J.Natn.Sci.Foundation Sri Lanka 2019 47 (3): 271 - 284
DOI: http://dx.doi.org/10.4038/jnsfsr.v47i3.9276

RESEARCH ARTICLE

Mapping ilmenite deposit in Pulmudai, Sri Lanka using a hyperspectral imaging-based surface mineral mapping method

E.M.M.B. Ekanayake[1*], S.S.P. Vithana[2], E.M.H.E.B. Ekanayake[1], A.R.M.A.N. Rathnayake[1], A.M.R. Abeysekara[1], T.S.J. Oorloff[1], H.M.V.R. Herath[1], G.M.R.I. Godaliyadda[1], M.P.B. Ekanayake[1] and A. Senaratne[3]

[1] *Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya.*
[2] *Sri Lanka Technological Campus, Ingiriya Road, Padukka.*
[3] *Department of Geology, Faculty of Science, University of Peradeniya, Peradeniya.*

Submitted: 25 October 2018; Revised: 24 January 2019; Accepted: 28 March 2019

Abstract: Mineral detection using remote sensing techniques is important since it saves the time and effort of carrying out manual land surveys. In this paper a novel algorithm, which can be used to detect ilmenite using hyperspectral image analysis is discussed. To investigate this task, a hyperspectral image obtained from the Earth Observing-1 (EO-1) satellite's Hyperion sensor was used. In the proposed algorithm, first, principal component analysis (PCA) was used for dimensionality reduction and an Euclidean distance-based method was used to extract the pixels containing soil. Thereafter, lab spectral data of typical ilmenite deposits were considered as the reference and a correlation factor analysis was carried out to determine the soil pixels, which are most likely to contain ilmenite and most unlikely to contain ilmenite. Using these two sets of pixels, a training set was constructed to apply Fisher's discriminant analysis (FDA) in order to separate the dataset into two distinct classes – ilmenite and non-ilmenite. Based on the spectral similarity, each pixel of the image was classified under one of these classes. This paper also introduces a probability-based approach to obtain results that are more accurate. A probability density function was designed considering the spatial distribution of the mineral. Thereafter, classification was done considering the probability measure as well. Lab tests performed on the soil samples collected from the locations, which were detected by the algorithm validate that the algorithm is accurate.

Keywords: Correlation coefficient, data engineering, Fisher's discriminant analysis, principal component analysis, probability-based approach, remote sensing.

INTRODUCTION

A hyperspectral image contains information corresponding to a large number of continuous wavelengths. This facilitates the analysis and detection of intricate details of a given image. As the existence of a particular surface mineral results in a different composition of soil, compared to that of the places where the respective surface mineral is unavailable, image-processing techniques can be used to detect possible deposits of surface minerals. The use of satellite images is a non-intrusive and economical way for surface mineral detection as it avoids strenuous land surveys and soil testing. Since hyperspectral data contains richer information in the form of contiguous spectral bands compared to multispectral images, the detection process is more accurate and reliable.

A number of papers have been published in the recent past, proposing different methods of determining possible surface mineral deposits using hyperspectral imagery. Applicability of hyperspectral data to determine trends in hydrothermal alteration intensity in and around the Izok Lake Volcanogenic Massive Sulfide (VMS) deposit in Northern Canada has been discussed by Laakso *et al.* (2016). An efficient procedure for mineral mapping is discussed by Notesco *et al.* (2014), with a unique hyperspectral remote-sensing fingerprint in the

* Corresponding author (mevanekanayake@gmail.com; https://orcid.org/ 0000-0002-4768-5073)

This article is published under the Creative Commons CC-BY-ND License (http://creativecommons.org/licenses/by-nd/4.0/). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

longwave infrared spectral region enabling identification of the most abundant minerals in the continental crust - quartz and feldspars. The presence of copper in Basavanakote, Karnataka was found using hyperspectral remote sensing techniques by Aravinth & Roopa (2017). A new method has been introduced to extract features using mineral absorption by Zhao *et al.* (2017). Here, Reference Spectral Background Removal (RSBR) has been introduced into mineral absorption feature extraction from high vegetation density areas. Diagnostic absorption feature has the potential to be the key factor in mineral information extraction. Saralıoğlu *et al.* (2016) have attempted to explore minerals with hyperspectral image fusion. Due to the insufficient spatial resolution of current hyperspectral image sensors, a single pixel might include more than one mineral. This paper has revealed a method to overcome this problem by fusing such an image with a high-resolution image. Niranjan *et al.* (2016) have presented a study of minerals and vegetation in explored fields around the San Juan coal mines west of Farmington, New Mexico. Here, Minimum Noise Fraction (MNF) and Pure Pixel Index (PPI) methods are used for the extraction of endmember fraction, and spectral signature matching procedure is done with the United States Geological Survey (USGS) spectral library. Swamy *et al*. (2017) have carried out a dimension reduction based algorithm on complex wavelet filter bank and the mimetite mineral spectral signatures, and classified using the orthogonal subspace projection method. The desired mimetite spectral signature is segregated from the mixed pixels and estimated in the presence of two other minerals with the help of complex wavelet filter bank and spectral matching. Vithana *et al*. (2019) have exploited the relative proximity of spectral signatures among classes of remotely sensed hyperspectral images in order to generate an adaptive hierarchical structure for image classification. This enables a level-by-level optimisation for clustering at each stage of the hierarchy. Ekanayake *et al*. (2019) have proposed a novel method for hyperspectral unmixing, in which the fundamental notions of independent component analysis are utilised to improve the accuracy of the standard non-negative matrix factorisation algorithm. This method proves to be highly accurate with regard to hyperspectral image classification.

This paper presents a method of detecting possible ilmenite deposits in the Northeastern region of Sri Lanka, using concepts related to data engineering, mathematics, machine learning and signal processing, on a hyperspectral image of the respective region obtained by the Earth Observing-1 satellite's Hyperion sensor. The detection algorithm first deals with classifying the pixels in the region under its sub components. It was observed that the selected geographical region constitutes mainly of vegetation, water bodies, paddy fields, beaches with sand, and soil. Thus, three main classes, i.e. 'vegetation', 'water bodies' and 'sparse vegetation and soil' were defined. Here, the class 'sparse vegetation and soil' covers areas where there are paddy fields, soil and sand. Throughout the rest of this paper, we refer the 'sparse vegetation and soil' class as 'soil'. Thus, the primary intention is to classify the pixels in the selected image region under the relabelled classes, 'vegetation', 'water bodies' and 'soil'. Then, the pixels containing soil are extracted for further analysis in order to find out the regions containing ilmenite. A spectral signature for the composition of soil where ilmenite deposits exist was created using data available in spectral libraries. With the aid of the reference spectral signature created, a correlation analysis was carried out to select a training sample of pixels, which were then used to develop the detection algorithm. The results of the analysis were validated by a field visit, where soil samples were collected from several locations predicted by the algorithm. The lab results confirmed the presence of ilmenite in locations detected by the proposed method.

