Research Paper

Application of Remote Sensing and Machine Learning Methods for Coal Seam Distribution Mapping at Lower Benue Trough

Okeke Francis Ifeanyi[1], Ibrahim Adesina Adekunle[2], Echeonwu Emmanuel Chinyere[3]

[1]*(Department of Geoinformatics and Surveying, University of Nigeria, Enugu Campus, Enugu, Nigeria)*
[2]*(Department of Geoinformatics and Surveying, University of Nigeria, Enugu Campus, Enugu, Nigeria)*
[3]*(Machine Intelligence, African Institute for Mathematical Sciences, Rwanda)*
Corresponding Author: Ibrahim Adesina Adekunle

ABSTRACT : *Attempt was made to map coal seam area distribution based on remote sensing and machine learning techniques in this work through the use of Landsat 8 OLI at lower Benue trough, Nigeria. Techniques such as band ratio, Principal Component Analysis (PCA) and Crosta were applied in the former to map ferric iron and hydroxyl oxide mineral locations amenable to coal formations. Crosta method gave the best results of the three. The two: ferric iron and hydroxyl oxide image maps, resulting from that were combined to form a new image. Composite image of the three was rendered in RGB format, and Spectral Angle Mapper (SAM) classification was conducted. Machine learning method such as Support Vector Machine, Ensemble (AdaBoost and Gradient Descent) and Neural Network (mlp) were engaged to predict areas likely to contain coal seams based on surface feature characteristics. Surface coal mine areas were successfully mapped with SAM classification of the Crosta technique, while other areas suggestive of coal presence were successfully mapped with Machine learning predictive method, checked by using F1-score metric. Validation was carried out based on the knowledge of existing coal seam locations, exploitation sites, coal outcrop areas and ancillary data such as the mineral map, geological map, historical information of the area, including past field campaign information and site visit. NDVI, lineament and temperature maps were also produced. Envi 5.1, Erdas Imagine 2015, ArcMap 10.3, paint and Python API were employed for remote sensing image processing, ASCII data handling, map production, map presentation and machine learning programming interface respectively.*
KEYWORDS: *Coal, Remote sensing, Crosta technique, Alteration maps, Spectral Angle Mapper, Machine learning.*
Received 18 December, 2020; Accepted 02 January, 2021 © The author(s) 2021.
Published with open access at www.questjournals.org

I. INTRODUCTION

Coal is a fossil fuel used as a source of energy. It has between 25% and 95% carbon, with the balance from Hydrogen, Sulphur, oxygen and nitrogen in varying degrees. Other chemical constituents such as silica, aluminium, iron, magnesium, and little quantity of calcium and sodium can be readily observed. The composition essentially depends on the quality, source, and combustion condition [1]. Silver is a volatile constituent which separates at very high temperature [2]. A thorough Heat value, Atomic absorption (Wet Chemical Analysis), Optical Emission Spectrographic, X-ray Fluorescence, and Neuron Activation Analyses respectively can yield 7, 35, 7, 2 and 19 different components of inorganic elements and oxides according to [3]. Coal fly ash minerals consist of three types of components [4], which are crystalline minerals, unburnt carbon particles and non-crystalline minerals. Quarts, Mullite, cristobalite, magnetite and hematite are regarded as the most common crystalline phases. These are essentially ferric iron minerals common in higher rank coals. Others that may be present include oxides and hydroxyl minerals such as lime (CaO) portlandite ($Ca(OH)_2$), periclase (MgO), Calcite and calcium aluminate minerals from calcium-rich, lower coal seams [4].

Coal is found abundantly in many parts of the world, including Nigeria. The first discovery in Nigeria was made in 1909 at Udi in Enugu State by the Mineral Survey of Southern Protectorate established by the British government in 1904, primarily for oil search. Another discovery was made at Ogbete drift mine shortly

after. The success at these areas and others rallied the formation of Nigeria Coal Corporation (NCC) in 1950. NCC activities then began to spread across the entire Benue trough which comprises Upper, Middle and Lower Benue troughs. Anambra basin within the Lower Benue trough contains the richest and the most economically viable coal in Nigeria which is Bituminous coal. The basin, subdivided into seven mining districts for administrative purpose has Kogi coal district as one of them. Unfortunately, exploration, and exploitation activities which took off on a good foothold began to move in a downward trend sooner than later in the 1950s when oil discovery was made. The challenge was compounded during the civil war between 1967 and 1970 because all coalfields were abandoned. Immediately after the war, the Nigeria Railway Corporation (NRC) and the National Electric Power Authority (NEPA), the two major users of coal, diverted to use of diesel and gas respectively. As a result, coal exploration and exploitation collapsed. Before this period, only two out of seven areas in Kogi coal district were worked on, and even to a small degree. The huge gap which remains to be covered has therefore been left unfilled in coal exploration, in Kogi district.

In the past, the method of exploration engaged was crude and inefficient. It was often carried out on a point-by-point investigation which was prone to delay, inefficiency, gross high-cost due to strenuous, and low technological base. Satellite Remote Sensing techniques which have the capacity to delineate minerals, supported by ancillary data such as geological, geochemical, geophysical, historical, aeromagnetic etc., offer a better alternative for large area coverage. In areas where coal already exists, predictive method of a survey which involves machine learning can be suggested. Results from the two techniques can be integrated using Geographic information System to delineate coal seam potential sites, for further confirmation.

Mineral mapping and, specifically coal, has been redefined to entail evaluation of new, or existing deposits to determine quality, quantity, and geological activities which can affect both its exploitation and use [5]. The reflectance properties of coal were regarded to be very low (5 – 30% lower) compared to adjacent rocks [6]. Coal outcrops have special dark hue and bedded texture in the visible band of electromagnetic spectrum, with a high-heat thermal abnormality, only in the daytime thermal infrared images. Night time heat abnormality takes place only at coal fire area. Combination of visible bands, and thermal infrared, therefore, can be helpful in coal seam discovery wherever geological formation approves such. Chemical constituents of coal favours vegetation growth so, exposed areas along faults, weathered soil, and eroded areas will be more suitable for vegetation growth and can be looked out for. Investigating the topography of the area will therefore be very essential couple with the spatial distribution of vegetation. TM, ETM, ASTER, SPOT, and Rapid eye or IRS-1C are often used as the information source of remote sensing to carry out the work.

[7] Successfully applied Landsat TM and ETM+ data together with Radar satellite data in obtaining spatial trial dataset for prospecting coal seam which was validated through field test across Western Mozambique, Eastern Kalimantan, Northern Mozambique, Bangladesh, Central Myanmar, the eastern region of India and southern Laos. It is important to explore any available materials and ideas from all related professionals such as Surveyors, Mining Engineers, Geologists, Computer Engineers, Excavation Engineers as well as multi-systems (temporal, spatial, angular, spectral, sensors), land use, vegetation, temperature, etc. for a successful exploration. Field mapping, interpretation of air photos and satellite images, airborne and ground geophysical surveys data such as gravity, magnetic and seismic can be helpful [8]. These can be followed by ground-truth and drilling campaign to determine features of associated rock, seam thickness, quality and structure, and groundwater distribution for mine and environmental sustainability planning and management.

Ability to map earth's objects including minerals through digital processing of reflected, emitted or backscattered electromagnetic radiation from earth's features has been revolutionized. Useful results are adequately achieved and in time, through the interplay of spatial, spectral, temporal, and radiometric resolutions of satellite images. Landsat 8 OLI-TIRS is the latest in the category of image data provided free by the United States Geological Survey (USGS) at a moderate 30m spatial resolution except panchromatic band 8 and the two thermal bands 10 and 11 which are respectively 15m and 120m out of the total number of 11 bands. Whereas bands 1 to 7 are optical, band 6 is short wave infrared (SWIR). Optical bands provide avenues to view earth's objects in multispectral form by combining three different image bands of different wavelengths. Infrared bands allow discrimination of vegetation while the thermal portion allows heat, and temperature discrimination which can be used to separate coal seam from surrounding rocks.

