeling of Hyperspectral Images for Oil Spills Classification

Madonna Said, Monica Hany, Monica Magdy, Omar Saleh, Maha Sayed, Yomna M.I. Hassan, Ayman Nabil
Misr International University, Faculty of Computer Science

*Abstract*—The constant increase in oil demand caused a huge loss in the form of oil spills during the process of exporting the product, which leads to an increase in pollution, especially in the marine environment. This research assists in providing a solution for this problem through modern technology by detecting oil spills using satellite imagery, more specifically hyperspectral images (HSI). The obtained dataset from the AVIRIS satellite is considered raw data, which leads to the availability of a vast amount of unlabeled data. This was one of the main reasons to propose a method to classify the HSI by automatically labeling the raw data first through unsupervised K-means clustering. The automatically labeled HSI is used to train various classifiers, that are Support Vector Machine (SVM), Random Forest (RF), and K-nearest neighbor (K-NN), to accomplish the optimal accuracy to be comparable with another research accuracy. In addition, the results of the first region of interest (ROI) indicate that the SVM with RBF kernel obtains 99.89% with principle component analysis (PCA) and 99.86% without the PCA, which revealed better accuracy than RF and the K-NN, while in the second ROI the RF obtained 99.9% with PCA and 99.91% without the PCA, better than K-NN and SVM. The region of interests selected lies within the Gulf of Mexico area. This area was selected based on the frequency of usage in previous research in detecting oil spills.

*Keywords—Oil spills; hyperspectral imagery; unlabeled data; k-means cluster; classification*

## I. INTRODUCTION

Within the last century, Crude oil became the most demanded mineral in the global industry as it supports more than 40% of the global energy needs [1]. Accordingly, the world had increased the exporting rate to obtain more oil despite the amount of lost oil during the exporting phase in the form of spills or wells discharges. Oil spills occur when a liquid petroleum hydrocarbon is released into the environment, which could lead to the leakage of 4.5 million tons of oil in the marine or ocean water [2]. Oil spill pollution can cause several natural disasters such as preventing the sufficient amount of sunlight to penetrate the ocean surface and reducing the dissolving level of oxygen, and increase the threat of extinction of different kinds of animals in the marine environment due to reproductive rate may be slow and long term recovery will last longer than usual [3] [4]. Not to mention its harm to plants life. The huge amount of this oil loss leads to significant economic decline.

Several source of data type have been used in order to assist in locating the oil spills such as hyperspectral images and multi-spectral images which are different types of remote sensing data. It's found that the hyperspectral images (HSI) "Represented in Fig. 1." are the most used as it includes one continuous spectrum which is used in measuring each pixel and provide the ability to distinguish different objects [5], mentioning that the standard is that the spectral resolution is given in nano meters (nm) or wave number and more than 100 bands in various intervals of 5-10 nm throughout the visible light to infrared spectrum are frequently present in its content. [6] [7]. The mentioned datatype ensures that certain minor but important features could be detected.

Not only obtaining the data was burdensome process, another challenge faced was of acquiring unlabelled data. After a lot of research, to locate the oil slicks using the hyperspectral images, the research performed several steps such as pre-processing, classifiers, and machine learning techniques. The mentioned steps are python-based code with the aid of ArcGIS software. The classifiers require labelled data to run. The labelled data comes with a tag, the absence of this tag results in getting unlabelled data. The tag is the less obvious bit and it only depends on the context of the problem in which the system try to solve, and normally the prediction of a feature is based on other features[9].

Supervised and unsupervised machine learning algorithms, have shown significant results in acquiring information from huge datasets. Supervised learning is an algorithm that generalizes information from known data with signal or labelled examples such that the algorithm can be tested by recognizing new data [10]. Unsupervised learning is the process of dividing data into groups using automated methods for unlabelled or categorized data. The absence of labelled data for the learning algorithm can sometimes be useful since it allows the program to search back for patterns that were not previously examined [11]. In case of presence of few amounts of labelled data semi-supervised learning is specified for this process. This technique [12] solves the problem of missing data by making use of large amount unlabelled data and the few labelled ones' to create better classification.