METHODOLOGY

Dataset

The hyperspectral image that was used in this analysis was captured by the Hyperion sensor (USGS, EO1H1410532005260110PU_SGS_01) attached to the NASA's Earth Observing - 1 (EO - 1) satellite. The Hyperion sensor has the capability of resolving spectral bands from 0.4 µm to 2.5 µm - belonging to visible, near infrared and short wave infrared regions of the electromagnetic spectrum, with a spatial resolution of 30 m and a swath width of 7.5 km. The Hyperion sensor provides for two grating image spectrometers with detailed spectral mapping across all channels with high radiometric accuracy. These two spectrometers, VNIR (Visible and Near-Infrared) and SWIR (Short Wave Infrared) had been used to obtain the hyperspectral images consisting of 242 bands spanning across a range of wavelengths, from 355.59 – 2577.08 nm (Griffin, 2005).

As seen in Figure 1(a), the hyperspectral image subjected to analysis in this paper is a strip along the Northeastern region of Sri Lanka, which covers a geographical area of approximately 7.5×100 km^2. Its true colour version is shown in Figure 1(b). The geographical coordinates of the image are as follows:

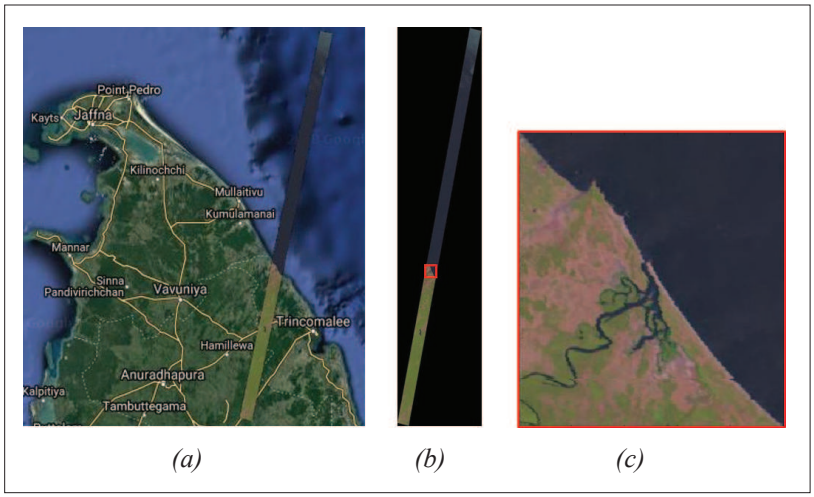


Figure 1: (a) The geographical location of the hyperspectral image strip; (b) true colour version of the hyperspectral image strip (the area chosen for analysis is marked by a red square); (c) the area chosen for analysis - image size 273 × 251 pixels

- Upper left corner -10.146810° N, 81.251940° E
- Upper right corner -10.133729° N, 81.319533° E
- Lower Left corner - 8.1372430° N, 80.818856° E
- Lower right corner - 8.1245150° N, 80.886088° E

The respective hyperspectral image strip was obtained from the USGS database (USGS, EO1H1410532005260110PU_SGS_01). The Level 1Gst (L1Gst) product was used since all radiometric and systematic geometric corrections are performed in the L1Gst product. In addition, a 90-meter Digital Elevation Model (DEM) has been employed in this particular product. The image was available in the form of 242 monochrome Geo 'TIFF' images belonging to each of the 242 spectral bands. The size of the image was 7481 × 1851 pixels and each pixel covered an area of 30 × 30 m^2. The data included in the images were radiance values and were processed by a scaling factor of 40 for the images obtained through the VNIR camera in the range of wavelengths from 400 – 1400 nm (bands 1–70), and by a scaling factor of 80 for the images obtained by the SWIR camera in the range of wavelengths from 900 – 1700 nm (bands 71–242). There were 44 uncalibrated bands among the 242 bands (USGS, EO1H1410532005260110PU_SGS_01).

In order to reduce the time of computation, algorithms were performed only on a selected region - the region displayed in Figure 1(c), which spanned from the rows 5627 to 5801 of the original image.

Preprocessing

Preprocessing is an essential procedure that should be done in remote sensing applications. It plays an important role, being a stage done prior to any form of data analysis and hence is important in hyperspectral image analysis as well. It essentially prepares and structures the dataset to a form, which could be input directly to the algorithm used, in a much efficient and effective manner. Simply, it converts the dataset to a more qualitative and productive form.

Conversion from radiance to reflectance

Hyperspectral image analysis uses reflectance values since it is the property, which varies with the chemical composition of materials. In addition, reduction in between-scene variability can be achieved through a normalisation for solar irradiance. Thus, spectral radiance values were converted to reflectance values by using the following equation (USGS, EO1H1410532005260110PU_SGS_01). The strength of the conversion from radiance to reflectance was acceptable when considering the mean spectral signatures of the three classes as in Figure 2(a) for the spectral bands with high signal to noise ratio (SNR).

$$\rho = \frac{\pi . L_{\lambda} . d^{2}}{ESUN_{\lambda} . cos(\Theta_{s})} \quad ...(1)$$

Where,

ρ = unit less planetary reflectance

L_λ = spectral radiance at the sensor's aperture
d = earth-sun distance in astronomical units
$ESUN_\lambda$ = mean solar exoatmospheric irradiance
Θ_s = solar zenith angle

Spectral radiance at the sensor's aperture are the values already given in the images. Due to the scaling done by USGS when processing the image, it is necessary to rescale them back to their original values. Therefore, the radiance values should be divided by 40 for bands 8–57 and by 80 for bands 77–224 (USGS, 2018). Earth-sun distance in astronomical units for a set of days of the year is given in USGS, (2018). The distance applicable to the day that the respective image was captured can be found by performing a linear interpolation. Mean solar exoatmospheric irradiances relevant to each band is given in USGS, (2018). The solar zenith angle based on location, date and time can be found in the metadata sheets available along with the hyperspectral image from USGS. The solar zenith angle was assumed constant throughout the region of interest because the image spans across a relatively small geographical region and has been captured within a small period of time (4 min 38 s). The image-dependent parameters used for equation (1) are shown in Table 1. Other band-dependent parameters such as mean solar exoatmospheric irradiance could be found at USGS, (2018).