Coal outcrop can be revealed by indication of spectral reflectance as the lowest against other rocks in coal-bearing areas [9]. It has a peak reflectance between the regions of 0.8 and 0.95 micrometre in infrared which decreases between region 0.95 and 1.10 micrometre, which is distinguishable in spectral reflectance. Image transformation activities such as image ratioing can reduce the effect of vegetation, seasonal effects, slope, and illumination angle on the image to aid discrimination of hidden objects not properly revealed in an image; Principal Component Analysis (PCA) reduces the dimension of the details and their repetition by presenting only important discrete and non-correlating features to delineate alteration zones; Crosta technique reduces the likelihood of mapping certain objects while aiding measurement of others by reducing the number of spectral bands input for principal components. It is regarded as a special form of PCA.

Remote sensing image processing and analyses involve the use of image processing software which include but not limited to IlWIS, Erdas Imagine, and Envi. The choice of which to apply is dictated by the strengths and weaknesses of an individual. No one software is a fit-for-all. Selection can also be based on performances and, or ability to automate some procedures which are backed by requisite statistical tests. Apart from digital processing such as radiometric correction which affects the brightness of an image at the sensor, atmospheric correction which eliminates intervening atmospheric effects of electromagnetic radiation received also at the sensor, possible image transformation by software and classification of image pixels which contain mineral can be carried out. Results of activities engaged at this level lead to the production of classified image layers to further be used for ground truth verification and in conjunction with other ancillary data such as elevation, geologic, gravity etc.

Classification methods have continued to advance from the rudimentary statistical-based techniques to artificial intelligence system [10]. The assumptions such as the correlation matrices and class density function the former rely on are not actually true in real situations and they contribute to its abandonment for advanced methods (Paul, 2013). Advanced classification techniques found in machine Learning such as the Neural Networks (NN), Support Vector Machine (SVM), Decision Trees (DT), Ensemble and their variants have been discussed by various authors such as [11],[12],[13],[14],[15],[16],[17]. On the note are their robustness, high accuracy, attainment of fuzzy class, non-parametric, ability to attain reliable result with little training sample, ability to handle subpixels categorization, directional data such as slope without stratification and, ability to integrate multi-specific and ancillary data. The outlined merits are functions of the chosen classification type for, no single classification method is all-encompassing [18]. The challenge which experts encounter is the availability of the software architecture to handle advanced classification. Most existing stable software which contained a few, is commercial and in most cases the choice cannot be made because they are to handle specifics. The method by which accuracies of the algorithms are assessed may not be understood and sometimes limited. One other major challenge with them is the restriction in flexibility and the inability of the users to tune the functions to their needs. This work also demonstrates the applicability of machine learning in coal seam area forecast.

Machine learning is an automated detection of meaningful patterns in data [19],[20]. Ensemble Learning offers the opportunity to combine models and earn their merits at once. Good understanding of calculus, linear algebra and probability are essential to navigate its well-discussed theory. Machine learning has three main phases which are the data preparation, learning and prediction [21] as seen in figure 1.0. It is used in solving two major problems- classification and regression for both supervised and unsupervised applications. Classification, $f : X \mapsto [0, 1, 2, 3, ... N]$ defines discrete output ($0, 1 ... to\ N$) which represents the most likely class an input X belongs, while regression $f : X \mapsto [-1.6, 2.2, 4.3, ...]$ connotes a continuous output of mapping N dimensional input vector X to an M directional vector y. Apart from Supervised learning which requires that features are linked to classes and, unsupervised learning which does not, there are also Semi-supervised, Reinforcement, and Deep Learning [2]. Machine learning techniques have been applied in many fields of human endeavours such as in facial recognition, spam mail separation, Autonomous Vehicle Driving (AVD), medical diagnosis (MD), Stock Exchange profiteering, marketing, sales, feature extraction, etc. to mention but a few.

For a supervised learning case, the algorithm defines each instance in a scenario with respect to the class each belongs to using a defined probability function. The contribution of each of the features can be evaluated during the preprocessing stage by the analyst so that features with less or no indicative contribution can be removed. Scaling and smoothing of the features also take place at this stage. A decision boundary will be created between the classes. There are possibilities of several decision boundaries but choosing the best is a task that requires understanding the dictates and type of the data one is dealing with, from simple to complex. It will involve selection of the best of available algorithms or its variants which should give the widest separating distance (hyper plane) between the classes. In that case a model which leads to the least error will be desired as result, and this will involve tuning the parameters of the algorithm being used to minimize some errors. The fear is that a model may fit input data with the least or no error but when used to predict unknown set of data, it may fail woefully. This results in what is called overfitting as a result of overtraining of training data. In the other way round, a very simple decision boundary, if used, may also lead to underfitting. What is therefore done is to strike a balance by applying a less complex decision boundary even if it does not perfectly remove the error to the desired level. In some cases, however, simple decision boundary may be most preferred as highlighted by "Occam's razor". The next stage is to generalize the decision rule for all the instances and when that is

accomplished, a new model which will be used to predict unknown instances will be realized. Generalization is the ability of a resulting model to correctly predict a new set of unseen data.

Performance tests are carried out on the generalized model and in doing so, the train data should not be used. The train data is already known to the predictor because it took place in the learning process, it will bias the test result. A test data is created for the purpose instead, especially when there are substantial training data. A random split of the data can be carried out to avoid skewness. While a part is used as the training data, the remainder will be used as a test. In a situation where the training data is small, the best recommended practice is to use K-fold cross-validation. K represents the number of partitions the training data is divided randomly. While K-1 part of the data is used as the training, the remaining one part is used as the test and the performance is noted. The test part is returned and the process is repeated for another part. This continues until each one part in the K-fold is used. The performances are summed, and the mean computed. The mean value will represent the average performance. As good as the method is applauded; consideration is given to the computation cost regarding timing and memory space. Receiver Operating Curve (ROC) and F1-score computed from Confusion matrix can be realised to view the performance. The entire training process and the test for unseen data can be represented in a flow diagram in figure 1.0.
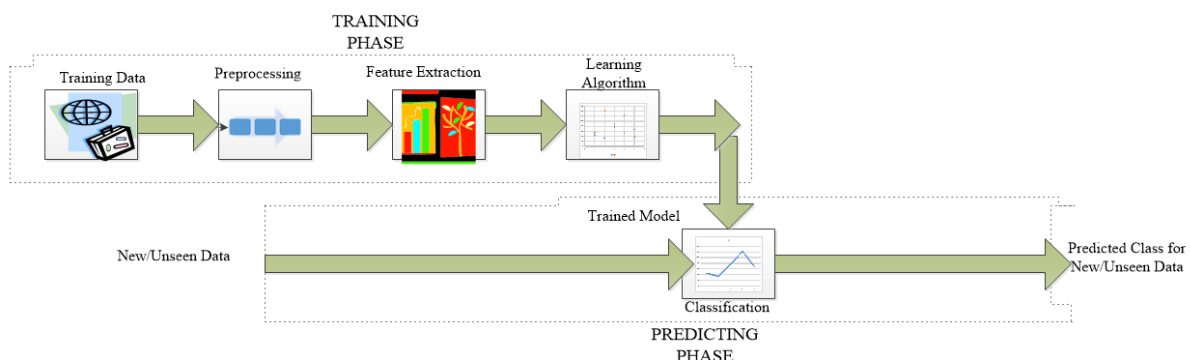

Figure 1.0. Machine Learning and Predicting Process

The entire process in figure 1.0 is automated in machine learning through any of stable and reliable Application Programming Interface, API. It is to note that these programming architecture are basically grouped into Commercial, Opensource and Commercial-with-Opensource editions [20]. Notable among the first group are MATLAB, Amazon Machine Learning, Angoss Knowledge studio, Databrics, Neural Designer, Neuro Solutions, Oracle Data Mining, Rcase etc; the second group has Scikit Learn with Python as the platform, R Programming, Weka, Open CV, ND4JwithLearning4J, Open NN, Orange, MLPack, H2O, OpenNN, Yooreekaetc, etc. The last group has KNIME and OpenMiner.

Understanding of the process requires some form of the inner workings of the entire system thus, for Supervised learning under consideration, we have:

For a given N number of instances in a training data set, the features and the target are expected to read [23]

$$
\begin{cases}
(x_1, y_1) \\
\quad . \\
\quad . \\
\quad . \\
(x_N, y_N
\end{cases}
$$
……………………………………………………………………………………..… 1

Where $x \ and \ y$ = the features and the target (class)

The learning algorithm will seek a function $g \ or \ f$ to model y with respect to the x in X input space and the Y output space, thus:

$$g : X \rightarrow Y$$
……………………………………………………………………………….…..2

g represents an element within the possible number of functions in the G space.