In this paper, our principle purpose is to classify the unlabelled hyperspectral data by proposing the K-means clustering algorithm then processing on it various classifiers. To summarise, the research's important points are divided into five categories that the researchers want to achieve:

*1)* Pre-processing the HSI on ArcGIS.
*2)* PCA is used to reduce dimensionality.
*3)* Labelling the HSI the unlabelled data by K-means clusters.
*4)* Classifying image using SVM, RF and k-NN.
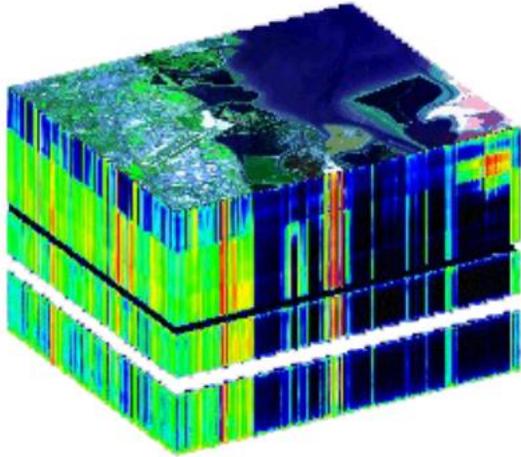*5)* Comparing the results with PCA and without PCA.

Fig. 1.    Hyperspectral Image [8].

Techniques used in other areas of research are experimented in the field of HSI analysis. The main contribution of this paper following these objectives is utilizing the concept of applying k-means clustering with automatic selection of the most suitable number of clusters. The automated clusters are then used as a labelling technique within the field of oil spills detection using hyper-spectral imagery. This technique has been earlier used in [13]. The only other clustering for labelling technique used for oil spills was [6] by the C-DPMM clustering algorithm, which we followed by testing a different technique (K-means) to experiment its efficiency, especially it is considered a simpler technique.

Another contribution of the paper is presented the exact details about the experiment including the exact data and tools utilized, which has not been presented in such details in other publications.

The remainder of the paper is laid out as follows. Section2 gives a brief overview of related works that focuses on the task of classifying pixels into various oil types. Section 3 describes the dataset used in the system. In Section 4, it mentions our methodology by discussing extensively our mains steps in obtaining our main goal, which is classifying the HIS image. The results after that are presented in Section 5. Finally, in Section 6, conclusions are reached.

## II.    RELATED WORK

Several systems focused on the task of classifying pixels into various oil types according to their spatial and spectral characteristics. Various techniques have been applied with numerous data types and structure. This section represents the most prominent work conducted in the fields related to the research at hand. In this section, it focuses on five main topics pertaining to this research. These topics cover a wide spectrum of research conducted within the areas of image processing and satellite imagery analysis. These topics are ordered as following:

- Labelling unlabelled data using k-means based clustering.

- Classification of HSI using SVM.

- Classification of HSI using DT.

- Classification of HSI using neural networks.

- Classification of HSI using neural networks.

- Usage of other classifiers.

### A.  Labelling using K-mean based Cluster

J. Xie and C. Wang [13] presented in their paper clustering as an unsupervised learning problem that uses unlabelled data to distinguish between groups of objects based on their features, which tries to group pixels with comparable spectral information into the same class. They collected a large number of synthetic random unlabelled datasets as part of the experiment. They experimented with various clustering techniques (presented in Table I), however they mainly focused on the k-means clustering algorithm as it doesn't need ground truth to initiate it. Then to minimize the number of clusters loaded by the k-means algorithm when searching the optimal number of clusters based on the k-d tree by extracting it from a plotted graph [14][15]. To classify the data they finally used the SVM to train the extracted new labelled data using several kernel functions. They compared the three methods which are (a) the fast global k-means clustering method, (b) the fast k-means clustering based on k-d trees, and (c) the global k-means clustering to choose the K-clusters. Following this, various kernels types are used to divide the data for the SVM, as shown in Table II. The results of the experiments on synthetically randomly created data sets show that the clustered SVM is very efficient and effective in categorizing completely unlabelled datasets. This inspired us to follow their methodology within our application domain.

TABLE I.        CSVM RESULTS [13]

| Clustering Algorithm | Kernel Function | Classification Accuracy | Time |
|---|---|---|---|
| a | Linear Polynomial rbf | 0.986 1 1 | 0.407 0.360 0.328 |
| b | Linear Polynomial rbf | 0.9836 1 1 | 0.265 0.203 0.172 |
| c | Linear Polynomial rbf | 0.9839 1 1 | 0.500 0.531 0.437 |

TABLE II.        RESULTS PRESENTED IN [21]

|  |  | Sensitivity | Specificity | Precision | Time (s) |
|---|---|---|---|---|---|
| SVM | RBF Linear Sigmoid | 88.59 75.17 64.78 | 98.84 97.80 96.36 | 90.07 68.74 59.46 | 262.52 291.08 529.15 |
| RF |  | 93.83 | 99.32 | 95.35 | 269.33 |
| DA | Linear Diag-Linear | 78.92 65.46 | 78.92 65.46 | 96.84 94.16 | 8.12 3.5 |
| KNN | K=1 K=3 k=5 k=7 | 91.97 88.41 86.98 86.40 | 99.01 98.61 98.46 98.39 | 91.75 89.51 88.75 88.62 | 3435.95 3521.30 3714.08 3886.08 |

