

Boise State University

ScholarWorks

Geosciences Faculty Publications and
Presentations

Department of Geosciences

2-2024

Geotechnical Characterisation of Coal Spoil Piles Using High-Resolution Optical and Multispectral Data: A Machine Learning Approach

Sureka Thiruchittampalam
University of New South Wales

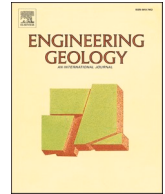
Bikram Pratap Banerjee
University of Southern Queensland

Nancy F. Glenn
Boise State University

Simit Raval
University of New South Wales

Publication Information

Thiruchittampalam, Sureka; Banerjee, Bikram Pratap; Glenn, Nancy F.; and Raval, Simit. (2024). "Geotechnical Characterisation of Coal Spoil Piles Using High-Resolution Optical and Multispectral Data: A Machine Learning Approach". *Engineering Geology*, 329, 107406. <https://doi.org/10.1016/j.enggeo.2024.107406>



Geotechnical characterisation of coal spoil piles using high-resolution optical and multispectral data: A machine learning approach

Sureka Thiruchittampalam^{a,b}, Bikram Pratap Banerjee^c, Nancy F. Glenn^d, Simit Raval^{a,*}

^a School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia

^b Department of Earth Resources Engineering, University of Moratuwa, Moratuwa 01400, Sri Lanka

^c School of Surveying and Built Environment, University of Southern Queensland, Toowoomba, QLD 4350, Australia

^d Department of Geosciences, Boise State University, Boise, ID, USA

ARTICLE INFO

Keywords:

Object-based image analysis
Morphology-based segmentation
Waste materials
Mine dump
High-resolution UAV images
Shear strength parameters

ABSTRACT

Geotechnical characterisation of spoil piles has traditionally relied on the expertise of field specialists, which can be both hazardous and time-consuming. Although unmanned aerial vehicles (UAV) show promise as a remote sensing tool in various applications; accurately segmenting and classifying very high-resolution remote sensing images of heterogeneous terrains, such as mining spoil piles with irregular morphologies, presents significant challenges. The proposed method adopts a robust approach that combines morphology-based segmentation, as well as spectral, textural, structural, and statistical feature extraction techniques to overcome the difficulties associated with spoil pile characterisation. Additionally, it incorporates minimum redundancy maximum relevance (mRMR) based feature selection and machine learning-based classification. This automated characterisation will serve as a proactive tool for dump stability assessment, providing crucial data for improved stability models and contributing to a greener and more responsible mining industry.

1. Introduction

Coal mining, including the construction of coal spoil dumps, poses significant environmental and safety challenges when not managed properly. The geotechnical characterisation of coal spoil is essential for monitoring and evaluating environmental impacts and safety concerns, as well as guiding effective land restoration efforts (Zevgolis et al., 2021). This data also supports policymaking, raises awareness about coal mining effects, and facilitates informed decision-making. However, the irregular deposition processes of spoil and inadequate compaction during placement can introduce complexities and uncertainties in spoil dump behavior (Masoudian et al., 2019).

Analysing the spatial arrangement of characterised coal spoil is crucial, but it's challenging due to variations in properties between mines, influenced by factors like ore type and mining techniques (Lottermoser, 2007). Traditional manual field methods for spoil characterisation are time-consuming and hazardous. To address these issues, unmanned aerial vehicles (UAVs) provide a promising solution for characterising mine materials (Yang et al., 2023).

Selecting the right sensor for a specific task in mining requires

considering spectral and spatial resolution. For visual inspection, mapping, and providing useful visual information, red-green-blue (RGB) sensors are commonly used (Sinaice et al., 2022; Yang et al., 2023). Multispectral sensors with expanded spectral coverage (beyond RGB) are relevant for a range of mining applications and more advanced image analysis. In addition, the calibration of multispectral images captured helps mitigate the challenges posed by variations in reflectance caused by topographic changes.

In the context of spoil characterisation, selecting appropriate sensors involves understanding the challenges associated with each step of the methodical workflow. Initially, this involves recognising the environmental complexities, notably the irregular topography of heterogeneous mine spoil piles. These irregularities can impact the accuracy of classification in very high-resolution remote sensing images (Aryaguna and Danoedoro, 2016; Michel et al., 2012). Furthermore, capturing low-resolution images may lead to the loss of important textural information. Therefore, careful consideration of these factors is necessary when selecting sensors for spoil characterisation to ensure accurate and comprehensive analysis.

Employing appropriate sensors in conjunction with object-based

* Corresponding author.

E-mail addresses: s.thiruchittampalam@unsw.edu.au (S. Thiruchittampalam), bikram.banerjee@unisq.edu.au (B.P. Banerjee), nancyglenn@boisestate.edu (N.F. Glenn), simit@unsw.edu.au (S. Raval).

<https://doi.org/10.1016/j.enggeo.2024.107406>

Received 14 August 2023; Received in revised form 4 January 2024; Accepted 5 January 2024

Available online 6 January 2024

0013-7952/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

image analysis (OBIA) offers an effective approach for spatially mapping features of interest. OBIA, as demonstrated by Asthana et al. (2020), Keyport et al. (2018), excels in mapping rugged terrains similar to spoil environments. The segmentation-based methodology groups pixels into meaningful objects and accounts for spatial relationships between similar pixels. The integration of ancillary data, such as digital surface models (DSM), further improves object boundary definition, even in challenging terrain.

Notably, few studies (Thiruchittampalam et al., 2023a; Thiruchittampalam et al., 2023b) have explored the potential of UAV-based imagery and sensitive sensors for monitoring mine dump spoil piles. Thiruchittampalam et al. (2023b) focuses on pixel-level coal spoil characterisation via UAV-based optical remote sensing, utilising structural descriptors and machine learning for in-depth analysis. This study also emphasises the importance of future work on OBIA for more accurate results. Furthermore, Thiruchittampalam et al. (2023a)'s evaluation of segmentation algorithms in spoil pile delineation benefits the mining industry, setting the stage for OBIA.

To this end, this study sets two primary objectives. Firstly, it aims to expand upon existing research by comparing the effectiveness of optical RGB and multispectral sensors, utilised in conjunction with UAVs, for characterising materials in undulating terrains, particularly in spoil pile environments. One of the main goals of this study is to investigate the capabilities of RGB and an expanded multispectral sensor which offers calibrated reflectance signatures with a higher number of bands. Both sensors were employed to assess their comparative performance. Secondly, this study seeks to explore segmentation and classification techniques specifically tailored to spoil characterisation, taking into account the distinctive challenges presented by the undulating topography of mining spoil piles. It aims to comprehensively compare the classification performance of optical RGB and multispectral sensors, with the objective of determining the most cost-effective and efficient option for identifying spoil piles. Additionally, the study aims to investigate the application of OBIA for geotechnical characterisation of spoil pile. By leveraging this analytical approach, the research aims to extract valuable insights and provide a more efficient and accurate mapping of material piles characterised by uneven surfaces and sporadically clustered debris.

The outcomes of this study hold significant potential for the mining and civil industry for characterising similar environments. The research findings are expected to facilitate enhanced decision-making in mine management and environmental monitoring by advancing UAV-based remote sensing techniques for mining applications.

2. Material and methods

2.1. Methodological overview

The workflow implemented to characterise dump piles in the study area is shown in Fig. 1. Initially, ground truth data was collected, which serves as both the training and testing dataset for evaluating classifier performance in subsequent stages. Subsequently, an orthomosaic was generated by utilising geotagged RGB and multispectral images obtained through UAV data acquisition. Features were extracted from these orthomosaics for further analysis. Concurrently, segments were produced from the digital surface model (DSM) using a Voronoi-based segmentation algorithm. Within each segment, zonal statistics of features derived from the orthomosaics were calculated. These features were then ranked using the minimum redundancy maximum relevance (mRMR) algorithm. Finally, a selected set of features underwent machine learning based analysis to determine the best-performing algorithm for each attribute.

2.2. Study site and UAV data acquisition

The study area covers 62,567 m² of paddock-dumped coal spoils at a mine dump site in the Sydney basin of New South Wales, Australia (Fig. 2). This site is characterised by dump trucks unloading the spoil onto a confined neighboring field, resulting in the creation of ridges and valleys within the environment. Over time, the accumulated spoil forms a unique topography that exhibits a series of ridges and valleys, with varying elevations and slopes influenced by the dumped materials and the dumping method employed (Fig. 2(d)).

The aerial images of the fields were captured using a DJI M300 RTK drone (SZ DJI Technology Co., Ltd., Shenzhen, China). The UAV was equipped with a Micasense Altum multispectral camera (MicaSense,

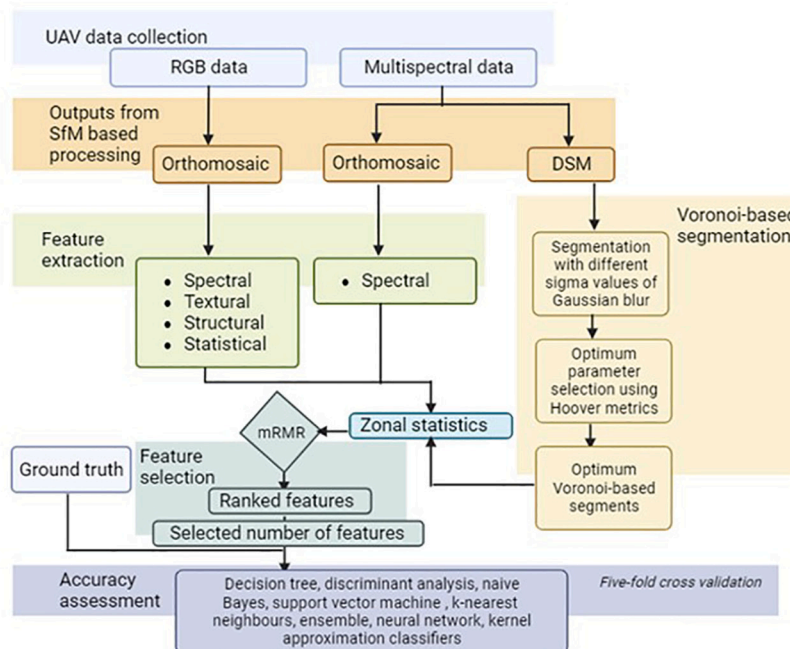


Fig. 1. Workflow of object-oriented classification of attributes related to coal spoil characterisation.

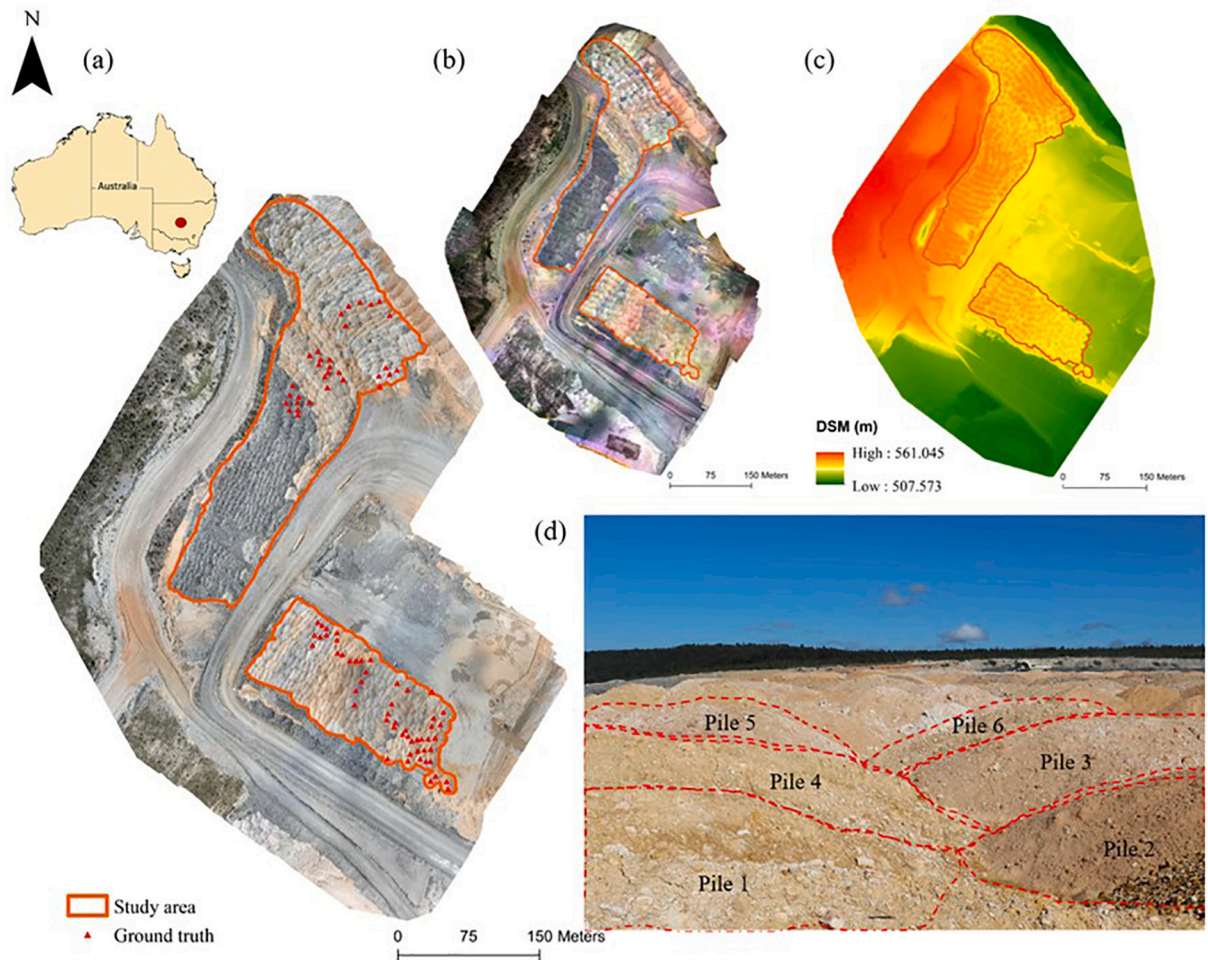


Fig. 2. (a) Location of the selected mine site in New South Wales, Australia, and distribution of ground truth sample points within the RGB orthomosaic from Zenmuse P1 RGB camera images, (b) true color composite of the orthomosaic from Micasense Altum-PT sensor images, (c) digital surface model (DSM) derived from Micasense Altum-PT sensor images, (d) image of spoil piles (outlined with red dotted line) captured using hand-held camera which shows patterns of ridges and valleys indicative of distinct topographical features. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Inc., Seattle, WA, USA), and a DJI Zenmuse P1 optical camera. Both sensor payloads facilitated a comprehensive data collection for an in-depth analysis of the spoil piles and the surrounding environment. The data collection was performed around local solar noon.