In other words:

$$f: X \times Y \to \mathbb{R}$$
....3

$$g(x) = arg\,max_y\ f(x,y)$$
....4

With g giving the highest score of function f within scoring function space, F.

If the learning algorithm gives a probabilistic model, g takes the form of conditional probabilistic model

$$g(x) = P(y|x)$$
....5

Or y will take the form of a joint probability model

$$f(x,y) = P(x,y)$$
....6

Whereas Naïve Bayes and Linear Discriminant Analysis are modelled with joint probability, logistic regression is a conditional probability model.

Both the Empirical risk minimization and Structural risk minimization are employed in selecting $f\ and\ g$. The former seeks a function that best fits the training data while the latter, Structural risk minimization with penalty function controls the bias–variance tradeoff.

The performance of the fitting function on the training data is measured by Loss function expressed as

$$L : Y \times Y \to \mathbb{R}^{\geq 0}$$
....7

For the scenario (x_i, y_i) under consideration, the loss function predicting the class value $\hat{y}$ will be: $(y_i, \hat{y})$. Where i = 1 to N.

The risk $R(g)$ of function g is defined as expected loss of g and, it is estimated from the training data as

$$R_{emp}(g) = \frac{1}{N} \sum_i L(y_i, g(x_i))$$
....8

Empirical risk minimization works is = n, a way to seek g with minimum loss function $R(g)$ in equation 8 which in essence can be arrived at by engaging optimization algorithm to find the g for a conditional probability distribution $P(y|x)$, and negative log likelihood loss function expressed as:

$$L(y, \hat{y}) = -\log P(y|x)$$
....9

As such, the risk minimization will be equivalent to maximum likelihood estimation. Empirical risk minimization leads to high variance and consequently poor generalization in a situation where there are many functions or the training data set is not large enough within the G space and this results in overfitting - a condition whereby the function fits well with the training set but perform woefully with test data set as a result of poor generalization of functions. When such happens, the structural risk minimization comes to rescue by importing a regularization penalty into the optimization.

There are varieties of regularization penalties that can be employed depending on the complexity of the functions. Examples include L_0 norm, L_1 norm, and L_2 norm expressed respectively as $\beta_j\,s$, $\sum_j |\beta_j|$, and $\sum_j \beta_j^2$.

For a linear function, $g(x) = \sum_{j=1}^{d} \beta_j x_j$. If the penalty is denoted as $C(g)$, the supervised learning optimization will find the function g that minimizes

$$J(g) = R_{emp}(g) + \lambda C(g)$$ ……………………………………………………………………………10

Where λ controls the bias-variance tradeoff. When $\lambda = 0$, the empirical risk minimization will have low bias and high variance. When λ becomes large, the learning algorithm will have a high bias, and low variance. Between the two extremes, the value of λ is selected empirically through cross-validation.

II. MATERIALS AND METHODS

Remote Sensing and machine learning techniques were employed in this work in information extraction leading to coal seam distribution map production. The result was validated using mineral, landcover maps, and ground truth.

2.1 STUDY AREA

The study area covers a part of Lower Benue trough with geographical coordinates 7.225164°N, 6.865464°E LL and 7.837803°N, 7.841064°E UR. The trough is linear and subdivided into Upper, Middle and Lower with no definitive boundaries. They are often distinguished by towns each contained. Benue trough is a sediment basin. It contains more than 5000m predominantly Aptian to Maastrichtian sediments. It is the geological region, apart from Sokoto Basin in the northwest, mid-Niger (Bida) basin and Dahomey embayment that accounts for presence of coal mineral in the country [24]. It has the largest deposit of coal whose coalfields (Enugu, Okaba, Ogboyoga, Orukpa, Lafia-Obi. Gombe and Chikala) fell into Turonian-Coniacian (which formed from swamp) and Campano-Maastrichtian (non-marine: paralic mamu formation that contains fine-grained sandstone, shale, muddy stones and sandy shale with coal seams).
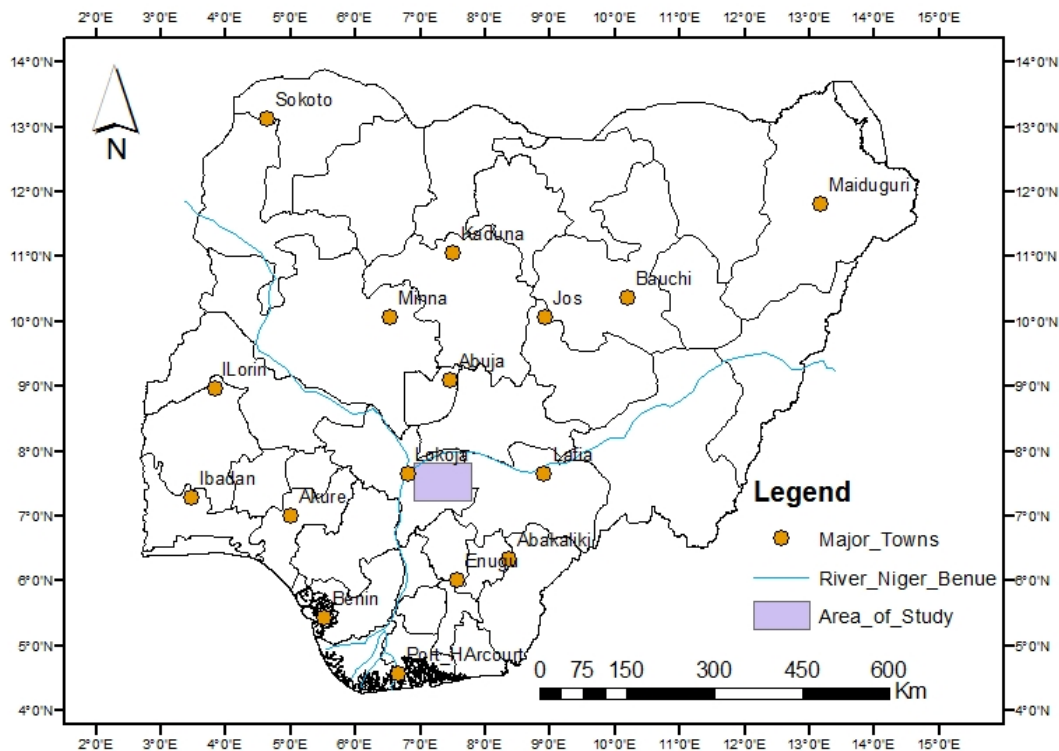

Figure 2. Map of Nigeria Showing the Study Area

2.2 METHODOLOGY

Two Landsat 8 image scenes covering the area were downloaded from the USGS site. Radiometric calibration was conducted to convert the Digital Number (DN) values to brightness values. Fast Line-of-sight Atmospheric Analysis of Spectra l Hypercubes (FLAASH) was used to perform the atmospheric correction. This was followed by statistics' computation in order to rescale the data value (normalization) to limits of zero and one. The two scenes processed separately were mosaicked, after which the study area was clipped. Principal Components Analysis was conducted on the mosaicked image. Eigenvector loadings (which provide information using magnitude and sign (+or-)) was examined in order to decide which of the principal component images will concentrate information directly related to the theoretical signature of specific features which are highlighted either in dark or light pixels. The image was verified for presence of noise, and consequently the bands that contain noise were discarded. Image ratios 4/2 and 6/7 were conducted to reveal Ferric iron- and Hydroxyl-bearing minerals locations. Special PCA (Crosta technique) of combined 2456 and 2567 bands were separately carried out. The products were analyzed appropriately. PC containing Ferric iron-bearing and Hydroxyl-bearing minerals were obtained and named as F and H images respectively. The two images were layer-stacked and named H+F image. Crosta H, H+F and F images were combined to form a composite image in Red, Blue and Green channels respectively. It is regarded as the most conventional way to show alteration result [25]. Spectral Angle Mapper classification was carried out using spectral library signature of features' Endmember. Validation of supposed coal mineral locations were carried out based on Google image coal site locations, identified coal outcrop areas and others based on mineral map.