## B. Classifications using SVM Approach

Sahar A. and Ali H. [16] proposed a system that uses HSI images to detect oil spills. The collected user data is AVIRIS data of the Gulf of Mexico and the Adriatic Sea. They applied four various classifiers, which are SVM, BE, MD, and parallelepiped classifiers to classify the pixels. SVM, BE, MD indicated higher classification accuracy than the parallelepiped approach. Where BE, MD, parallel-piped, and SVM reached 88.4423%, 94.6399%, 46.9012%, and 99.8325%, respectively in the first dataset and reached 96.5338%, 98.6135%, 62.9116%, and 100% respectively in the second dataset. The system architecture also explained that there are some pre-processing steps before classifying as determining Region of Interest and applying PCA. They were able to identify the existence of various types of oils (i.e. dark and light oils). They also separated the appearance of pixels identified as oil from other pixels that identified as other components such as water.

Dabbiru, Lalitha, et al. [17] performed data fusion of hyperspectral and SAR imaging at both levels of data and functionality, to improve target detection, a combined spatial-spectral analysis is obtained and analysed the fused data to combine with Support Vector Machine (SVM). The ground truth classes are composed of six different classes. The system started with feature extraction of HSI and SAR, HSI using PCA to reduce dimensions, and SAR using GLCM to extract features from SAR data in different spatial orientations. The results of the SVM classifier were tested for each combination giving the highest accuracies for the two classes: healthy vegetation and lightly oiled vegetation. While the researchers were searching for the best accuracy classification approach the traditional approaches as SVM became of non-interest and they are still searching and developing new approaches in this field.

## C. Classification using Decision Tree

Liu, Y. Li, P Chen and X. Zhu [18] proposed a system that used a Decision Tree (DT) to extract the information of oilspills based on minimum noise fraction (MNF) transformation.Before establishing the DT; the system used MNF to decreasethe redundant data and the noise of the image. The results forMNF-based decision tree classification can cluster the classesefficiently and distribute their ac-curacies. The study area ofthe system was the Gulf of Mexico which counted 69.16% ofthin oil film to approve as the dominant class, while very thinoil film, thick and medium thickness oil film counted 12.26%, 5.50%, and 6.22%, respectively.

## D. Classification using Neural Network

JF Yang, JH Wan, Y Ma, and J Zhang [19] utilized a deep convolutional neural network (DCNN) for oil spill detection accuracy of the sea surface oil and comparing with the traditional SVM, RF, and DBN methods. Based on different-scale features, the results were briefed on the increase in the numbers of the accuracy of DCNN to reach 85%, which is even greater than SVM, RF, and DBN methods. While the results based on spectral feature information of one level WT with low-frequency component produced the highest accuracy detection that reached 87.51%. The author in [20] presented a spatial and spectral features by trying to stack auto-encoder (SSAE) to cluster and classify oil slicks on the surface of the sea and comparing it with the classical SVM, BPNN, and SAE. SSAE is based on the SAE network taking into consideration spatial information during classification. The accuracy results were impressive as the SAE and SSAE approaches have reached71.43% and 73.97%, respectively, while SVM and BPNN reached 68.89% and 63.81%, respectively. After comparing the images of SAE and SSAE approaches we may begin to understand that the SSAE eliminated the scatters across the thick film regions. Classification result accuracy increased by 4% to overcome the SAE over-fitting problem. We also compared the original image and PCA + SSAE and decision tree (DT), but SSAE accuracy was much better.

## E. Classification using SVM, RF, DA and K-NN

In [21], comparison research of four supervised classifiers has been offered using HSI of different datasets obtained using AVIRIS sensor. They used the following classification methods which are SVM with different kernel types, Random Forest RF, Discriminant Analysis (DA) with two different kernels, and K-NN. They selected the more relevant bands from the utilized hyperspectral datasets using a mutual information-based filtering approach. The experimental results show that feature extraction as a pre-processing step of classification using mutual information is effective. Then, several measurements have been generated to evaluate each classifier overall accuracy. The demonstrative accuracy shown in Table II indicates that the SVM with RBF kernel performs well and appears to be the most effective as a supervised classifier for hyperspectral image classification, followed by RF and K-NN. The DA is the least excellent and performs badly in contrast to the other techniques.