Table 1: Parameters used to convert the image from radiance values to reflectance values

Parameter	Value
Earth-sun distance	1.0051 astronomical units
Solar zenith angle	27.52°

Standardisation of the dataset

The image chosen for analysis consists of reflectance values of spectral bands obtained for each pixel. Therefore, it is obvious that the dataset could be affected due to factors such as different lighting conditions, different atmospheric conditions, measurement errors, and noise. Therefore, the spectral data of identical pixels would have the possibility of being different, which would lead to inaccuracies in the analysis. In order to remove or minimise the effect of such factors that are not characteristics of the contents of the pixel, the dataset was standardised by zero meaning and normalising with respect to the standard deviation (Oorloff *et al*., 2017) using equations (2) to (5). The smiling effect was not taken into consideration. Solely addressing the biases with respect to mean and standard deviation gave accurate results. Thus, the influence of the smiling effect could be considered negligible.

$$data_mat_z(i,:) = data_mat(i,:) - \mu_i \quad ...(2)$$

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} data_{mat(i,j)} \quad ...(3)$$

$$data_{mat_N(i,:)} = \frac{data_{mat_z(i,:)}}{\sigma_i} \quad ...(4)$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[data_mat(i,j) - \mu_i]^2} \quad ...(5)$$

Where,
$data_mat_z(i,:)$ = ith pixel with zero mean spectral data
$data_mat(i,:)$ = ith pixel with original spectral data
μ_i = mean of the ith pixel's spectral data
$data_mat_N(i,:)$ = ith pixel with normalized spectral data
σ_i = standard deviation of the ith pixel's spectral data

Subtracting the mean of spectral data of each pixel from its original data was done in order to remove any biases in pixels. Normalising with respect to standard deviation was done with the intention of minimising the effect of power variation across the spectral bands of each pixel. After performing these calculations, a standardised dataset was obtained to be input to the algorithms developed.

Classification algorithm - sub component analysis

The first stage of the algorithm follows a similar approach to what is discussed by Ekanayake *et al.* (2018) and Vithana *et al*. (2019). A straightforward method, which considers the vector representation of spectral characteristics of a pixel, has been used. After removing the 44 uncalibrated bands, 198 properly calibrated bands were remaining so that a single pixel has a spectral signature representing reflectance values corresponding to 198 bands. Thus, a single pixel can be represented in a 198-dimensional space, which is impossible to illustrate yet mathematically explicable, where each axis represents a spectral band in reflectance units.

The intention of the initial stage was to classify the pixels in the selected image region under its underlying classes. Moreover, the algorithm discussed by Ekanayake *et al.* (2018) identifies the pixels that contain more than one component (mixed pixels) along with their percentages. Since this algorithm requires normalised data, the standardisation mentioned in the earlier section is essential.

Since a large amount of data introduces redundancy and complexity to the analysis, principal component analysis (PCA) (Tyo *et al*., 2003; Raiko *et al.,* 2007; Zabaiza *et al*., 2014) was used to transform the 198-dimensional space into a 20-dimensional space where the variation of the dataset was maximum. To find the transformation matrix in PCA, a training set of 60 pixels was chosen, 20 from each class - soil, vegetation and water bodies. The training set was selected with the aid of Google Earth Pro and the Image Processing toolbox available in MATLAB®. Equations (6) to (10) were used to implement PCA and transform the dataset on to a 20-dimensional space.

$$C = E\{(X_n - \mu)(X_n - \mu)^T\} \quad ...(6)$$

$$C = \frac{1}{N}\sum_{n=1}^{N}(X_n - \mu)(X_n - \mu)^T \quad ...(7)$$

$$CV = \lambda V \quad ...(8)$$

Where,

C = covariance matrix
E = expected value operation
X_n = vector representing the spectral band information of a pixel
μ = vector representing the mean spectral information of all pixels
N = number of pixels in the training sample
V = eigenvector of the covariance matrix
λ = eigenvalue of the covariance matrix

The eigenvalues and the eigenvectors of the covariance matrix were considered in order to construct the transformation matrix. The eigenvectors corresponding to the largest 20 eigenvalues of the covariance matrix were stacked as rows to form the transformation matrix. Thus, the transformation matrix is given by equation (9). Using this transformation matrix, all pixels of the image were transformed to the new reduced dimensional space using the transformation given by equation (10).

$$T = [V_1\, V_2\, V_3 \ldots V_{20}]^T \quad ...(9)$$

$$P1 = TP \quad ...(10)$$

Where,

$V_1, V_2, V_3 \ldots V_{20}$ = eigenvectors of the covariance matrix
T = transformation matrix
P = matrix representation of pixels (in columns) in the original space
$P1$ = matrix representation of pixels (in columns) in the reduced space

After applying PCA, each pixel can now be represented in a 20-dimensional space. Moreover, each pixel can be represented as a vector in a 20-dimensional space. In this 20-dimensional space, the direction of each vector characterises the spectral behaviour of the pixel. However, the magnitude of the vector could vary due to the effects of the environment, sensing equipment, etc. Therefore, the directions of the vectors were considered as the basis for the classification task. In order to normalise the magnitude, each vector was divided by its Euclidean norm using equation (11). Now all the pixels have a unit magnitude and as mentioned above, the spectral characteristics of each pixel are denoted by the direction of the vector representation in the 20-dimensional space.

$$\omega = \frac{w}{|w|} \quad ...(11)$$

Where,

ω = unit vector in the reduced space
w = original vector in the reduced space
$|w|$ = magnitude of the original vector in the reduced space

The reference spectral signatures belonging to each class - soil, vegetation and water bodies, were constructed by considering the means of the training set of each class, i.e. three reference vectors were constructed to represent soil, vegetation and water bodies by taking the 20 training samples of each class and calculating the mean vector. The Euclidean distance between each pixel vector in the image and the three reference vectors were calculated to form a distance vector for each pixel, which is given by,

$$\delta = [|\omega - \omega_s| \quad |\omega - \omega_v| \quad |\omega - \omega_w|]^T \quad ...(12)$$

Where,

δ = the distance vector
ω = unit vector representing a given pixel
ω_s = unit vector representing the reference spectral signature of soil
ω_v = unit vector representing the reference spectral signature of vegetation
ω_w = unit vector representing the reference spectral signature of water

Although the distance vector was calculated, the requirement is to obtain a measure of affinity. Thus, the reciprocal of the distance vector was considered and an affinity vector was calculated, which represented the percentage affinity of a given pixel to the reference spectral characteristics of the three classes using equation (13). The unmixed pixels were selected using a threshold for the largest affinity percentage as in equation (14).