Machine learning approach involved collection of train data and test data. In order to obtain train data, portion of the image where coal was known to exist was sampled and other portions where coal was not present were also obtained at random from within the enlarged processed image. These sampled data were converted into ASCII format. They were labelled appropriately into the classes they belong as coal (1) or non-coal (0). Part of these sets of data were set aside to use for test and validation. The accuracy of resulting models from the algorithms were also computed. The entire image which constituted the area of interest, AoI, converted to ASCII format was set aside as new data whose class is unknown and that which is to be classed by the accepted model derived from the training algorithms whose accuracies are good enough for use.

Erdas imagine 2015 was employed for pixels data conversion to ASCII. ArcMap10.3 was used for integration of ASCII data with ancillary data, before map production thereafter. Paint package was used for presentation. Python API was used for machine learning data training, model generation and prediction of coal seam areas in new data (AoI). Delimit, excel and notepad packages were used in formatting the pixels numerical data.

Algorithms which include Multilayer Perception (MLP), Extreme Learning Machine (ELM), Support Vector Machine (SVM) - Linear and Radial Basis Function and, Ensemble method (Bagging and Random Forest) were engaged.

III. RESULTS AND DISCUSSION

Principal Component eigenvalues graph plotted for the PCA conducted is displayed in figure 3a. This, with statistical computation results in table 1 revealed that PCs 1, 2 and 3 concentrated the features. PC1, 2 and 3 were composed in RGB format in figure 3b.

Table 1. PCA Statistical Results of Seven Bands.

| Input band	Band1	Band2	Band3	Band4	Band5	Band6	Band7	Eigenvalue
PC1	0.184652	0.200035	0.250406	0.386228	0.111752	0.608679	0.575357	68.594
PC2	0.104342	0.102377	0.150993	0.010646	0.918376	0.013726	-0.334839	16.052
PC3	0.450642	0.45229	0.437343	0.333348	-0.290541	-0.317775	-0.323374	13.165
PC4	0.104027	0.10256	0.09033	-0.185067	0.219005	-0.660617	0.672215	1.215
PC5	-0.325585	-0.30847	-0.124312	0.826381	0.091834	-0.303077	0.013897	0.761
PC6	-0.484976	-0.244903	0.824044	-0.149409	-0.05658	0.007083	-0.014055	0.185
PC7	0.632384	-0.76111	0.143414	-0.010753	-0.009255	0.004517	0.003483	0.028
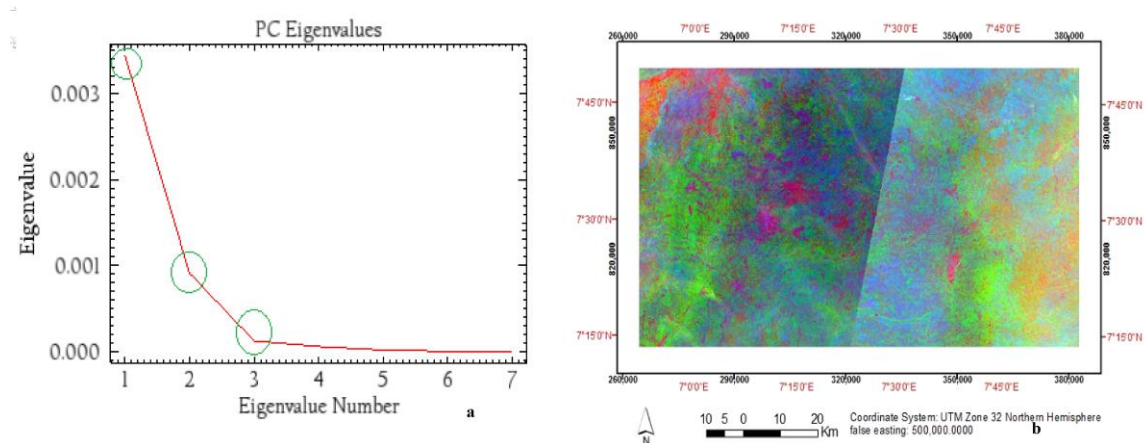
Figure 3. a: PC Eigenvalues Showing Principal PC-PC3 with Highest Variance Values. b: PC1, PC2 and PC3 Presented in RGB format

The first principal component (PC1) in table 1 has positive weightings of all bands and carries 68.594% followed by PC2 and PC3. They all accounted for the high percentage of the features in the image. PC4 has the highest magnitude in eigenvalue loadings (-0.660617) in band 6 and opposite sign with band 7 (0.672215), the second in magnitude. The PC image depicts the hydroxyl mineral as dark pixels. It was negated to depict hydroxyl mineral in bright pixels displayed in figure 4a. PC2 and PC3 are mixed features. Eigenvalue loadings of PC5 indicate that it is dominated by vegetation as a result of the highest magnitude indicated by band 4. PC6 has the highest eigenvalue loadings in band 3 followed by band 1 and, the two in opposite signs which therefore depict the ferric iron as bright pixels in figure 4b.
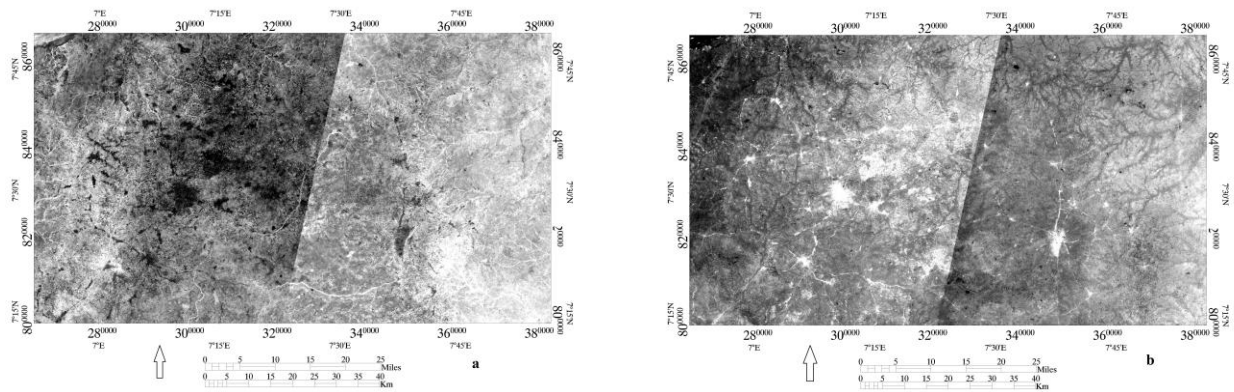

Figure 4. a: PC2 Showing Hydroxyl-bearing Minerals as Bright Pixels (image H). b: PC6 Showing Ferric Iron-bearing Minerals as Bright Pixels (image F).

PC2 and PC6 were added as image F+H. Composite image of H, F+H and F was formed and presented as figure 5. Please note the white pixels.
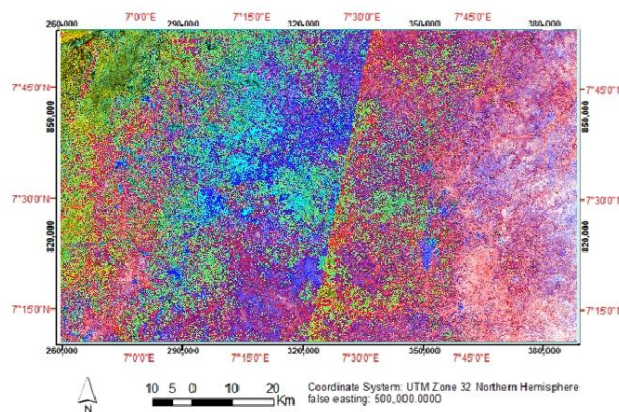