In summary, it is viewed that most prominent techniques rely heavily on the availability of training sets, ether already available or gathered manually. The existence of labelled data in itself is considered an issue. Therefore this system focus on how to automate the labelling process while achieving an acceptable accuracy.

## III. DATASET

Since it's concluded that the hyperspectral images are too expensive and time-consuming, the researchers figured to build their studies on very limited study areas as it is the most accurate for the detection of oil spill [16]. The hyperspectral image captures the area at a very large number of wavelengths and breaks the image down into tens of thousands of colours. AVIRIS remote sensing dataset which we are using is specialized in capturing the hyperspectral image, as it provides data of spatial and spectral features. This type of image is very large in size due to the huge information reflected from the ground features for each pixel in different wavelengths in bands. The visible RGB spectrum is used to determine the range of oil spills detected at each pixel, which ranges from 0.35 to 2.5 microns: In 224 bands, blue 0.4 microns, green 0.53 microns, and deep red 0.7 microns were used. The kind of oil thickness and the wavelength range of each band are significant characteristics of each band. Sample of the data is presented in Fig. 2.
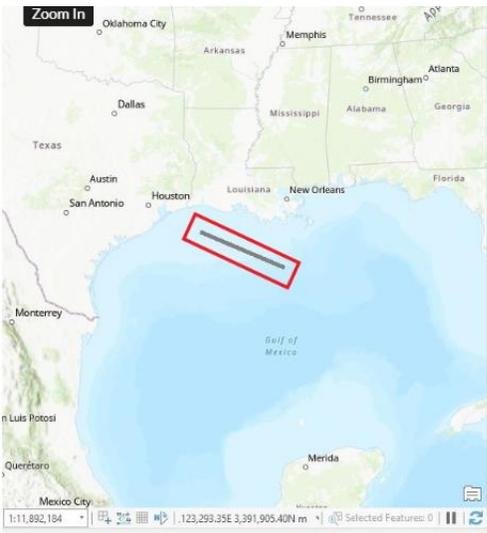
Fig. 2. Gulf of Mexico AVIRIS Flight Strip.

## A. Region of Interest

As our area of interest is "Gulf Of Mexico", we obtained the data using an AVIRIS camera which collects the data using flights drawn on the map, so from the nearest flight record, a 9.2 GB dataset was collected for the mentioned location in binary [22]. The dataset contains two important files, which are the header file that contains information about the image, and the second file is the Bip file of the image. To prepare our data to import to our platform we used to convert our data from binary format into four files the most important one is TIFF (Tag Image File Format), which has four types one of them is RGB which we need while dealing with our hyperspectral /multispectral images.TIFF files own the privilege of the option of owning a supplementary file called TFW which contains six rows of data [23].

For this research, we have downloaded data for two different flights. The utilized flights are:

- Dataset 1 (name: f100523t01p00r08, Id: f100523t01, Flight: 100524).

- Dataset 2 (name: f100525t01p00r05, Id: f100525t01, Flight: 100525).

## IV. METHODOLOGY

The proposed system is used to reduce the environmental disorders that are caused by oil spills, using hyperspectral remote sensing imagery. The obtained AVIRIS data is enhanced first using ArcGIS to be easily imported in python code. The extracted image is analysed by unsupervised classification to label the data by using a clustering algorithm which is implemented by python code. We used the labelled K-means image to be able to train our system with various classifiers and test its performance using the original enhanced image. Fig. 3 explains the steps of the system.

In the beginning, we worked on ArcGIS software [24] to prepare the AVIRIS dataset in order to efficiently reach our goal. Firstly, we imported the image in the ".bip" extension which in the geographic field, is one of three primary ways for

