



**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2020

DEVELOPING INNOVATIVE SPECTRAL AND MACHINE LEARNING METHODS FOR MINERAL AND LITHOLOGICAL CLASSIFICATION USING MULTI-SENSOR DATASETS

Chandan Kumar
Michigan Technological University, ckumar@mtu.edu

Copyright 2020 Chandan Kumar

Recommended Citation

Kumar, Chandan, "DEVELOPING INNOVATIVE SPECTRAL AND MACHINE LEARNING METHODS FOR MINERAL AND LITHOLOGICAL CLASSIFICATION USING MULTI-SENSOR DATASETS", Open Access Dissertation, Michigan Technological University, 2020.
<https://doi.org/10.37099/mtu.dc.etr/1120>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etr>



Part of the [Geological Engineering Commons](#)

DEVELOPING INNOVATIVE SPECTRAL AND MACHINE LEARNING
METHODS FOR MINERAL AND LITHOLOGICAL CLASSIFICATION USING
MULTI-SENSOR DATASETS

By

Chandan Kumar

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Geological Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2020

© 2020 Chandan Kumar

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Geological Engineering.

Department of Geological and Mining Engineering and Sciences

Dissertation Co-Advisor: *Dr. Snehamoy Chatterjee*

Dissertation Co-Advisor: *Dr. Thomas Oommen*

Committee Member: *Dr. Chad Deering*

Committee Member: *Dr. William H. Farrand*

Department Chair: *Dr. Aleksey Smirnov*

Table of Contents

List of figures	vi
List of tables.....	x
Preface.....	xiii
Acknowledgments.....	xv
List of abbreviations	xvi
Abstract.....	xviii
1 Introduction.....	1
1.1 Background	1
1.2 Statement of the problem	2
1.3 Significance and innovation	4
1.4 Research scope and objectives	5
1.5 Structure of the dissertation.....	5
2 Mapping hydrothermal alteration minerals using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski gold deposit area, India	7
2.1 Introduction	7
2.2 Geological setting of the area	9
2.3 Materials and methods.....	11
2.3.1 Airborne and field hyperspectral datasets.....	11
2.3.2 AVIRIS-NG hyperspectral image processing and classification.....	12
2.4 Results and discussion.....	14
2.4.1 Alteration minerals.....	14
2.4.2 Spectral similarity measures and derivation of optimum threshold..	16
2.4.3 Performance evaluation of SAM, SID, and SIDSAM _{tan} using RSDPW	20
2.4.4 Hydrothermal alteration mineral mapping.....	21
2.4.5 Accuracy assessment	22
2.5 Conclusions	25
3 New effective spectral matching measures for hyperspectral data analysis	27
3.1 Introduction	27
3.2 Hyperspectral datasets and methods.....	29

3.2.1	Hyperspectral datasets	29
3.2.2	Spectral matching measures.....	35
3.2.2.1	Existing spectral matching measures	35
3.2.2.2	Proposed spectral similarity measures	36
3.3	Computation of SMMs and RSDPW statistics.....	38
3.4	Results and discussion.....	39
3.4.1	Spectral similarity analysis	39
3.4.2	Performance evaluation of SMMs using RSDPW statistics	42
3.4.2.1	USGS spectra of minerals	42
3.4.2.2	USGS spectra of vegetation	45
3.4.2.3	USGS spectra of minerals and vegetation with added noise	47
3.4.2.4	USGS spectra of mixed minerals and mixed vegetation	49
3.4.2.5	Hydrothermal alteration minerals	55
3.4.3	Effectiveness of proposed SMMs	57
3.5	Conclusions	58
4	Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India.....	60
4.1	Introduction	60
4.2	Location and geological setting of the study area	62
4.3	Datasets and methods	64
4.3.1	Datasets	64
4.3.2	Methods.....	65
4.3.2.1	Preparation of high-resolution reference lithology map	65
4.3.2.2	Preparation of different input datasets using AVIRIS-NG hyperspectral data	66
4.3.2.3	Training data collection and spectral separability computation	68
4.3.2.4	Implementation of MLAs for automated lithological classification	68
4.4	Results and discussion.....	70
4.4.1	Spectral characteristics of rocks.....	70
4.4.2	High-resolution reference lithology map	72
4.4.3	Feature selection and extraction-based hyperspectral input datasets	73
4.4.4	Spectral separability analysis of hyperspectral input datasets	75
4.4.5	Comparative analysis of MLAs using different input datasets	78
4.4.6	Validation of generated lithological map using field observation	79
4.4.7	Impact of number and quality of training samples on MLAs	84
4.5	Conclusions	88

5	Multi-sensor datasets-based optimal integration of spectral, textural, and morphological characteristics of rocks for lithological classification using machine learning models	89
5.1	Introduction	89
5.2	Location and geological setting.....	92
5.3	Datasets	92
5.4	Methods.....	94
5.4.1	Data pre-processing	95
5.4.2	Preparation of reference lithology map.....	96
5.4.3	Derivation of different input features	97
5.4.3.1	Spectral and transformed spectral input features	97
5.4.3.2	Morphological features	97
5.4.3.3	Textural features	98
5.4.3.4	Training data collection and class separability computation	99
5.4.3.5	Variable importance measure	99
5.4.3.6	Machine learning algorithms	100
5.5	Results and discussion.....	102
5.5.1	Reference lithology map	102
5.5.2	Spectral characteristics of rocks.....	104
5.5.3	Spectral-intra-class variability	106
5.5.4	Variable importance and selection.....	107
5.5.5	Comparison of ML models using different input datasets.....	111
5.6	Conclusions	117
6	Conclusions and future work	118
	References.....	120
	Appendix A.....	135

List of figures

Figure 2.1. Location and lithological map of the studied area showing major rock types, major locations, and location of gold occurrences reported by Geological Survey of India.	10
Figure 2.2. A work flowchart showing methodology adopted in the processing and classification of the hyperspectral data.	12
Figure 2.3. Reflectance spectra of altered minerals such as (a) goethite, (b) chlorite, (c) chlorite at of mine site, (d) kaolinite, and (e) muscovite, showing their diagnostic absorption features. (f) Non-target spectra. NT ₁ : soil, NT ₂ : poorly exposed rock surface, NT ₃ : mixed pixel of vegetation and rocks, and NT ₄ : dry vegetation. Image, laboratory, and USGS mineral library spectra are denoted as blue, black, and red color, respectively.	15
Figure 2.4. Correctly classified pixels against the threshold value ranges used for each class to obtain the optimum threshold value for (a) SAM, (b) SID, and (c) SIDSAM _{tan} classification method.	19
Figure 2.5. Graphical representation of RSDPW of target minerals showing discrimination power of SAM, SID, and SIDSAM _{tan} . Goe: goethite, Chl: chlorite, Chl _{mine} : chlorite at the mine site, Kao: kaolinite, and Mus: muscovite.	21
Figure 2.6. (a) False Color Composite (FCC) of the AVIRIS-NG image with few major locations. The mineral map obtained using (b) SAM, (c) SID, (d) SIDSAM _{tan} . The zoomed view of SIDSAM _{tan} result at few locations showing the distribution of mapped minerals: (e) muscovite in the northern part of the area, (f) chlorite at mine site, (g) goethite and muscovite, and (h) kaolinite and chlorite in the southern part of the area.	23
Figure 3.1. The spectral curve of minerals obtained from the USGS mineral spectral library.	30
Figure 3.2. The spectral curve of vegetation obtained from the USGS vegetation spectral library.	30
Figure 3.3. The spectral curve of minerals of the USGS mineral spectral library was obtained by adding the 10% of random noise in the spectra.	31
Figure 3.4. The spectral curve of vegetation of USGS vegetation spectral library obtained by adding the 10% of random noise in the spectra.	32
Figure 3.5. Mixed spectra of minerals obtained by linear mixing of actual mineral class and other minerals with equal proportion (i.e., 50% of actual mineral class + 8.33% of other six minerals (i.e., 50%)) using USGS mineral spectral library.	32
Figure 3.6. Mixed spectra of vegetation obtained by linear mixing of actual vegetation class and other vegetation with equal proportion (i.e., 50% of actual vegetation class + 8.33% of other six vegetation (i.e., 50%)) using USGS vegetation spectral library.	33
Figure 3.7. Geographical location and geological setting of the area showing major lithological units and gold occurrences reported by the Geological Survey of India.	34
Figure 3.8. The spectra of hydrothermal alteration minerals obtained using AVIRIS-NG airborne hyperspectral data.	35

Figure 3.9. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the USGS spectra of minerals.	44
Figure 3.10. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the spectra of vegetation derived from the USGS spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rs: rabbitbrush, and Sb: saltbush.	46
Figure 3.11. Graphical representation of RSDPW statistics of different spectral similarity measures derived using 10% of random noise added in the spectra of minerals derived from the USGS spectral library.....	48
Figure 3.12. Graphical representation of RSDPW statistics of different spectral similarity measures derived using 10% of random noise added in the spectra of vegetation derived from the USGS spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rs: rabbitbrush, and Sb: saltbush.....	50
Figure 3.13. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the mixed spectra of minerals prepared by linear mixing of other classes with equal proportion (50% of actual mineral class + 8.33% of six other mineral classes (i.e., 50%)) using USGS mineral spectral library.....	51
Figure 3.14. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the mixed spectra of vegetation prepared by linear mixing of other classes with equal proportion (50% of actual vegetation class + 8.33% of six other vegetation classes (i.e., 50%)) using USGS vegetation spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rs: rabbitbrush, and Sb: saltbush.	53
Figure 3.15. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the spectra of hydrothermal alteration minerals of AVIRIS-NG hyperspectral data. Go: goethite, Ch: chlorite, Cm: chlorite at the mine site (a mixture of chlorite and mined materials), Ka: kaolinite, and Mu: muscovite.	56
Figure 4.1. Geographical location and geology map of the area showing major lithological units with a few locations and gold occurrences.	63
Figure 4.2. The methodology's workflow chart adopted for automated lithological mapping using the integrated spectral enhancement and ML methods.	65
Figure 4.3. Spectral curve of different rock types (a) acidic intrusive, (b) amphibolite, (c) granite, (d) metabasalt, (e) migmatite, (f) mine site, and (g) black soil derived from AVIRIS-NG reflectance image. [Arrow indicates the absorption band position]. ...	71
Figure 4.4. The spectral characteristics of different lithological units, including mine site and black soil. (a) the spectral curve of rock types derived from average training pixels of AVIRIS-NG hyperspectral data used in lithological classification, and (b) a box plot of overall average reflectance of different lithological units. [Ac: acidic intrusive, Am: amphibolite, Gr: granite, Me: metabasalt, Mg: migmatite, Ms: Mine site, and Bs: black soil].	72
Figure 4.5. The image enhanced products such as (a) FCC (R: 850 nm, G: 650 nm, B: 550 nm) of AVIRIS NG, (b) band combination of PCA (R: PC3, G: PC2, B: PC1) and (c) band combination of ICA (R: IC2, G: IC1, B: IC3) derived from calibrated ASTER data discriminating the litho-units of the area, (d) training data samples overlaid on the FCC of AVIRIS-NG data, (e) conventional geology map prepared from district	

resource map mapped at the scale of 1:2,50,000 by the geological survey of India, and (f) high-resolution reference lithology map derived from the band combination of PCA and ICA transformation using ASTER data.....	74
Figure 4.6. The accuracy statistics (a) overall accuracy and (b) kappa coefficient of JMIM bands using different MLAs.	76
Figure 4.7. Accuracy statistics (a) overall accuracy and (b) kappa coefficient of MLAs using different input datasets such as all spectral bands, SWIR spectral bands, first 20 PC transformed bands of higher variance, and JMIM based optimum bands.	79
Figure 4.8. (a) Reference lithology map derived from spectral enhancement products using ASTER data; and (b) lithological classification map obtained from SVM using JMIM based optimum bands of AVIRIS-NG hyperspectral data.....	80
Figure 4.9. Field photographs of different rock types were acquired during fieldwork in the study area at the validation points shown in Figure 4 (a). (a) exposure of granite in North-East of the Tupdoor area, (b) granite and amphibolite contact was taken from Wandli area, and (c) highly weathered metabasalt covered by black soil occurring in pedi-plain, which is flanked by amphibolite in North-West of Hosur area, (d) metabasalt dumped in Utti mine, (e) metabasalt, and (f) altered rock in Yetgal area.	85
Figure 4.10. Reduction in sample datasets by 15%, 30%, and 45% and their corresponding accuracy statistics of training and testing samples (a) OA (%) and (b) k of different MLAs.	86
Figure 4.11. Mislabeling in sample datasets by 15%, 30%, and 45% and their corresponding accuracy statistics of training and testing samples. (a) Overall accuracy (%) and (b) kappa coefficient of different MLAs.	87
Figure 5.1. (a) Location of the study area, (b) a false-color composite map prepared from ASTER data showing overall landcovers, (c) DEM showing the overall topography, and (d) geology map exhibits major rock types found in the study area.	93
Figure 5.2. The workflow chart of the overall methodology adopted for improved lithological mapping.	95
Figure 5.3. The enhanced spectral products are discriminating different rock types found in the study area. (a) band combination of PCA transformation (R: PC4, G: PC2, B: PC1), (b) band combination of ICA (R: IC4, G: IC3, B: IC1) derived from processed ASTER data, (c) reference lithology map, and (d) the geo-location of training samples overlaid on the false-color composite of ASTER data.	102
Figure 5.4. The zoomed view of PCA and ICA transformation enhanced spectral products discriminating different rock types found in the study area. 1: limestone, 2: sandstone, 3: Fe-bearing sandstone, 4: shale, 5: chert breccia, 6: metabasalt, 7: granite-gneiss, 8: ultramafic, 9: dike, and 10: water body.	103
Figure 5.5. Field photographs of a few rock types at the selected traverse captured during the field visit in the study area. (a) the sandstone of the Kaimpur group in the Vikrampur area, (b) granite of Bundelkhand granitoid complex captured in the Amroniya area, (c) shale of Semri group.....	104
Figure 5.6. The overall average spectra of different rock types derived from processed ASTER data. [Dashed line indicates the diagnostic absorption position]. Lm:	

limestone, Sd: sandstone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Cb: chert breccia, Mb: metabasalt, Gg: granite-gneiss, um: ultramafic rock, Dk: dike.....	105
Figure 5.7. A boxplot of overall average spectra of different rock types derived from the training dataset of processed ASTER data.....	106
Figure 5.8. Variable importance of spectral input features derived from processed ASTER data. B: spectral bands of VNIR and SWIR.	107
Figure 5.9. Variable importance of transformed spectral input features derived from processed ASTER spectral data using PCA and ICA transformation.	108
Figure 5.10. Variable importance of different morphological input features derived from PALSAR (P) and Sentinel-1 (S) data.....	109
Figure 5.11. SAR indices (a) P_8 and (b) S_1 derived from PALSAR and Sentinel-1 data respectively displaying the backscattered energy of different rock types found in the area.....	109
Figure 5.12. Variable importance of textural input features derived from DEM with an optimum window.	110
Figure 5.13. Textural variables such as (a) elevation, (b) correlation, and (c) entropy derived from the optimum window size using DEM illustrating their ability to discriminate different rock types found in the area.	111
Figure 5.14. The accuracy statistics of lithological classification obtained from different input features using SVM and RF models. A: spectral input features, B: spectral and transformed input features, C: spectral and morphological input features, D: spectral and textural input features, and E: optimal hybrid input features.....	112
Figure 5.15. The lithological classification was obtained from the optimal hybrid input data using RF and SVM. (a) Reference lithology map, (b) lithological classification map using RF, and (c) lithological classification map using SVM.	116