Figure 5. False Colour Composite for Images H, F+H and F

Figure 6a represents MNF result graph and figure 6b, the corresponding MNF image of PCs 1, 2 and 3 earlier shown in figure 3b.
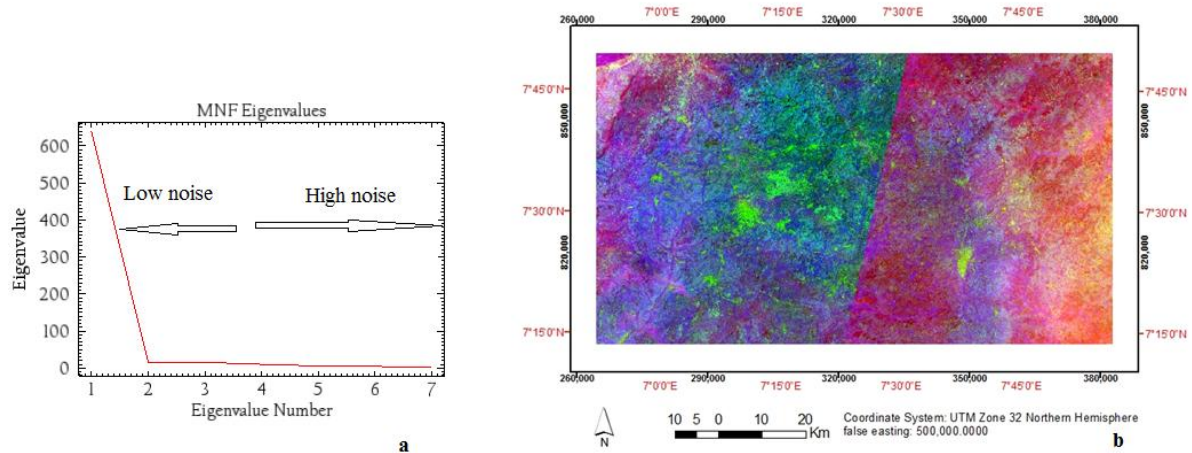

Figure 6. a: MNF Process Result Graph Carried out on PC1-PC3. b: RGB Colour Composite of MNF Image Depicting Anomaly in PC1, 2 and 3.

PCA is challenged with noise. Figure 7a and b depict ratio results for 4/2 and 6/7 which represent ferric iron- and hydroxyl-bearing minerals in bright pixels.
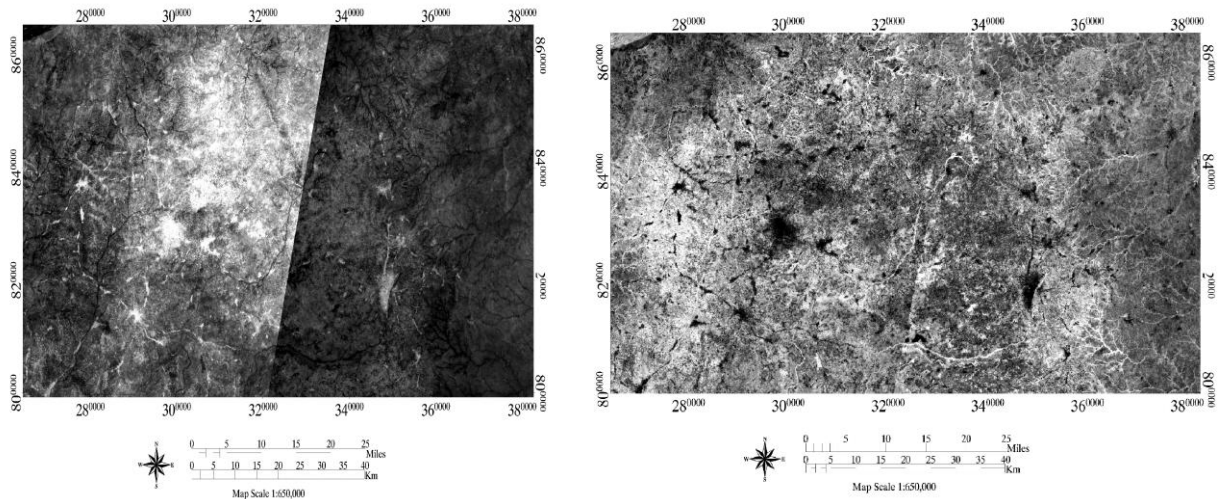

Figure 7. a: Ferric Iron Mineral (4/2) Altered Areas Shown in Bright Pixels. b. Hydroxyl Mineral (6/7) Altered Areas Shown in Bright Pixels.

The corresponding eigenvector matrix for Crosta 2456 and 2567 are presented in tables 2 and 3 respectively.

Table 2. Eigenvector Matrix for Bands 2456

Input Band	Band2	Band4	Band5	Band6	Eigenvalue
PC1	0.606406	0.621556	0.141653	0.475263	64.418
PC2	0.204540	0.158779	-0.947789	-0.186143	22.859
PC3	0.539311	0.019183	0.275820	-0.795424	12.114
PC4	0.547338	-0.766870	-0.074527	0.326767	0.609

In table 2, the eigenvector loadings for bands 2, 4, 5 and 6 with corresponding PCs are loaded. Band 7 has been deliberately avoided in order not to map hydroxyl minerals. In the interpretation, PC1 represents albedo; IR versus visible are in PC2 and PC3. Meanwhile, vegetation is more pronounced in PC3 while IR is more in PC2. PC4 eigenvector loadings are higher in magnitudes for bands 4 and 2 and are in opposite signs,

therefore, PC4 represents image where ferric iron minerals will be shown in dark pixels. To present the ferric iron minerals in bright pixels, then the PC4 was negated (figure 8a). This image is denoted as F image.

In table 3, Crosta 2567 is presented statistically. Bands 1 and 4 have deliberately been removed in order not to map iron oxide.

Table 3. Eigenvector Matrix for Bands 2567

Input Band	Band2	Band5	Band6	Band7	Eigenvalue
PC1	0.561846	0.608553	0.142109	0.542031	56.939
PC2	0.275463	-0.014637	-0.961048	-0.017132	24.084
PC3	0.378109	0.371412	0.11769	-0.839783	18.112
PC4	0.682263	-0.701072	0.20577	0.025959	0.865

Examining the eigenvalue loadings, PC1 represents albedo. In PC2, Vegetation is presented as dark pixels. Hydroxyl-bearing minerals' concentration is depicted in bright pixels in PC3, while PC4 is visible versus IR. The image (PC3) is denoted H, and it is shown in figure 8b.
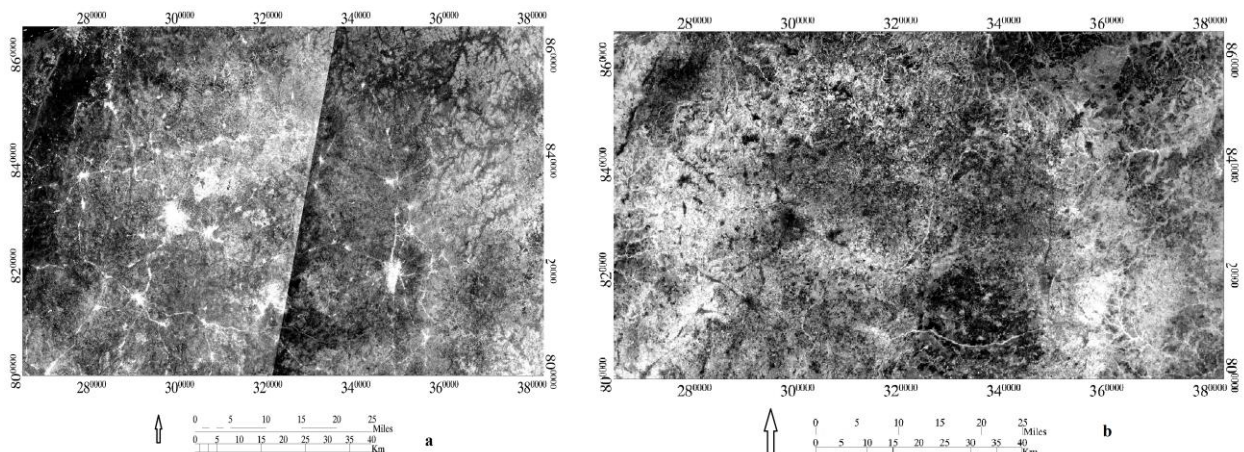

Figure 8. a: Negated PC4 of Crosta 2456 Depicting Ferric Iron-bearing Minerals as Bright Pixels (Image F). b: PC3 of Crosta 2567 Depicting Hydroxyl-bearing Minerals in Bright Pixels (Image H).