The Micasense Altum camera employed is capable of capturing images across six spectral ranges, including blue, green, red, red-edge, near-infrared, and thermal infrared, as indicated in Table 1. This extensive spectral coverage allows for a comprehensive analysis of the study area and the spoil piles. The Micasense Altum-PT sensor is

equipped with a downwelling light sensor (DLS) to account for variations in illumination across the area. This sensor measures ambient light conditions and the sun's angle for each spectral band, enabling the adjustment of illumination discrepancies during data capture. Several measures were implemented to ensure accurate and precise georeferencing. Firstly, a calibrated reflectance panel (CRP) image was captured at the beginning of each multispectral flight. This panel serves as a reference for radiometric calibration. Additionally, five ground control points (GCP) were strategically placed around the study area. These

Table 1
Specification of the sensors deployed in this study.

Sensor	Bands	Image width × height (pixels)	Sensor dimensions (mm)	Focal length (mm)	Field of view	Centre wavelength (nm)	Ground sampling distance (GSD) (cm)
Micasense Altum	Blue	2066 × 1544	7.121 × 5.327	8	48° HFOV × 36.8° VFOV	475	4.41 cm at 100 m altitude
	Green	2066 × 1544				560	
	Red	2066 × 1544				668	
	Red edge (RE)	2066 × 1544				717	
	Near Infrared (NIR)	2066 × 1544				842	
	Thermal Infrared (TIR)	320 × 256				3.840 × 3.072	
Zenmuse P1	Blue	8192 × 5460	35.9 × 24	35	53.630° HFOV × 36.960° VFOV	470	1.22 cm at 80 m altitude
	Green					550	
	Red					660	

GCPs were evenly spaced, and their coordinates were measured using an Emlid Reach RS2 multi-frequency global navigation satellite system (GNSS) receiver, operating in kinematic survey mode and connected to a networked transport of RTCM via internet protocol (NTRIP) correction service. The integration of this setup guarantees precise positioning of the GCPs, which is crucial for obtaining high-quality and accurately georeferenced aerial imagery.

The surveys were meticulously planned and executed using the DJI Pilot app (SZ DJI Technology Co., Ltd.). The UAV was operated at an altitude of 100 m above ground level (AGL) for the Micasense Altum camera, and 80 m for the Zenmuse P1 camera, ensuring comprehensive coverage with an 80% forward and side overlap. Both cameras were positioned at nadir (90°), and a single grid flight path was established. The automatic capture mode was employed for image acquisition by both sensors.

The DJI M300 RTK drone, serving as the platform for mounting the DJI Zenmuse P1 and Altum multispectral camera, exhibits robustness and power. It possesses physical dimensions of 81 cm × 67 cm × 43 cm and a mass of approximately 6.29 kg. Equipped with advanced navigation systems, including GPS, GLONASS, BeiDou, and Galileo, the drone offers reliable navigation and precise image tagging capabilities.

2.3. Conventional coal spoil characterisation approach and ground truth data collection

The ground truthing approach utilised for characterising coal spoil and collecting ground truth data is based on the work by Simmons and McManus (2004). The work developed the BHP Mitsubishi Alliance Coal (BMAC) spoil shear strength framework (Tables 2 and 3), which has gained widespread acceptance as a categorisation system. This framework relies on visual-tactile characteristics to assess shear strength parameters of coal mine spoil, eliminating the need for time-consuming laboratory tests.

Spoil characterisation involves consistently assigning attributes to the spoil and assigning weights to each attribute using standardised procedures. These attributes and weights are presented in Table 2. The “Predominant particle size” is determined through recognised visual methods, while “Consistency/Relative density” is evaluated based on the moist condition, using tactile procedures for cohesive or cohesionless materials. “Plasticity” is assessed either through a liquid limit test or visually. The “Fabric structure” is a spoil characteristic based on the distribution of particle sizes. It is divided into two components: the “Framework” and the “Matrix.” The framework consists of larger

Table 2

BHP Mitsubishi Alliance Coal (BMAC) framework spoil categories and their visual-tactile attributes (for current spoil strength) (adapted from Simmons and McManus, 2004).

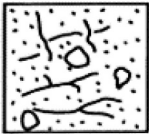
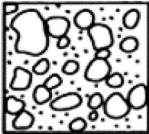
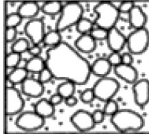
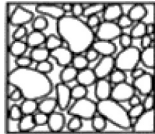
Category →	1	2	3	4	Weightage ←
Description	Fine-grained clay-rich high plasticity	Fine-grained low plasticity with larger clasts	Larger clasts with fine matrix, low plasticity	Large blocks, minor fines, minor slaking	
Predominant particle size	Clay	Sand	Gravel	Cobbles	11.6
Consistency: (Cohesive) Relative density: (Cohesionless)	Soft to Firm Loose	Stiff Medium Dense	Hard Dense	Hard packed	26.9
Structure					26.9
Liquid Limit	Matrix only High (>50)	Matrix supported Intermediate (35–50)	Framework supported Low (20–35)	Framework only Not Plastic (<20)	34.6

Table 3

BHP Mitsubishi Alliance Coal (BMAC) framework spoil categories and their shear strength parameters with mobilization modes (after Simmons and McManus, 2004).

Category →		1	2	3	4
Unsaturated	γ (kN/m ³)	18 (1)	18 (1)	18 (1)	18 (1)
	c' (kPa)	20 (10)	30 (15)	50 (15)	50 (15)
	φ' (deg)	25 (2.5)	28 (3)	30 (2)	35 (2.5)
	γ (kN/m ³)	20 (1)	20 (1)	20 (1)	20 (1)
	c' (kPa)	0 (0)	15 (7.5)	20 (10)	0 (0)
Saturated	φ' (deg)	18 (3)	23 (2.5)	25 (2.5)	30 (1.5)
	φ' (deg)	18 (1.5)	18 (1.5)	18 (1.5)	28 (2)
	c' = 0 kPa				

particles that transmit forces within the spoil and form a rigid structure upon contact. The matrix comprises smaller particles that fill the gaps between the framework particles. The BMAC categories are determined by summing the relative weights of each attribute using the following equation, and the spoil is assigned to the category with the highest overall weight.

$$\%likelihood\ of\ category\ i = \sum_{w_{attribute}=0\ or\ w_{attribute}} W_{Particle\ size\ distribution} + W_{Relative\ density} + W_{Fabric\ structure} + W_{Plasticity}$$

where, *i* is category (1, 2, 3, 4). *w_{attribute}* is weightage of particle size distribution, relative density, fabric structure and plasticity under category *i* (Table 2).

For each attribute, a weightage is allocated to a specific category that corresponds to the spoil, whereas all other categories within the same attribute are assigned a weightage of zero. The % likelihood is calculated for all categories, and the final BMAC category is determined as the category with the highest overall attribute weightage.

The framework, summarised in Table 2, classifies spoil into one of the four categories and provides peak shear strength parameters, cohesion (*c'*) and friction angle (*φ'*) for three strength mobilization modes, corresponding to unsaturated, saturated, and remoulded conditions (Table 3). This approach enabled efficient and practical geotechnical characterisation of spoil piles, providing valuable ground truth data for

the analysis and validation of remote sensing data obtained from aerial imagery.

In this study, a total of 82 piles located within a dump area spanning 62,567 m² were systematically categorised based on their variability and similarity. The distribution of groundtruths for each category across attributes is given in Fig. 3. The dump area consisted of coal spoil piles that were deposited in a paddock-like configuration. Each pile was then characterised according to the BMAC framework, and their spatial coordinates were measured using an Emlid Reach RS2 multi-frequency GNSS receiver. Lithology has been included in the ground truth, alongside the BMAC category, due to the recognition of its significance in influencing various BMAC attributes in several studies (Andrade et al., 2011; Bishwal et al., 2017). The selection of ground truths for the classification process, was based on two main criteria. Firstly, the accessibility of a pile was considered, ensuring that only piles that could

be readily accessed were included in the study. Secondly, the differentiation of the pile groups from the rest of the pile environment was taken into account. Within each pile group, a subset of piles with unique and distinguishable characteristics, setting them apart from the neighboring piles, was selected as ground truths. To establish similarity among the piles, attributes were assigned to those piles that displayed comparable features. Piles with similar nature, as determined through visual observation and analysis, were labelled with the same attributes. This labelling process enabled the creation of categories of piles that shared common characteristics, facilitating further analysis and interpretation of the data.

2.4. UAV Data processing

The raw images obtained from the UAV mission, which included

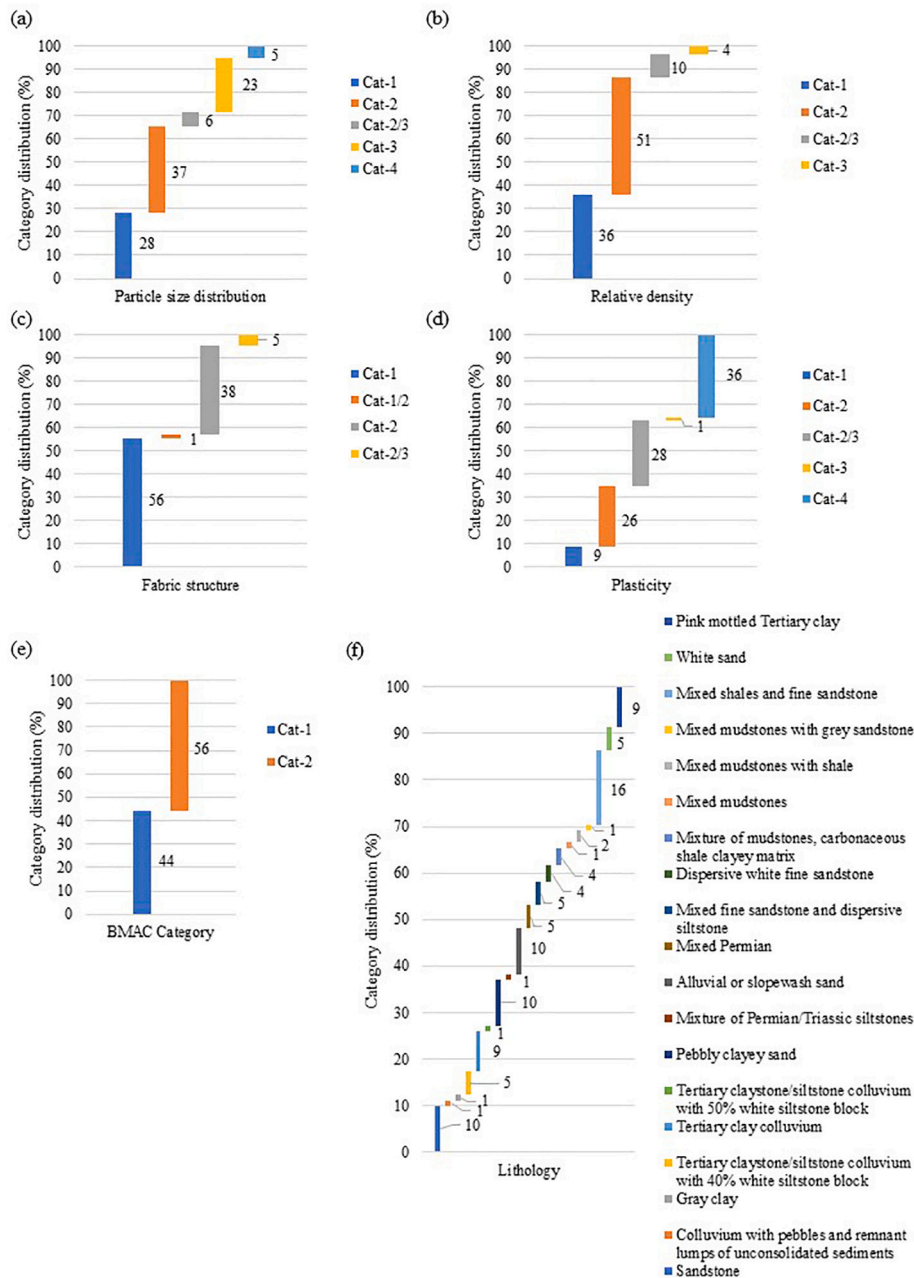


Fig. 3. Distribution of groundtruths for each category across attributes: (a) particle size distribution, (b) relative density, (c) fabric structure, (d) plasticity, (e) BMAC category and (f) lithology.

both optical and multispectral images, were processed using Pix4Dmapper (Pix4D SA, Lausanne, Switzerland), a structure from motion (SfM)-based photogrammetric stitching package. The image sets consisted of 838 optical and 1175 multispectral images and mosaicked in respective datasets.

The processing workflow involved an alignment process to identify common features in overlapping images, which helped estimate the camera's position and orientation for each image. This ensured the accurate alignment of the images, which is crucial for generating precise models. After alignment, Pix4Dmapper conducted camera calibration to estimate the intrinsic and extrinsic parameters of the camera. Intrinsic parameters, such as focal length, lens distortion, and principal point, and extrinsic parameters, including camera position and orientation in the scene, were determined. GCPs were utilised to refine the camera positioning and orientation data, serving as reference points to enhance the model's accuracy. Pix4Dmapper employed the GCPs to correct for camera position and orientation errors, thus facilitating georeferencing of the resulting orthomosaic and DSM.