List of tables

Table 2.1. Key characteristics of AVIRIS-NG and AVIRIS-Classic sensor (Thorpe et al., 2016). VNIR: Visible Near-Infrared; SWIR: Shortwave Infrared.....	11
Table 2.2. Spectral parameters such position (in nm) and depth of prominent diagnostic absorption feature of detected minerals (depth is in arbitrary map units).	16
Table 2.3. Spectral similarity matrix of target and non-targets using SAM (in radians), SID (SD), and SIDSAM _{tan} (SS) (in radians). Go: goethite, Cl: chlorite, C _m : chlorite at the mine site, Ka: kaolinite, Mu: muscovite, NT _(1 to 4) : non-target spectra. Values marked in bold indicate the minimum spectral angle.	18
Table 2.4. The optimum threshold value was obtained using the statistical and proposed threshold selection method to get the mineral classification map using SAM (in radians), SID, and SIDSAM _{tan} (in radians). SAM _{SM1} and SAM _{SM2} indicate the possible optimum threshold derived using Schwarz and Staenz (2001) (SM ₁) and Hecker et al. (2008) (SM ₂) statistical methods.	19
Table 2.5. RSDPW of SAM, SID, and SIDSAM _{tan} similarity measures for target and non-target classes using linearly mixed spectrum as a reference. Goe: goethite, Chl: chlorite, Chl _{mine} : chlorite at the mine site, Kao: kaolinite, and Mus: muscovite. NT ₍₁₋₄₎ : non-target spectra.	20
Table 2.6. Spatial distribution of hydrothermal alteration minerals derived from the SAM, SID, and SIDSAM _{tan} using a spectral similarity-based optimum threshold value. ..	22
Table 2.7. Confusion matrix of the classified image using SAM, SID, and SIDSAM _{tan} . Goe: goethite, Chl: chlorite, Chl _{mine} : chlorite at the mine site, Kao: kaolinite, and Mus: muscovite, OC: other classes/non-targets. UA: user accuracy (%), PA: producer accuracy (%), OA: overall accuracy (%), k: kappa coefficient.	24
Table 3.1. Spectral similarity matrix of USGS spectra of minerals using Spectral Angle Mapper (SAM) (in radians). Mixture: linearly mixed spectrum of mineral with equal proportion.....	40
Table 3.2. Spectral similarity matrix of USGS spectra of vegetation using SAM (in radians). Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rb: rabbit brush, Sb: saltbrush, Ms: linearly mixed spectrum of vegetation with equal proportion.	41
Table 3.3. Spectral similarity matrix of USGS mineral spectra after adding 10% of random noise using SAM (in radians).	41
Table 3.4. Spectral similarity matrix of USGS vegetation spectra after adding 10% of random noise using SAM.....	41
Table 3.5. Spectral similarity matrix of mixed spectra of minerals (USGS spectral library) prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e., 50%)) using SAM (units are in radians).....	41
Table 3.6. Spectral similarity matrix of mixed spectra of vegetation (USGS spectral library) prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of different classes (i.e., 50%)) using SAM (in radians).....	42
Table 3.7. Spectral similarity matrix of hydrothermal alteration minerals using SAM (in radians).....	42

Table 3.8. RSDPW statistics of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for USGS spectra of minerals.	44
Table 3.9. RSDPW of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for USGS spectra of vegetation. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rb: rabbitbrush, Sb: saltbush.	46
Table 3.10. RSDPW statistics of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for USGS minerals spectra after adding 10% of random noise.	48
Table 3.11. RSDPW statistics of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for USGS vegetation spectra after adding 10% of random noise.	50
Table 3.12. RSDPW of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for mixed spectra of minerals prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e. 50%)) using linear spectral mixing.	52
Table 3.13. RSDPW statistics of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for mixed spectra of vegetation (prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e., 50%)) using the linearly mixed spectrum of vegetation as a reference.	54
Table 3.14. RSDPW statistics of SAM, SID, SIDSAM _{tan} , DSSC, KJSSC, and KJDSSC _{tan} for hydrothermal alteration minerals using the linearly mixed spectrum of all minerals with equal proportion as a reference.	56
Table 4.1. Specifications of AVIRIS- NG sensor (Thorpe et al., 2016). VNIR: Visible Near Infrared, and SWIR: Shortwave Infrared.	64
Table 4.2. Description of the lithological units, area, number of training and testing samples (Pixels of AVIRIS-NG data) used in the lithological classification using MLAs.	75
Table 4.3. Spectral dissimilarity matrix computed using J-M distance of input datasets such as all spectral bands, SWIR bands, JMIM, and PC based optimum bands. Ai: acidic intrusive, Am: Amphibolite, Gr: Granite, Me: Metabasalt, Mi: Migmatite, Ms: Mine site, Bs: Black soil.	77
Table 4.4. Confusion matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using SWIR spectral bands. LDA: linear discriminant analysis, RF: random forest, SVM: support vector machine. Ai: acidic intrusive, Am: Amphibolite, Gr: Granite, Me: Metabasalt, Mi: Migmatite, Ms: Mine site, BS: Black soil. Row: row total, Col: column total, UA: user accuracy, PA: producer accuracy.	80
Table 4.5. Confusion matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using all spectral bands of VNIR and SWIR.	81
Table 4.6. Confusion Matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using first 20 PC bands.	82
Table 4.7. Confusion Matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using JMIM based optimum bands.	83
Table 4.8. Reduction in training datasets by 15%, 30%, and 45%, and their impact on the performance of different MLA. OA: overall accuracy and k: kappa coefficient.	86
Table 4.9. Mislabeling in training datasets by 15%, 30%, and 45% and their impact on the performance of different MLA. OA: overall accuracy (%) and k: kappa coefficient.	87

Table 5.1. A brief sensor specification of different datasets used in this study.....	93
Table 5.2. Description of indices prepared using PALSAR (P: P ₁₋₁₂) and Sentinel-1 (S: S ₁₋₁₂) data to derive the surface's morphological features. HH: horizontal transmitting and horizontal receiving, HV: horizontal transmitting and vertical receiving, VV: vertical transmitting and vertical receiving, and VH: vertical transmitting and horizontal receiving.	98
Table 5.3. A summary of lithological units, areal extent, number of training, and testing datasets used in lithological classification using ML models.	99
Table 5.4. Spectral dissimilarity matrix of overall average reflectance of training datasets derived from processed ASTER data using the J-M distance measure. Lm: limestone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Mb: metabasalt, Gg: granite-gneiss, Um: ultramafic rock, and Dk: dike.	106
Table 5.5. Textural features derived from optimum window size using DEM.	110
Table 5.6. Summary of different input features extracted from multi-sensor datasets and the optimum number of variables used in lithological classification.	111
Table 5.7. The accuracy statistics of lithological classification are derived from spectral input data and integrated spectral and transformed spectral input datasets using SVM and RF models. Lm: limestone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Mb: metabasalt, Gg: granite-gneiss, Um: ultramafic rock, and Dk: dike. OA: overall accuracy (%), k: kappa coefficient, UA: user accuracy (%), PA: producer accuracy (%).	114
Table 5.8. The accuracy statistics of lithological classification derived from integrated spectral and morphological input dataset, and integrated spectral and textural input dataset using SVM and RF models.	114
Table 5.9. The accuracy statistics of lithological classification are derived from the optimal hybrid input dataset using SVM and RF models.	115

Preface

Chapters 2 and 4 are composed of published research articles that are accessible through peer-reviewed journals. Chapter 3 is composed of an accepted research article. Chapter 5 is composed of a submitted article for publication in a peer-reviewed journal.

Chapter 2: Kumar, C., Chatterjee, S., & Oommen, T. (2020). Mapping hydrothermal alteration minerals using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski gold deposit area, India. *International Journal of Remote Sensing*, 41(2), 794-812. doi: <https://doi.org/10.1080/01431161.2019.1648906>

Authors' contributions: Kumar developed the research idea, processed and analyzed the data, derived the results, and wrote the manuscript. Chatterjee and Oommen provided guidance and revisions throughout the manuscript.

Permission: Reproduction of this article falls under the non-commercial use, and written permission is not required. <https://authorservices.taylorandfrancis.com/copyright-and-you/>.

Chapter 3: Kumar, C., Chatterjee, S., Oommen, T., & Guha, A. (2020). Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *International Journal of Applied Earth Observation and Geoinformation*, 86, 102006. doi: <https://doi.org/10.1016/j.jag.2019.102006>

Authors' contributions: Kumar developed the research idea, processed and analyzed the data, derived the results, and prepared the manuscript. Chatterjee and Oommen provided guidance and revisions throughout the manuscript. Guha provided the field data and suggestions to improve the manuscript.

Permission: Reproduction of this article falls under the non-commercial use, and written permission is not required. <https://www.elsevier.com/about/policies/copyright#Author-rights>.

Chapter 4: Kumar, C., Chatterjee, S., Oommen, T., & Guha, A. (2020). New effective spectral matching measures for hyperspectral data analysis. Accepted in *International Journal of Remote Sensing*, Taylor and Francis.

Authors' contribution: Kumar developed the research idea, processed and analyzed the data, derived the results, and drafted the manuscript. Chatterjee and Oommen guided throughout the manuscript. Guha provided suggestions to improve the manuscript.

Permission: Reproduction of this article falls under the non-commercial use, and written permission is not required. <https://authorservices.taylorandfrancis.com/copyright-and-you/>.

Chapter 5: Kumar, C., Chatterjee, S., Oommen, T., Guha, A., & Mukherjee, A. (2020). Multi-sensor datasets-based optimal integration of spectral, textural and morphological characteristics of rocks for lithological classification using machine learning models. Submitted in *Geocarto International*, Taylor & Francis.

Authors' contributions: Kumar developed the research idea, processed and analyzed the data, derived the results, and wrote the manuscript. Chatterjee and Oommen provided guidance and revision throughout the manuscript. Guha provided technical expertise in preparing the lithology map and suggestions to improve the manuscript. Mukherjee provided the field data and suggestions to improve the manuscript.

Permission: Reproduction of this article falls under the non-commercial use and written permission is not required.
<https://authorservices.taylorandfrancis.com/copyright-and-you/>.

Acknowledgments

I would love to take this opportunity to present my honest gratitude to everyone in my family for their invaluable support and motivation to make this possible. They continued backing me at every move in my life, which cannot be express in words. The conversation with my mother and father has always boosted my energy and confidence to tackle every hurdle. Their love, care, and support never let me feel alone and far from home. I am fortunate to have such parents. I want to thank my wife (Nidhi) for her love and dedicated support to continue my work. I would also express my gratitude to Mammy and Papa Jee for their blessing. I would also like to thank my sisters (Puja and Ranju) and brother-in-law (Vishal Jee and Rahul Jee) for their encouragement and support. I would also like to thank Rajan bhaiya for his advice in the early stage of my academic career.

I want to express my sincere gratitude to my Ph.D. supervisors (Dr. Chatterjee and Dr. Oommen) for their continuous support and motivation to complete my Ph.D. successfully. It would not have been possible without their friendly nature, dedicated time, inspiration, and support. I want to thank my committee members (Dr. Farrand and Dr. Deering) for their encouragement and guidance. I would also like to thank Dr. Guha for providing field data and valuable suggestions. I would love to thank Dr. Shetty and Dr. Champati ray for their support and motivation in applying for Ph.D. fellowships during my master's program. I also express my gratitude to Dr. Raval and Dr. Sakthivel for their help and encouragement to pursue higher studies abroad. I would also like to thank Dr. Patnayak and Dr. Mipun, for their support. I would love to thank my friend Abhishek for his encouragement and help in applying for the Ph.D. fellowship at Michigan Tech.

I would take the opportunity of expressing my sincere gratitude to GMES for providing a graduate research assistantship during my Ph.D. through NASA grant available with my advisors. I also thank Grad school for awarding me with the project incentive award (2019) and the Fall 2020 finishing fellowship. I would also like to thank the Graduate Student Government for awarding me a few travel grants (2018 & 19). I express my sincere thanks to the current and previous chairs of GMES (Dr. Smirnov and Dr. Geirke). I would also like to thank Brittany and Carol for their help throughout my Ph.D.

I am happy to thank my friends Komal, Rajeev, Amol, Abhishek, Upendra, Sushree, Jordi, Udit, Apurva. Abhijeet, Parth, and others for their support during my Ph.D. I am thankful to Komal for her encouragement and proofreading of my research papers. I also want to thank my cricket and badminton group for offering me an opportunity to get involved in sports and recreational activities to keep me fit and energetic.

I am fortunate to have the grace of God because none of these would have been possible without his mercy. Finally, I am thankful to everyone who provided valuable support, encouragement, and motivation to succeed in my research.