The combined F + H image will have both the ferric iron and the hydroxyl minerals concentration map represented in bright pixels as shown in figure 8a.
Crosta H, H+F and F images are combined to form a composite image in Red, Blue and Green channels respectively to return the image in figure 9b[25].
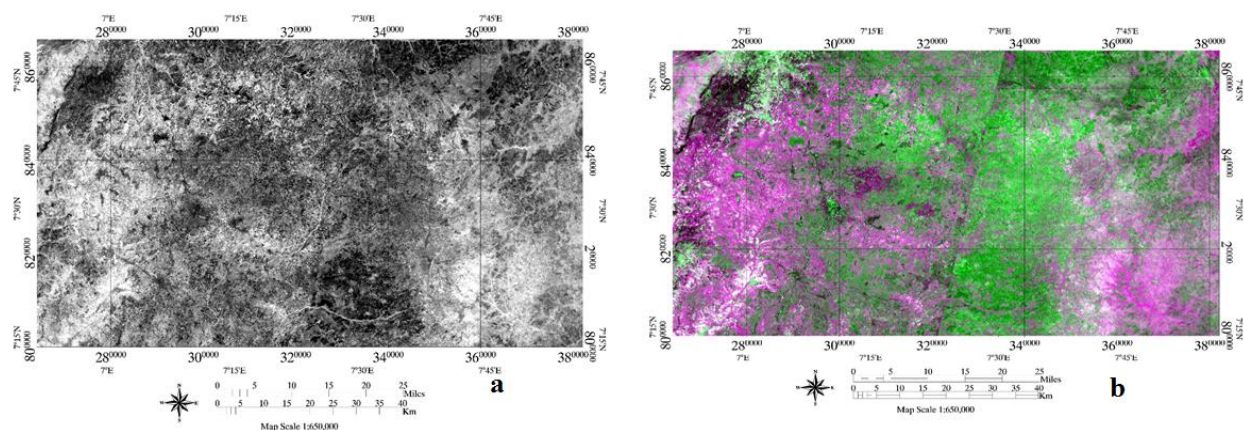

Figure 9. a: Added Iron Oxide Plus Hydroxyl-bearing Minerals (image F+H) Areas Depicted as Bright Pixels. b: Composite Image of Crosta H, F+H and F Depicting Hydroxyl-bearing Minerals as Magenta and Iron Oxide as Green.

The bright white pixels in figure 9b represent interesting areas. The whitish pixels indicate where both alteration types are overlapping and they are the most active areas where the supposed minerals are either already being exploited or are typically concentrated (venting points or outcrops for minerals) [25]. It was noted that the white patches (soil tainting or contamination as a result of mineral/material beneath) is well spread across (at the south west and south east). Spectral Angle Mapper classification was carried out to designate these areas. The result is shown in figure 10a. Overlay of known coal exploitation areas marked on georeferenced Google Earth image was made on classified image to validate the result as shown in figure 9b.
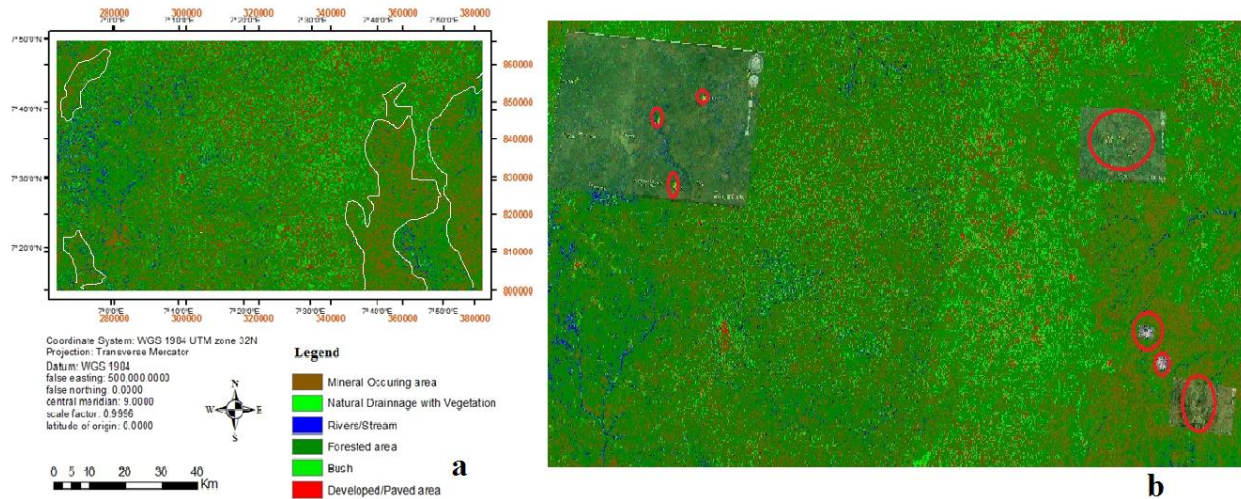

Figure 10. a: Spectral Angle Mapper Classification to Show Mineral Altered Area. b: Validated Area with Known Coal and Outcrops' Locations Marked in Red

Figure 11 shows product of SAM classification of identified surface features including the surface coal exploitation locations.
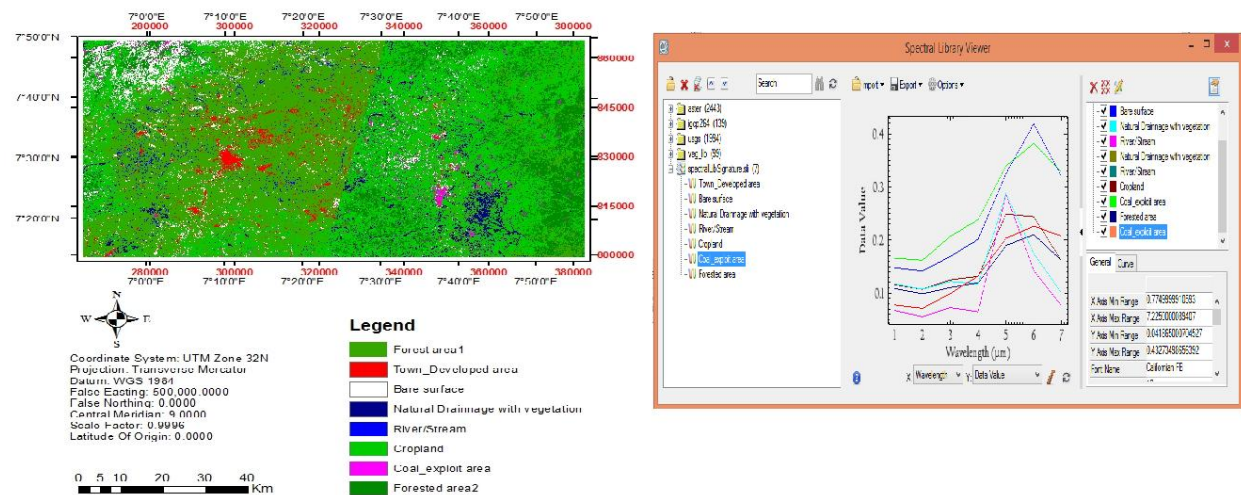

Figure 11. Spectral Angle Mapper Classification of PCA (PCs 1,2 and 3) Image Showing Locations of Surface Coal Exploitation Areas.

The following figures (figure 12a-h) show the results of various machine learning algorithms engaged. Areas predicted to contain area are shown in grey while non-coal areas are in black. Metrics which show the summary of attained precision of results is equally presented together with confusion matrix for each learner.
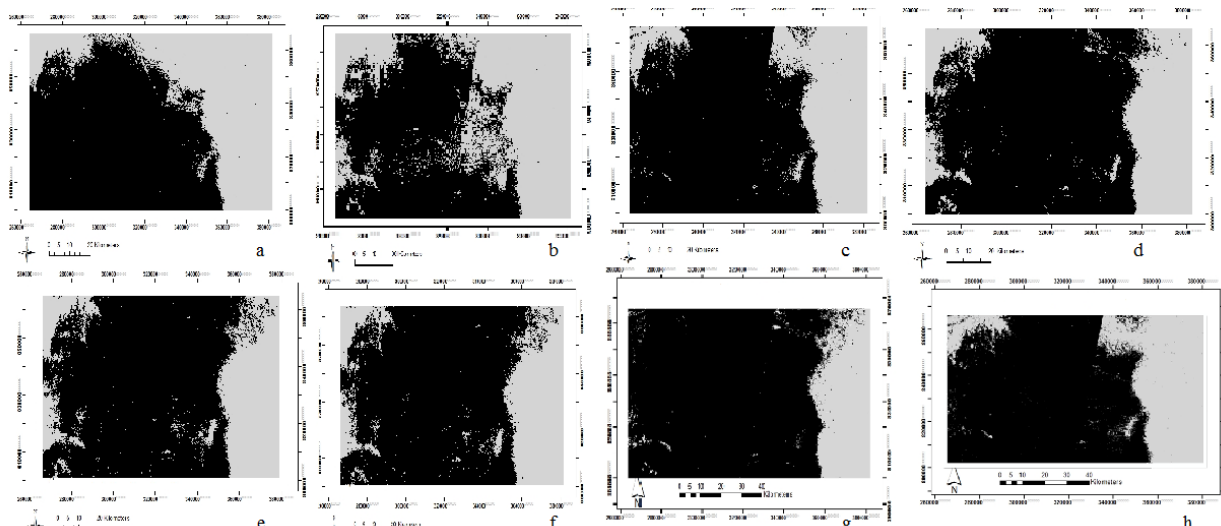