encoding picture data for multiband raster images. Since the imported image is in binary format, which is a black HSI, we assigned the RGB of the whole image. Furthermore, we enhanced the image by using different methods as histogram equalization and percent-clip. At the end of the process, we took a clip after picking the ROI of the image and save it as a ".tif" extension to reduce its size [25]. These steps are presented in Fig. 4.
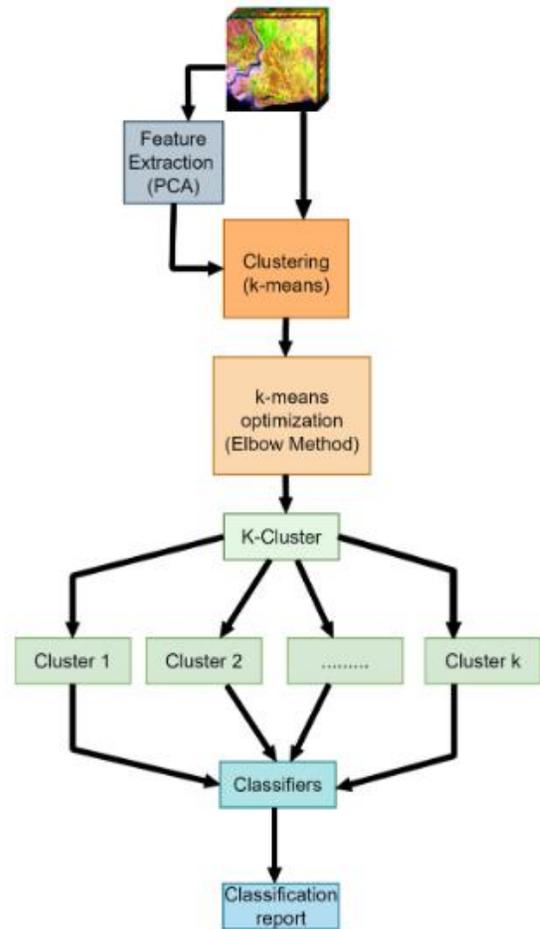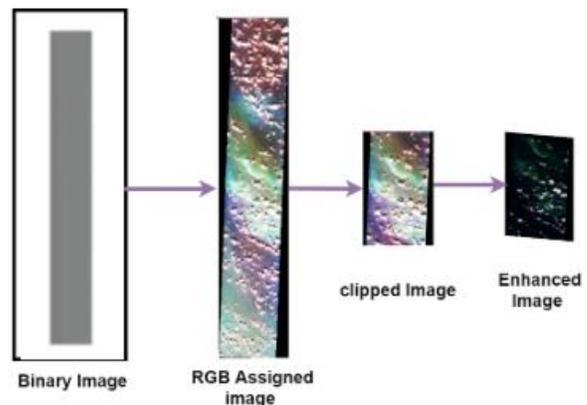


Fig. 3. Flow Chart.



Fig. 4. ArcGIS Steps.

Initially, we extracted the features of the hyperspectral clipped image using PCA. Accordingly, reduce the redundant features. In order not to accidentally lose the image's features in this process, we compared the extracted classification results using PCA and without using PCA. The hyperspectral image is composed of three dimensions, we converted it from a 3D-cube image to 2D image preserving its important features in each band by the concatenation of (x, y) dimensions.

### A. K-means Clustering

The second step is using the k-means unsupervised cluster for labelling our data. The output of the 2d array is taken as an input for the k-means cluster [26]. To determine the number of clusters, we optimized k-means by elbow method using mean and median methods as follows in equation 1, 2 and 3:

$$Odd_{ME} = \frac{\frac{X_n}{2} + \frac{X_n}{2}}{2} \qquad (1)$$

$$Even_{ME} = \frac{X_{n+1}}{2} \qquad (2)$$

$$Median_{ME} = \frac{MinVal + MaxVal}{2} \qquad (3)$$

The unsupervised classification is considered an effective method for unlabelled data, K-means cluster was the optimal choice to be used as a labelling technique in pixel by pixel labelling. This technique separates the dataset into distinct groups based on how closely the data points are together, and these groups are referred to as "clusters". Euclidean distances used to determine how close data points are to each other. To implement this technique, one data point is assigned per cluster, known as the "centroid". Then, based on the cluster with the closest mean value, each point is assigned to it. After all of the points have been assigned to clusters, the means are calculated again, and the procedure is repeated until the mean square error of two consecutive processes is the same. The K-means cluster contains the input data into clusters that contain image features [27].

### B. Classifications

The final step is the classification which is done by trainingdifferent ROI hyperspectral imagery, using three classifiers(SVM, K-NN, RF), and testing their performance.

*1) Support vector machine classifier:* The SVM tries toobtain the optimal hyper-planes between different points ofclasses by the selection of the largest gap between pointsand reducing the error using the optimal hyper-plane to avoidthe occurrence of over fitting. The different kernel functionsare used in different hyper-planes as well. The radial basis function in our system which is used according to [27]. The RBF kernel is considered the best among the Polynomial and Linear kernels, when it comes to classifying hyperspectral images, due to the high number of layers within the image. Polynomial and Linear kernels appeared to be time-consumingand low efficiency while using hyperspectral images.