$$\gamma = \begin{bmatrix} \dfrac{\frac{1}{|\omega-\omega_S|}}{\frac{1}{|\omega-\omega_S|}+\frac{1}{|\omega-\omega_V|}+\frac{1}{|\omega-\omega_W|}} \\ \dfrac{\frac{1}{|\omega-\omega_V|}}{\frac{1}{|\omega-\omega_S|}+\frac{1}{|\omega-\omega_V|}+\frac{1}{|\omega-\omega_W|}} \\ \dfrac{\frac{1}{|\omega-\omega_W|}}{\frac{1}{|\omega-\omega_S|}+\frac{1}{|\omega-\omega_V|}+\frac{1}{|\omega-\omega_W|}} \end{bmatrix} \quad ...(13)$$

$$\tau = max(\gamma) \quad ...(14)$$

Where,

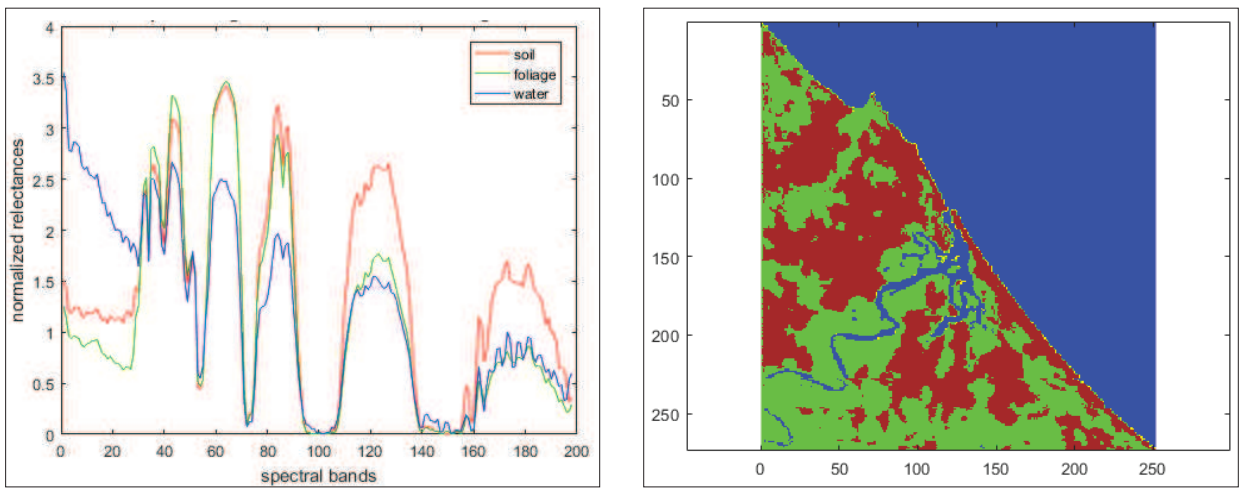
γ = percentage affinity vector

τ = maximum affinity percentage of the affinity vector and the other terms are the same as before.

The classification criteria used were,

- If 50 %, the considered pixel was classified under the class, which has the maximum affinity.
- If 50 %, the pixel was labelled as a 'mixed pixel' and the percentage of each end member (soil, vegetation and water) was obtained by the percentage affinity vector.

Figure 2(b) shows the classified image according to the above criteria. It is clearly seen that these results show an accurate correlation with the true colour version of the image depicted in Figure 1(c). Using the above criteria, the pure unmixed soil pixels were determined. These pixels were noted in order to perform the next step of the algorithm.



(a) (b)

Figure 2: (a) The spectra of the three reference vectors created (b); the classified hyperspectral image under soil (brown), vegetation (green), water bodies (blue) and mixed pixels (yellow)

Identification of probable ilmenite deposits

The coastline along the Northeastern region of Sri Lanka around Pulmudai is rich in mineral sand, especially ilmenite. With the intention of identifying probable ilmenite deposits using hyperspectral images obtained by the Hyperion sensor of the Earth Observing - 1 satellite, an algorithm was developed using correlation coefficient analysis and Fisher's discriminant analysis (FDA). The accuracy of the results was further enhanced by utilising a probability-based approach dependent on the spatial distribution of the ilmenite deposit.

Correlation analysis

According to Premaratne & Rowson (2003), the mineral percentages of the soil in the coast around Pulmudai, which is our target region of analysis is as in Table 2.

Then, a reference signature, which is shown in Figure 3, was constructed to represent the soil, which probably would consist of ilmenite. The construction of the reference spectral signature was done by obtaining the lab spectral signatures of the constituents from USGS spectral library (Kokaly *et al*., 2017) and by weighing

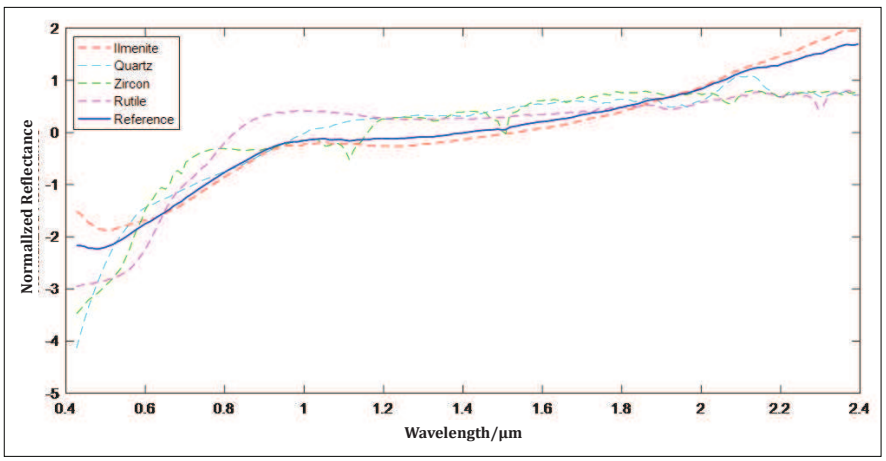
each of the spectra by their respective percentages depicted in Table 2, after normalising to remove any biases in the signatures.

Table 2: Mineral percentages in the Northeastern coast

Mineral	Percentage (%)
Ilmenite	71
Quartz	10
Zircon	10
Rutile	8
Sillimanite	1

However, the lab spectral data generally has a higher SNR compared to the hyperspectral images obtained by the Hyperion sensor. The SNR of the Hyperion sensor depicted in Figure 3(b) shows that the SNR fluctuates with the wavelength whilst having significantly low SNR values at certain ranges of wavelengths. As the data in the regions of low SNR values would not yield useful information and would distort the results, the data in the regions of SNR values below 50 were removed from both the Hyperion dataset as well as the reference signature created.