List of abbreviations

ALI	- Advanced Land Imager
ALOS	- Advanced Land Observing Satellite
ASD	- Analytical Spectral Devices
ASTER	- Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVIRIS	- Airborne Visible/Infrared Imaging Spectrometer
AVRIS-NG	- Airborne Visible/Infrared Imaging Spectrometer- Next Generation
CARET	- Classification and Regression Training
CEM	- Constrained Energy Minimization
CM	- Confusion Matrix
DEM	- Digital Elevation Model
DSC	- Dice Similarity Coefficient
DSSC	- Dice Spectral Similarity Coefficient
ED	- Euclidean Distance
ENVI	- Environment for Visualizing Images
EO	- Earth Observation
ERS	- European Remote Sensing
ESA	- European Space Agency
FCC	- False Color Composite
FLAASH	- Fast Line-of-sight Atmospheric Analysis of Hypercubes
FPPI	- Fast Pixel Purity Index
GIS	- Geographic Information System
GRD	- Ground Range Detected
GSI	- Geological Survey of India
HH	- Horizontal Transmitting and Horizontal Receiving
HV	- Horizontal Transmitting and Vertical Receiving
IC	- Independent Component
ICA	- Independent Component Analysis
ISRO	- Indian Space Research Organization
JM	- Jeffries-Matusita
JMIM	- Joint Mutual Information Maximization
JPL	- Jet Propulsion Laboratory
KJSC	- Kumar-Johnson Similarity Coefficient
KJSCC	- Kumar-Johnson Spectral Similarity Coefficient
LDA	- Linear Discriminant Analysis
ML	- Machine Learning
MLA	- Machine Learning Algorithm
MNF	- Minimum Noise Fraction
NASA	- National Aeronautics Space Administration

NIR	- Near Infrared
NJMIM	- Normalized Joint Mutual Information Maximization
OA	- Overall Accuracy
OOB	- Out of Bag
OSP	- Orthogonal Subspace Organization
PA	- Producer Accuracy
PALSAR	- Phased Array type L-band Synthetic Aperture Radar
PC	- Principal Component
PCA	- Principal Component Analysis
PU	- Producer Accuracy
RF	- Random Forest
RSDPW	- Relative Spectral Discrimination Power
SAM	- Spectral Angle Mapper
SAR	- Synthetic Aperture Radar
SCM	- Spectral Correlation Mapper
SFF	- Spectral Feature Fitting
SID	- Spectral Information Divergence
SMM	- Spectral Mapping/Matching Method/Measure
SNAP	- Sentinel Application Platform
SNR	- Signal to Noise Ratio
SVM	- Support Vector Machine
SWIR	- Shortwave Infrared
TCIMF	- Target Constrained Interference Matched Filter
TIR	- Thermal Infrared
UA	- User Accuracy
USGS	- United States Geological Survey
UTM	- Universal Transverse Mercator
VH	- Vertical Transmitting and Horizontal Receiving
VNIR	- Visible Near Infrared
VV	- Vertical Transmitting and Vertical Receiving

Abstract

The sustainable exploration of mineral resources plays a significant role in the economic development of any nation. The lithological maps and surface mineral distribution can be vital baseline data to narrow down the geochemical and geophysical analysis potential areas. This study developed innovative spectral and Machine Learning (ML) methods for mineral and lithological classification. Multi-sensor datasets such as Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Advanced Land Observing (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR), Sentinel-1, and Digital Elevation Model (DEM) were utilized. The study mapped the hydrothermal alteration minerals derived from Spectral Mapping Methods (SMMs), including Spectral Angle Mapper (SAM), Spectral Information Divergence (SID), and SIDSAM_{tan} using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski area (India). The SIDSAM_{tan} outperforms SID and SAM in mineral mapping. A spectral similarity matrix of target and non-target classes based optimum threshold selection was developed to implement the SMMs successfully. Three new effective SMMs such as Dice Spectral Similarity Coefficient (DSSC), Kumar-Johnson Spectral Similarity Coefficient (KJSSC), and their hybrid, i.e., KJDSSC_{tan} has been proposed, which outperforms the existing SMMs (i.e., SAM, SID, and SIDSAM_{tan}) in spectral discrimination of spectrally similar minerals. The developed optimum threshold selection and proposed SMMs are recommended for accurate mineral mapping using hyperspectral data. An integrated spectral enhancement and ML methods have been developed to perform automated lithological classification using AVIRIS-NG hyperspectral data. The Support Vector Machine (SVM) outperforms the Random Forest (RF) and Linear Discriminant Analysis (LDA) in lithological classification. The performance of SVM also shows the least sensitivity to the number and uncertainty of training datasets. This study proposed a multi-sensor datasets-based optimal integration of spectral, morphological, and textural characteristics of rocks for accurate lithological classification using ML models. Different input features, such as (a) spectral, (b) spectral and transformed spectral, (c) spectral and morphological, (d) spectral and textural, and (e) optimum hybrid, were evaluated for lithological classification. The developed approach has been assessed in the Chattarpur area (India) consists of similar spectral characteristics and poorly exposed rocks, weathered, and partially vegetated terrain. The optimal hybrid input features outperform other input features to accurately classify different rock types using the SVM and RF models, which is ~15% higher than as obtained using spectral input features alone. The developed integrated approach of spectral enhancement and ML algorithms, and a multi-sensor datasets-based optimal integration of spectral, morphological, and textural characteristics of rocks, are recommended for accurate lithological classification. The developed methods can be effectively utilized in other remote sensing applications, such as vegetation/forest mapping and soil classification.

1 Introduction

1.1 Background

Remote sensing has coined as the most promising tool for mineral prospecting and geological exploration with their wide range of sensors functional within different parts of the electromagnetic spectrum (Bishop et al., 2011; Pour & Hashim, 2011; Farifteh et al., 2013; Farooq & Govil, 2014; Carrino et al., 2018). Geological targets like minerals and rocks have wavelength-specific diagnostic spectral signatures produced due to electronic and vibrational processes (Clark et al., 1993; Clark, 1999; Clark et al., 2003). The spectral signatures of rocks and minerals have been extensively used for their identification, discrimination, and classification (Hunt, 1977; Kruse, 1988; Kruse et al., 1993; Kruse, 1996; Crosta et al., 1998; Kruse, 1998; Kruse et al., 2003; Van der Meer, 2004; Kruse et al., 2006; Van der Meer, 2006; Bishop et al., 2011; Van der Meer & De Jong, 2011; Kruse, 2012; Van der Meer et al., 2012). The accurate identification, discrimination, and classification of hydrothermal alteration minerals along with accurate lithological maps are essential in successful ore prospecting and mineral exploration (Crosta et al., 1998; Sabins, 1999; Rowan & Mars, 2003; Rowan et al., 2005; Van der Meer et al., 2012; Carranza & Laborte, 2015; Carrino et al., 2018; Bhattacharya et al., 2019b; Guha et al., 2020). A wide range of remote sensors operates in various parts of the electromagnetic spectrum, including optical and microwave regions that can be effectively used to discriminate different rock types accurately to produce accurate lithological maps (Lillesand et al., 2015). Optical remote sensing, including both multispectral and hyperspectral sensors, have been widely exploited in minerals and rock types mapping (Kruse, 1998; Kruse et al., 2003; Ninomiya, 2004; Gomez et al., 2005; Ninomiya et al., 2005; Kruse et al., 2006; Bedini, 2009; Goetz, 2009; Bedini, 2011; Kruse, 2012; Zhang et al., 2013; Kumar et al., 2014; Zhang & Li, 2014; Kumar et al., 2015; Kumar et al., 2020a; Kumar et al., 2020b). The hyperspectral sensors measure the reflected or emitted energy of the surface in many narrow and contiguous spectral bands. In contrast, multispectral sensors are usually limited to a fewer number of spectral bands. Hyperspectral sensors efficiently capture the subtle variation in the geological targets caused due to the crystal structure, grain size, moisture content, and chemical composition, which is limited to multispectral sensors (Clark, 1999; Van der Meer, 2004; Van der Meer & De Jong, 2011; Van der Meer et al., 2012). Although hyperspectral sensors are fascinating in capturing geological targets' spectral signatures, their availability is usually limited. Hyperspectral datasets generated from the laboratory, field and airborne/spaceborne platforms have been extensively used in the identification of various groups of minerals such as clays, carbonates, and silicates (Goetz et al., 1985; Kruse, 1988; Clark et al., 1993; Clark, 1999; Bierwirth et al., 2002; Clark et al., 2003; Goetz, 2009; Kruse, 2012; Zhang et al., 2013; Zadeh et al., 2014). Furthermore, the airborne hyperspectral sensors provide a better scope to capture a more significant number of spectrally pure targets than spaceborne sensors due to their higher spectral and spatial resolution and higher signal-to-noise ratio (Kruse et al., 2003; Van der Meer, 2004). Unlike the optical remote sensors, the microwave remote sensors operate at different frequencies, and polarization can penetrate through clouds and vegetation to derive rocks' surface

characteristics (Schmullius & Evans, 1997). The imaging microwave remote sensor or Synthetic Aperture Radar (SAR) measures the surface backscattered energy, which is mainly controlled by surface roughness and dielectric constant (Hajnsek et al., 2003). The utilization of multi-sensor datasets of different wavelength regions provides an exciting opportunity to derive rocks' surface characteristics for accurate lithological discrimination and classification.

1.2 Statement of the problem

The sustainable exploration of mineral resources plays a significant role in the economic development of any nation. The accurate spatial distribution of minerals and high-resolution lithological maps are essential in successful mineral and geological exploration. The high spectral and spatial resolution and high signal-to-noise ratio of airborne hyperspectral sensors provide an ideal scope to analyze the minute variation in geological targets' characteristics. Although, the application of remotely sensed datasets in the minerals and lithological mapping is not new. The optimization of existing and development of new methods can facilitate their successful exploitation. The classical approach of mineral classification is Spectral Mapping/Matching Methods (SMMs) such as Spectral Angle Mapper (SAM) (Kruse et al., 1993) and Spectral Information Divergence (SID) (Chang, 1999, 2000). The main limitation of most SMMs is that they require a threshold value to obtain the final classification result, which is challenging and usually a trial and error approach. This approach becomes more uncertain when there is a lack of comprehensive field/lab data, and visual interpretation of target classes is difficult (Shanmugam & SrinivasaPerumal, 2014; Kumar et al., 2020a). The implementation of hybrid SMM (i.e., SIDSAM_{tan}) and their comparison with other SMMs have not been evaluated extensively for mineral mapping using high-resolution hyperspectral data. Furthermore, most SMMs do not perform satisfactory discrimination of spectrally similar minerals due to their less effective in dealing with subtle variations and outliers present in hyperspectral datasets.

Lithological classification is challenging, particularly in a complex geological setting; rocks exhibit similar mineral composition and vegetated and poor rock exposure (Jones et al., 2004). The homogeneity in major mineral composition of different rock types and weathered surface cover make their spectral signatures less distinctive and illustrative to be used as primary information for accurate discrimination and lithological classification. A few recent studies demonstrated that integrating information derived from multi-sensor datasets could improve the intra-rock separability and classification accuracy (Othman & Gloaguen, 2014; Wei et al., 2016; Masoumi et al., 2017; Othman & Gloaguen, 2017). The utilization of multi-sensor datasets has been appreciated in several remote sensing applications, including land-use/landcover classification (Jin et al., 2008; Dong et al., 2012), vegetation studies (Franklin et al., 2000; Seifi Majdar & Ghassemian, 2017), urban mapping (Dian et al., 2015), and lithological classification (Othman & Gloaguen, 2014; Masoumi et al., 2017; Othman & Gloaguen, 2017).

The distribution of surface minerals coupled with accurate litho-contacts can be a vital baseline data to narrow down the area of geochemical and geophysical anomalies. The spectral enhancement techniques such as Principal Component Analysis (PCA) (Amer et al., 2010), Independent Component Analysis (ICA) (Kumar et al., 2015; Pour et al., 2019), and band ratios (Gad & Kusky, 2007) have been widely used to enhance the spectral contrast of different rock types. The common limitation of these techniques is that they do not produce a labeled classification output and require manual demarcation. The PCA and ICA outputs usually need comprehensive field data in interpretation and become fuzzier in the weathered and vegetated surface (Kumar et al., 2020b).

In recent decades, Machine Learning Algorithms (MLAs) have shown great success over the conventional classification methods for successful classification of geological targets (Thompson et al., 2001; Waske et al., 2009; Baykan & Yilmaz, 2010; Waske et al., 2010; Abedi et al., 2012; Wu et al., 2013; Cracknell & Reading, 2014; Othman & Gloaguen, 2014; Carranza & Laborte, 2015; Lary et al., 2015; Rodriguez-Galiano et al., 2015; Othman & Gloaguen, 2017). MLAs utilize artificial intelligence derived from input datasets by analyzing their structure, characteristics, and pattern with minimal human involvement (Michie et al., 1994; Zuo, 2017; Mohri et al., 2018). MLAs' successful utilization depends on several factors, such as the quality and quantity of training dataset, selection of optimum variables, and training parameters (Michie et al., 1994; Cortes et al., 1995; Batista et al., 2004; Zuo, 2017; Mohri et al., 2018). The feature/variable selection and extraction become more critical when one deals with hyperspectral data or data consist of a higher number of variables (Hall & Smith, 1999; Yu & Liu, 2003; Archibald & Fann, 2007; Waske et al., 2010; Bennasar et al., 2015). The feature selection method reduces the dimensionality of data by selecting uncorrelated variables to increase the class separability; whereas, the feature extraction method transforms the existing features into a new low-dimensional feature space (Guyon et al., 2004; Hoque et al., 2014; Ren et al., 2014). The feature selection method's advantage over feature extraction is that the feature subset does not lose its physical importance and interpretation. However, there is a wide range of feature selection methods available in the literature, but the suitability and performance of different feature selection methods in lithological classification are rarely explored (Wu et al., 2013). The selection of the training parameter of Machine Learning (ML) models is another crucial aspect of optimum classification results (Bergstra et al., 2011). The trial and error approach-based selection of training parameters are prone to over-or under-fitting the ML models and can produce uncertain classification results (Bergstra et al., 2011). Although the utilization of MLAs in the lithological classification is not new, to the best of my knowledge, there is no single study considered to address all three factors of ML such as quality and quantity of training datasets and optimization parameters together to develop ML models for accurate lithological mapping using hyperspectral data. This study also integrates the spectral enhancement techniques and ML model for lithological classification. Furthermore, the novelty of SAR is well recognized in the geological structural mapping (Pour & Hashim, 2014b, 2015; Guha et al., 2020) but remains less explored in rock types classification (Tan et al., 2009). Rare studies have illustrated the utility of SAR datasets for lithological classification (Tan et al., 2009; Wang et al., 2018). To the best of my knowledge, the potential of multi-sensor datasets of different wavelength

regions such as optical and SAR has not been explored in the lithological mapping using ML models.

1.3 Significance and innovation

The selection of the optimum threshold needed in most SMMs is crucial in successful hyperspectral data analysis. The biased and randomly selected threshold value can significantly influence hyperspectral data's ability to map the geological targets accurately. The spectral similarity matrix of target and non-target classes based optimum threshold selection method has been developed for accurate classification of minerals using the hyperspectral dataset to address this limitation. The information of non-target classes has been successfully used in sub-pixel SMMs to reduce the influence of interference. The proposed threshold selection method would reduce the uncertainty in selecting the threshold required to obtain the final classification using SMMs. The developed threshold method has been compared and evaluated using the two statistical techniques suggested by Schwarz and Staenz (2001) and Hecker et al. (2008). The proposed method's novelty is that it does not require extensive field/lab data to select the optimum threshold and can be effectively applied for any SMMs. The existing SMMs (such as SAM and SID) do not perform good discrimination of spectrally similar minerals. Three new effective SMMs, such as Dice Spectral Similarity Coefficient (DSSC), Kumar-Johnson Spectral Similarity Coefficient (KJSSC), and their hybrid, i.e., $KJDSSC_{tan}$, have been proposed for successful hyperspectral data analysis. The novelty of the proposed SMMs has been compared and evaluated with existing SMMs using a wide range of hyperspectral datasets acquired under laboratory and real atmospheric conditions. The spectral discrimination ability of different SMMs have been further assessed using synthesized mixed spectra, and spectra with added random noise component. The developed SMMs are not only applicable to spectral discrimination. Still, they can also be effectively utilized in suitable band selection, endmember extraction, class separability, clustering, and image classification for successful hyperspectral data analysis.

The accurate lithological maps are essential in understanding the association of minerals and hydrothermal alteration for ore prospecting. MLAs' success in the lithological classification depends on several factors, including the quality and quantity of training datasets, suitable variables, and training parameters. This research has developed a method of obtaining relevant training datasets and variable selection with optimum training parameters in poorly exposed and weathered geological sites for lithological classification using ML models. The utility of PCA and ICA transformation for preparing the suitable reference lithology map has been illustrated to obtain an appropriate training dataset. An optimum feature selection method has been implemented to remove the correlated and less important variables to improve the model generalization and complexity. The performance of three MLAs, such as Support Vector Machine (SVM), Random Forest (RF), and Linear Discriminant Analysis (LDA), have been evaluated using different input datasets derived from the hyperspectral dataset.