Following camera calibration and GCP integration, Pix4Dmapper generated the final orthomosaic and DSM. For multispectral data, Pix4Dmapper included radiometric calibration tools. The raw multispectral bands were radiometrically calibrated within Pix4Dmapper's inbuilt workflow using MicaSense's CRP. This calibration process enabled the acquisition of absolute reflectance values from the multispectral images, ensuring consistency and comparability across different bands. Accurate analysis of multispectral data heavily relies on this calibration step. The resulting orthomosaics, both multispectral and RGB, had ground sampling Distances (GSD) of 4.41 cm and 1.22 cm, respectively. Finally, orthomosaics and DSMs were generated with a resolution of 5 cm for the multispectral data and 0.5 cm for the optical RGB data, as illustrated in Fig. 2.

2.5. Segmentation

We aimed to use object-based classification approaches, which are more suitable for identifying heterogeneous spoil piles, instead of pixel-based classification, which is less effective for this task. To this end, we developed a segmentation method specifically tailored for mining spoil piles with irregular-shaped blobs. Our segmentation approach was based on Voronoi tessellation and employed a series of processes: noise reduction, seed point detection, background seed point removal, and pile polygonisation. To address the challenges posed by the irregular debris distribution within the spoil piles, we applied a noise reduction technique called Gaussian blurring (Nusantika et al., 2021). This technique helped smooth out small irregularities and noise in the DSMs, improving the accuracy of the segmentation. Careful parameter selection was crucial to avoid blurring or distorting the shapes of the blobs.

Seed point detection was a critical first step in our Voronoi-based segmentation approach (Li et al., 2020). We used local maxima detection, which involved scanning the DSM pixel by pixel and comparing the elevation values of each pixel with its surrounding pixels within a defined neighbourhood. Pixels with higher elevation values than their neighbours were considered local maxima and selected as seed points for segmentation. Thresholding played a vital role in distinguishing between the background (valley between spoil piles) and the spoil piles (blobs) during the segmentation process. Adjustments were necessary to account for the variability in elevation values caused by the presence of irregularly shaped blobs. We applied Otsu's thresholding method (Otsu, 1979) to accurately differentiate between the background and pile foreground. Following the thresholding step, background seed points were eliminated through a binary operation, and the pile polygon areas were drawn based on Voronoi tessellation computed from the remaining seed points. By optimising the segmentation algorithm through adjustments to the sigma values of the Gaussian blur, we obtained a shapefile containing precise segmented objects for object-based image analysis.

To determine the optimal segmentation with the best parameter

setting, specifically the sigma values of the Gaussian blur during the smoothing phase of the segmentation algorithm, the Hoover metrics (Hoover et al., 1996) were employed. These metrics utilise measures based on the degree of area overlap between segments and can be extended to account for partial segment matching by employing a threshold. These metrics aim to identify under-segmentation and over-segmentation errors compared to 232 manually generated groundtruth segments. In other words, the metrics enable the identification of perfect, over-segmented segments (where a single pile is split into multiple segments), under-segmented (where multiple piles are inadequately partitioned), and missed.

To assess the performance of the segmentation with different sigma values during the smoothing phase, an overlapping threshold of 0.5 was utilised. Additionally, this threshold was used to evaluate how effectively the segmentation algorithms performed. Four scores were computed to evaluate the segmentation algorithms: the correct detection score, the over-segmentation score, the under-segmentation score, and the missing detection score.

2.6. Feature extraction

After the process of segmentation, the next step involved accurately representing each cluster of pixels to facilitate the precise classification of the respective object. Various features were utilised to describe the characteristics of objects and their associations with one another. In this study, these features can be categorised into four groups: spectral, textural, structural (edge), and statistical features. These features pertain to the relationship of objects within the segments and their relationship to the overall image. It is possible to extract each of these features from every generated object.

Table 4 presents the features employed in this study. Spectral features pertain to the spectral properties of an image, encompassing reflectance and radiance. They are used to analyse the spectral characteristics of objects, such as their color (Belgiu and Drăguț, 2014). Textural features describe the texture and patterns of objects, including roughness, coarseness, and directionality (Ma et al., 2015). Structural features encompass the size, shape, orientation, and arrangement of objects, involving lines, edges, and shapes (Grinias et al., 2016). Statistical features involve quantitative measures that describe the distribution and pattern of pixel intensities or colors in an image. These distinct feature types provide complementary information, enabling a comprehensive understanding of the image's content and context.

The multispectral data has been transformed to obtain the spectral features i.e., band ratio and normalised difference spectral indices. On the other hand, the RGB data has been converted to obtain spectral, textural, structural, and statistical features. The textural, structural, and statistical features have been derived from individual bands (i.e., red, green and blue) of the high-resolution RGB orthomosaic (spatial resolution of 0.5 cm) to ensure that as much information as possible is extracted. A detailed account of the extracted features used in this study is described and listed in Supplementary Information. A total of 564 features were extracted by determining zonal statistics (mean and standard deviation) of features within each segment generated from Voronoi-based approach. However, after eliminating those features that contained null values, the number of retained features was reduced to 516. Of these 516 features, 72 were generated from multispectral data while the remaining 444 features were generated from RGB data.

2.7. Feature selection

The inclusion of a large number of input features offers a comprehensive range of dimensions for training machine learning models within the method. However, high dimensionality can lead to inefficiency and prolonged convergence time for the models (Hu et al., 2019). To address this issue, feature selection was implemented to identify and select only the spoil pile features that are relevant to the

Table 4
Features extracted from optical and multispectral data.

	Feature category	Features	Parameters [Refer Supplementary for further details]	Number of features
	Spectral	Red, Green, Blue, Red-Green ratio, Green-Blue ratio, Red-Blue ratio		12
	Textural	Gabor texture features, Haralick features [Energy, Entropy, Correlation, Inverse Difference Moment, Inertia, Cluster Shade, Cluster Prominence, Haralick Correlation]	<i>Gabor</i> : θ values of $0, \pi/4$; σ values of 1,3; λ values of $0, \pi/4, \pi/2, 3\pi/4$, and γ values of 0.05, 0.5. <i>Haralick</i> : kernel sizes - $(3 \times 3), (5 \times 5), (7 \times 7), (9 \times 9), (11 \times 11)$.	432
RGB data	Structural (Edge)	Canny, Prewitt, Roberts, Scharr, Sobel	<i>Sobel, Prewitt and Scharr</i> use 3×3 kernel and <i>Roberts</i> edge operators use 2×2 kernel. <i>Canny</i> : Minimum and maximum values for double thresholding are chosen at 100 and 200, respectively.	30
	Statistical	Gaussian, Median	<i>Gaussian</i> : σ values of 3 and 7. <i>Median</i> : 3×3 kernel	18
Multispectral data	Spectral	Red, Green, Blue, RE, NIR, TIR, Red-Green ratio, Red-Blue ratio, Red-RE ratio, Red-NIR ratio, Red-TIR ratio, Green-Blue ratio, Green-RE ratio, Green-NIR ratio, Green-TIR ratio, Blue-RE ratio, Blue-NIR ratio, Blue-TIR ratio, RE-NIR ratio, RE-TIR ratio, NIR-TIR ratio, Normalised red-green difference index, Normalised red-blue difference index, Normalised red-RE difference index, Normalised red-NIR difference index, Normalised red-TIR difference index, Normalised green-blue difference index, Normalised green-RE difference index, Normalised green-NIR difference index, Normalised green-TIR difference index, Normalised blue-RE difference index, Normalised blue-NIR difference index, Normalised blue-TIR difference index, Normalised RE-NIR difference index, Normalised RE-TIR difference index, Normalised NIR-TIR difference index		72

overall spoil pile classification model. The mRMR algorithm (Peng et al., 2005) was employed as a feature selection technique to filter out noisy and redundant features from the high-dimensional spectral, textural, structural, and statistical features of spoil environment images. This algorithm effectively mitigates overfitting, enhances model performance, and conserves computational resources (Gopika and Meena, 2018). By eliminating unnecessary or redundant features, feature selection ensures that the selected subset of features is optimised for accuracy.

Given the presence of 516 features in this particular study, the mRMR algorithm (Hanchuan et al., 2005) was used to identify the optimal subset of features that yielded the highest level of accuracy for the model. The algorithm achieves this by selecting a subset of features that exhibit low similarity to each other while displaying a high correlation with the response variable (Zhang et al., 2012).

To evaluate the similarity between two features, X and Y, the mRMR algorithm assesses each feature individually in relation to the entire dataset and calculates their mutual information using the formula $I(X, Y)$.

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} (x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

where, $p(x, y)$ is joint probability distribution of X and Y variables, and $p(x), p(y)$ are marginal probability distribution of X and Y respectively.

Maximum relevancy ($maxV$) and minimum redundancy ($minW$) conditions to determine optimal set of features are defined as follows:

$$maxV, V = \frac{1}{|S|} \sum_{F_i \in S} I(F_i, C)$$

$$minW, W = \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j)$$

where, S is selected subset of features, F_i, F_j are elements of subset S, and C is response variable (i.e., class label of particle size distribution, relative density, fabric structure, plasticity and BMAC category).

The value of the mutual information quotient (MIQ) is then calculated to rank the features. A subset of features that are maximally relevant and least redundant is produced by the mRMR algorithm by choosing features with the highest MIQ values.

$$MIQ_{F_i} = \frac{V_{F_i}}{W_{F_i}}$$

where, relevancy and redundancy of a feature are denoted by V_{F_i} and W_{F_i} .

In accordance with principles of relevance and redundancy, all 516 features generated in this study have been ranked. The objective was to identify a compact and effective set of features, primarily for the purpose of reducing computational demands. The drop in the importance score reflects the algorithm's confidence in feature selection. A substantial drop implies confidence in selecting the most important predictor, while minor drops indicate that differences in predictor importance are not significant. Hence, features were selected up to the point where a significant drop was observed.

2.8. Machine learning algorithms

A range of machine learning algorithms were utilised to evaluate and compare their performance in the classification of spoil piles using UAV imagery. The adoption of multiple algorithms is crucial to assess their accuracy in relation to the specific classification task and dataset, as no single algorithm universally excels in all classification scenarios. Through the comparative analysis of various algorithms, the study aimed to identify the most effective algorithm for accurately classifying spoil piles. Consequently, a selection of classification models listed in Table 5 was employed.

In the context of decision tree, the number of splits in a decision tree analysis defines the level of complexity in the tree structure. Higher number splits improve categorisation accuracy, but also cast a higher chance of overfitting, vice-versa. Hence, an optimal number of split that provides a trade-off between accuracy and resource requirement produces the optimal decision tree model. This study used fine, medium and

coarse trees with maximum splits 100, 20, 4, respectively. Gini's diversity index, which estimates the likelihood of misclassifying a randomly selected element from the set based on the distribution of labels in the subset, was used as a split criterion to split the data effectively at each node (Zambon et al., 2006).

Under discriminant analysis, decision boundaries are defined based on the assumed type of relationship between features and response variables. Here, feature vectors are assumed to have Gaussian distribution. Linear discriminant analysis presumes a linear relationship and equal covariance matrices for each class whereas, quadratic discriminant analysis presumes non-linear relationship and separate covariance matrices for each class to define decision boundaries (Tharwat, 2016). Both types of models were deployed in this investigation.

Naive Bayes classifier is a probabilistic model that assumes conditional independence of features (Jahromi and Taheri, 2017). Gaussian naive Bayes and kernel naive Bayes are two types of naive Bayes classifiers used in this analysis. Gaussian naive Bayes assumes features are normally distributed and calculates the likelihood of assigning to the classes using Bayes theorem (Kamel et al., 2019), whereas kernel naive Bayes transfers input feature space into high dimensional feature space to divide classes easily using linear hyperplane (Lee et al., 2012).

Support vector machine (SVM) is an algorithm that defines optimal boundaries that separates classes by maximising the margin between the classes. Based on the type of function used to differentiate classes, orderly linear, quadratic, cubic, or Gaussian function, SVM is labelled as linear, quadratic, cubic, or Gaussian SVM (Anthony et al., 2008). Additionally, the kernel scale that smooths the decision boundary subclassifies the Gaussian SVM as fine, medium, and coarse (Son et al., 2022; Song et al., 2002). In this study, linear, quadratic, cubic, fine Gaussian (kernel scale: 5.7), medium Gaussian (kernel scale: 23) and coarse Gaussian (kernel scale: 91) were employed.

k-nearest neighbours (kNN) uses the most common class in the nearest feature space to assign a class value to an unknown data point (Singh et al., 2017). Based on the number of neighbours, kNN which uses Euclidian distance metric is further divided into fine, medium, and coarse kNN (in this study 1, 10, and 100 neighbours are used, respectively). Cosine kNN and cubic kNN uses cosine and Minkowski (cubic) distance metric to determine the similarity of data points. Weighted kNN uses square inverse distance weight and Euclidian distance metric to assess data similarity (Maghari, 2018). In this work, the number of neighbours was set at 10 for the cosine, cubic, and weighted kNN models.

Ensemble algorithms categorise unclassified data by voting on the

predictions produced by a set of classifiers (Dietterich, 2000; Fu et al., 2022). The ensemble algorithms - boosted tree, bagged tree, and RUS-Boosted tree - use decision trees as the basic classifier. Boosted trees use weak decision trees and integrate their outcomes, bagged trees train random samples independently using several decision trees, and the final prediction is made by voting on the outcomes of each algorithm, and RUSBoosted is a boosted tree derivative that uses random under-sampling to balance the class distribution in the data (Singh and Ranga, 2021). In this work, maximum number of splits and number of learners were kept as 20 and 30 for boosted, bagged, and RUSBoosted trees. Subspace discriminant and subspace kNN employ random subspace algorithm (Ashour et al., 2018; Rashid et al., 2021) to increase the accuracy of their base learners (i.e., discriminant and kNN, respectively). The number of learners and subspace dimensions for these two classifiers were set at 30 and 258, respectively.