*2) K-Nearest Neighbour:k-NN approach in particular,* usesnearest neighbor (NN) classifiers which are one of the mostbasic and yet effective classification criteria, and they areextensively utilized in practice. This approach is consideredsupervised neural network classification. A training set ofpattern vectors from such a class is provided for each class asa set of sample models. When classifying an unknown vector,its k nearest neighbors are discovered among all prototypevectors, and the class label is determined using a majority rule.The value of k should be odd to avoid conflicts on class overlapregions. Although this rule is basic and straightforward, it hasa low error rate in practice. For this study we estimated ModelEvaluation for k to 7.

*3) Random forest:* The random forest classifier is mainlycomposed of several tree classifiers, each one of the trees isgenerated using a random vector obtained independently fromthe input vector. Each tree gives a vote for the most convenientposition to classify. In our system we estimated the numberof trials in the RF by 100, applied the idea of grouping thesimilar features in classes, and used these classes as subsetswhich are chosen randomly to train the data.For all the classifiers, the training and testing methods areassigned to 67:33% respectively in order to reach differentaccuracies. We compared the results to adjust the best classi-fication method.

## V. RESULT

In this study, the data preparation stage is feature extraction by labelling the data using a clustering algorithm. Two HSI from the "Gulf of Mexico" were used to validate the proposed method. The experimental results using the two datasets are summarized to be labelled by the k-mean clusters on HIS demonstrate that the SVM and RF techniques perform close and better accuracies done by the K-NN algorithm. To be able to train the classifiers using labelled data we use k-means cluster for that. Each image contains a different number of clusters used for the process according to the elbow method we discover that the optimal number of k to be able to label the data according to the extracted result. As shown in Fig. 5 and 6, the graphs represent the k value as in the first dataset it contains 6 k-clusters and the second dataset is 4k-clusters. The right figure represents the generating labelled data using the optimal clusters.
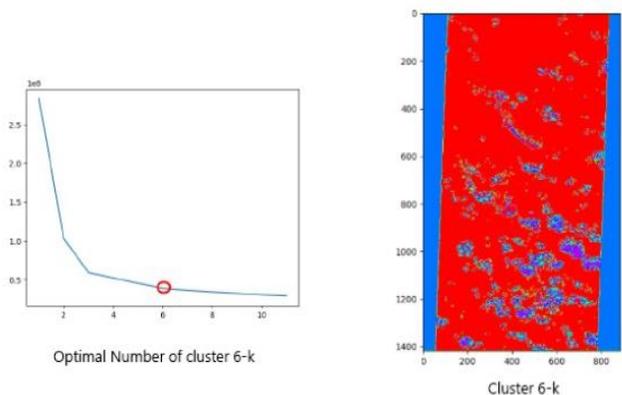


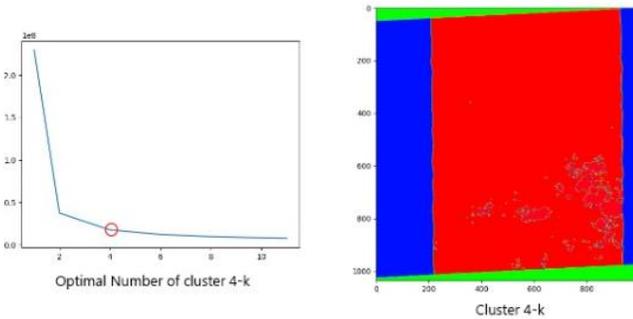Fig. 5. Elbow Method applied to Dataset 1.

Fig. 6.   Elbow Method applied to Dataset 2.

The obtained labelled data are classified with three different classifications, which are SVM, RF, and K-NN as mentioned before. In discovering the best way for the automated detection of oil spills a comparison is demonstrated between the results of the different classifiers. A confusion matrix is utilized to determine the accuracy of each classification. Table III and Table IV exhibit the confusion matrices for the first and the second datasets respectively. For the chosen dataset, it can be observed that the majority of classification algorithms work effectively and have an accuracy rate of more than 90% without the presence of overfitting, which was validated by comparing the output results from the training and testing accuracies over the two utilized datasets. It has been demonstrated that the SVM with the RBF kernel, RF, and K-NN achieved the best classification results for 67 selected pixels was training and 33 selected pixels as testing. The accuracy has been calculated by using ready-made python functional measurement criteria that are mainly used in similar systems:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (4)$$