Furthermore, the impact of the number and quality of training datasets on MLAs' performance has been analyzed. The classical lithological classification is performed using spectral data. However, spectral data may not be the ideal primary data for rock types classification of similar mineral composition. A multi-sensor data-based integration of spectral, textural, and morphological characteristics of rocks has been developed for lithological classification using ML models. Different input features such as spectral, transformed spectral (PCA and ICA), textural, morphological, and their optimal hybrid have been evaluated for their novelty in lithological classification using ML models. The applicability of developed lithological classification approaches using hyperspectral and multi-sensor datasets is not limited to a specific geological setting. Still, it can be efficiently applied in other regions for successful lithological mapping.

1.4 Research scope and objectives

This research aims to develop innovative methods for discrimination and classification of minerals and rock types/lithological units. Two sites, i.e., Hutti-Maski and Chattarpur area situated in India, have been explored for hydrothermal alteration minerals and lithological mapping using various remotely sensed datasets such as high-resolution Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Advanced Land Observing (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR), Sentinel-1, and Digital Elevation Model (DEM). Several SMMs and MLAs have been evaluated for their novelty in minerals and lithological classification. The developed methods are not only applicable in geological remote sensing applications. They can be effectively utilized in other applications of remote sensing such as vegetation/forest mapping and soil classification. The objectives of this research are the following:

- i. To evaluate the utility of AVIRIS-NG hyperspectral data for hydrothermal mineral mapping and to develop an optimum threshold selection for SMMs.
- ii. To develop effective SMMs for accurate discrimination of minerals using hyperspectral data.
- iii. To develop an integrated approach of spectral enhancement techniques and ML models for lithological classification using AVIRIS-NG hyperspectral data.
- iv. To develop multi-sensor datasets based optimal integration of spectral, morphological, and textural characteristics of rocks for lithological classification using ML models.

1.5 Structure of the dissertation

The content of this dissertation is structured into five chapters. Chapter 1 provides a state of the art of remote sensing in minerals and lithological classification. This chapter consists of background, statement of the problem, significance and innovation, and research scope and objectives. Chapter 2 addresses the first objective of the dissertation. It evaluates the utility of high-resolution AVIRIS-NG hyperspectral data for hydrothermal alteration

mineral mapping in the greenstone belt of the Hutti-Maski area using different SMMs. This chapter also develops the spectral similarity matrix of target and non-target classes based optimum threshold selection for SMMs. Chapter 3 addresses the second objective of the dissertation. It develops three new effective SMMs (such as DSSC, KJSSC, and $KJDSSC_{tan}$) to accurately discriminate spectrally similar materials using hyperspectral data. This chapter also evaluates the performance of existing and newly proposed SMMs using a wide range of hyperspectral datasets of minerals and vegetation species acquired under laboratory and real atmospheric conditions. Chapter 4 addresses the third objective of the dissertation. It develops an integrated spectral enhancement and ML method for accurate lithological classification. This chapter also presents a comparative analysis of different input features derived by applying feature selection and extraction techniques on hyperspectral data using MLAs. This chapter also highlights the impact of the number of samples and mislabelling or uncertainty on different MLAs' performance. Chapter 5 addresses the fifth objective of the dissertation. It develops multi-sensor datasets based optimal integration of spectral, morphological, and textural characteristics of rocks for lithological classification using ML models. This chapter illustrates the novelty of integrating multi-sensor datasets of optical and microwave regions for accurate lithological classification using SVM and RF models.

2 Mapping hydrothermal alteration minerals using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski gold deposit area, India

(This material has been published in the “*International Journal of Remote Sensing, Taylor and Francis.*” Please refer to page XIII for the original paper link).

Abstract: The present study exploits high-resolution hyperspectral imagery acquired by the Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) sensor from the Hutti-Maski gold deposit area, India, to map hydrothermal alteration minerals. The study area is a volcanic dominated late Archean greenstone belt that hosts major gold mineralization in the Eastern Dharwar Craton of southern India. The study encompasses pre-processing, spectral, and spatial image reduction using Minimum Noise Fraction (MNF) and Fast Pixel Purity Index (FPPI), followed by endmember extraction using n -dimensional visualizer and the United States Geological Survey (USGS) mineral spectral library. Image derived endmembers such as goethite, chlorite, chlorite at the mine site (chlorite mixed with mined materials), kaolinite, and muscovite were subsequently used in spectral mapping methods such as Spectral Angle Mapper (SAM), Spectral Information Divergence (SID) and its hybrid, i.e., SIDSAM_{tan}. The spectral similarity matrix of the target and non-target-based method has been proposed to find the possible optimum threshold needed to obtain a mineral map using spectral mapping methods. Relative Spectral Discrimination Power (RSDPW) and Confusion Matrix (CM) have been used to evaluate the performance of SAM, SID, and SIDSAM_{tan}. The RSDPW and CM illustrate that the SIDSAM_{tan} benefits from the unique characteristics of SAM and SID to achieve better discrimination capability. The Overall Accuracy (OA) and kappa coefficient (κ) of SAM, SID, and SIDSAM_{tan} were computed using 900 random validation points and obtained 90% (OA) and 0.88 (κ), 91.4% and 0.90, and 94.4% and 0.93, respectively. The obtained mineral map demonstrates that the northern portion of the area mainly consists of muscovite, whereas the southern part is marked by chlorite, goethite, muscovite, and kaolinite, indicating the propylitic alteration. Most of these minerals are associated with altered metavolcanic rocks and migmatite.

2.1 Introduction

The development of hyperspectral sensors has been an exciting innovation in the field of remote sensing. Hyperspectral sensors measure the reflected energy of terrestrial and extra-terrestrial surfaces in many narrow and contiguous spectral bands. The continuous spectral curve of each pixel of a hyperspectral image can be derived for material identification, discrimination, and quantification (Kruse, 1988, 1998; Van der Meer & De Jong, 2011; Zadeh et al., 2014). Hyperspectral datasets generated from the laboratory, field, and airborne/spaceborne platforms have been extensively used in the identification of groups of minerals such as clays, carbonates, and silicates (Goetz et al., 1985; Clark, 1999; Bierwirth et al., 2002; Kruse et al., 2006; Plaza et al., 2009; Kruse, 2012; Zadeh et al.,

2014). The broad wavelength coverage and high spectral resolution enable the hyperspectral mode of investigation to detect and discriminate even subtle variations in the minerals caused by the crystal structure, grain size, moisture content, and chemical composition (Clark, 1999). The interpretation of hyperspectral data usually involves the inspection of the spectral signatures with spectral libraries' help. Spectral parameters such as absorption position, area, depth, width, and symmetry are essential to identify and characterize the minerals detected from hyperspectral data (Magendran & Sanjeevi, 2014; Awad et al., 2018; van der Meer et al., 2018).

Hyperspectral based identification and mapping of hydrothermal alteration zones have been extensively used as a proxy among the geological community to investigate several metals and economic deposits such as iron, copper, lead, zinc, and gold (Kruse, 1988; Bishop et al., 2011; Pour & Hashim, 2011; Farifteh et al., 2013; Farooq & Govil, 2014; Carrino et al., 2018). A typical hyperspectral data processing incorporates image calibration/atmospheric correction, spectral and spatial data reduction, endmember extraction, and classification, followed by accuracy assessment (Boardman, 1998). The final product of hyperspectral image analysis results in distribution and abundance/fraction maps by deploying pixel and sub-pixel spectral classification, respectively. Some of the most common mapping methods in the literature are Jeffries-Matusita (JM) distance (Richards & Richards, 1999), Spectral Angle Mapper (SAM) (Kruse et al., 1993), Spectral Information Divergences (SID) (Chang, 1999), and Spectral Correlation Mapper (SCM) (De Carvalho & Meneses, 2000) that have been deployed on many hyperspectral datasets to map a variety of target materials successfully. Studies have also been focused on the development of hybrid algorithms by incorporating different similarity measures into a single measure such as SIDSAM_{tan} (Du et al., 2004), SIDSCM_{tan} (Naresh Kumar et al., 2011), and JMSAM (Padma & Sanjeevi, 2014a, 2014b).

Previous studies such as Du et al. (2004); Naresh Kumar et al. (2011); Padma and Sanjeevi (2014a, 2014b); and Adep et al. (2016) demonstrated that the hybrid methods perform superior to their standard version. Most of these spectral mapping methods require a threshold to obtain the distribution map, which is subjective and challenging to determine (Asadzadeh & de Souza Filho, 2016). The different threshold produces different distribution maps with varying accuracy. The most common thresholding procedure for these classifiers is the trial and error approach, resulting in over-or under-prediction (mapping) of target classes.

An important aspect of hyperspectral image analysis is evaluating the performance of mapping methods and accuracy assessment of distribution/classified maps. The most commonly used method is the computation of Overall Accuracy (OA) and the kappa coefficient (k) from the confusion matrix (Congalton, 1991). Chang (1999) has introduced a performance measure statistics for spectral mapping methods called a Relative Spectral Discriminatory Power (RSDPW) and intriguingly adopted in recent literature (Du et al., 2004; Naresh Kumar et al., 2011; Adep et al., 2016). The accuracy assessment of mineral maps has been carried out by the various approaches and can be categorized as field/laboratory analysis and Virtual Verification (VV) (King & Clark, 2000). The

field/laboratory-based evaluation is usually carried by a geological field investigation and the in-situ/lab spectral measurements of rock samples/exposure followed by petrographic and geochemical analysis. The VV method uses high-resolution image information to evaluate the classification results and can be an effective alternative where field data is limited (King & Clark, 2000). This method has been successfully adopted in recent studies to evaluate the hyperspectral classification results (Kumar et al., 2014; Molan et al., 2014; Adep et al., 2016).

Despite the high gold potential, the Hutti-Maski area in Karnataka, India, has not been investigated using hyperspectral remotely sensed data to improve the spatial information of alteration minerals and its relation to orogenic gold potential zones. In this study, high resolution airborne hyperspectral data acquired by recently developed Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) sensor has been used to map the hydrothermal alteration minerals using the most widely explored spectral mapping methods such as SAM, SID, and its hybrid, i.e., SIDSAM_{tan}. These methods' performance has been evaluated based on RSDPW, OA, and *k* with the application of an optimum method of thresholding based on the spectral similarity of the target and non-target classes.

2.2 Geological setting of the area

The present study has been conducted on the part of Hutti-Maski Schist Belt, situated between 16°1'12" to 16°24'0" N and 76°36'0" to 76°51'0" E, with a spatial extent of about 232 km². Figure 2.1 illustrates the geographical location and main lithological units of the area. The lithological map has been prepared from the district resource map of the Geological Survey of India (GSI) (scale: 1:2,50,000). The region of Hutti-Maski is a volcanic-dominated late Archean greenstone belt located in the eastern Dharwar Craton, Southern India. The belt hosts several gold deposits such as Hutti, Utti, and Hira-Buddini gold mines (Mishra et al., 2005). The Hutti Mine is the largest operational gold mine in India and has a proven reserve of >120 t Au, with an average grade of 4.42 g t⁻¹ Au (Hazarika et al., 2015). The structural setting, mineralization, and fluid composition of gold deposits demonstrate various similarities to typical greenstone-hosted mesothermal lode type or orogenic gold deposits (Mishra et al., 2005). It consists of a series of parallel N-S trending, around 1 km long and 2 to 10 m thick shear zones, which are best developed in coarse-grained amphibolite adjacent to biotite and sulfide-rich alteration zones (Mishra & Pal, 2008; Hazarika et al., 2015). The belt mainly consists of highly deformed basic meta-volcanic rocks (>90%) with minor intercalated felsic units and subordinate clastic metasedimentary rocks (Rogers et al., 2007). The dominating pillow bearing basaltic rocks are followed in abundance by acidic to intermediate volcanic rocks and quartz/rhyolite porphyry. The primary alteration process in the area is retrograde, saussuritization, and propylitization. Chloritization and silicification are common in the outer and central parts, respectively. The retrograde alteration has been observed by the alteration of biotite to chlorite and further chlorite to sericite. Propylitic alteration is evidenced by the assemblage of carbonate, epidote and chlorite (Pour & Hashim, 2011). Saussuritization is indicated by

the alteration of plagioclase to chlorite and epidote. The silicification alteration is evidenced by fine-grained aggregates of quartz. Many of the gold-bearing host rocks are rich in Fe content. Sulfidation of wall rocks causes the destabilization of the bi-sulphide complex causing the precipitation of gold (Pirajno, 2009). Most of the previous studies in this area are mainly focused on the geochemical analysis of sub-surface altered rock samples of known gold mineralized zone in the vicinity of Hutti, Uti, and Hira-Buddini mines and lack comprehensive information of surface alteration minerals in a spatial domain.

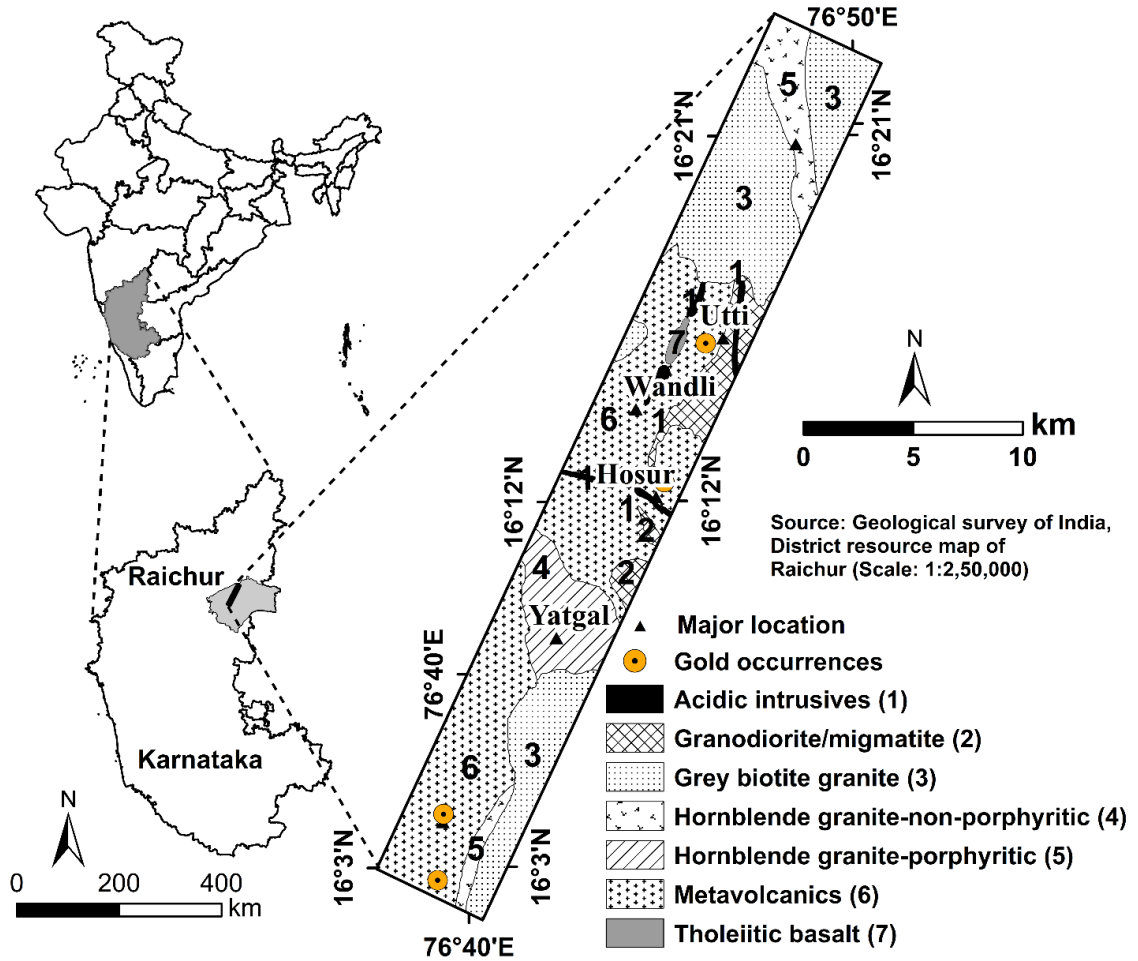


Figure 2.1. Location and lithological map of the studied area showing major rock types, major locations, and location of gold occurrences reported by Geological Survey of India.