Ilmenite has unique spectral characteristics, which is explicitly seen even in compounds consisting of ilmenite. This can be clearly seen when comparing the spectral signatures of various types of minerals available in the USGS spectral library (Kokaly *et al*., 2017). Hence, in the process of obtaining a training set of pixels for ilmenite deposits in the region, first, the pixels containing soil were isolated using the classification algorithm



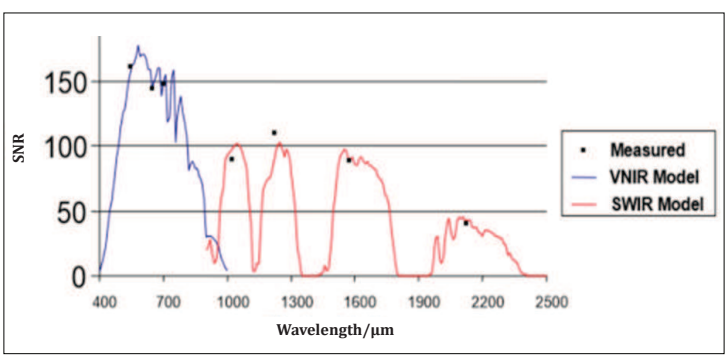


Figure 3 : (a) Reference spectral signature of the ilmenite deposits and its component; (b) signal to noise ratio (SNR) of the hyperion sensor

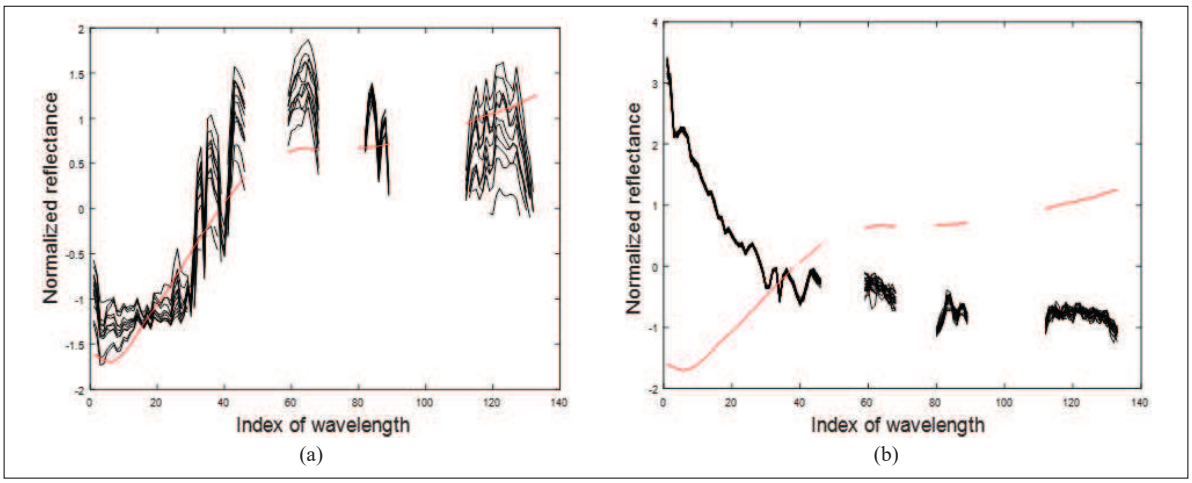


Figure 4: Spectral signatures of the training sets of (a) ilmenite deposits; (b) non-ilmenite soil (the reference signature is marked in red)

discussed in the earlier section of this paper. Then, a correlation coefficient analysis was carried out between the created reference spectral signature and the pure soil pixels classified by the aforementioned algorithm, using equation (15).

$$\rho_c(i) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{data_{mat_N(i,:)} - \mu_i}{\sigma_i}\right)\left(\frac{ref_N - \mu_{ref}}{\sigma_{ref}}\right) \quad ...(15)$$

Where,

$data_mat_N$ = normalised matrix of which the rows correspond to the spectral data of the soil pixel

$\rho_c(i)$ = correlation coefficient between the reference spectrum and the spectrum of the ith soil pixel

ref_N = normalised reference spectrum

μ_i, σ_i = mean and standard deviation of the ith soil pixel

μ_{ref} = mean of the reference spectrum

σ_{ref} = standard deviation of the reference spectrum

K = index of the spectral bands

n = number of spectral bands

Thereafter, based on the correlation values obtained, 20 soil pixels that were highly correlated with the reference spectral signature were taken as the training set of ilmenite deposits, and the 20 least correlated pixels were taken as the training set for non-ilmenite deposits. The spectral signatures of the training sets are shown in Figure 4, and it could be seen that the training sets of ilmenite deposits and non-ilmenite soil pixels are distinct.

Transforming to a new space

Increasing the separability of classes would ease classification in the process of identifying the probable ilmenite deposits. Hence, Fisher's discriminant analysis was used to transpose the pixels on to a new space in which the separability of the two classes - ilmenite and non-ilmenite, is increased. Fisher' discriminant analysis (Imani & Ghassemian, 2015; Sugiyama, 2016; 2017) was applied on the aforesaid training set using equations (16) to (19). The idea of using Fisher's discriminant analysis is to transform the dataset into a new space where the scatter between the two classes is maximised while the scatter within a class is minimised. The equations below were used to construct the Fisher's transformation matrix.

$$S_b = \sum_{i=1}^{l} n_i(\mu_i - \mu)(\mu_i - \mu)^T \quad ...(16)$$

$$S_w = \sum_{i=1}^{l}\sum_{j\in i}(x_j - \mu_i)(x_j - \mu_i)^T \quad ...(17)$$

$$\mu_i = \frac{\sum_{j\in i} x_j}{n_i} \quad ...(18)$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \quad ...(19)$$

Where,

S_b = between-class scatter matrix
μ = vector representing mean spectral information of all pixels
μ_i = vector representing mean spectral information of the pixels of class i
n_i = number of training pixels in class i
l = number of classes
S_w = within-class scatter matrix
x_j = vector representing the spectral information of a pixel in the training class i
x_i = vector representing the spectral information of a pixel in the training set
n = total number of training pixels

Since it is required to minimise the within-class scatter and increase the between-class scatter, the transformation matrix was constructed by considering the eigenvalues and eigenvectors of the Fisher matrix given by equation (20). Using these eigenvectors, data were transformed into a new space with reduced dimensionality using equations (21) and (22).

$$S_w{}^{-1} S_b \varphi = \lambda \varphi \quad ...(20)$$

$$\Gamma = [\varphi_1\ \varphi_{2.....}\varphi_n]^T \quad ...(21)$$

$$V_2 = \Gamma V_1 \quad ...(22)$$

Where,

λ = eigenvalues of the Fisher matrix $S_w{}^{-1} S_b$,

φ = eigenvectors of the Fisher matrix $S_w{}^{-1} S_b$

Γ = transformation matrix

V_1 = matrix representing all the pixels in original space

V_2 = matrix representing the pixels in the new space

$\varphi_1\ \varphi_{2.....}\varphi_n$ = real eigenvectors of the Fisher matrix

Subsequently, all soil pixels were transformed into the new space using the aforementioned transformation matrix. This transformation resulted in the increase of the scatter between the classes of ilmenite and non-ilmenite while reducing the scatter or in other words contracting each of the clusters of ilmenite and non-ilmenite. Finally, each pixel vector was normalised with respect to its magnitude using equation (23).