*1) First dataset results:* The first dataset performance of SVM with RBF kernel and PCA was 99.89% and without PCA was 99.86%, then the RF classification results with the PCA was 99.78% and without the PCA was 99.85%, while the results for the K-NN with the PCA was 99.77% and without the PCA was 99.78%. After observing the classification results, it was concluded that the SVM has the highest accuracy results followed by the RF classification which is considered slightly better than the K-NN. The achieved results were close to what was mentioned in [21] as the SVM, RF, and K-NN classifiers were achieving the best accuracies respectively, which is the same conclusion that was reached in this dataset. 2) Second dataset results: The second dataset classification results are completely different from the first. The RF performed the best results with and without PCA to achieve99.90% and 99.91% respectively, then the K-NN performance was unexpected to be better than SVM and achieve high results as it reached 99.89% and 99.88% respectively with and without PCA, while the SVM which performed the best in the first dataset, performed the worst in the second dataset to attain 98.31% and 98.17% respectively with and without PCA. Classifiers' accuracy results are

relatively close to each other with a slight difference between them, they performed unexpectedly in the two datasets. The SVM performed the best in the first dataset and the worst in the second dataset which determines that choosing the optimal hyperplane in the second dataset was not accurate. Furthermore, the three classifiers' accuracy before and after PCA are relatively similar. Although processing time without PCA is quite long. PCA was used to reduce the dimensionality of datasets before the classification procedure, and eventually enhance the runtime [16]. The difference in runtime is shown in Table V by specific platform of OS version(windows 10.0.19042) with Processor Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz, 4 Core(s), 2112 Mhz, 8 Logical Processor(s). There is a lack of clarity in the comparative results with the similar systems since they did not mention the exact AVIRIS flight nor the source of the ground truth data. In conclusion, the SVM, K-NN, and RF techniques are accurate in classifying hyperspectral datasets from the Gulf of Mexico AVIRIS flights.

TABLE III.     ACCURACY OF THE FIRST DATASET – GULF OF MEXICO

| Classifier | With PCA | Without PCA |
|---|---|---|
| SVM | 99.89% | 99.86% |
| Random Forest | 99.78% | 99.85% |
| K-NN | 99.77% | 99.78% |

TABLE IV.     ACCURACY OF THE SECOND DATASET – GULF OF MEXICO

| Classifier | With PCA | Without PCA |
|---|---|---|
| SVM | 98.31% | 98.17% |
| Random Forest | 99.90% | 99.91% |
| K-NN | 99.89% | 99.88% |

TABLE V.     CLASSIFICATION RUNTIME

| | Classifier | Time Taken in Hours |
|---|---|---|
| Without PCA | SVM | 10-12 |
| | RF | 5-8 |
| | KNN | 12-15 |
| With PCA | SVM | 10-11 |
| | RF | 2-3 |
| | KNN | 11-13 |

VI.   CONCLUSION AND FUTURE WORK

The hyperspectral imagery typically generates a massive amount of data, which makes the computing expense and associated classification work might be challenging. The paper demonstrates several solutions and methods to efficiently and accurately detect oil spills. However, labelling the HSI is high expensive, and time-consuming the clustering technique by the-means was used for labelling our datasets. The k-mean labels were classified using supervised classifications to train and test the model according to the original HSI which performed high accuracies using the SVM, RF, and K-NN

approaches. Futuristically, it is expected to test the system on various other datasets to ensure the generalization of the technique used. The hyperspectral imaging accuracies are commonly high, thus we can attempt using other clustering techniques that are expected to perform better than the k-means clustering due to its simplicity, to perform a more accurate and efficient system.

REFERENCES

[1] Maizar Rahman. "Oil and gas: the engine of the world economy". In: Presentation as OPEC Secretary General at the Tenth International Financial and Economic Forum, Vienna, Austria. 2004, pp. 10–11.

[2] Best Ordinioha and Seiyefa Brisibe. "The human health implications of crude oil spills in the Niger delta, Nigeria: An interpretation of published studies". In: Nigerian medical journal: journal of the Nigeria Medical Association54.1 (2013), p. 10.

[3] Oil spill — Definition, Causes, Effects, List, & Facts —Britannica. https : / / www. britannica . com / science / oil -spill.

[4] Yifan Zhang, Ting Wang, Shaohui Mei, et al. "Subpixel Mapping of Hyperspectral Images Using a Labelled -Unlabelled Hybrid Endmember Library and Abundance Optimization". In:IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing13(2020), pp. 5036–5047.

[5] Rohit Sharma. What is Clustering and Different Types of Clustering Methods — up Grad blog. https: / / www.upgrad . com / blog / clustering - and - types - of - clustering -methods/#.

[6] Hao Wu and Saurabh Prasad. "Semi-supervised deep learning using pseudo labels for hyperspectral image classification". In:IEEE Transactions on Image Pro-cessing27.3 (2017), pp. 1259–1270.