Neural networks resemble biological neural networks and use interconnected nodes to determine the relationship between features and final predictions (Bishop, 1994). Based on the number of neurons in a layer, a neural network can be classified as: narrow, medium, and wide (Khan et al., 2022). This study employed only one layer with the sizes of 10, 25 and 100, respectively. All of these classifiers in this study use ReLU as their activation function, with a cap of 1000 iterations. According to the number of connected layers, neural networks can be divided into bilayered (i.e., two fully connected layers) and trilayered categories (i.e., three fully connected layers) (Khan et al., 2022). Both of these neural networks were used in this investigation, with a layer size of 10 for each, ReLU serving as their activation function, and a maximum of 1000 iterations. The input data was standardised in all types of neural networks.

Kernel approximation algorithms project the data that cannot be linearly separated into a linearly separable high dimensional space (Tariq et al., 2023). This study employed SVM and logistic regression as kernel functions for this data projection. The number of expansion dimension, regularisation strength and kernel scale parameters were set to auto to determine optimal parameters for the given data. Maximum iteration limit was kept at 1000 for both algorithms.

2.9. Accuracy assessment

The attributes of ground truth were merged with the segments generated using Voronoi-based segmentation, based on the corresponding ground truth points in the segmented polygon. These polygons, along with their attributes and features, were utilised in the training and testing processes.

Data partitioning, specifically the separation of training and test sets, has a significant impact on the performance of models. To ensure that a model can effectively generalise to unseen data, rather than relying on a single training and test dataset, k-fold cross validation is employed (Lyons et al., 2018). In the present study, a five-fold cross validation approach was utilised. This technique involves dividing the dataset into five equal parts, and subsequently training and testing the model five times. During each iteration, a portion of the dataset amounting to 20% (16 data points) is allocated for testing, whereas the remaining four portions, constituting 80% (66 data points), are employed for training. This process is repeated five times, with each portion of the dataset serving as the test set exactly once. The results obtained from the five iterations are then averaged to provide an overall assessment of the model's performance.

The assessment of predictive capabilities for distinct models on the coal spoil attributes involved the utilisation of evaluation metrics, including accuracy, precision, and recall, alongside the examination of confusion matrices. Overall accuracy served as a fundamental metric, quantifying the ratio of accurately predicted classifications to the total count of classified data.

Table 5
Classifiers used in this study.

Classifiers	Types of deployed classifiers	Reference
Decision tree	Fine, medium, and coarse	(Loussaief and Abdelkrim, 2018)
Discriminant analysis	Linear and quadratic discriminant	(Arabameri and Pourghasemi, 2019)
Naive Bayes	Gaussian and kernel naive Bayes	(McCann and Lowe, 2012)
Support vector machine (SVM)	Linear SVM, quadratic SVM, cubic SVM, fine Gaussian SVM, medium Gaussian SVM, coarse Gaussian SVM	(Loussaief and Abdelkrim, 2018)
k-nearest neighbours (kNN)	Fine kNN, medium kNN, coarse kNN, cosine kNN, cubic kNN, weighted kNN	(Loussaief and Abdelkrim, 2018)
Ensemble	Boosted Trees, bagged Trees, subspace discriminant, subspace KNN, RUSBoosted Trees	(Arboleda, 2019)
Neural network	Narrow neural network, medium neural network, wide neural network, bilayered neural network, trilayered neural network	(Corenblit et al., 2023)
Kernel approximation	SVM kernel, logistic regression kernel	(Tien Bui et al., 2016)

$$\text{Overall accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

In this context, true negative (TN) represents the count of instances accurately identified as negatives, true positive (TP) signifies the count of instances correctly identified as positives, false negative (FN) denotes the count of instances erroneously classified as negatives, and false positive (FP) indicates the count of instances erroneously classified as positives.

In addition, per-class accuracy was determined using precision, recall, and F1-score.

Precision quantifies the ratio of correct positive identifications, indicating the probability that the predicted class for the spoil category truly belongs to that particular class.

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

Recall determines the ratio of true positives among all actual positives that were accurately identified.

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

F1-score represents the harmonic average of precision and recall.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In order to evaluate the accuracy of models when utilising features derived from both RGB and multispectral data, five-fold cross validation was employed. Subsequently, an assessment of the accuracy of the selected subset features using mRMR algorithm was conducted.

3. Results

3.1. Segmentation

Table 6 displays the segmentation parameters associated with the Hoover metrics scores obtained from Voronoi-based segmentations. The Hoover scores, namely R_c , R_f , R_a , and R_m , represent the metrics for correct detection, over-segmentation, under-segmentation, and missed detection, respectively. The correct detection score gauges the proportion of spoil pile segments accurately identified in the image. A higher detection score indicates superior performance in this regard. The over-segmentation score quantifies the percentage of detected piles that are segmented into multiple parts. Conversely, the under-segmentation score reflects the percentage of piles that are incompletely segmented. The missed detection score indicates the percentage of piles that remain undetected. Lower scores in over-segmentation, under-segmentation, and missed detection signify better performance in accurately detecting all objects present in the image.

The results demonstrate that the Voronoi-based segmentation achieves optimal performance when utilising a sigma value of 12 for Gaussian blurring. When the sigma value is set to 12, both correct detection and over-segmentation of piles yield a Hoover score of 0.77. Fig. 4 visually illustrates the output generated by the Voronoi-based segmentation algorithm, showcasing precise delineation of individual spoil piles in both the DSM (Fig. 4(b)) and the true color composite of the orthomosaic (Fig. 4(b)). Fig. 4(d) displays a zoomed view of a subarea within the dumpsite, highlighting the ground truth segment, Voronoi-based segments, and the visual representation of Hoover scores. These results were obtained using images captured by the Micasense Altum-PT sensor.

3.2. Feature selection

The 516 features in the dataset were ranked based on their importance, considering their relevance to the response variable, i.e. particle size distribution, consistency/relative density, fabric structure, liquid

limit, BMAC category, and lithology. Additionally, the redundancy of each feature in the dataset was taken into account with a measure of mRMR.

A substantial drop suggests the algorithm's strong confidence in choosing the most crucial predictor, while minor drops indicate that differences in predictor importance are not notable.

Fig. 5 illustrates the outcomes of the mRMR algorithm. Specifically, considering particle size distribution, a significant drop in importance scores (6.95×10^{-3} to 2.58×10^{-13}) occurs after the 462nd feature. For relative density, the importance scores exhibit a significant decrease after the 419th feature. In the case of fabric structure, the initial four features display minimal score variation, followed by a sudden drop in importance scores beyond that point. In terms of plasticity, a noteworthy drop in importance scores emerges after the 445th feature. Regarding the BMAC category, there is an abrupt decrease in the importance score after the 337th feature. Lastly, in the context of lithology, a substantial decrease in importance score is observed after the 12th feature.

The results of the mRMR algorithm, including the top 20 features for each attribute and their corresponding confidence of feature selection, are presented in Fig. 6.

The results indicate that Haralick features, Gabor texture features, and spectral features derived from multispectral bands are ranked as the topmost important features for the classification of all the attributes.

3.3. Accuracy assessment

3.3.1. Comparison of accuracy of features derived from RGB and multispectral data

The best-performing machine learning algorithm and its corresponding relative overall accuracy for features derived from optical RGB, multispectral, and a combination of both data sources are illustrated in Fig. 7. The results demonstrate that input features obtained from RGB data achieve the highest overall accuracy for classifying particle size distribution (fine kNN - 90.1%), relative density (wide neural network - 87.7%), and fabric structure (wide neural network - 95.1%). On the other hand, input features derived from multispectral data result in the highest overall accuracy for plasticity (fine kNN - 90.4%).

For the BMAC category, the highest overall accuracy (92.6%) is attained when features are combined from both RGB and multispectral data using the ensemble (subspace discriminant) algorithm. The use of solely RGB data also yields a comparable accuracy for the BMAC category. In the case of lithology, the ensemble (subspace discriminant) algorithm using features combined from both RGB and multispectral data achieves the best overall accuracy (80.2%).

When comparing algorithms using RGB data as input, they exhibit 2.4%, 3.7%, and 6.2% higher accuracy in particle size distribution, relative density, and fabric structure, respectively, compared to algorithms using features from both RGB and multispectral data. Conversely, algorithms utilising features derived from multispectral data as input show 7.7% higher accuracy for plasticity than those using features from both RGB and multispectral data combined. For the BMAC category and lithology, algorithms using features from both RGB and multispectral data as input outperformed other approaches.

Table 6

Hoover metrics scores for corresponding parameters of Voronoi-based segmentation.

Gaussian (Sigma)	R_c	R_f	R_a	R_m
9	0.651	0.109	0.015	0.025
10	0.646	0.124	0.020	0.022
11	0.659	0.102	0.027	0.025
12	0.681	0.090	0.021	0.021
13	0.656	0.095	0.037	0.034
14	0.642	0.106	0.041	0.033
15	0.634	0.101	0.052	0.036

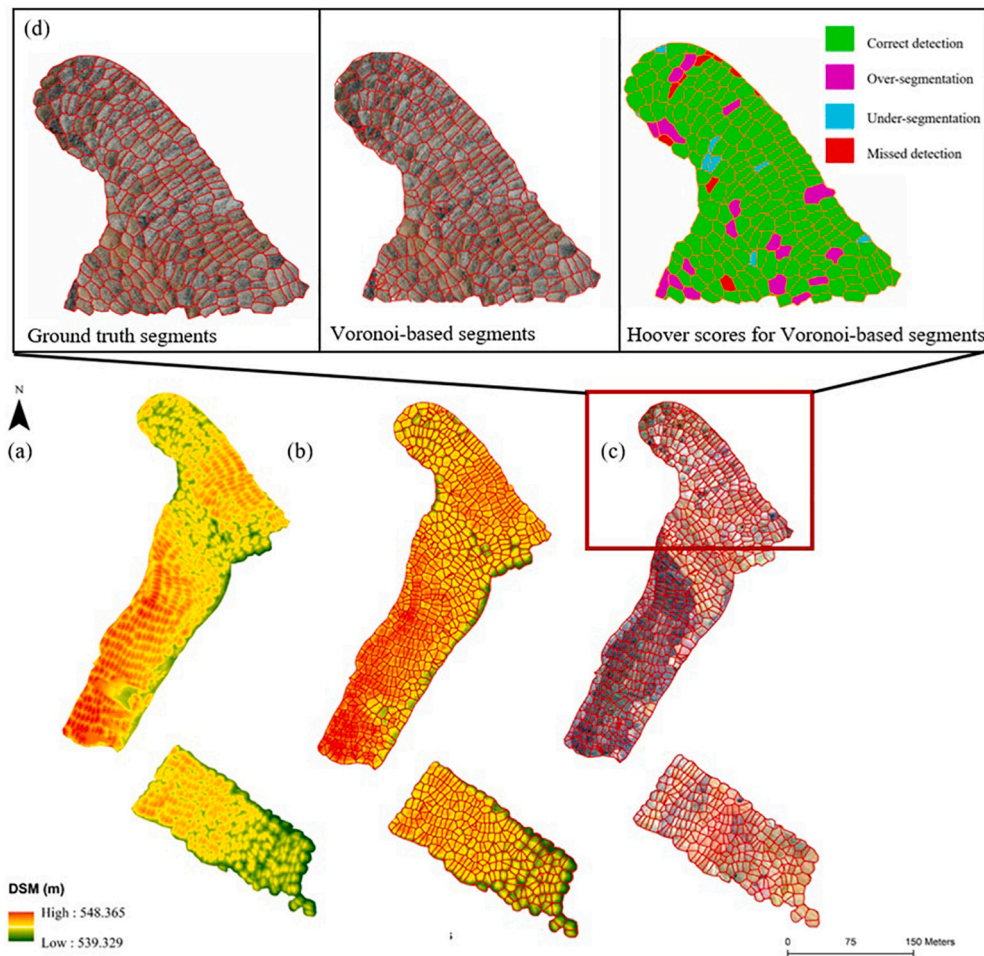


Fig. 4. (a) Digital surface model (DSM) of the selected area in the dump site revealing the morphological variation of piles, (b) Voronoi-based segments derived from the DSM, (c) segments overlaid on the true color composite of the orthomosaic obtained from Micasense Altum-PT sensor images, demonstrating effective segmentation, (d) zoomed view of ground truth segments, Voronoi-based segments and a visual representation of Hoover scores.

Thematic maps created using all features derived from RGB and multispectral data as input for the best machine learning algorithms, representing various relevant attributes are presented in Fig. 8. In the study area, the thematic maps reveal each attribute's distribution of specific categories. The results show that there are more spoil piles falling into category 3 for particle size distribution, category 2 for relative density, category 1 for fabric structure, category 2/3 for plasticity, and category 2 for BMAC category. Additionally, the lithology in the study area is characterised by a mix of shales and fine sandstones.

An evaluation of the algorithm's performance using per-class accuracy metrics such as precision, recall, and F1-score are given in Fig. 9. Analysing the F1-scores for different categories reveals the algorithm's performance in each specific category. Category 2 demonstrates high F1-scores in particle size distribution (F1-score of 0.92), relative density (high F1-score of 0.89), and BMAC category (F1-score of 0.91). For fabric structure, category 1 exhibits a high F1-score of 0.91, while category 4 performs well in terms of plasticity (F1-score of 0.88). Regarding lithology, the mixed Permian and dispersive white fine sandstone category achieves a perfect F1-score of 1.

It is worth noting that both fabric structure category 1/2 and plasticity category 3, yield an F1-score of 0. Additionally, categories such as colluvium with pebbles and remnant lumps of unconsolidated sediments, gray clay, tertiary claystone/siltstone colluvium, a mixture of Permian/Triassic siltstones, mixed mudstones, mixed mudstones with shale, and mixed mudstones with gray sandstone all result in an F1-score of 0.