2.3 Materials and methods

2.3.1 Airborne and field hyperspectral datasets

The AVIRIS-NG sensor has acquired the airborne hyperspectral data by recording the reflected energy in 425 spectral channels in the wavelength range of 380 to 2510 nm at 5 nm and 7.5 m spectral and spatial resolution, respectively. The AVIRIS-NG sensor has various improvements in resolution, uncertainty, and signal-to-noise ratio over the previous AVIRIS-Classic (Table 2.1). The enhanced specifications of AVIRIS-NG provide a better opportunity to accurately identify, discriminate, and quantify geological and other target materials. The hyperspectral data has been acquired under the collaborative campaign of the Indian Space Research Organization (ISRO) and National Aeronautics and Space Administration (NASA) in the years of 2015/16 and processed as level 1 (L1: at sensor radiance), and level 2 (L2: surface reflectance) products with geometric correction applied on both the levels. The spectral curve of rock samples (referred hereafter as lab spectra) collected from the area has been derived using Analytical Spectral Device (ASD) FieldSpec4[®] spectroradiometer in the wavelength range of 350–2500 nm at the sampling interval of 1.4 nm for 350–1000 nm and 2 nm for 1000–2500 nm. A detailed description of the ASD spectroradiometer can be accessed at this website (<http://www.asdi.com>). The software packages such as Environment for Visualizing Images (ENVI)[®] 5.4 and Aeronautical Reconnaissance Coverage Geographical Information System (ArcGIS)[®] 10.5 have been used for image processing and the preparation of spatial maps, respectively.

Table 2.1. Key characteristics of AVIRIS-NG and AVIRIS-Classic sensor (Thorpe et al., 2016). VNIR: Visible Near-Infrared; SWIR: Shortwave Infrared.

Characteristic	AVIRIS-NG	AVIRIS-Classic
Spectral range	380 to 2510 nm	380 to 2500 nm
Spectral channels	425	224
Spectral resolution	5 nm	10 nm
Calibration	± 0.3 nm	± 0.3 nm
Signal to Noise Ratio (SNR)	>2000 for VNIR, >800 for SWIR	>1000 for VNIR, >400 for SWIR
Accuracy	95% (<5 % uncertainty)	90% (<10 % uncertainty)
Sample distance	0.3 to 20 m	4 to 20 m
Radiometric resolution	14 bits	10 and 12 bits

2.3.2 AVIRIS-NG hyperspectral image processing and classification

The level 2 (L2) processed (atmospheric and geometrically corrected) AVIRIS-NG data consists of 425 spectral bands that were subjected to the removal of uncalibrated/noisy bands. Two hundred ninety-five spectral bands have been found suitable, whereas remaining bands such as 1–15, 146–158, 188–227, 275–328, and 421–425 were discarded due to the significant noise. These 295 bands were utilized in a typical hyperspectral data processing scheme such as spectral and spatial image reduction, endmember extraction, image classification followed by accuracy assessment (Kruse, 1998) to obtain the distribution of hydrothermal alteration minerals. A work flowchart of the methodology adopted in the study is illustrated in Figure 2.2.

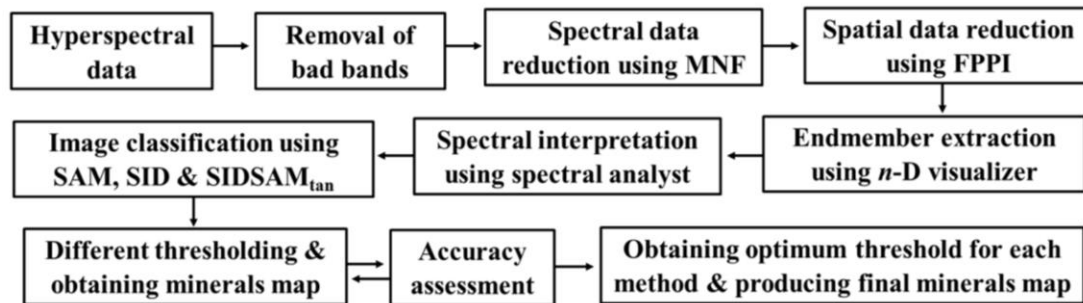


Figure 2.2. A work flowchart showing methodology adopted in the processing and classification of the hyperspectral data.

Minimum Noise Fraction (MNF) has been used to segregate the noise and reduce the dimensionality of the data (Green et al., 1988). The MNF plot shows the variance of spectral data captured by each transformed MNF bands and has been used to select the number of optimum MNF bands for the extraction of spectrally pure pixels. The first 34 MNFs marked by higher variance, less noise, and redundancy have been used to compute the spectrally pure pixels using the Fast Pixel Purity Index (FPPI) (Plaza & Chang, 2005) with iterations of 30,000. Different iterations such as 15,000, 20,000, 30,000 and 35,000 have been tested. The final selection of the number of iterations has been made based on the plot of FPPI. The plot begins to level off as it approaches an iteration of 30,000, which indicates that each subsequent iteration fails to find new pure pixels. These pure pixels have been used as the primary input to the n -dimensional (n -D) visualizer tool of ENVI for the endmember extraction. The n -D tool has been used interactively to rotate the data cloud (pure pixels) in image dimension space and extract the corner pixels representing better characteristics to be an endmember candidate. To preserve the subtle spectral features of an endmember, no more than 2 to 4 pixels have been clubbed together.

Further, these endmembers have been subjected to spectral interpretation and identification using the United States Geological Survey (USGS) mineral spectral library in conjunction with the Spectral Analyst tool of ENVI. Continuum removed spectra were also used to interpret the absorption features related to different minerals. Continuum removal normalizes the reflectance spectra to highlight the individual absorption features for better

spectral comparison (Clark & Roush, 1984). The USGS library spectra were resampled at the spectral resolution of the hyperspectral image before making any comparison. The endmembers that showed the highest match with the minerals of USGS spectra were considered as a final endmember for that particular class.

Further, image spectra were also compared with lab spectra (resampled at image spectral resolution) derived from the rock samples of major lithological units of the area. Spectral parameters such as absorption position and depth of each mineral's diagnostic absorption feature for image, lab, and USGS library spectra have also been computed for comprehensive spectral comparison and inspection using 'hsdar' (Lehnert et al., 2016) package available in 'R' software. The main purpose of such a comparison was to understand the characteristics of spectral signatures quantitatively.

SAM, SID, and its hybrid version, i.e., SIDSAM_{tan}, have been employed to map the alteration minerals. The SAM is a deterministic approach that measures the angle between the image spectra and endmember spectra and treats them as a vector in a space with dimensionality equal to the number of spectral bands (Kruse et al., 1993). The SID is derived from the concept of divergence in information theory, which measures the divergence of probabilistic behaviors between the image spectra and reference spectra (Chang, 1999). SAM suppresses the effects of shadows (Kruse et al., 1993), while SID solves the uncertainty introduced by unknown interference common in the hyperspectral dataset (Chang, 1999). The SIDSAM_{tan} is a product of SID and SAM with tangent (tan) trigonometric function (i.e., $SID \times \tan(SAM)$) (Du et al., 2004). The hybrid approach of SAM and SID incorporates the unique characteristics of both measures to improve the discrimination capability.

To measure the capability of the spectral matching methods in terms of the power of discriminating one spectrum from another relative to a reference spectrum, RSDPW statistics were used. Higher the RSDPW, the better the discrimination ability of a method. Unlike the similarity value produced by spectral matching methods (such as SAM and SID), the value of RSDPW is comparable for different methods. The computation of RSDPW is performed with respect to a reference spectrum, which is usually a linearly mixed spectrum of each class or a spectrum, which has some similarity with the target spectra (Chang, 1999; Naresh Kumar et al., 2011). In this study, the RSDPW was computed using a linear mixture of the target and non-target classes as a reference spectrum. Non-targets have been included in the linear mixing as some non-target spectra show considerable similarity with target spectra.

These classification algorithms' performance also depends on intra-class variability, the complexity of the area, and the threshold factor used to generate the distribution map. These classifiers' performance can be improved by incorporating information of non-target/undesired classes for a better separation of the target with spectrally similar non-target classes. The information of non-target has been successfully used in sub-pixel classification methods such as Orthogonal Subspace Projection (OSP) (Harsanyi & Chang, 1994), Constrained Energy Minimization (CEM) (Chang, 2000), and Target Constrained

Interference Minimized Filter (TCIMF) (Ren & Chang, 2000) in order to reduce the false positives.

In this study, the non-target information was extracted from the spectra of poorly exposed rock surfaces, soil, and mixed pixel of vegetation and rocks. The spectral similarity of target minerals and non-targets has been computed for SAM, SID, and SIDSAM_{tan} results using “proxy” (Meyer & Buchta, 2009; Meyer, 2018) and “resemble” (Ramirez-Lopez & Stevens, 2016) packages in R software. The similarity matrix was further used to obtain possible optimal threshold value needed to produce a mineralogical map from these classifiers. The detailed procedure adopted to choose the optimum threshold has been discussed in section 2.4.2. Statistical methods of threshold selection for SAM suggested by Schwarz and Staenz (2001) (SM₁), and Hecker et al. (2008) (SM₂) have been used to compare the reliability of the proposed method for optimum threshold selection. Schwarz and Staenz (2001) use the mean and standard deviation of the rule image of the SAM. An arbitrary parameter (i.e., m_{user} (ranges from 1 to 3)) given in the Equation (2.1) for obtaining the threshold value (T). In this study, m_{user} equals to 1 as a value greater than 1 yields a negative threshold, which may not be applicable. Hecker et al. (2008) suggested to use the 25th percentile value of the rule images of each class as a threshold value for the SAM.

$$T = mean - m_{user} \times standrad\ deviation \quad (2.1)$$

VV (King & Clark, 2000), RSDPW (Chang, 1999), confusion matrix, and calculation of k and OA (Congalton, 1991) have been used to access the classification accuracy of images.

2.4 Results and discussion

2.4.1 Alteration minerals

Goethite, chlorite, chlorite from the mine site (a mixture of chlorite and mined materials), kaolinite, and muscovite (Figure 2.3) have been extracted from the pure pixels of the AVIRIS-NG hyperspectral image as endmembers. The spectral parameters such as absorption position and depth of the diagnostic absorption feature of detected minerals from the image, lab, and USGS library spectra (Table 2.2) have been used in the spectral interpretation. The wavelength position of absorption (Figure 2.3 and Table 2.2) shows a slight difference in these three sources due to variation in the crystal structure, grain size, and presence of secondary minerals or impurities (Clark, 1999). It is also worth considering that the image spectrum is a representative of its pixel size (i.e., 7.5 m in this case) usually, a mixed spectrum, whereas USGS library spectra typically collected from the majorly pure mineral samples. The detected minerals represent the alteration process in the area. The spectral curves of goethite (Figure 2.3 (a)) show the diagnostic absorption feature at 600 and 900 nm due to ferric iron (Fe³⁺) (Pontual et al., 2008). The goethite has also been detected in the spectral curve of kaolinite (Figure 2.3 (d)) and muscovite (Figure 2.3 (e)). Chlorite shows its diagnostic absorption feature at 2250 and 2330 nm due to Fe-OH and Mg-OH molecules, respectively (Figure 2.3 (b) and (c)) (Clark, 1999).

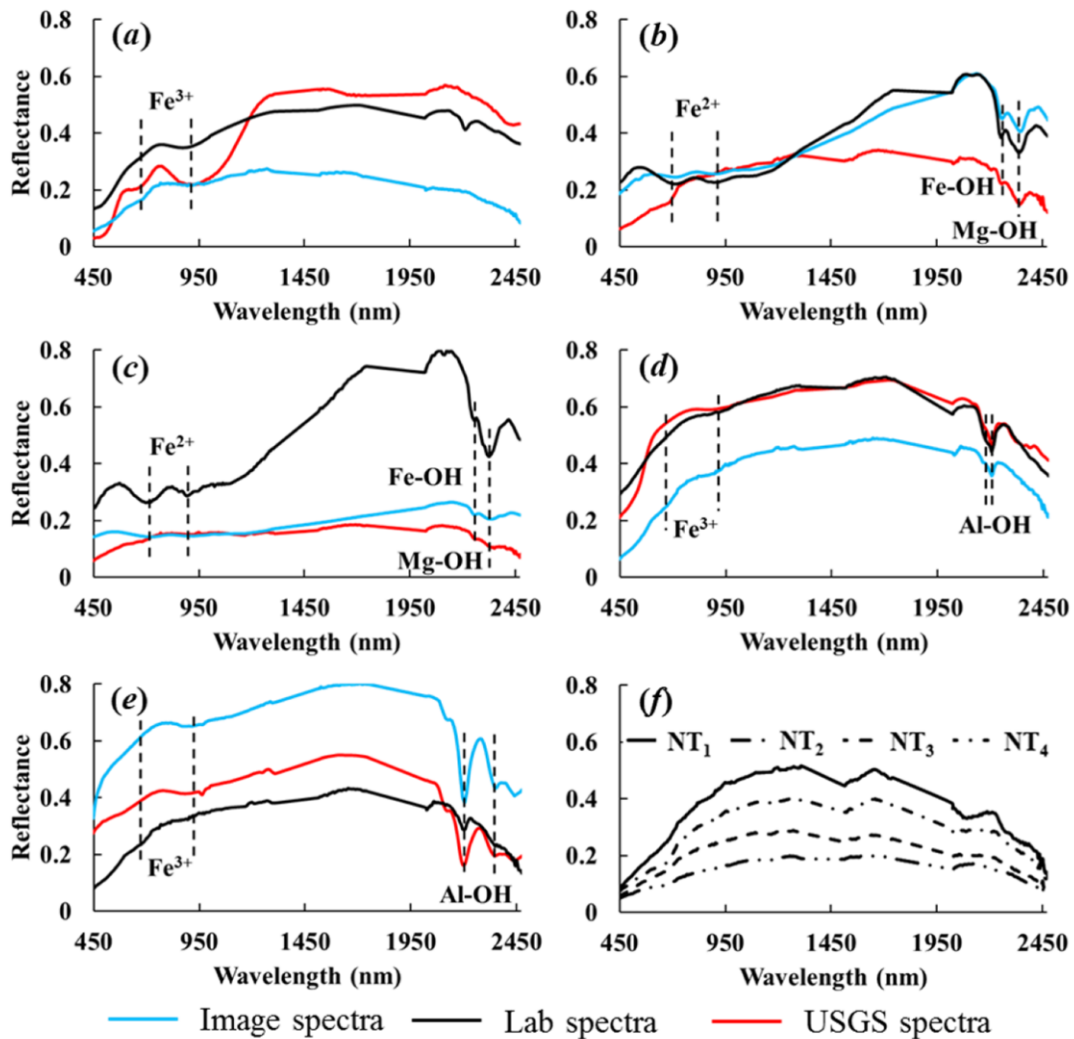


Figure 2.3. Reflectance spectra of altered minerals such as (a) goethite, (b) chlorite, (c) chlorite at of mine site, (d) kaolinite, and (e) muscovite, showing their diagnostic absorption features. (f) Non-target spectra. NT₁: soil, NT₂: poorly exposed rock surface, NT₃: mixed pixel of vegetation and rocks, and NT₄: dry vegetation. Image, laboratory, and USGS mineral library spectra are denoted as blue, black, and red color, respectively.