[7] Merv Fingas. "The challenges of remotely measuring oil slick thickness". In: Remote sensing10.2 (2018),p. 319.

[8] ARSET - Hyperspectral Data for Land and Coastal Systems - NASA Applied Science. https://appliedsciences.nasa . gov / join - mission / training / english / arset -hyperspectral-data-land-and-coastal-systems.

[9] G Andreoli, B Bulgarelli, B Hosgood, et al. "Hyperspectral analysis of oil and oil-impacted soils forremote sensing purposes". In:European Commission Joint Research Centre: Luxembourg36 (2007).

[10] Michael W Berry, Azlinah Mohamed, and Bee WahYap. Supervised and unsupervised learning for data science. Springer, 2019.

[11] Ramadass Sathya, Annamma Abraham, et al. "Comparison of supervised and unsupervised learning algorithms for pattern classification". In: International Journal of Advanced Research in Artificial Intelligence2.2 (2013),pp. 34–38.

[12] Xiaojin Jerry Zhu. "Semi-supervised learning literature survey". In: (2005).

[13] Juanying Xie, Chunxia Wang, Yan Zhang, et al. "Clustering support vector machines for unlabelled data classification". In: 2009 International Conference on Test and Measurement. Vol. 2. IEEE. 2009, pp. 34–38.

[14] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek."The global k-means clustering algorithm". In: Patternrecognition36.2 (2003), pp. 451–461.

[15] Stephen J Redmond and Conor Heneghan. "A method for initialising the K-means clustering algorithm using kd-trees". In:Pattern recognition letters28.8 (2007),pp. 965–973.

[16] Sahar A El Rahman, Wateen A Aliady, and Nada IAlrashed. "Supervised classification approaches to an-alyze hyperspectral dataset". In: International Journal of Image, Graphics and Signal Processing (IJIGSP)7.5(2015), pp. 42–48.

[17] Lalitha Dabbiru, Sathishkumar Samiappan, Rodrigo AANobrega, et al. "Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico". In:2015 IEEE international geoscience and remote sensing symposium (IGARSS).IEEE. 2015, pp. 1901–1904.

[18] Bingxin Liu, Ying Li, Peng Chen, et al. "Extraction of oil spill information using decision tree based minimum noise fraction transform". In: Journal of the Indian Society of Remote Sensing44.3 (2016), pp. 421–426.

[19] Jun-Fang Yang, Jian-Hua Wan, Yi Ma, et al. "Oil spill hyperspectral remote sensing detection based on DCNN with multi-scale features". In: Journal of Coastal Re-search90.SI (2019), pp. 332–339.

[20] Bingxin Liu, Qiang Zhang, LI Ying, et al. "Spatial–spectral jointed stacked auto-encoder-based deep learning for oil slick extraction from hyperspectral images".In:Journal of the Indian Society of Remote Sensing47.12 (2019), pp. 1989–1997.

[21] Hasna Nhaila, Asma Elmaizi, Elkebir Sarhrouni, et al." Supervised classification methods applied to airborne hyperspectral images: comparative study using mutual information". In:Procedia Computer Science148(2019). The Second International Conference on Intelligent Computing in Datasciences, ICDS2018, pp. 97–106.ISSN: 1877-0509.DOI: https : / / doi . org / 10 . 1016 / j . procs . 2019 . 01 . 013.URL: https://www.sciencedirect.com/science/article/pii/S1877050919300134.

[22] Nasa.JPL - AVIRIS Data Portal https://aviris.jpl.nasa.gov/dataportal/.

[23] File info.TIFF File Extension - What is a .tiff file and how do I open it?. https://fileinfo.com/extension/tiff.

[24] Esri. Download, install, and authorize—ArcGIS Pro —Documentation. https://pro.arcgis.com/en/pro-  app/2.6/get-started/install-and-sign-in-to-arcgis-pro.htm.

[25] Jian Guo Liu and Philippa J Mason. Image processing and GIS for remote sensing: techniques and applications. John Wiley & Sons, 2016.

[26] Edy Umargono, Jatmiko Endro Suseno, and SK Vin-censius Gunawan. "K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula". In: The 2nd International Seminar on Science and Technology (ISSTEC 2019). Atlantis Press. 2020, pp. 121–129.

[27] Sameer Ranjan, Deepak Ranjan Nayak, KallepalliSatish Kumar, et al. "Hyperspectral image classification: A k-means clustering based approach". In: 2017 4thInternational Conference on Advanced Computing and Communication Systems (ICACCS). 2017, pp. 1–7.DOI:10.1109/ICACCS.2017.8014707.