3.3.2. Accuracy assessment of selected subset features from mRMR

The overall accuracy (Fig. 10), utilising mRMR-selected features (Section 3.2), plays a crucial role in assessing the performance of classification models. Specifically, an overall accuracy of 87.7% was achieved for particle size distribution through the utilisation of the ensemble (subspace discriminant) model in conjunction with a selection of 462 optimal mRMR features. In contrast, the relative density attained an overall accuracy of 87.7%, employing a set of 419 selected features identified by the medium neural network algorithm. For the fabric structure attribute, an overall accuracy of 82.7% was realised, guided by the weighted kNN approach, which made use of a minimal subset of 4 selected features. Plasticity achieved an accuracy of 87.7%, through wide neural network algorithm, incorporating a collection of 445 selected features. Similarly, the BMAC category attribute demonstrated an accuracy at 96.3%, relying on the linear discriminant method, with 337 best selected features. In contrast, the attribute of lithology exhibited a lower overall accuracy of 66.7% while employing the ensemble (bagged trees) approach with 12 selected features.

In the context of the per-class evaluation metrics, Fig. 11 depicts precision, recall, and F1-score metrics for various categories within distinct attributes. Notably, the particle size distribution exhibited varying performance across its categories, with Cat-2 displaying a particularly high F1-score of 0.93. However, Cat-2/3 and Cat-4 demonstrated perfect precision but a comparatively lower recall and F1-score. The relative density also showcased noteworthy results, with Cat-1 achieving a precision of 0.93 and an F1-score of 0.91, reflecting

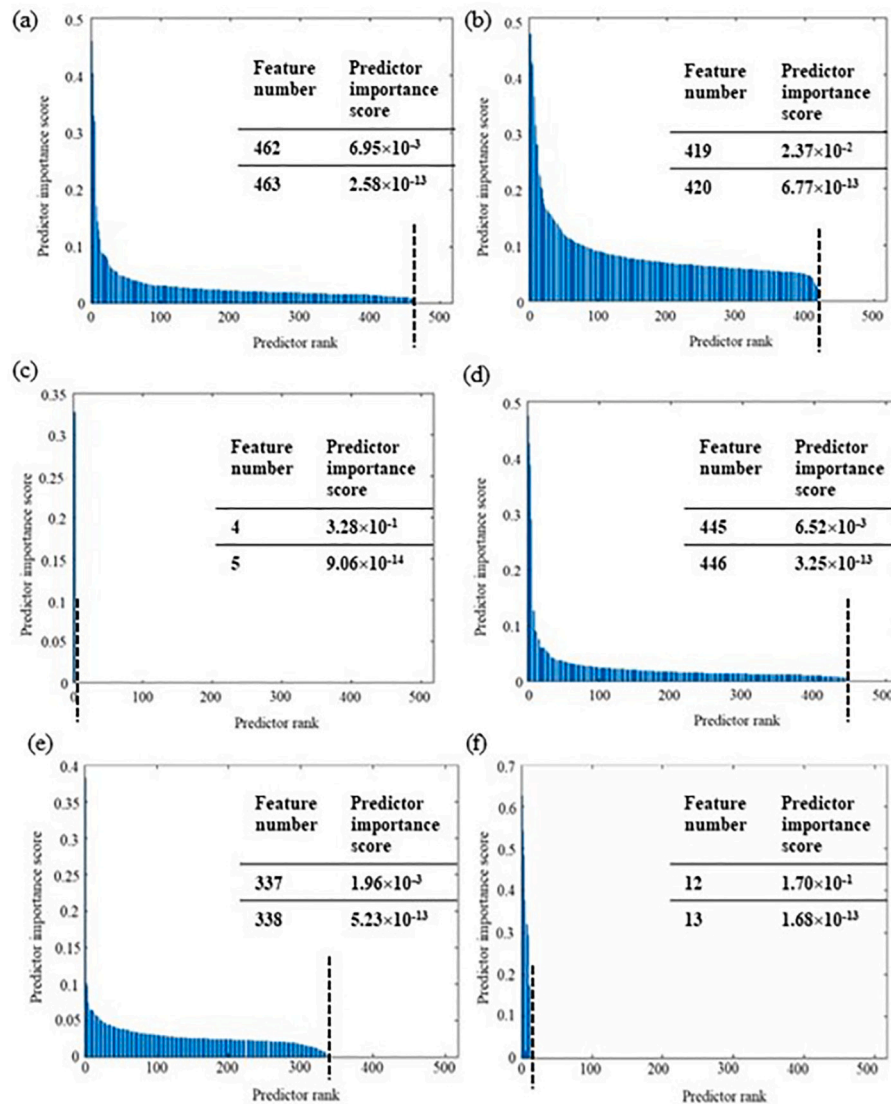


Fig. 5. An evaluation of feature importance using minimum redundancy maximum relevancy algorithm. All 516 features were ranked in descending order of importance when the response variable is (a) particle size distribution, (b) relative density, (c) fabric structure, (d) plasticity, (e) BMAC category and (f) lithology.

strong overall performance. Conversely, F1-score of fabric structure illustrated a distinction between the performance of Cat-1 and Cat-1/2 categories, where the latter exhibited negligible values due to a lack of data. The plasticity further highlighted the need for precision-recall balance, as Cat-1 displayed perfect precision but reduced recall (0.86), while Cat-3 exhibited negligible value. The BMAC category showed F1-scores of 0.96 and 0.97 for both the categories. The lithology encompassed a diverse range of categories, with notable performance observed in sandstone achieving F1-score of 0.94.

4. Discussion

This study presents a novel workflow that introduces the application of an object-based approach coupled with machine learning to characterise spoil piles with rough morphology, specifically for geotechnical applications. The proposed methodology aims to streamline and automate the analysis and characterisation of spoil piles, offering a more efficient and reliable alternative to traditional manual approaches. The methodology was tailored to suit the unique conditions of the site, taking into account factors such as the distribution of specific data (including distinct lithologies exclusive to the area) and the utilisation of images captured under optimal conditions (e.g., sunny day) to ensure

accurate categorisation across all classifications. While the primary objective was to determine the final BMAC category, which holds significant importance for the industry, this study also examined other attributes related to the BMAC category. This approach provides additional insights into the variability of these attributes, even when different materials fall under the same BMAC category. By leveraging object-based techniques, this study opens up a promising avenue for effectively assessing and understanding the geotechnical properties of spoil piles. It paves the way for improved geotechnical analysis and decision-making processes, offering new possibilities for advancements in this field.

4.1. Segmentation

Object-based classification methods offer distinct advantages over pixel-based classification, such as mitigating the influence of variations in acquisition parameters and delivering realistic and geometrically precise spatial feature mapping (Blaschke, 2010). Nonetheless, segmenting high-resolution remote sensing images of mining spoil piles with diverse terrains poses significant challenges. A primary hurdle lies in devising algorithms and parameters that consistently yield stable segmentation performance for all objects and locations in the image

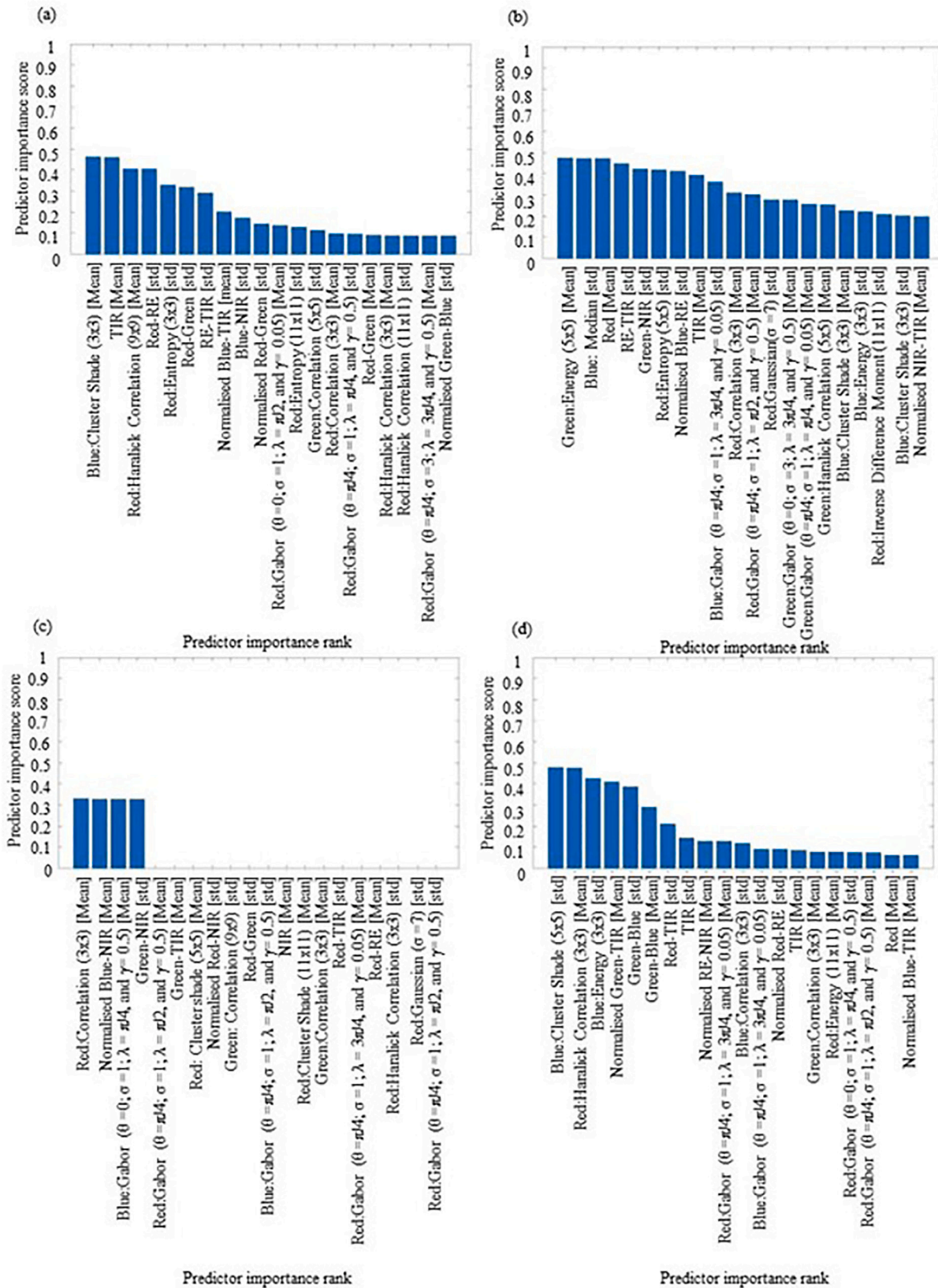


Fig. 6. Plot showing the top 20 most important features for classification of (a) particle size distribution, (b) relative density, (c) fabric structure, (d) plasticity, (e) BMAC category and (f) lithology according to minimum redundancy maximum relevancy algorithm.

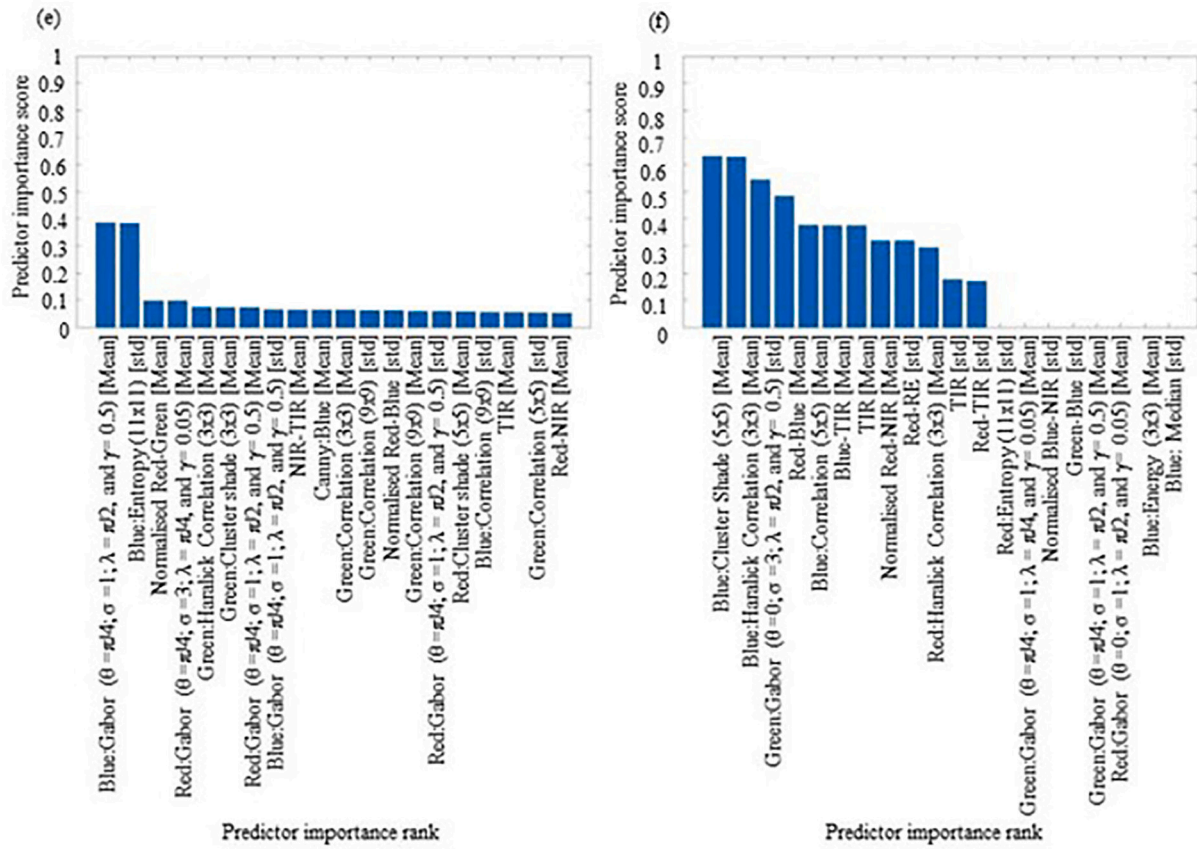


Fig. 6. (continued).