The spectral curve of pure chlorite (Figure 2.3 (b)) and chlorite at the mine site (Figure 2.3 (c)) also show small absorption features at 600 and 900 nm due to ferrous iron (Pontual et al., 2008). The spectra of chlorite at the mine site (Figure 3 (c)) preserve the diagnostic absorption feature the same as pure chlorite but are less intense. Although the spectra were derived from the mixture of mined materials, the spectral resolution of the image is suitable to detect the presence of chlorite. The spectral plots of kaolinite show their diagnostic absorption feature referred to as doublet centered at 2165 and 2200 nm caused due to Al-

OH vibrational process (Figure 2.3 (d)) (Clark, 1999; Pontual et al., 2008). A subtle absorption feature can be noticed at 660 and 900 nm due to ferric iron (Pontual et al., 2008). The spectral plots of muscovite show its diagnostic absorption feature at 2200 nm due to Al-OH vibrational process (Figure 2.3 (e)). Muscovite has two other absorption features at 2240 and 2340 nm derived from Al-OH molecules (Pontual et al., 2008). The main difference between kaolinite and muscovite spectra is doublet in kaolinite (at 2165 and 2200 nm) and a single absorption feature of 2200 nm in the muscovite.

Table 2.2. Spectral parameters such position (in nm) and depth of prominent diagnostic absorption feature of detected minerals (depth is in arbitrary map units).

	Goethite			Chlorite			Chlorite (mine site)		
	Image	Lab	Library	Image	Lab	Library	Image	Lab	Library
Position	917	972	947	2340	2445	2335	2335	2325	2320
Depth	0.09	0.1	0.41	0.25	0.23	0.32	0.16	0.15	0.34
	Kaolinite			Muscovite					
	Image	Lab	Library	Image	Lab	Library			
Position	2204	2204	2204	2199	2199	2199			
Depth	0.18	0.14	0.18	0.18	0.30	0.34			

2.4.2 Spectral similarity measures and derivation of optimum threshold

Spectral similarity among target minerals and non-target classes has been computed using SAM, SID, and SIDSAM_{tan} results (Table 2.3) to understand the intra-class variability and further derive the optimum threshold value for each method. The similarity matrix illustrates that some target spectra have higher similarity with non-target spectra of soil (NT₁), poorly exposed rock surface (NT₂), the mixed pixel of vegetation and rocks (NT₃), and dry vegetation (NT₄). For example: if we observe the row of goethite mineral (Table 2.3), the chlorite has a higher similarity (i.e., 0.117) in the target group, but NT₃ of the non-target group is even much similar (i.e., 0.071) than chlorite. This means that there are high chances of inclusion of NT₃ and chlorite as false positives while classifying goethite. A similar situation can be seen in the case of chlorite where kaolinite has higher similarity (i.e., 0.080) in the target group, but NT₂ shows more similarity (0.078) than kaolinite that means while classifying chlorite, there are higher chances of the inclusion of NT₂ and kaolinite as false positives. Other target classes such as chlorite (at mine site), kaolinite, and muscovite have fewer chances of getting affected by non-targets due to higher spectral dissimilarities. Still, at the same time, spectrally similar target classes may cause misclassification. Similar inferences can be made from the similarity matrix computed using SID and SIDSAM_{tan} (Table 2.3).

The spectral similarity matrix of the target and non-target classes have been further used to find the possible optimum threshold for each classification method. Obtaining an optimum threshold is crucial and a challenging task to utilize these classification methods effectively. The spectral similarity matrix incorporates intra-class variability and non-

target spectral similarity information to find an optimum threshold for producing classification output. To obtain the possible optimum threshold for each class, the minimum similarity value (Table 2.3) of that class has been used as an initial value and subsequently, different ranges have been tested to obtain an optimum threshold for that class (Figure 2.4).

To explain the procedure adopted to find the optimum threshold, let's consider the example of the goethite class. The minimum similarity value in the row of goethite is 0.070 of NT₃ (non-target class) (Table 2.3). Any values greater than or equal to 0.070 would increase the chances of inclusion of NT₃ as goethite in the classification output. The possible optimum threshold value would be less than 0.070 for the goethite class. It is evident that a smaller threshold (i.e., spectral angle or divergence) would yield higher accuracy to detect/map that class, but at the same time, it would also cause to leave out many pixels of that class. An optimum threshold would be the one that includes the maximum number of pixels of that class and, at the same time, have less false positives. Based on this condition, different value ranges less than 0.070, such as 0.070–0.060, 0.060–0.050, 0.050–0.045, and 0.045–0.040, have been evaluated.

To evaluate each range, 150 random pixels were selected to extract the spectra from the calibrated hyperspectral image. These extracted spectra were then inspected visually against the given endmember used in the classification. We count the correctly matched spectra out of 150 spectra to compute its accuracy. The threshold range that produces an accuracy of $\geq 85\%$ was considered as a possible optimum threshold in this study. It is also possible that a much smaller threshold may produce better accuracy, but it would also cause to leave out more pixels, which may result in under predictive classification (Schwarz & Staenz, 2001). Similarly, different ranges of the threshold for other target classes such as chlorite, chlorite at the mine site, muscovite, and kaolinite have been evaluated to find the possible optimum threshold for each class. The same approach has been used for SID and SIDSAM_{tan} to obtain the optimum threshold for each class.

The statistical methods (i.e., SM₁ and SM₂) failed to provide an optimum threshold value needed to produce a distribution map of the target classes. For example, SM₁ and SM₂ produce 0.038 and 0.083, respectively, as a threshold for the goethite to be used for classification using SAM (Table 2.4). Figure 2.4 (a) of the goethite class clearly illustrates that values greater than 0.05 could result in a poor classification. However, the threshold value 0.038 is close to the threshold value obtained by the proposed method, which would yield high accuracy but may also leave out many truly classified pixels due to a smaller angle. Furthermore, the spectral similarity value of SAM (Table 2.3) clearly illustrates that the possible threshold to classify goethite using SAM should be less than 0.07 as greater values would increase the chances of the inclusion of non-targets in the classification. Thus, the threshold value of 0.083 obtained by SM₂ cannot be an excellent choice to use as a threshold for the goethite class. SM₁ also does not show success in getting less biased threshold values. The threshold value of goethite and kaolinite resulting under predictive, whereas other classes are over predictive classification. Figure 2.4 (a) clearly illustrates that threshold values obtained by the statistical methods would result in poor accuracy

classification maps of hydrothermal alteration minerals. The optimum threshold for each classification method for all the minerals classes is presented in Table 2.4 and Figure 2.4, which was used to obtain the mineral map.

The spectral similarity matrix of target and non-target classes can be a practical approach to optimum threshold selection to obtain accurate classification results using SMMs. The consideration of non-target information appeared to be an effective approach to reduce the uncertainty of threshold selection. This approach further provides information about the classes that are more prone to misclassification. Another advantage of the proposed threshold selection is that it does not require extensive lab/field data to ensure the selected threshold's reliability. However, the prerequisite for the successful utilization of proposed threshold selection is that the user needs to be familiar with the non-target classes that are spectrally similar and mixed with target classes.

Table 2.3. Spectral similarity matrix of target and non-targets using SAM (in radians), SID (SD), and SIDSAM_{tan} (SS) (in radians). Go: goethite, Cl: chlorite, C_m: chlorite at the mine site, Ka: kaolinite, Mu: muscovite, NT_(1 to 4): non-target spectra. Values marked in bold indicate the minimum spectral angle.

SM	Go	Cl	C _m	Ka	Mu	NT ₁	NT ₂	NT ₃	NT ₄
Go	--	0.117	0.137	0.124	0.194	0.083	0.089	0.071	0.084
Cl	0.117	--	0.134	0.080	0.113	0.128	0.078	0.131	0.106
C _m	0.137	0.135	--	0.143	0.162	0.191	0.137	0.177	0.179
Ka	0.124	0.080	0.143	--	0.101	0.149	0.189	0.151	0.125
Mu	0.194	0.113	0.162	0.101	--	0.218	0.144	0.223	0.195
SD	Go	Cl	C _m	Ka	Mu	NT ₁	NT ₂	NT ₃	NT ₄
Go	--	0.016	0.021	0.019	0.039	0.008	0.009	0.005	0.009
Cl	0.016	--	0.021	0.009	0.039	0.017	0.008	0.019	0.011
C _m	0.021	0.026	--	0.033	0.035	0.042	0.021	0.035	0.039
Ka	0.019	0.009	0.033	--	0.010	0.026	0.013	0.028	0.018
Mu	0.039	0.013	0.035	0.010	--	0.048	0.022	0.051	0.038
SS	Go	Cl	C _m	Ka	Mu	NT ₁	NT ₂	NT ₃	NT ₄
Go	--	0.0018	0.0029	0.0024	0.0077	0.0006	0.0008	0.0004	0.0007
Cl	0.0018	--	0.0029	0.0007	0.0044	0.0022	0.0006	0.0025	0.0012
C _m	0.0029	0.0036	--	0.0048	0.0056	0.0082	0.0030	0.0063	0.0070
Ka	0.0024	0.0007	0.0048	--	0.0010	0.0038	0.0026	0.0043	0.0023
Mu	0.0077	0.0015	0.0056	0.0010	--	0.0107	0.0032	0.0116	0.0074

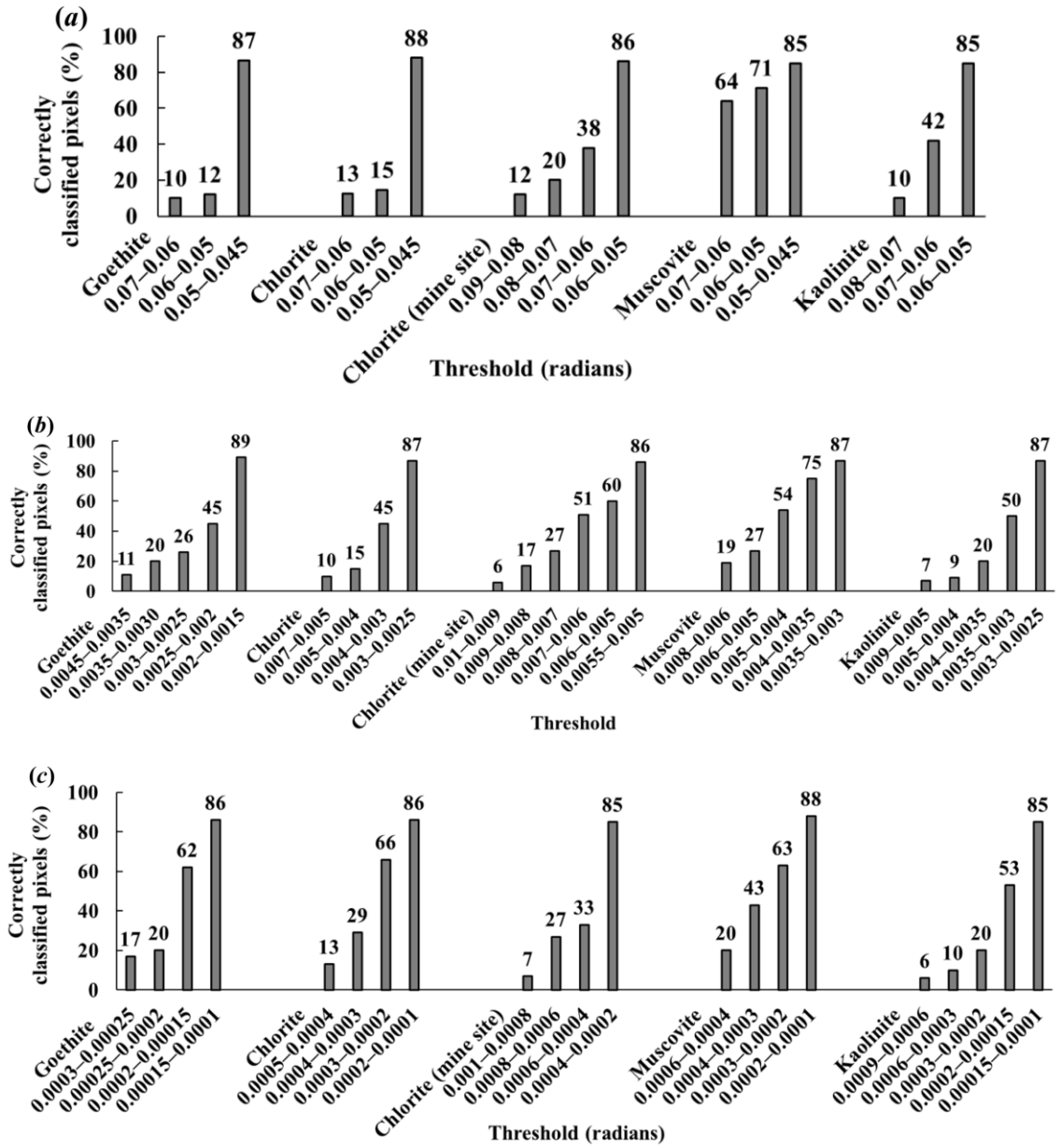


Figure 2.4. Correctly classified pixels against the threshold value ranges used for each class to obtain the optimum threshold value for (a) SAM, (b) SID, and (c) SIDSAM_{tan} classification method.

Table 2.4. The optimum threshold value was obtained using the statistical and proposed threshold selection method to get the mineral classification map using SAM (in radians), SID, and SIDSAM_{tan} (in radians). SAM_{SM1} and SAM_{SM2} indicate the possible optimum threshold derived using Schwarz and Staenz (2001) (SM₁) and Hecker et al. (2008) (SM₂) statistical methods.