$$x_N = \frac{x}{|x|} \quad ...(23)$$

Where,

x = a pixel vector
$|x|$ = magnitude
x_N = calculated normalised pixel vector

In order to classify the pixels into the two classes, the spectral similarity of each pixel with the mean spectral signature of the training set of each class was calculated using equation (24).

$$\alpha_{ik} = \frac{1}{||data_mat_N(i,:) - mean_{ref_k}||} \quad ...(24)$$

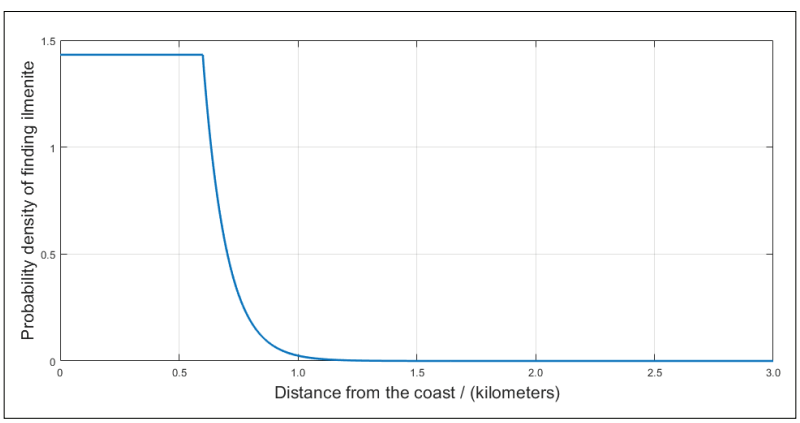


Figure 5: Probability density function of the probability of finding ilmenite against the distance measured from the coast

Where,

α_{ik} = spectral similarity of the ith pixel's spectral signature and the kth class (k = 1 for ilmenite and k = 2 for non-ilmenite)

$data_mat_N(i,:)$ = the normalized spectral signature of the ith pixel

$mean_ref_k$ = mean reference spectral signature of the training set of the kth class

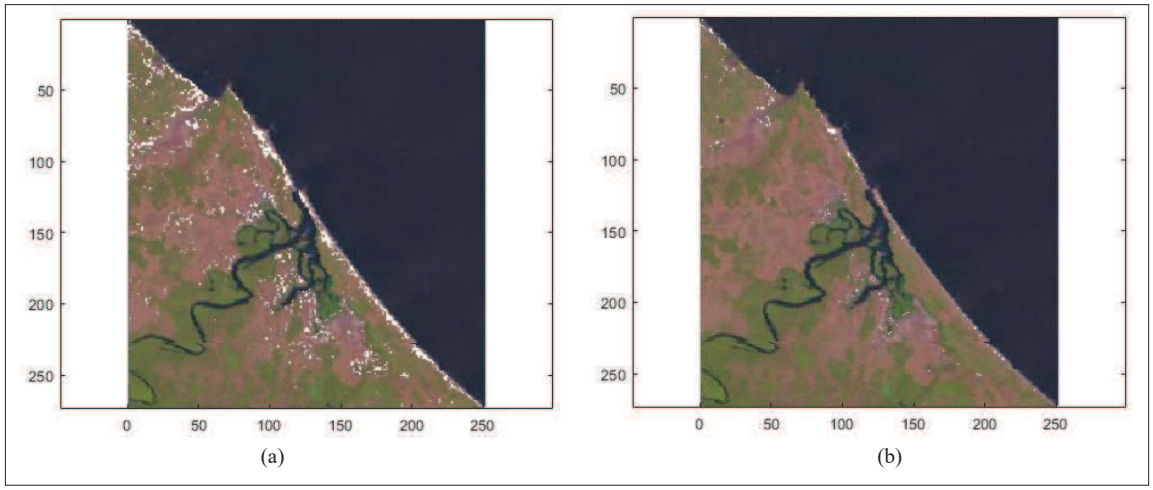


Figure 6: Recreated map of the region identifying probable ilmenite deposits (a) based on the spectral similarity measure, α (identified pixels are depicted in white); (b) based on the novel probability-based approach considering the joint measure, β (identified pixels are depicted in white)

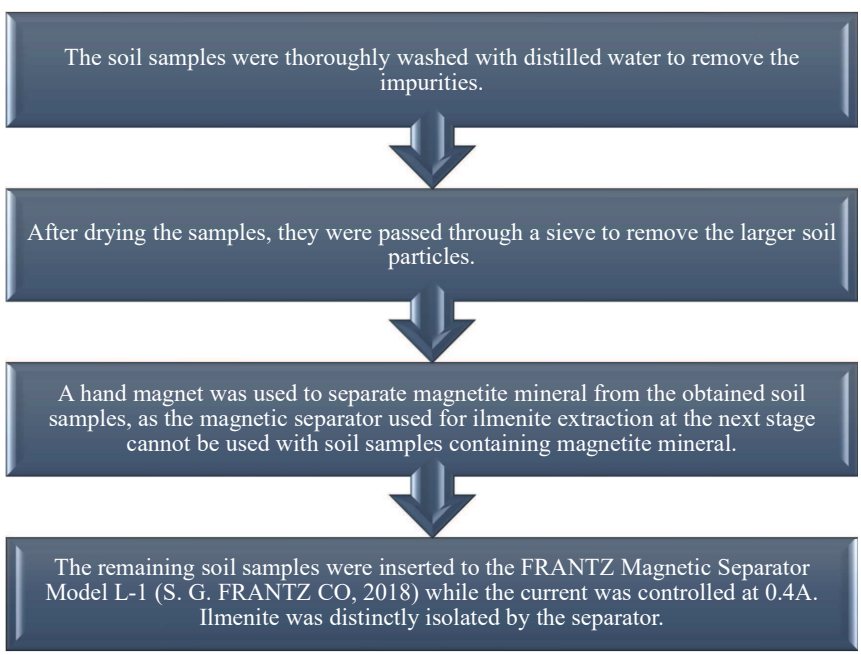


Figure 7: Block diagram of steps followed to extract ilmenite from the soil samples

Finally, each pixel was labelled as ilmenite or non-ilmenite based on the spectral similarity measure, α. If the spectral similarity between a certain pixel and the mean reference spectral signature of ilmenite class was higher than the spectral similarity between that particular pixel and the mean reference spectral signature of non-ilmenite class, then that pixel was classified under 'ilmenite', and *vice versa*.

Probability-based approach

In the Pulmudai region, ilmenite has a larger probability to be found in the coastal area. Statistically, there is a larger probability within 600 m from the coast, and from there onwards, the probability of finding ilmenite falls gradually. Thus, a probability density function was approximated in such a way that the probability of finding ilmenite was constant until 600m. From there onwards, the probability was approximated to fall exponentially as shown in Figure 5. The spatial information used to generate the probability density function can be found in Overstreet (1972).