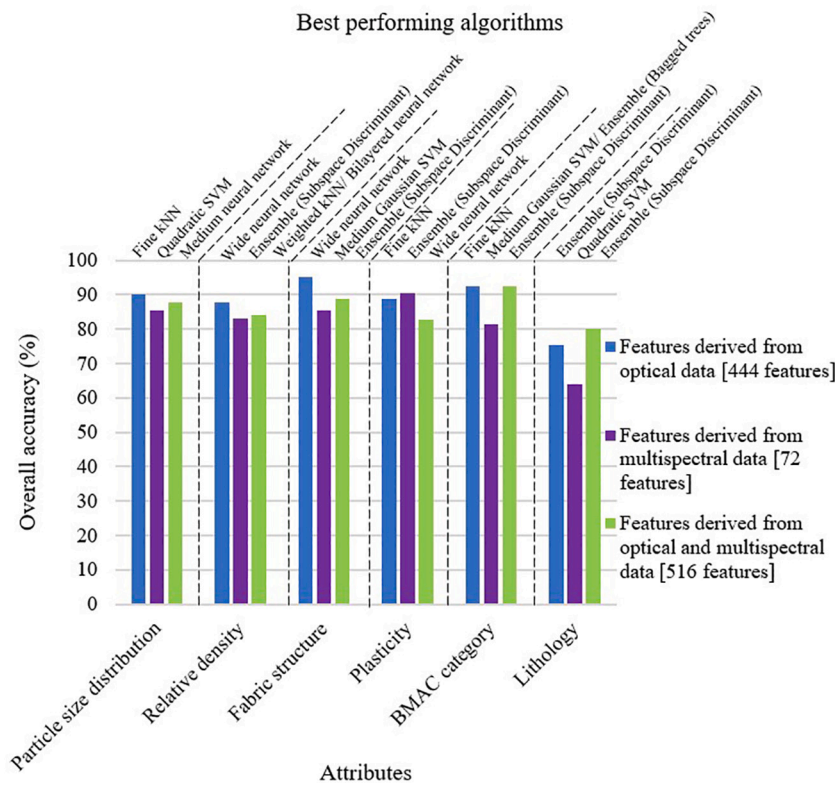


Fig. 7. Overall accuracy of best performing machine learning algorithms for attributes related to BMAC framework when input features are derived from optical (RGB) data, multispectral data and combination of both.

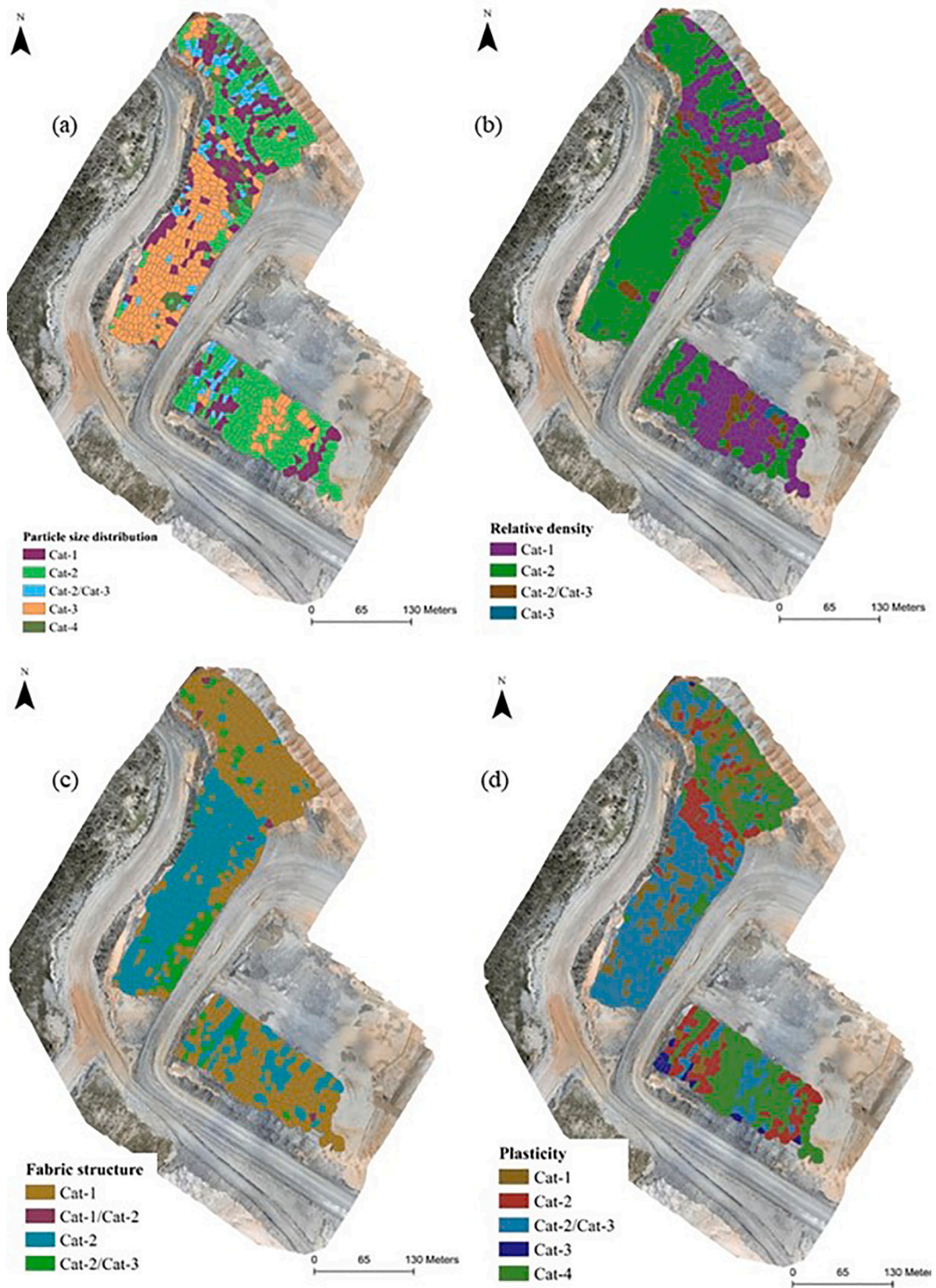


Fig. 8. Object-based classifications using all features derived from both RGB and multispectral data as input for: (a) particle size distribution when deploying medium neural network, (b) relative density when deploying bilayered neural network, (c) fabric structure when deploying ensemble (subspace discriminant), (d) plasticity when deploying wide neural network, (e) BMAC category when deploying ensemble (subspace discriminant), and (f) lithology when deploying ensemble (subspace discriminant).

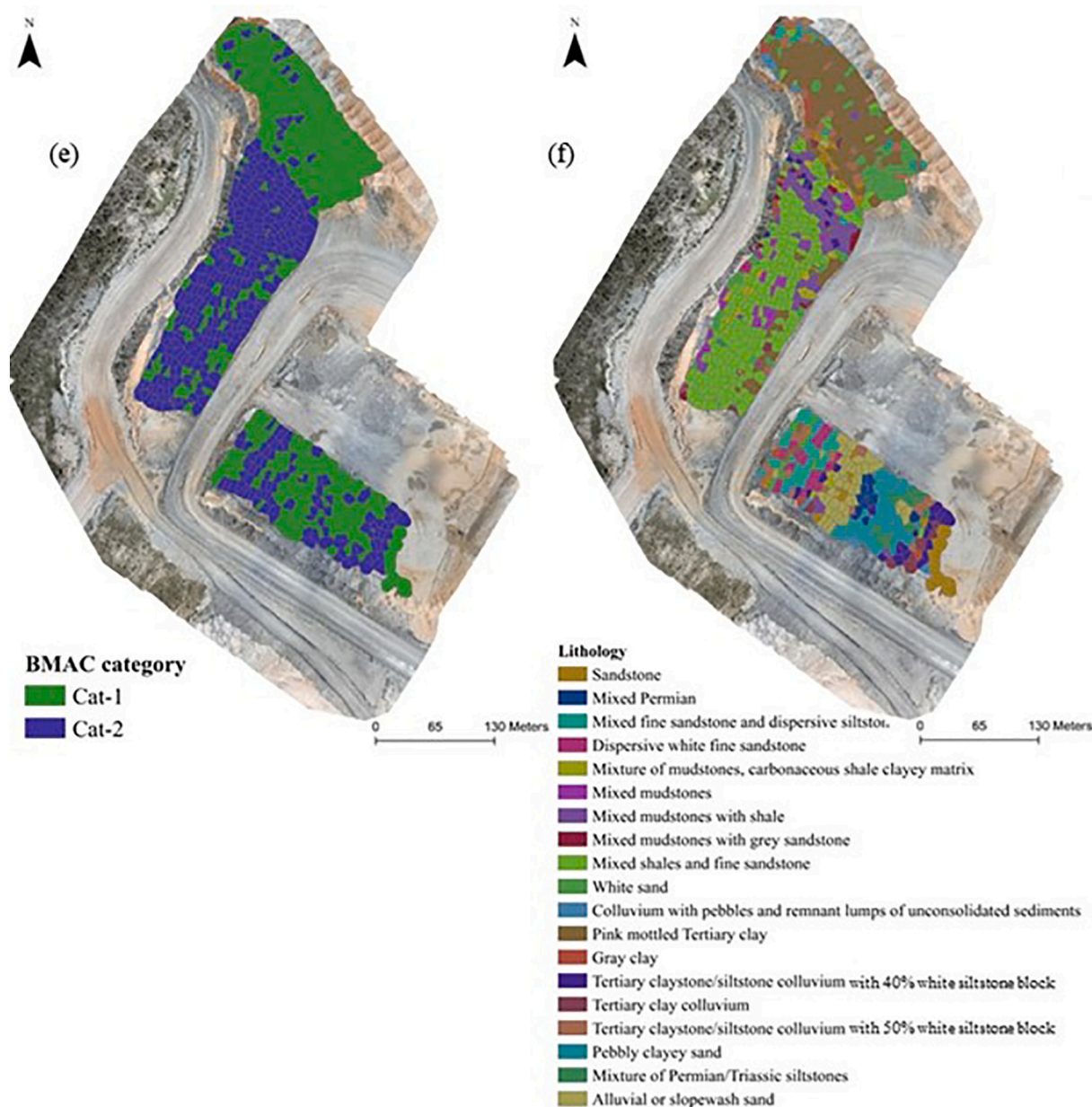


Fig. 8. (continued).

(Hassanein et al., 2018). Moreover, segmentation quality is subjective, depending on the user's definition of objects of interest and their level of granularity (Thiruchittampalam et al., 2023a). Noteworthy segmentation shortcomings encompass over-segmentation, where the object of interest fragments into multiple pieces, and under-segmentation, where the object of interest merges with its surroundings. These issues directly affect classification accuracy, underscoring the pivotal role of spoil pile segmentation in OBIA. In our study, we adopted a Voronoi-based segmentation approach relying on morphology on DSM. This choice was motivated by the observation that morphology-based segmentation outperforms color-based methods, which are sensitive to external factors like sun angle and shadows (Thiruchittampalam et al., 2023a). Since DSM is less affected by these external factors, morphology-based segmentation proves more robust for undulating pile terrains. Nevertheless, the irregular-shaped blobs in DSMs posed segmentation challenges. To address this, we applied Gaussian blurring as a pre-processing step to smooth out minor irregularities and noise. Careful parameter selection

was crucial to prevent excessive blurring or distortion of blob shapes (Jones et al., 2005). Seed point detection, a fundamental aspect of the Voronoi-based segmentation approach, relied on local maxima detection. This method facilitated the identification of potential seed points for segmentation based on elevation values. Adjustments to the thresholding step were necessary to accurately distinguish between the irregular-shaped blobs and the background. The study found that optimal segmentation could be achieved by combining a Voronoi-based segmentation method with a Gaussian filter. The choice of the sigma value played a pivotal role in the segmentation results, particularly in noise reduction. A sigma value below a specific threshold failed to effectively differentiate individual pile maxima from significant noise in the images. Conversely, a sigma value above a certain threshold resulted in blurry images, making it challenging to accurately separate one pile from another during local maxima identification. Through meticulous experimentation, we determined that selecting an appropriate sigma value, in this case, 12, provided the best segmentation results. This value