Figure 12. a SVM Linear. b, SVM Polynomial- 3rd Degree. c, SVM – Radial Basis Function (rbf). d Ensemble – Bagging. e Ensemble Random Forest. f Ensemble – Gradient Boosting. g Multi-Layer Perception (MLP) Classifier Using Backpropagation (Linear). h Multi-Layer Perception (MLP) Classifier Using Backpropagation Radial Basis Function (RBF)

The metrics relating to each of the results in figure12 a-f are as follows

Table 4. Metrics of the Results

SVM_Linear. training score:= 0.9897265447952863						**SVM_Polynomial**					
	precision	recall	f1-score	support			precision	recall	f1-score	support	
0	1.00	1.00	1.00	1403		0	0.94	1.00	0.97	1403	
1	0.99	1.00	1.00	804		1	1.00	0.89	0.94	804	
accuracy			1.00	2207					0.96	2207	
macro avg.	1.00	1.00	1.00	2207		macro avg.	0.97	0.94	0.95	2207	
weighted avg.	1.00	1.00	1.00	2207		weighted avg.	0.96	0.96	0.96	2207	
[[1402 1] [1 803]]						[[1403 0] [90 714]]					
SVM_Rbf training score:= 0.9984891976260124						**Ensemble Bagging** training score:= 0.9897265447952863					
	precision	recall	f1-score	support			precision	recall	f1-score	support	
0	1.00	1.00	1.00	1403		0	1.00	0.99	1.00	1403	
1	1.00	1.00	1.00	804		1	0.99	0.99	0.99	804	
accuracy			1.00	2207		accuracy			0.99	2207	
macro avg.	1.00	1.00	1.00	2207		macro avg.	0.99	0.99	0.99	2207	
weighted avg.	1.00	1.00	1.00	2207		weighted avg.	0.99	0.99	0.99	2207	
[[1403 0] [0 804]]						[[1395 8] [5 799]]					
Ensemble Random Forest Classifier training score:= 0.9995467593292038						**Ensemble Gradient Boost** training score:= 0.9897265447952863					
	precision	recall	f1-score	support			precision	recall	f1-score	support	
0	1.00	1.00	1.00	1403		0	0.99	0.99	0.99	1403	
1	0.99	1.00	1.00	804		1	0.98	0.99	0.99	804	
accuracy			1.00	2207		accuracy			0.99	2207	
macro avg.	1.00	1.00	1.00	2207		macro avg.	0.99	0.99	0.99	2207	
weighted avg.	1.00	1.00	1.00	2207		weighted avg.	0.99	0.99	0.99	2207	
[[1398 5] [3 801]]						[[1389 14] [7 797]]					
ANN_mlp						**ANN_rbf**					
	precision	recall	f1-score	support			precision	recall	f1-score	support	

	precision	recall	f1-score	support			precision	recall	f1-score	support
0	1.00	1.00	1.00	1403		0	1.00	1.00	1.00	1403
1	1.00	1.00	1.00	804		1	0.99	1.00	1.00	804
accuracy			1.00	2207		accuracy			1.00	2207
macro avg.	1.00	1.00	1.00	2207		macro avg.	1.00	1.00	1.00	2207
weighted avg.	1.00	1.00	1.00	2207		weighted avg.	1.00	1.00	1.00	2207
[[1403 0] [1 803]] average precision score: 0.9992093226662339						[[1403 0] [0 804]] training score:= 0.9999712345678743				

Map a in figure 14 depicts the NDVI of the area rendered in colour, the b part is the lineament distribution map. The metadata shows that the satellite made its passage for sensor imaging around 9 a.m. The derived temperature map, c shows the temperature distribution of the area at the time. Coal colour varies from brown to black. Dark objects absorb more heat, convert it to thermal energy and, emit it as a heat more than other objects. The linear density map, d shows that coal formation at the area is at the low linear density area.
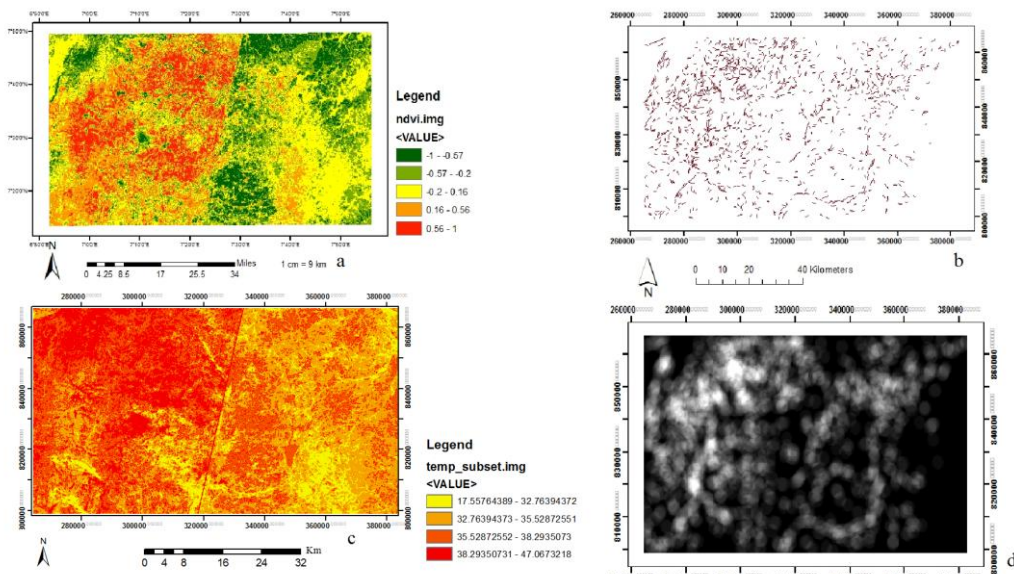

Figure 14. a, Derived NDVI to depict vegetation cover and Vigour. b, Derived Lineament for the AOI. c, Derived Feature Temperature in Celsius Degrees (^{o}C). d, Derived Lineament Density of the AOI

Figure 15a yellow rectangle defines the study area boundary in a Google Earth image. The markers are identified coal locations from available ancillary data and site visits. Figure 15b is Google Earth image visualization of the area as of 2017. Coal activities were very common and well-spread around the eastern flank. These areas, for example, figure 16a and 16b were not seen on Google earth image as at earlier site visitation in 2017 because, the image available to us then was 2014 G.E. The blue peg point in figure 14b is shown in figure 16a, but nothing was seen on the Google (2014) image data as the Google reference predates the commencement of the exploitation of some places. This location was about to be exploited when the site was visited. The overburden to coal in figures 16a and b were already removed which means the raw coal was being stood upon, with a veil of topsoil. The two most-visibly-seen exploited areas (encircled in red dots) in figure 15d are part of the areas included in the training data used in machine learning process, apart from some coal outcrop points. Other areas known to appear on the image are faint. They further serve as validation points. Figures 15e and f depict zoomed-in Google Earth image of coal outcrop locations on the north-east area of figure 15c (2019 Google image).