Using this novel probability-based method, the classification was done while considering the probability of finding ilmenite in each pixel as well. Thus, it is expected to obtain results with more bias towards the coastal region of the geographical area and less bias as the distance increases from the coast.

In order to classify the pixels in the region into the two classes, using this novel probability-based approach, a similar method was used as earlier, but with a slight modification. This time, the probability to find ilmenite at each pixel was also considered. First, as in equation (24), the spectral similarity measure, α was calculated. Then equation (25) was used to calculate the novel joint measure.

$$\beta_{ik} = P_i \alpha_{ik} \quad ...(25)$$

Where,

$\beta_{ik} =$ A joint measure of the spectral similarity of the ith pixel's spectral signature and the kth class (k = 1 for ilmenite and k = 2 for non-ilmenite) and the probability of finding ilmenite on the region denoted by the ith pixel

$\alpha_{ik} =$ spectral similarity of the ith pixel's spectral signature and the kth class (k = 1 for ilmenite and k = 2 for non-ilmenite)

$P_i =$ Probability of finding ilmenite on the region denoted by the ith pixel

Finally, each pixel was labelled as ilmenite or non-ilmenite based on the joint measure, β. If the joint measure between a certain pixel and the mean reference spectral signature of the class, ilmenite was higher than the joint measure between that particular pixel and the mean reference spectral signature of the class, non-ilmenite, then that pixel was classified under 'ilmenite', and *vice-versa*.

RESULTS AND DISCUSSION

The results generated using spectral similarity measure, α was mapped as shown in Figure 6(a). The results generated using the novel probability-based algorithm considering the joint measure, β was mapped as shown in Figure 6(b). As expected, the probability-based approach based on the spatial distribution of the ilmenite deposit has given more bias towards the coastal region of the geographical area and less bias as the distance increases from the coast. Thus, the results obtained from the probability-based approach were taken into consideration for validation.

In order to check whether the identified pixels actually contained ilmenite, soil samples were collected from a few locations recognised by the probability-based algorithm. A total of eight locations were selected from the map in Figure 6(b), and their soil samples were taken for lab testing. The steps depicted in Figure 7 were followed during the testing process.

The weight measurements taken during the testing were used to calculate the ilmenite percentage of each sample. The results are shown in Table 3.

The areas, which were not identified to exist ilmenite from the probability-based algorithm showed no visual observations of existence of ilmenite. Thus, soil samples from those areas were not taken for laboratory testing.

Hyperspectral imaging is an emerging technique used for feature extraction and classification using hyperspectral image data. This stands out from other image processing techniques due to its ability to identify fine features of an image, which is not detectable by other image processing techniques. This is mainly used in land cover mapping, facial recognition, agriculture, and military applications.

Table 3: Results of the lab testing on the soil samples collected from the locations near Pulmudai

Sample No.	Coordinates	Snapshot	Ilmenite percentage from lab testing
1	8.890481° N, 81.031436° E		0.59 %
2	8.891395° N, 81.030864° E		1.53 %
3	8.891396° N, 81.028555° E		12.27 %
4	8.895059° N, 81.028557° E		10.73 %
5	8.895081° N, 81.009026° E		1.95 %
6	8.894307° N, 81.011228° E		3.80 %
7	8.8921107° N, 81.0159079° N		5.70 %
8	8.897387° N, 81.005462° E		8.17 %

In this paper, the main focus was to develop a universal algorithm to detect ilmenite mineral based on the unique spectral characteristics of the areas where the particular mineral is located. Based on that, an algorithm was built and the lab testing proves that the locations detected by the novel algorithm actually contained ilmenite.

The classification algorithm used in classifying the pixels containing soil was a combination of PCA and comparison of affinity based on Euclidean distance. The classification process resulted in an image similar to the true RGB image. It could further be concluded that the part of the image of a strip in Northeastern Sri Lanka is mainly covered by soil, followed by foliage. Water bodies are much less in the region of consideration.

In the second part of the algorithm, the soil only pixels were solely taken into consideration. A reference spectral signature for ilmenite deposits was created using lab data and it was compared with the pixel signatures of the image. Using a correlation coefficient analysis, a training set was formed by including the highest correlated and the lowest correlated soil only pixels. Then Fisher's discriminant analysis was performed to further separate the two classes – ilmenite, and non-ilmenite and the classification was done by comparing the spectral similarity of each pixel with the mean spectral signatures of the two classes. In order to further enhance the results, a probability-based approach was utilised. A probability density function was designed to determine the probability of finding ilmenite with the distance

measured from the coast. Results were generated taking the probability of finding ilmenite into account as well. Eight locations were chosen from those results and soil samples were collected for lab testing, which ultimately proved that all those eight locations contained ilmenite.

A mineral sands deposit, which is suitable for mineral extraction typically has a heavy mineral grade (HM grade) ranging from 0.5 % to above 20 %. It was observed that all eight locations considered for testing contained ilmenite within that range. Some locations contained very high relative percentages of ilmenite (sample no. 3 – 12.27 % and sample no. 4 – 10.73 % as seen in Table 3). These amounts are suitable for the extraction of the mineral for commercial use. Thus, this indicates that the algorithm established gives accurate results in finding probable ilmenite deposits, which could be used for commercial purposes.

Prior evidence suggested of an ilmenite deposit in the Pulmudai region of Sri Lanka. Thus, the study was based on that particular region only. In addition, due to the lack of hyperspectral databases, the number of satellite hyperspectral images covering the areas in Sri Lanka are minimal. This prevented us from looking into the adjacent areas for the possibility of the existence of ilmenite.

CONCLUSION

The objective of developing an algorithm to detect ilmenite deposits in Pulmudai region using hyperspectral images is successful. This novel algorithm, which is based on feature reduction and classification methods, signal processing techniques and statistical and mathematical tools, proves to give highly accurate results. Lab testing performed on the collected soil samples further validates the results.

Acknowledgement

The authors are grateful to the USGS for providing hyperspectral images from the EO-1 satellite's Hyperion sensor, and relevant documents for standard preprocessing of the images. The authors also acknowledge the Department of Geology, University of Peradeniya and Dr N.W.B. Balasuriya of the Department for making arrangements and providing equipment to perform lab testing on the soil samples to extract ilmenite.