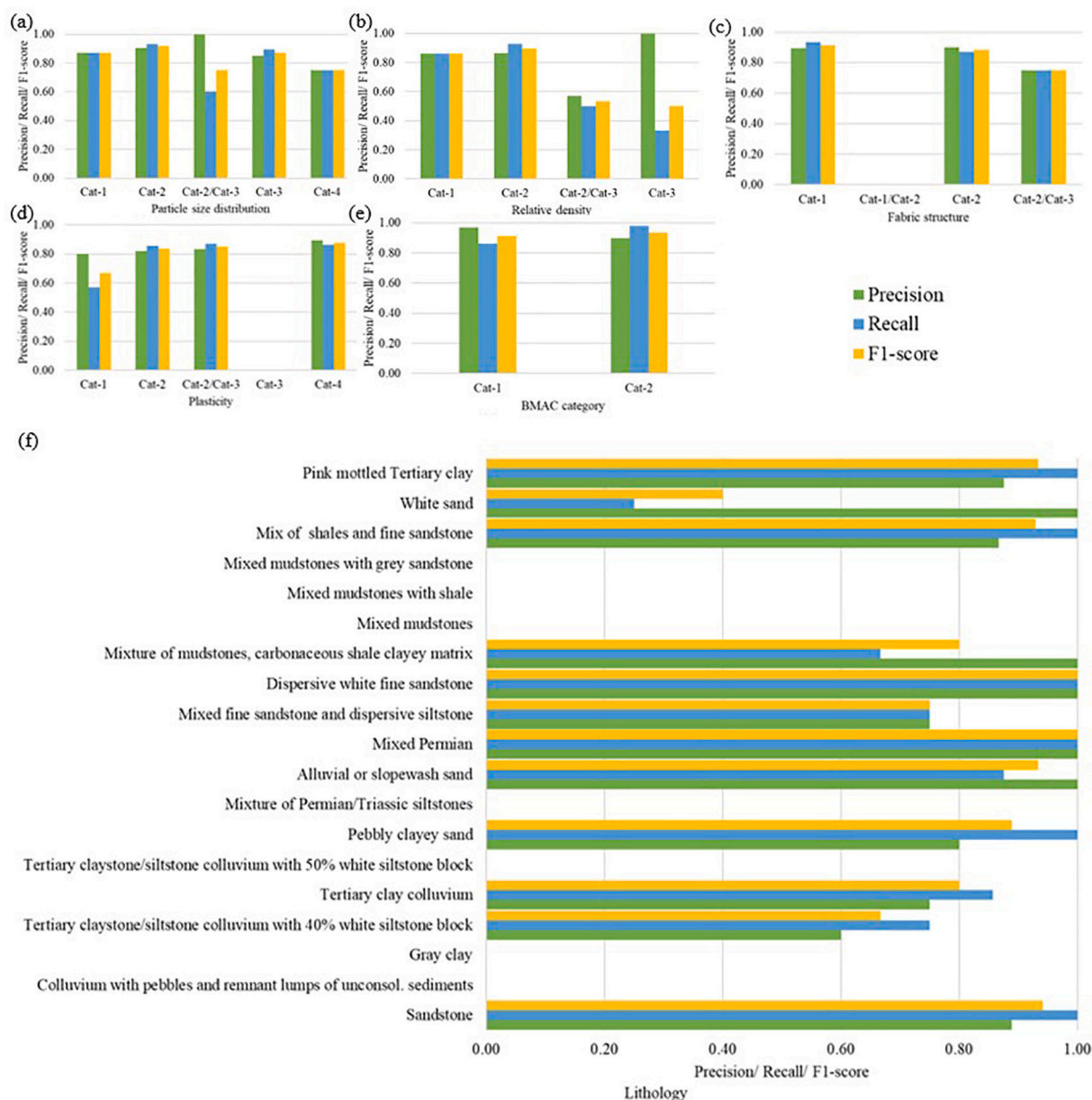


Fig. 9. Precision, recall and F-Score of classifications, respectively for categories of (a) particle size distribution when deploying medium neural network, (b) relative density when deploying bilayered neural network, (c) fabric structure when deploying ensemble (subspace discriminant), (d) plasticity when deploying wide neural network, (e) BMAC category when deploying ensemble (subspace discriminant), and (f) lithology when deploying ensemble (subspace discriminant). It is noteworthy, rare classes in field are unrecognised.

struck a balance between noise reduction and maintaining clear boundaries for individual spoil piles, ensuring precise segmentation and subsequent analysis. Thiruchittampalam et al. (2023a) provides additional background on the performance of segmentation of coal spoil pile and a comprehensive analysis of the Hoover analysis.

4.2. Feature extraction and selection

Performance evaluation of a classification machine learning algorithm, along with relative overall accuracy for a given feature set, underscores the crucial role of feature selection in achieving high attribute-specific accuracy. This emphasises the need for a precise feature combination to optimise classification performance while minimising computational demand.

The study underscores the significance of feature selection by identifying optimal feature sets for each attribute, leading to the improved

classification accuracy (Zhou et al., 2018). Notably, the study reveals that the nature of the attribute being classified and scale influences the choice of features.

Texture within an image can vary in scale. Different window sizes enable capturing texture information at various scales. For example, smaller windows highlight micro-texture, like fine grain or small patterns, while larger windows focus on macro-texture, such as larger structures. To accommodate this, Haralick features were extracted using window sizes like 3×3 , 5×5 , 7×7 , 9×9 , and 11×11 and Gabor features, resulting in 432 textural features, totalling 516 features along with spectral, statistical and structural features.

Feature selection involved the mRMR algorithm as illustrated in Fig. 5. mRMR's feature selection depends on various factors, including response variable variance and feature-response variable relationships. Higher response variable variance or weak feature-response variable correlation leads to the selection of more features, as seen in particle size

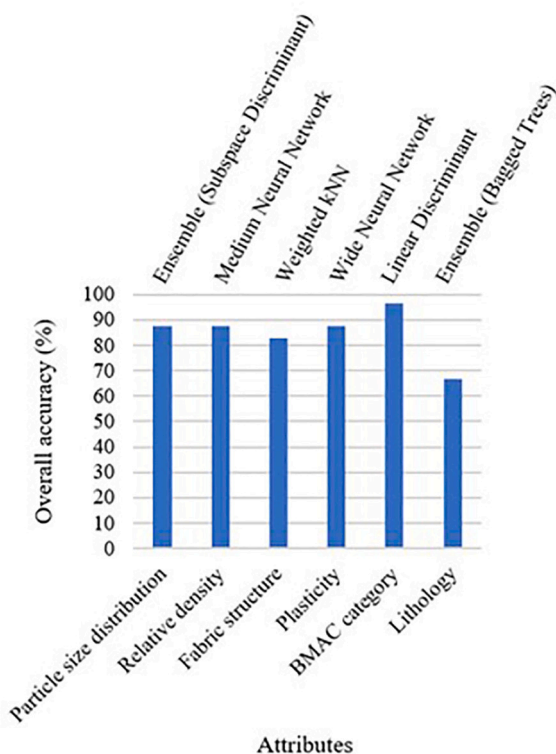


Fig. 10. Overall accuracy of best performing machine learning algorithms for attributes related to BMAC framework, employing feature selection via the mRMR algorithm.

distribution (462 features). Conversely, strong correlations or multicollinearity led to fewer features, as observed in fabric structure (four features). These observations highlight mRMR's adaptive nature, adjusting feature selection based on specific response variable characteristics (Wang et al., 2022) to enhance attribute characterisation and prediction accuracy. Future work could focus on extracting more robust features to reduce computational demands. Nevertheless, this study provides an initial framework for feature extraction and selection in spoil characterisation.

The study's findings highlight the importance of multispectral bands, Haralick features, and Gabor texture features for classifying attributes, likely influenced by the attributes' dependence on mineralogy and spatial arrangement of particles. Rankings of these features vary with attributes. For example, Gabor texture features dominate in BMAC classification, while other attributes favour Haralick texture features. This variance underscores how different attributes rely on these features, and at least one multispectral characteristic among the top four predictors emphasises the importance of multispectral data in accurate spoil attribute classification.

4.3. Accuracy assessment

Multispectral data is vital for accurate material classification based on unique spectral signatures, while high spatial resolution RGB data excels in capturing visual attributes like texture and shape. Specifically, plasticity benefits most from multispectral data, whereas other attributes perform best with RGB data, underscoring their reliance on particle spatial arrangement. The study's use of diverse spatial resolution features from both RGB (0.5 cm) and multispectral sensors (5 cm) enhances categorisation of specific attributes. Notably, the BMAC category and lithology exhibit superior accuracy while using combined features, emphasising the potential of multi-resolution image analysis.

The analysis reveals varying feature requirements for optimal accuracy across attributes. Some attributes need fewer features for high

accuracy, while others require more. For instance, particle size distribution attains peak accuracy with the top 462 ranked features, while fabric structure achieves this with just the top 4 ranked features. Feature subset selection enhances accuracy by addressing the "curse of dimensionality," reducing noise, and improving model efficiency. However, feature selection is data-dependent, and sometimes all features are necessary for the best performance, so it should be done carefully, considering the specific dataset and problem domain. In this study, feature subsets from the mRMR algorithm provided the highest accuracy, except for fabric structure and lithology.

Per-class accuracy is a crucial performance metric for evaluating the effectiveness of machine learning models in multi-class classification tasks (Nasiri et al., 2023). However, when dealing with imbalanced datasets, it is essential to consider additional evaluation metrics to gain a comprehensive understanding of model performance. In multi-class classification, precision, recall, and F1-score are commonly used metrics, providing insights into the performance of the model for each individual class. Mine spoil classes often exhibit imbalances, leading to underrepresentation of less common classes due to model optimisation for overall error rates. Some classes may still show an F1-score of 0 after feature selection, highlighting the need for addressing class imbalance with techniques like oversampling or undersampling. However, these techniques should be used cautiously, considering potential overfitting, especially with limited data in the study.

4.4. Study contribution

In this study, the primary focus is on the application of machine learning algorithms to classify spoil based on the BMAC spoil categorisation framework. The research presents a comprehensive methodology for spoil classification utilising RGB and multispectral datasets, which can be extended to other regions with similar spoil categories. The findings demonstrate that incorporating textural, statistical, and structural information extracted from the RGB and multispectral data enhances the accuracy of the algorithms in generating spoil categorisation maps. The integration of an object-based classification approach, combined with machine learning algorithms, provides an effective system for expert-driven spoil classification. This approach not only improves the accuracy of the classification process but also offers practical implications for various applications such as stability analysis, mapping factors for mine dump design and construction, and land-use planning in areas with rough morphology and similar spoil classes. The results of this study contribute to the field by demonstrating the feasibility and benefits of utilising machine learning techniques in the classification of spoil. The proposed methodology can aid decision-making processes in geotechnical applications and enhance the understanding of spoil characteristics in similar contexts.

5. Conclusions and further study

Mining activities generate large volumes of spoil, which can pose significant environmental and safety risks if not properly managed. Accurate and efficient geotechnical characterisation of spoil piles is crucial to mitigate design and structural stability issues. Traditionally, this task has relied on field experts who face hazardous conditions during the manual field assessment. However, technological advancements, such as UAVs and SfM photogrammetry, offer a promising solution to automate spoil characterisation, reducing the time and bias associated with manual methods. Taking advantage of these technological advancements, this study proposes an object-based approach that leverages high-resolution RGB and multispectral data to characterise the geotechnical properties of spoil piles. The methodology involves a combination of Voronoi-based segmentation, spectral analysis, textural analysis, structural analysis, statistical analysis, feature extraction techniques, mRMR based feature selection, and machine learning-based classification.

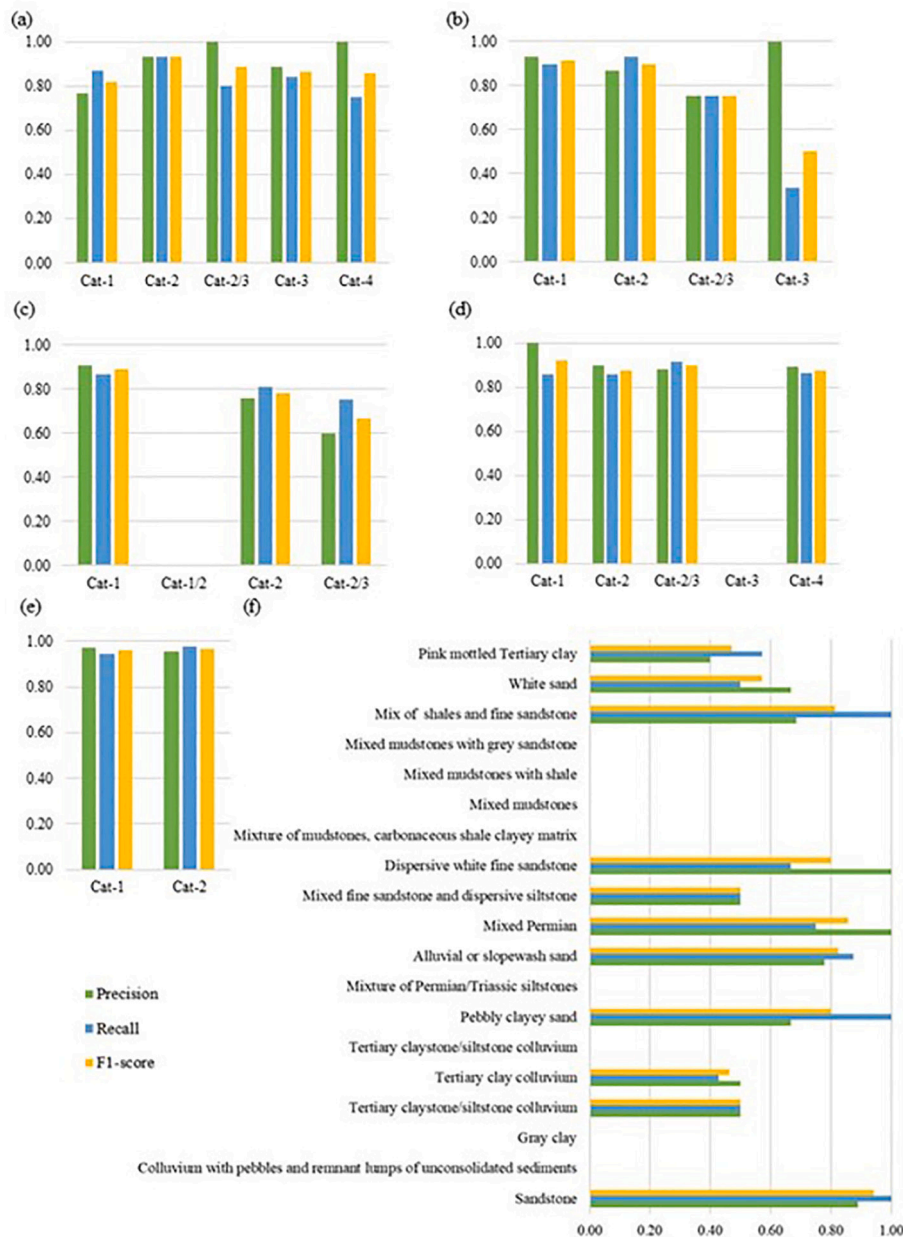


Fig. 11. Precision, recall and F1-Score of classifications, respectively for categories of (a) particle size distribution, (b) relative density, (c) fabric structure, (d) plasticity, (e) BMAC category, and (f) lithology. It is noteworthy, rare classes in field are unrecognised.