Visitation to the sites showed that some areas were exploited in the past, and have since been abandoned. New places were being opened for exploitation by companies. Local exploitation activities were also taking place at various sites at micro levels especially where there are coal outcrops, and abandoned sites, scattered around the area.
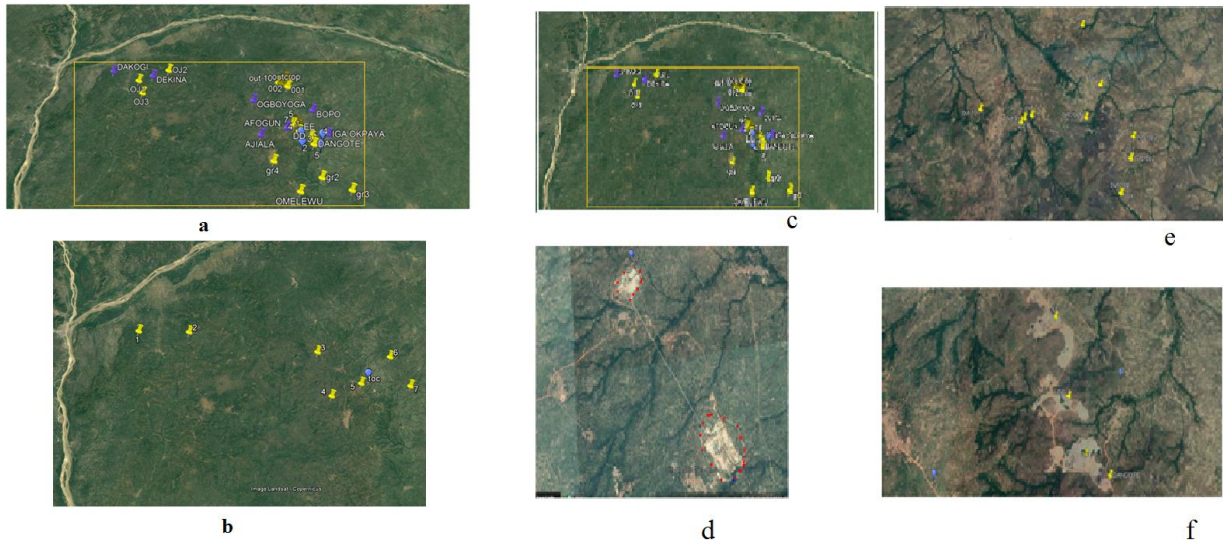
Figure 15: Google Earth Image Showing Area of Interest (In the Pink Box), the Surrounding Area and Rivers Niger and Benue. a, Recent Google Earth Image Covering the Study Area. c, Coal Outcrop Locations on the Northeast of c. d, Exploited Areas in Red Rings and, e, Exploited Areas as At July, 2019.

Figure 16. a, Coal Exploitation Site, Showing Machineries and Excavation (Overburden Removal) Done to Coal Point. b, New Expansion Site for Coal Exploitation, Showing Machineries and Excavation to Top-of-Coal after Removal of Earth's Overburden Material.

IV. CONCLUSION

Spectral response of various earth's features like coal can be applied in remote sensing for mapping, particularly in large areas. The method is cheap but requires experts' touch. Surface Coal site mapping has been demonstrated to be feasible with Crosta technique. Prediction of coal seam areas can also be achieved with machine learning. Areas predicted should however be validated with all available ancillary data and in-situ verification.

REFERENCES

[1]. Kusuma, G. J., Shimada, H., Sasaoka, T., Matsui, K., Nugraha, C., Gautama, R. S., & Sulistianto, B. (2012). *An Evaluation on the Physical and Chemical Composition of Coal Combustion Ash and Its Co-Placement with Coal-Mine Waste Rock*. *2012*(July), 589–596.

[2]. Maurice, J. B., & Richard, V. W. (1917). The Volatilefile:///C:/my jobs/UNN/PROF/Coal Journals/Coal Chemical physical analysis.pdf Constituents of Coal. *Journal of the Chemical Society, Transactions*.

[3]. Virgil, A. T., Jack, H., Medlin, S. L. C., & Ronald, W. S. (1982). *Chemical Analyses and Physical Properties of 12 Coal Samples from the Pocahontas Field, Tazewell County, Virginia, and McDowell County, West Virginia*.

[4]. Isabel, S.-R., & Colin, R. W. (2008). *Applied Coal Petrology*.

[5]. Kang, L. (2002). Coal, Oil Shale, Natural Bitumen, Heavy Oil and Peat. *Coal Exploration and Mining*, *I*. - www.eolss.net/sample-chapters/c08/e3-04-01-02.pdf

[6]. Tan, K., Wan, Y., Sun, S., Kuang, J., & Chen, X. (2011). Application Research on Coal Prospecting with Remote Sensing. *Procedia*

Earth and Planetary Science, *3*, 103–112. https://doi.org/10.1016/j.proeps.2011.09.071

[7].   Bhattacharyya, M. (2013). Comprehensive Solutions for Modern Coal Mining Through Applied Remote Sensing and Integrated GIS. *14th Esri India User Conference*, 1–5.

[8].   Ward, C. R. (2008). Coal Exploration and Mining Geology. In *Encyclopedia of Life Support Systems, Vol. V* (Vol. V).

[9].   Kuenzer, C., Hecker, C., Zhang, J., Wessling, S., & Wagner, W. (2008). The potential of multidiurnal MODIS thermal band data for coal fire detection. *International Journal of Remote Sensing*, *29*(3), 923–944. https://doi.org/10.1080/01431160701352147

[10].   Kiang, M. Y. (2003). *A comparative assessment of classification methods*. *35*, 441–454. https://doi.org/10.1016/S0167-9236(02)00110-0

[11].   Christos, S., & Dimitrios, S. (n.d.). *NEURAL NETWORKS*. www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/...

[12].   Ei–Helly, M., A., Rafea, S., Ei–Gamal, & Whab, R. A. E. (2003). Integrating diagnostic expert system with image processing via loosely coupled technique. In Central Laboratory for Agricultureal Expert Systems (CLAES). *IICAI*, .1182–1195).

[13].   Friedl, M. A., & Brodley, C. E. (1997). Decision Tree Classification of Land Cover from Remotely Sensed Data. *Sensing of Environment.*, *61*, 399–409.

[14].   Malpica, J. A. (n.d.). *Consequences of the Hughes Phenomenon on some Classification*

[15].   Mather, M. P., & Paul, M. (2001). Decision Tree Based Classification of Remotely Sensed Data. *22nd Asian Conference on Remote Sensing*.

[16].   Wiley, I. P. L. (2013). *Advances in Remote Sensing and GIS Analysis* (M. A. Peter & J. T. Nicholas (eds.)). Radha offset, Delhi.

[17].   Persello, C. (2010). *Advanced Techniques for the Classification of Very High Resolution and Hyperspectral Remote Sensing Images*. DISI University of Trento.

[18].   Lu, D., & Weng, Q. (2007). *Review article A survey of image classification methods and techniques for improving classification performance*. *28*(5), 823–870. https://doi.org/10.1080/01431160600746456

[19].   Ben-David, S., & Shalev-Shwartz, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning Please

[20].   Tom, M. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.

[21].   Gillian, N., & Environment, R. (2013). IAP Gesture Recognition Workshop Session 1 : Gesture Recognition & Machine Learning Fundamentals. *My Research*.

[22].   Alpaydin, E. (2010). *I2Ml* (Second). MIT Press. https://doi.org/10.1016/j.neuroimage.2010.11.004

[23].   Alpadyin, E. (2004). *Introduction to Machine Learning the Wikipedia Guide*. MIT Press.

[24].   Fatoye, F. B., & Gideon, Y. B. (2013). Appraisal of the Economic Geology of Nigerian Coal Resources. *Journal of Environmental and Health Science*, *3*(11), 25–32.

[25].   Teledetección, M., Caso, E. L., Volcán, D. E. L., Costa, T., Bragado-massa, E., Marchamalo, M., Rejas, J. G., Bonatti, J., Martínez-frías, J., Martínez, R., Ingeniería, D. De, Madrid, P. De, Instituto, I., Técnica, N. De, Universidad, C., Rica, D. C., Geociencias, I. De, Csic-ucm, I., & Geológicas, F. D. C. (2014). *Monitoring Hydrothermal Alteration in Active Volcanoes Using Remote Sensing: The Case of Turrialba Volcano (Costa Rica) Centro de Investigación en Ciencias Atómicas, Nucleares y Moleculares*. 69–82. https://doi.org/10.15517/rgac.v51i1.16848