Minerals	Statistical methods		Proposed method		
	SAM _{SM1}	SAM _{SM2}	SAM	SID	SIDSAM _{tan}
Goethite	0.38	0.083	0.045	0.0020	0.00015
Chlorite	0.18	0.070	0.050	0.0030	0.00020
Chlorite (Mine site)	0.098	0.146	0.060	0.0055	0.00040
Kaolinite	0.022	0.071	0.060	0.0030	0.00015
Muscovite	0.066	0.116	0.050	0.0035	0.00020

2.4.3 Performance evaluation of SAM, SID, and SIDSAM_{tan} using RSDPW

The similarity values of SAM, SID, and SIDSAM_{tan} are not comparable, as these measures use different theories and matching criteria. To evaluate the performance of SAM, SID, and SIDSAM_{tan}, RSDPW was used (Table 2.5). The computation of RSDPW is performed with respect to a reference spectrum, usually a linearly mixed spectrum of each class or a spectrum similar to the target spectra (Chang, 1999; Naresh Kumar et al., 2011). We computed the RSDPW by using a linear mixture of the target (15%) and non-target (6.25%) classes as a reference spectrum (Table 2.5). Non-targets have also been included in the linear mixing to obtain the reference spectrum, as some non-target spectra show considerable similarity with target spectra. A graphical representation of the different pairs of target classes has been presented in Figure 2.5 to illustrate and compare the performance of SAM, SID, and SIDSAM_{tan}.

Table 2.5 and Figure 2.5 illustrate that the SIDSAM_{tan} outperforms the SID and SAM. The RSDPW value of SIDSAM_{tan} is around two and three times higher than SID and SAM while discriminating most of the classes. It should also be noticed that the classes have more spectral dissimilarities, with other classes yield much higher RSDPW than the less dissimilar classes. The discrimination power of SID is better than SAM for most of the classes. The RSDPW also illustrates that SID and SAM complement each other to improve the hybrid measure's discrimination power. The discrimination power of SID is less than SAM while discriminating goethite-kaolinite and kaolinite-muscovite minerals but hybrid measures display a better discrimination power.

Table 2.5. RSDPW of SAM, SID, and SIDSAM_{tan} similarity measures for target and non-target classes using linearly mixed spectrum as a reference. Goe: goethite, Chl: chlorite, Chlmine: chlorite at the mine site, Kao: kaolinite, and Mus: muscovite. NT₍₁₋₄₎: non-target spectra.

SAM	Chl	Chl _{mine}	Kao	Mus	NT ₁	NT ₂	NT ₃	NT ₄
Goe	1.820	1.416	1.513	1.350	1.249	1.857	1.259	1.010
Chl	--	2.577	1.203	2.457	2.274	1.020	2.292	1.803
Chl _{mine}		--	2.141	1.049	1.133	2.629	1.124	1.429
Kao			--	2.042	1.890	1.228	1.905	1.498

Mus				--	1.081	2.507	1.072	1.363
SID	Chl	Chl _{mine}	Kao	Mus	NT ₁	NT ₂	NT ₃	NT ₄
Goe	2.704	4.352	1.184	1.370	1.203	1.172	1.374	1.246
Chl	--	11.769	3.201	3.705	3.253	2.307	3.715	2.170
Chl _{mine}		--	3.677	3.177	3.618	5.102	3.168	5.424
Kao			--	1.158	1.016	1.387	1.161	1.475
Mus				--	1.139	1.606	1.003	1.707
SIDSAM_{tan}	Chl	Chl _{mine}	Kao	Mus	NT ₁	NT ₂	NT ₃	NT ₄
Goe	4.931	6.177	1.887	1.853	1.505	2.181	1.733	1.258
Chl	--	30.458	3.853	9.138	7.420	2.261	8.544	3.919
Chl _{mine}		--	7.905	3.333	4.105	13.471	3.565	7.772
Kao			--	2.372	1.926	1.704	2.217	1.017
Mus				--	1.232	4.042	1.070	2.332

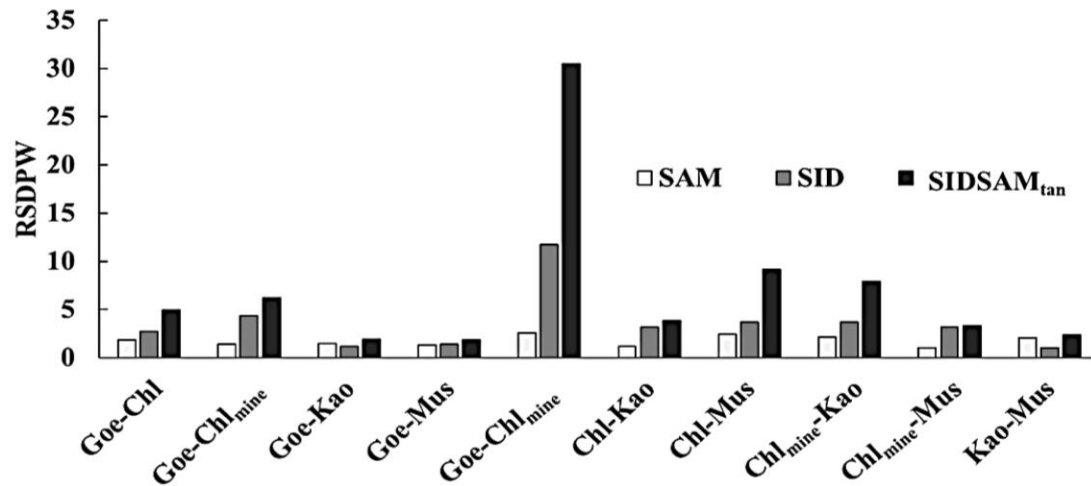


Figure 2.5. Graphical representation of RSDPW of target minerals showing discrimination power of SAM, SID, and SIDSAM_{tan}. Goe: goethite, Chl: chlorite, Chl_{mine}: chlorite at the mine site, Kao: kaolinite, and Mus: muscovite.

2.4.4 Hydrothermal alteration mineral mapping

Spectral mapping methods such as SAM, SID, and SIDSAM_{tan} have been employed on the hyperspectral image using the optimum threshold values derived by this study's proposed method (Table 2.4) to obtain the spatial distribution of the hydrothermal alteration minerals (Figure 2.6). The false-color composite of the area with a few major locations is shown in Figure 2.6 (a). The alteration minerals found in the area are goethite, chlorite, chlorite at the mine site, kaolinite, and muscovite. Table 2.6 illustrates the spatial extent of detected minerals derived by spectral mapping methods. The mineral map derived using SAM (Figure 2.6 (b)) and SID (Figure 2.6 (c)) show less distribution of altered minerals as compared with SIDSAM_{tan} (Figure 2.6 (d)).

The northern portion of the area mainly consists of muscovite associated with grey biotite granite and hornblende granite (non-porphyritic) rock types (Figure 2.6 (e)). The chlorite at the mine site shows the least spatial extent compared to other minerals, as it is mainly found at the mine site (Figure 2.6 (f)). Goethite and chlorite are comparatively less than kaolinite and muscovite. At a few locations, kaolinite, chlorite, muscovite, and goethite are associated with the lithological boundary of grey biotite granite and hornblende granite. The southern area's major lithological units marked by propylitic alteration affect metavolcanic rocks, hornblende granites (porphyritic), granodiorite/migmatite, display the association of kaolinite, chlorite, goethite, and muscovite. Goethite is found in the central and southern parts of the area (Figure 2.6 (g)) in the metavolcanic rocks and granodiorite. The southernmost part of the site mainly consists of kaolinite and chlorite (Figure 2.6 (h)).

Altered metavolcanic rocks have been the most important rock to explore the formation of the gold, according to Mishra et al. (2005); Mishra and Pal (2008); Hazarika et al. (2015). Hazarika et al. (2015) performed the geochemical and petrographic analysis of tourmaline minerals from altered rocks of the subsurface from the mine sites as proximal to the gold-bearing veins of these orogenic gold deposits. Tourmaline growth preceded the sulfide phases until the final sulfidation of wall rocks resulted in gold precipitation. The metavolcanic rocks host tourmaline in association with biotite-muscovite foliation, plagioclase, and chlorite minerals. The distribution of chlorite and muscovite obtained in this study can be used as a proxy to tourmaline (here, not spectrally investigated) in the area as it shows some association with these minerals (Hazarika et al., 2015).

Table 2.6. Spatial distribution of hydrothermal alteration minerals derived from the SAM, SID, and SIDSAM_{tan} using a spectral similarity-based optimum threshold value.

Minerals	Spatial extent of alteration minerals (km²)		
	SAM	SID	SIDSAM _{tan}
Goethite	0.60	0.74	2.50
Chlorite	2.50	3.88	11.04
Chlorite (Mine site)	0.10	0.16	0.24
Kaolinite	10.59	24.63	30.88
Muscovite	9.55	11.26	11.83

2.4.5 Accuracy assessment

A virtual verification approach has been used to assess the accuracy of classified images obtained from SAM, SID, and SIDSAM_{tan} due to a limited number of samples and a lack of geochemical analysis. One hundred fifty spectral curves of each class (including target and non-target classes) were randomly extracted from the calibrated hyperspectral image, and then their diagnostics spectral characteristics have been inspected against their corresponding endmembers used in the classification.

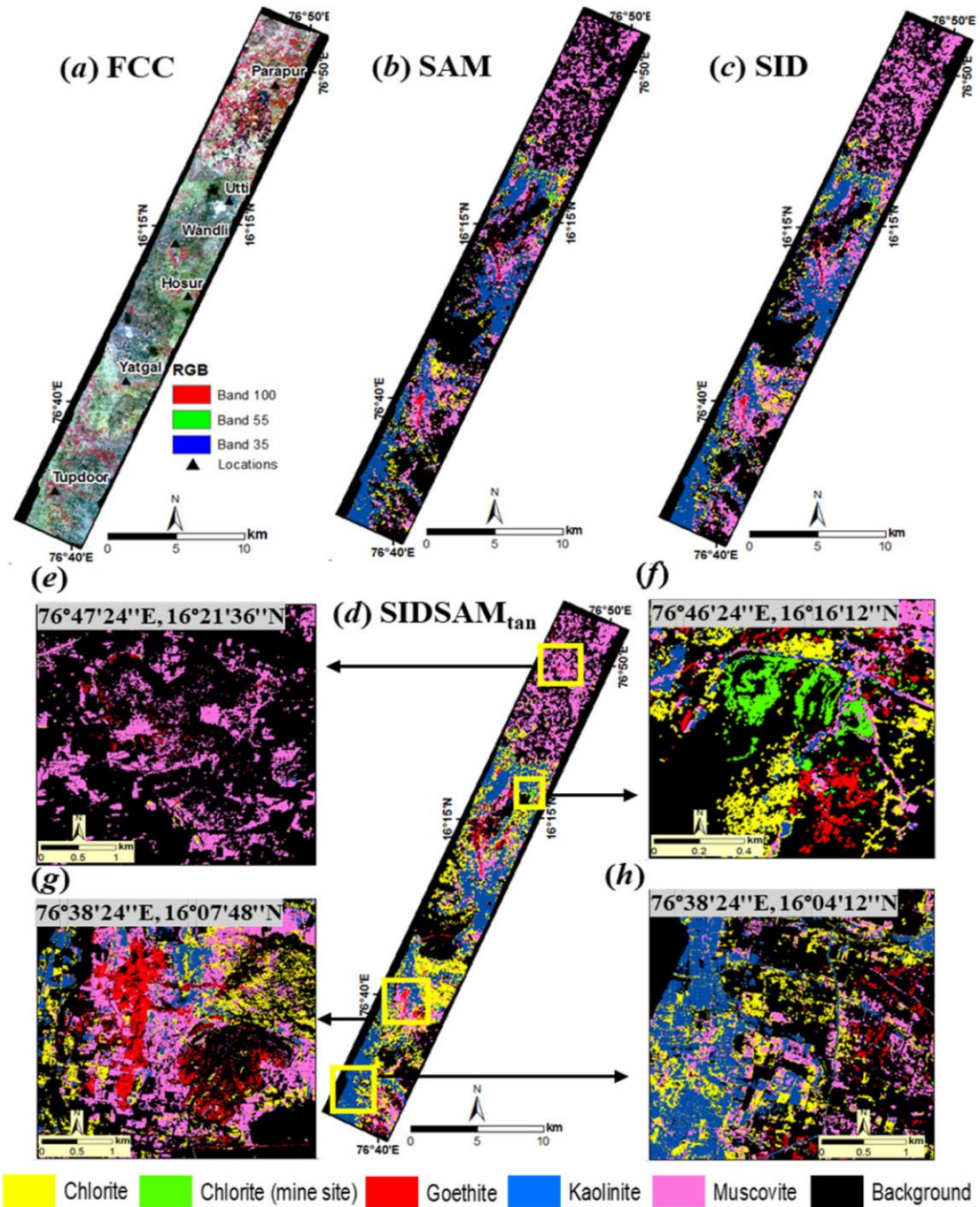


Figure 2.6. (a) False Color Composite (FCC) of the AVIRIS-NG image with few major locations. The mineral map obtained using (b) SAM, (c) SID, (d) SIDSAM_{tan}. The zoomed view of SIDSAM_{tan} result at few locations showing the distribution of mapped minerals: (e) muscovite in the northern part of the area, (f) chlorite at mine site, (g) goethite and muscovite, and (h) kaolinite and chlorite in the southern part of the area.

Table 2.7. Confusion matrix of the classified image using SAM, SID, and SIDSAM_{tan}. Goe: goethite, Chl: chlorite, Chl_{mine}: chlorite at the mine site, Kao: kaolinite, and Mus: muscovite, OC: other classes/non-targets. UA: user accuracy (%), PA: producer accuracy (%), OA: overall accuracy (%), k: kappa coefficient.