REFERENCES

Aravinth J. & Roopa S. (2017). Identifying traces of copper in Basavakote, Karnataka using hyperspectral remote sensing. *International Conference on Technological Advancements in Power and Energy (TAP Energy),* Kollam, pp. 1– 6. DOI: https://doi.org/10.1109/TAPENERGY.2017.8397231

EarthExplorer-Metadata (2018). *Full Display of EO1H 1410532005260110PU_SGS_01.* Available at https://*earthexplorer.usgs.gov/metadata/1854/EO1H 14105320052 60110PU_ SGS_01/, Accessed 6 October 2018.*

Ekanayake E.M.M.B., Ekanayake E.M.H.E.B., Rathnayake A.R.M.A.N., Vithana S.S.P., Herath H.M.V.R., Godaliyadda G.M.R.I. & Ekanayake M.P.B. (2018). A semi-supervised algorithm to map major vegetation zones using satellite hyperspectral data. *9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS),* Amsterdam, The Netherlands, pp. 1– 5. DOI: https://doi.org/10.1109/WHISPERS.2018.8747025

Ekanayake E.M.M.B., Rathnayake Bhathiya, Ekanayake E.M.H.E.B., Rathnayake A.R.M.A.N., Herath H.M.V.R., Godaliyadda G.M.R.I. & Ekanayake M.P.B. (2019). Enhanced hyperspectral unmixing *via* non-negative matrix factorization incorporating the end member independence. *2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Yokohama, Japan.

Griffin M.K., Hsu S.M., Burke H.K., Orloff S.M., Upham C.A. & Misra B. (2005). *Examples of EO-1 Hyperion Data Analysis*. Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, USA.

Imani M. & Ghassemian H. (2015). Feature space discriminant analysis for hyperspectral data feature reduction. *ISPRS Journal of Photogrammetry and Remote Sensing* **102**: 1–13. DOI: https://doi.org/10.1016/j.isprsjprs.2014.12.024

Kokaly R.F. *et al*. (11 authors) (2017). *USGS Spectral Library Version 7. U.S. Geological Survey Data Series*. Available at *https://pubs.er.usgs.gov/publication/ds1035*, Accessed 6 October 2018. DOI: https://doi.org/10.3133/ds1035

Laakso K., Peter J.M., Rivard B. & Gloaguen R. (2016). Combined hyperspectral and lithogeochemical estimation of alteration intensities in a volcanogenic massive sulfide deposit hydrothermal system: a case study from Northern Canada. *8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Los Angeles, USA, pp. 1–5. DOI: https://doi.org/10.1109/WHISPERS.2016.8071707

Niranjan S.S., Chaube N. & Sarup J. (2016). Mapping of the carnallite mineral and sagebrush vegetation plant by using hyperspectral remote sensing and USGS spectral library. *8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Los Angeles, USA, pp. 1–5. DOI: https://doi.org/10.1109/WHISPERS.2016.8071778

Notesco G., Ben Dor E. & Brook A. (2014). Mineral mapping of makhtesh ramon in Israel using hyperspectral remote sensing day and night LWIR images. *6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS),* Lausanne, Switzerland, pp. 1–4. DOI: https://doi.org/10.1109/WHISPERS.2014.8077538

Oorloff T.S.J., Abeysekara A.M.R., Vithana S.S.P., Rupasinghe R.A.A., Herath H.M.V.R., Godaliyadda G.M.R.I. & Ekanayake M.P.B. (2017). Spectral-spatial hybrid mechanism for feature detection using spectral correlation. *IEEE International Conference on Industrial and Information Systems (ICIIS),* Peradeniya, pp. 1–6. DOI: https://doi.org/10.1109/ICIINFS.2017.8300371

Overstreet W.C. (1972). *Ilmenite Beach Placer, Pulmudai, Ceylon and Ilmenite Resources of the Pleistocene and Holocene Formations of Ceylon*. U.S. Geological Survey-Open File Report. pp. 7–10. U.S. Geological Survey, Virginia, USA. DOI: https://doi.org/10.3133/ofr72277

Premaratne W.A.P.J. & Rowson N.A. (2003). The processing of beach sand from Sri Lanka for the recovery of titanium using magnetic separation. *Physical Separation in Science and Engineering* **12**(1): 13–22. DOI: https://doi.org/10.1080/1478647031000101232

Raiko T., Ilin A. & Karhunen J. (2007). Principal component analysis for large-scale problems with lots of missing values. *European Conference on Machine Learning (ECML 2007),* Warsaw, Poland, pp. 691–698. DOI: https://doi.org/10.1007/978–3–540–74958–5_69

S.G. Frantz Co. (2018). *Magnetic Separators-L-1 Isodynamic Laboratory Separator*. Available at *http://rivashost.com/subdomains/sgf/index.php/laboratory–separators/l–1–isodynamic–separators*, accessed 6 October 2018.

Saralıoğlu E., Görmüş T. & Güngör O. (2016). Mineral exploration with hyperspectral image fusion. *24th Signal Processing and Communication Application Conference (SIU),* Zonguldak, Turkey, pp. 1281–1284. DOI: https://doi.org/10.1109/SIU.2016.7495981

Sugiyama M. (2016). Local fisher discriminant analysis for supervised dimensionality reduction. *23rd International Conference on Machine Learning,* Pittsburgh, Pennsylvania, USA, pp. 905–912.

Sugiyama M. (2017). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research* **8**: 1027–1061.

Swamy S., Asutkar S.M. & Asutkar G.M. (2017). Remote sensing HSI classification and estimation of MIMETITE mineral spectral signatures from ISRO, India. *International Conference on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, India, pp. 1095–1099. DOI: https://doi.org/10.1109/ICOEI.2017.8300880

Tyo J.S., Konsolakis A., Diersen D.I. & Olsen R.C. (2003). Principal-components-based display strategy for spectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* **41**(3): 708–718. DOI: https://doi.org/10.1109/TGRS.2003.808879

USGS (2018). *Earth Explorer-Home*. Available at *https://earthexplorer.usgs.gov/*, Accessed 6 October 2018.

USGS, EO1H1410532005260110PU_SGS_01. *Earth Observing-1*. Available at *https://eo1.usgs.gov/*, Accessed 6 October 2018.

Vithana S.S.P., Ekanayake E.M.M.B., Ekanayake E.M.H.E.B., Rathnayake A.R.M.A.N., Jayatilaka G.C., Herath H.M.V.R., Godaliyadda G.M.R.I. & Ekanayake M.P.B. (2019). Adaptive hierarchical clustering for hyperspectral image classification: umbrella clustering. *Journal of Spectral Imaging* **8**: Article ID a11. DOI: https://doi.org/10.1255/jsi.2019.a11

Zabaiza J., Ren J., Yang M., Zhang Yi., Wang J., Marshall S. & Han J. (2014). Novel folded-PCA for improved feature extraction and data reduction with HSI and SAR in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* **93**: 112–122. DOI: https://doi.org/10.1016/j.isprsjprs.2014.04.006

Zhao H., Zhang L., Zhao X., Yang H., Yang K., Zhang X., Wang S. & Sun H. (2016). A new method of mineral absorption feature extraction from vegetation covered area. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* Beijing, China, pp. 5437–5440. DOI: https://doi.org/10.1109/IGARSS.2016.7730416