In this study, we conducted comprehensive experiments to investigate the characterisation of spoil piles using the BMAC framework with RGB and multispectral sensors. We found that applying a Voronoi-based segmentation method with a Gaussian filter of 12 sigma value during the preprocessing stage of DSM yielded optimal segmentation for accurately characterising the geotechnical properties of spoil piles. Our research revealed that multispectral data features are essential for identifying specific materials based on their spectral signatures, while features derived from RGB data with high spatial resolution excel in capturing visual information such as texture and shape. To reduce the computational load, we employed the mRMR algorithm for feature selection among the 516 generated features. The results highlighted the significance of multispectral features, Haralick features, and Gabor texture features in characterising all the attributes of spoil piles. However, the optimal combination and number of features varied depending on the specific attribute under examination, reflecting the intricate relationship between features and attributes.

Based on the successful outcomes of our study, we recommend that future research endeavours adopt the methodologies outlined in this paper to explore and evaluate materials with different geological and geotechnical characteristics. Furthermore, future studies could investigate advanced deep-learning techniques in conjunction with larger datasets. Incorporating very high-resolution satellite imagery such as, Maxar satellite constellation, into the existing methodology can open up new opportunities for detailed spoil categorisation. The increased spatial and temporal information available through such high-resolution satellite imagery can provide valuable insights over the long term. When implemented, this automated characterisation holds the potential to become a powerful tool for proactive dump stability assessment, revolutionizing the field of waste material management. By continuously gathering and analysing crucial data, it will offer invaluable insights into dump stability, allowing for early detection of potential issues and enabling timely preventive measures. Moreover, this cutting-edge technology will serve as a foundation for improved stability models,

advancing our understanding of dump behavior and enhancing the safety and efficiency of mining operations. In conclusion, the integration of this innovative characterisation system will undoubtedly pave the way for a safer and more sustainable approach to dump stability management, contributing to a greener, more responsible mining industry.

Funding

This work was supported by the Australian Coal Industry's Research Program - ACARP [Grant number C29048].

CRedit authorship contribution statement

Sureka Thiruchittampalam: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bikram Pratap Banerjee:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Nancy F. Glenn:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation. **Simit Raval:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.enggeo.2024.107406>.

References

- Andrade, F.A., Al-Qureshi, H.A., Hotza, D., 2011. Measuring the plasticity of clays: a review. *Appl. Clay Sci.* 51, 1–7. <https://doi.org/10.1016/j.clay.2010.10.028>.
- Anthony, G., Gregg, H., Tshilidzi, M., 2008. An SVM multiclassifier approach to land cover mapping. *ASPRS 2008. Annual Convention 28. April-2, May, Portland, Oregon*.
- Arabameri, A., Pourghasemi, H.R., 2019. 13 - spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R. In: Pourghasemi, H.R., Gokceoglu, C. (Eds.), *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier, pp. 299–321.
- Arboleda, E.R., 2019. Comparing performances of data mining algorithms for classification of green coffee beans. *Int. J. Eng. Adv. Technol.* 8, 1563–1567.
- Aryaguna, P.A., Danoedoro, P., 2016. Comparison effectiveness of pixel based classification and object based classification using high resolution image in floristic composition mapping (study case: Gunung Tidar Magelang city). In: *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, p. 012042.
- Ashour, A.S., Guo, Y., Hawas, A.R., Xu, G., 2018. Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images. *Health Inform. Sci. Syst.* 6, 1–10.
- Asthana, H., Vishwakarma, C.A., Singh, P., Kumar, P., Rena, V., Mukherjee, S., 2020. Comparative analysis of pixel and object based classification approach for rapid landslide delineation with the aid of open source tools in Garhwal Himalaya. *J. Geol. Soc. India* 96, 65–72. <https://doi.org/10.1007/s12594-020-1505-1>.
- Belgiu, M., Drăguț, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS J. Photogramm. Remote Sens.* 96, 67–75.
- Bishop, C.M., 1994. Neural networks and their applications. *Rev. Sci. Instrum.* 65, 1803–1832.
- Bishwal, M., Sen, P., Jawed, M., 2017. Characterization of shear strength properties of spoil dump based on their constituent material. *Int. J. Appl. Eng. Res.* 12, 8590–8594.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16.
- Corenblit, D., Decaux, O., Delmotte, S., Tournazet, J.-P., Arrignon, F., André, M.-F., Darrozes, J., Davies, N.S., Julien, F., Otto, T., 2023. Signatures of life detected in images of rocks using Neural Network analysis demonstrate new potential for searching for biosignatures on the surface of Mars. *Astrobiology* 23, 308–326.
- Dietterich, T.G., 2000. Ensemble methods in machine learning. In: *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1*. Springer, pp. 1–15.
- Fu, B., He, X., Yao, H., Liang, Y., Deng, T., He, H., Fan, D., Lan, G., He, W., 2022. Comparison of RFE-DL and stacking ensemble learning algorithms for classifying mangrove species on UAV multispectral images. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102890 <https://doi.org/10.1016/j.jag.2022.102890>.
- Gopika, N., Meena, K., 2018. Correlation based feature selection algorithm for machine learning. In: *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*. IEEE, pp. 692–695.
- Grnias, I., Panagiotakis, C., Tziritas, G., 2016. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* 122, 145–166.
- Hanchuan, P., Fuhui, L., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. <https://doi.org/10.1109/TPAMI.2005.159>.
- Hassanein, M., Lari, Z., El-Sheimy, N., 2018. A New vegetation segmentation approach for cropped fields based on threshold detection from hue histograms. *Sensors* 18, 1253.
- Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldfog, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A., Fisher, R.B., 1996. An experimental comparison of range image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 673–689.
- Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P., Wu, W., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *Int. J. Appl. Earth Obs. Geoinf.* 80, 218–229. <https://doi.org/10.1016/j.jag.2019.04.014>.
- Jahromi, A.H., Taheri, M., 2017. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In: *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, pp. 209–212.
- Jones, T.R., Carpenter, A., Golland, P., 2005. Voronoi-based segmentation of cells on image manifolds. In: *International Workshop on Computer Vision for Biomedical Image Applications*. Springer, pp. 535–543.
- Kamel, H., Abdulah, D., Al-Tuwaijari, J.M., 2019. Cancer classification using Gaussian Naive Bayes Algorithm. *Int. Eng. Conf. (IEC) 2019*, 165–170.
- Keyport, R.N., Oommen, T., Martha, T.R., Sajinkumar, K.S., Gierke, J.S., 2018. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinf.* 64, 1–11. <https://doi.org/10.1016/j.jag.2017.08.015>.
- Khan, M.U., Samer, S., Alshehri, M.D., Baloch, N.K., Khan, H., Hussain, F., Kim, S.W., Zikria, Y.B., 2022. Artificial neural network-based cardiovascular disease prediction using spectral features. *Comput. Electr. Eng.* 101, 108094 <https://doi.org/10.1016/j.compeleceng.2022.108094>.
- Lee, L.H., Wan, C.H., Rajkumar, R., Isa, D., 2012. An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. *Appl. Intell.* 37, 80–99.
- Li, X., Chen, J., Zhao, L., Guo, S., Sun, L., Zhao, X., 2020. Adaptive distance-weighted voronoi tessellation for remote sensing image segmentation. *Remote Sens.* 12, 4115.
- Lottermoser, B., 2007. *Mine Wastes*. Springer.
- Loussaief, S., Abdelkrim, A., 2018. Deep learning vs. bag of features in machine learning for image classification. In: *2018 International Conference on Advanced Systems and Electric Technologies (ICASET)*. IEEE, pp. 6–10.
- Lyons, M.B., Keith, D.A., Phinn, S.R., Mason, T.J., Elith, J., 2018. A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sens. Environ.* 208, 145–153.
- Ma, L., Cheng, L., Li, M., Liu, Y., Ma, X., 2015. Training set size, scale, and features in geographic object-based image analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS J. Photogramm. Remote Sens.* 102, 14–27.
- Maghari, A., 2018. Prediction of student's performance using modified KNN classifiers. In: *Alfere, S.S., Maghari, A.Y. (Eds.), Prediction of Student's Performance Using Modified KNN Classifiers. In The First International Conference on Engineering and Future Technology (ICEFT 2018)*, pp. 143–150.
- Masoudian, M.S., Zevgolts, I.E., Deliveris, A.V., Marshall, A.M., Heron, C.M., Koukousas, N.C., 2019. Stability and characterisation of spoil heaps in European surface lignite mines: a state-of-the-art review in light of new data. *Environ. Earth Sci.* 78, 1–18.
- McCann, S., Lowe, D.G., 2012. Local naive Bayes nearest neighbor for image classification. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3650–3656.
- Michel, J., Grizonnet, M., Canévet, O., 2012. Supervised re-segmentation for very high-resolution satellite images. In: *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 68–71.
- Nasiri, V., Beloiu, M., Asghar Darvishsefat, A., Griess, V.C., Maftai, C., Waser, L.T., 2023. Mapping tree species composition in a Caspian temperate mixed forest based on spectral-temporal metrics and machine learning. *Int. J. Appl. Earth Obs. Geoinf.* 116, 103154 <https://doi.org/10.1016/j.jag.2022.103154>.
- Nusantika, N.R., Hu, X., Xiao, J., 2021. Improvement canny edge detection for the UAV icing monitoring of transmission line icing. In: *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, pp. 1838–1843.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9, 62–66.

- Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238.
- Rashid, M., Bari, B.S., Hasan, M.J., Razman, M.A.M., Musa, R.M., Ab Nasir, A.F., Majeed, A.P.A., 2021. The classification of motor imagery response: an accuracy enhancement through the ensemble of random subspace k-NN. *PeerJ Comp. Sci.* 7, e374.
- Simmons, J.V., McManus, D.A., 2004. Shear strength framework for design of dumped spoil slopes for open pit coal mines. In: *Advances in geotechnical engineering: The Skempton conference: Proceedings of a three day conference on advances in geotechnical engineering, organised by the Institution of Civil Engineers and held at the Royal Geographical Society, London, UK, on 29–31 March 2004*. Thomas Telford Publishing, pp. 981–991.
- Sinaice, B.B., Owada, N., Ikeda, H., Toriya, H., Bagai, Z., Shemang, E., Adachi, T., Kawamura, Y., 2022. Spectral angle mapping and AI methods applied in automatic identification of placer deposit magnetite using multispectral camera mounted on UAV. *Minerals* 12. <https://doi.org/10.3390/min12020268>.
- Singh, P., Ranga, V., 2021. Attack and intrusion detection in cloud computing using an ensemble learning approach. *Int. J. Inf. Technol.* 13, 565–571.
- Singh, A., Halgamuge, M.N., Lakshminathan, R., 2017. Impact of different data types on classifier performance of random forest, naive Bayes, and k-nearest neighbors algorithms. *Int. J. Adv. Comput. Sci. Appl.* 8.
- Son, Y.-S., Noh, S.-G., Bang, E.-S., Kim, K.-E., Cho, S.-J., Baik, H., 2022. Ground-based visible–near infrared hyperspectral imaging for monitoring cliff weathering of a volcanic island in Dokdo, South Korea. *Eng. Geol.* 309, 106854 <https://doi.org/10.1016/j.enggeo.2022.106854>.
- Song, Q., Hu, W., Xie, W., 2002. Robust support vector machine with bullet hole image classification. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 32, 440–448.
- Tariq, A., Jiango, Y., Li, Q., Gao, J., Lu, L., Soufan, W., Almutairi, K.F., Habib-Ur-Rahman, M., 2023. Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive Bayes tree models with optical remote sensing data. *Heliyon* 9, e13212. <https://doi.org/10.1016/j.heliyon.2023.e13212>.
- Tharwat, A., 2016. Linear vs. quadratic discriminant analysis classifier: a tutorial. *Int. J. Appl. Patt. Recognit.* 3, 145–180.
- Thiruchittampalam, S., Banerjee, B.P., Singh, S.K., Glenn, N.F., Raval, S., 2023a. Evaluation of segmentation methods for spoil pile delineation using UAV images. In: *IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2450–2453.
- Thiruchittampalam, S., Singh, S.K., Banerjee, B.P., Glenn, N.F., Raval, S., 2023b. Spoil characterisation using UAV-based optical remote sensing in coal mine dumps. *Int. J. Coal Sci. Technol.* 10, 65. <https://doi.org/10.1007/s40789-023-00622-4>.
- Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13, 361–378.
- Wang, S., Zhou, K., Wang, J., Zhao, J., 2022. Identifying and mapping alteration minerals using HySpex airborne hyperspectral data and random forest algorithm. *Front. Earth Sci.* 10, 871529.
- Yang, P., Esmaeili, K., Goodfellow, S., Ordóñez Calderón, J.C., 2023. Mine pit Wall Geological Mapping using UAV-Based RGB Imaging and Unsupervised Learning. *Remote Sens.* 15 <https://doi.org/10.3390/rs15061641>.
- Zambon, M., Lawrence, R., Bunn, A., Powell, S., 2006. Effect of alternative splitting rules on image processing using classification tree analysis. *Photogramm. Eng. Remote. Sens.* 72, 25–30.
- Zevgolis, I.E., Theocharis, A.I., Koukoulas, N.C., 2021. Geotechnical engineering perspectives: challenges and solutions in the transition to a post-lignite era. *Mater. Proc.* 5, 17.
- Zhang, Z., Khow, C., Liu, J., Cheung, Y., Aung, T., 2012. Automatic glaucoma diagnosis with mRMR-based feature selection. *J. Biomet. Biostat.* 5 7, 2.
- Zhou, Y., Zhang, R., Wang, S., Wang, F., 2018. Feature selection Method based on High-Resolution Remote Sensing Images and the effect of Sensitive Features on Classification Accuracy. *Sensors* 18, 2013.