SAM	Goe	Chl	Chl _{mine}	Kao	Mus	OC	Row	UA	PA
Goe	132	6	0	0	1	11	150	88.0	89.8
Chl	3	135	0	2	1	9	150	90.0	84.4
Chl _{mine}	0	3	140	0	1	6	150	93.3	98.6
Kao	0	7	0	139	0	4	150	92.7	92.7
Mus	0	4	0	5	140	1	150	93.3	95.9
OC	12	5	2	4	3	124	150	82.7	80.0
Column	147	160	142	150	146	155	900		
OA = 90%, k = 0.88									
SID	Goe	Chl	Chl _{mine}	Kao	Mus	OC	Row	UA	PA
Goe	136	4	0	2	2	6	150	90.7	92.5
Chl	3	139	0	3	1	4	150	92.7	86.9
Chl _{mine}	0	1	138	4	2	5	150	92.0	97.2
Kao	1	6	0	140	2	1	150	93.3	90.3
Mus	0	8	0	2	140	0	150	93.3	93.3
OC	7	2	4	4	3	130	150	86.7	89.0
Column	147	160	142	155	150	146	900		
OA = 91.40%, k = 0.90									
SIDSAM_{tan}	Goe	Chl	Chl _{mine}	Kao	Mus	OC	Row	UA	PA
Goe	139	3	0	2	1	6	150	92.7	97.2
Chl	0	140	0	6	0	4	150	93.3	87.0
Chl _{mine}	0	2	145	1	2	0	150	96.7	98.6
Kao	1	6	0	143	0	0	150	95.3	93.5
Mus	0	7	0	0	143	0	150	95.3	97.3
OC	3	3	2	1	1	140	150	93.3	93.3
Column	143	161	147	153	147	150	900		
OA = 94.40%, k = 0.93									

The spectral curve that matches the corresponding endmember was counted as correctly classified, and if the spectral curve does not correspond with the respective endmember, then it has been compared with other endmembers, including other classes. There have been several occasions when a pixel shows spectral similarity with multiple endmembers due to the spectral mixing; in such case, a pixel that exhibits higher similarity with endmember has been considered. The accuracy measures such as User Accuracy (UA), Producer Accuracy (PU), OA, and k have been computed for the assessment of classification maps (Table 2.7). The non-target classes such as NT₁, NT₂, NT₃, and NT₄

has been clubbed together and named as other class (OC) in the confusion matrix for simplicity.

The accuracy measures illustrate that the SIDSAM_{tan} performs superior (OA: 94.40% and k: 0.93) than the SID (OA: 91.40% and k: 0.90) and SAM (OA: 90% and k: 0.88). Minerals that have more spectral dissimilarities yield higher accuracy than spectrally similar classes. Goethite has spectral similarity with chlorite and NT₃ (non-target/other class). The effects of the spectral similarity can be easily seen in the confusion matrix for all the methods. Chlorite and OC cause more false positives than other classes. A similar scenario can be observed while classifying the chlorite where goethite and other classes cause more false positives. However, SAM yields false positive than SID and hybrid measures. Chlorite at mine site displays more dissimilarity with other class hence yields higher accuracy in all the classification methods. It has been noticed that all the methods produce high accuracy that can be due to the employed optimum threshold factor obtained from the similarity matrix. The threshold factor is suitable in considering the intra-class variability and spectral dissimilarities with non-targets to yield good accuracy.

2.5 Conclusions

- The present study exploited AVIRIS-NG hyperspectral data to map the hydrothermal alteration minerals such as goethite, chlorite, kaolinite, and muscovite using spectral mapping methods such as SAM, SID, and SIDSAM_{tan} in the Hutti-Maski area of Karnataka. The spectral parameters such as absorption depth and wavelength position of the diagnostic absorption feature in conjunction with the USGS mineral spectral library has been used for the comprehensive spectral interpretation of image and rock samples of the studied area.
- The spectral similarity matrix of the target and non-target-based thresholding appears as a promising method to find an optimum threshold for the SMMs, as demonstrated in this study. The proposed threshold selection approach illustrates that the non-target information in the SMMs can improve the accuracy of these classifiers. Another advantage of this method is that the user may not need extensive field data, which is the primary constraint to exploit hyperspectral data in inaccessible locations or where field data/instrument is limited.
- RSDPW statistics and the confusion matrix have been used to evaluate the performance of SAM, SID, and SIDSAM_{tan}. It was observed that on a few occasions, SAM outperforms SID and vice-versa, but the hybrid method always shows great success over both the approaches. Furthermore, the capability to map the number of pixels of each class by these methods illustrate that SIDSAM_{tan} is superior to SAM and SID. The minerals have more intra-class variability, such as muscovite, chlorite of the mine site, and kaolinite, produce higher accuracy than the minerals that have less intra-class variability, such as goethite and chlorite.
- Identified minerals primarily belong to propylitic hydrothermal alteration. The northern portion of the area mainly consists of muscovite, whereas the southern part shows propylitic alteration characterized by the association of kaolinite, chlorite,

goethite, and muscovite. Most of these minerals are associated with metavolcanics and migmatite. The hydrothermal alteration mineral map obtained in this study can improve the understanding of the association of these minerals with tourmaline and gold mineralization in the area.

3 New effective spectral matching measures for hyperspectral data analysis

(This material has been accepted for publication in the “*International Journal of Remote Sensing, Taylor and Francis.*” Please refer to page XIII for the original paper link).

Abstract: The successful implementation of Spectral Matching Measures (SMMs) often plays a crucial role in material discrimination and classification using hyperspectral dataset. The commonly exploited SMMs, such as Spectral Angle Mapper (SAM), Spectral Information Divergence (SID), and their hybrid, i.e., SIDSAM_{tan}, show limited discrimination power while discriminating spectrally similar targets. This study presents three new effective SMMs named Dice Spectral Similarity Coefficient (DSSC), Kumar-Johnson Spectral Similarity Coefficient (KJSSC), and a hybrid of DSSC and KJSSC, i.e., KJDSSC_{tan}, for accurate discrimination of spectrally similar materials. A wide range of hyperspectral datasets of minerals and vegetation acquired under laboratory and real atmospheric conditions was used to compare and evaluate the performance of newly proposed and existing SMMs using Relative Spectral Discrimination Power (RSDPW) statistics. We also assessed the discrimination ability of the proposed and existing SMMs using spectra of selected minerals and vegetation with an added component of random noise and linearly synthesized mixed spectra. An in-depth comparison and evaluation of different SMMs demonstrated that the discrimination power of the proposed SMMs is significantly higher than existing SMMs. The proposed SMMs also outperform existing SMMs when discriminating spectra of minerals and vegetation with added noise and their linearly synthesized mixed counterpart. The KJSSC and DSSC show similar efficiency in discriminating spectra of minerals and vegetation; whereas, their hybrid measure, i.e., KJDSSC_{tan} shows much higher spectral discrimination ability. Therefore, the newly proposed hybrid measure, i.e., KJDSSC_{tan} is recommended over existing SMMs for successful material discrimination using hyperspectral data.

3.1 Introduction

The hyperspectral sensors measure the reflected or emitted electromagnetic radiation from the surface in many narrow and contiguous spectral channels to extract the materials' spectral signature (Goetz et al., 1985; Kruse, 1998). The material's spectral signature has been extensively utilized to study mineral composition, moisture content, and other physical properties (Cloutis, 1996; Clark, 1999). Hyperspectral remote sensing has emerged as the most promising tool for material detection, identification, discrimination, and quantification of earth and planetary surface (Colarusso et al., 1998; Clark et al., 2003; Goetz, 2009; Van der Meer et al., 2012; Bioucas-Dias et al., 2013).

The classical approach of hyperspectral data-based material identification follows the principle of matching the image spectra with a spectral library of known materials (Van der Meer, 2006; Clark et al., 2007). Therefore, the successful utilization of any hyperspectral data highly depends on the discrimination power of Spectral Matching

Measures (SMMs) to discriminate the spectrally similar targets (Homayouni & Roux, 2004; Keshava, 2004; Robila & Gershman, 2005). Several SMMs were used to analyze the spectral similarity between reference or library and target spectra of vegetation (Govender et al., 2007; Thenkabail & Lyon, 2016), soil (Ben-Dor et al., 2002; Govender et al., 2007), water (Govender et al., 2007), minerals, and rocks (Van der Meer, 2004; Van der Meer et al., 2012; Kumar et al., 2014; Kumar et al., 2020a).

The SMMs can be grouped into deterministic and stochastic measures. The deterministic measures are based on geometrical and physical aspects; whereas, stochastic measures deal with the probability distribution of spectra or pixel vectors (Vishnu et al., 2013). Some of the most popular deterministic SMMs are Euclidean Distance (ED) (Gower, 1985), Spectral Feature Fitting (SFF) (Clark et al., 1993), Spectral Angle Mapper (SAM) (Kruse et al., 1993), and Spectral Correlation Mapper (SCM) (Robila & Gershman, 2005). The widely employed stochastic SMMs include Jeffries Matusita (J-M) distance (Richards & Richards, 1999), Spectral Information Divergence (SID) (Chang, 2000), and many others (Shanmugam & SrinivasaPerumal, 2014).

Each method has some advantages and limitations over other methods. To minimize the limitation of individual measures, significant effort has been given in hybridizing multiple measures together to improve their discrimination ability (Du et al., 2004; Naresh Kumar et al., 2011; Nidamanuri & Zbell, 2011; Padma & Sanjeevi, 2014a, 2014b). Some of the well-exploited hybrid SMMs are SIDSAM_{tan} (Du et al., 2004), SIDSCM (Naresh Kumar et al., 2011), Normalized Spectral Similarity Score (NS3) (Nidamanuri & Zbell, 2011), and JM-SAM (Padma & Sanjeevi, 2014a). The primary purpose of the hybridization of multiple measures is to take each method's unique advantage and overcome their limitations to improve spectral discrimination power. The hybrid measures take the intra-class and inter-class variability into account more efficiently than the individual method (Shanmugam & SrinivasaPerumal, 2014; Padma & Sanjeevi, 2014b; Ding et al., 2015) in spectral discrimination measures.

The similarity or dissimilarity values obtained using various SMMs cannot be compared directly, as these methods use different theories and measurement units in computing similarities. Chang (2000) has proposed the Relative Spectral Discrimination Power (RSDPW) statistics to compare and evaluate the discrimination power of different SMMs. Unlike the similarity values obtained by different SMMs, the value of RSDPW can be used in the performance evaluation of SMMs (Chang, 1999, 2000, 2003; Du et al., 2004; Van der Meer, 2006; Naresh Kumar et al., 2011; Adep et al., 2016; Kumar et al., 2020a).

Although various SMMs are found in the literature, there is a continual demand to develop optimal SMMs for accurate material discrimination and classification (Shanmugam & SrinivasaPerumal, 2014; Souza et al., 2017). The most commonly employed SMMs are SAM and SID due to their simple computation and good discrimination ability. The SAM may not perform satisfactory discrimination of two spectra if the spectral angle between them is very small (i.e., it may not be able to accurately discriminate two very similar spectra); whereas, the SID does not consider the geometrical features of the spectra (Naresh

Kumar et al., 2011; Padma & Sanjeevi, 2014b; Friedrich et al., 2016). There are few studies available in the literature to assess the discrimination ability of SMMs using spectrally similar targets and the development of new measures for accurate discrimination of spectrally similar materials (Van der Meer, 2006; Naresh Kumar et al., 2011). To address this research gap, we propose three new effective SMMs, such as Dice Spectral Similarity Coefficient (DSSC), Kumar-Johnson Spectral Similarity Coefficient (KJSSC), and a hybrid of DSSC and KJSSC, named as $KJDSSC_{tan}$, for successful discrimination of spectrally similar materials. The DSSC and KJSSC can be derived by deploying the Dice Similarity Coefficient and Kumar-Johnson Similarity Coefficient (KJSC), respectively. It is worth mentioning here that the KJSC and DSC were successfully employed in the field of biological, medical, and computer sciences (Hubálek et al., 1995; LaPara et al., 2002; Murguía & Villaseñor, 2003; Cha, 2007; Sampat et al., 2009; Baum et al., 2010; Ardjmand et al., 2014; Ayseldeen et al., 2015), but remains unexplored in the field of hyperspectral remote sensing for matching the similarity of spectra. A comparative analysis of the proposed and existing SMMs has been carried out to evaluate their discrimination ability based on RSDPW statistics using seven different hyperspectral datasets such as (a) USGS spectra of minerals, (b) USGS spectra of vegetation, (c) USGS spectra of minerals with added noise component, (d) USGS spectra of vegetation with the added noise component (e) linearly synthesized mixed minerals spectra of USGS, (f) linearly synthesized mixed vegetation spectra of USGS, and (g) spectra of hydrothermal alteration minerals derived from airborne hyperspectral data.

3.2 Hyperspectral datasets and methods

3.2.1 Hyperspectral datasets

The hyperspectral data acquired under laboratory and real atmospheric conditions were used to compare and evaluate the performance of commonly adopted and newly proposed SMMs. The spectra of minerals (i.e., alunite, calcite, hematite, jarosite, kaolinite, and mixed spectrum (i.e., derived using linear spectral mixing of all minerals with equal proportion)) (Figure 3.1), and vegetation spectra (i.e., black brush, blue spruce, cheatgrass, juniper bush, rabbitbrush, salt brush, and mixed spectrum (i.e., derived using linear spectral mixing of all vegetation species with equal proportion)) (Figure 3.2) are obtained from the United States Geological Survey (USGS) spectral library within the wavelength range of 400 to 2560 nm. The USGS spectral library contains the spectra of several materials, including minerals, rocks, soils, vegetation, man-made materials, and others. These spectra have been commonly used as reference spectra for the spectroscopic interpretation and analysis of laboratory, field, and aerial or space-borne hyperspectral data. The advantage of using the USGS spectra of materials is that it contains comprehensive information of the sample, including companion description, photographs, spectral measurement setting, and quality or impurity information of the samples (Kokaly et al., 2017).

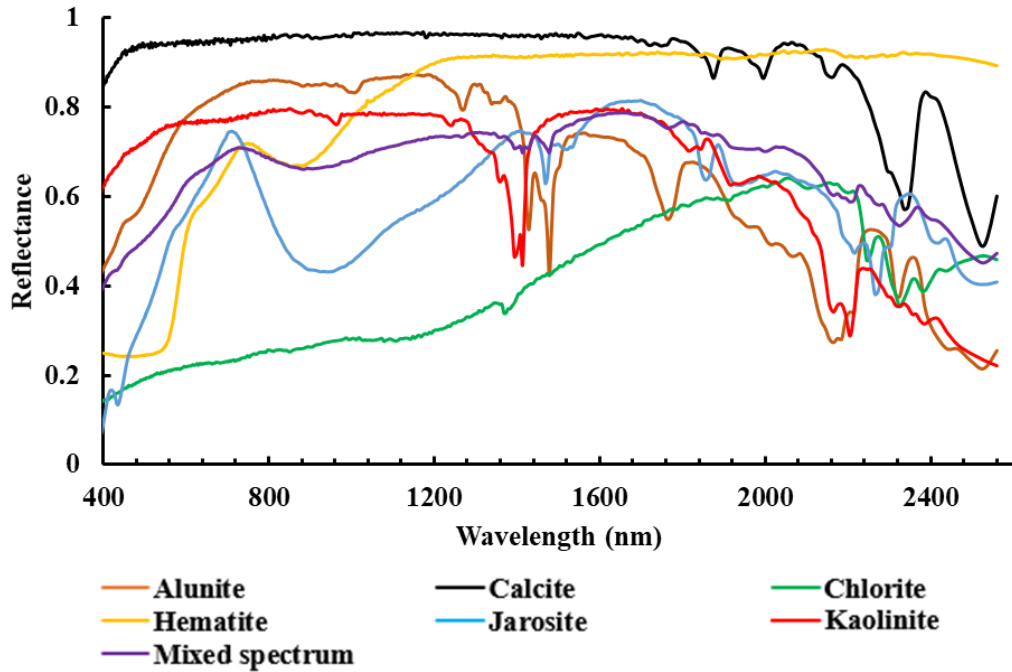


Figure 3.1. The spectral curve of minerals obtained from the USGS mineral spectral library.

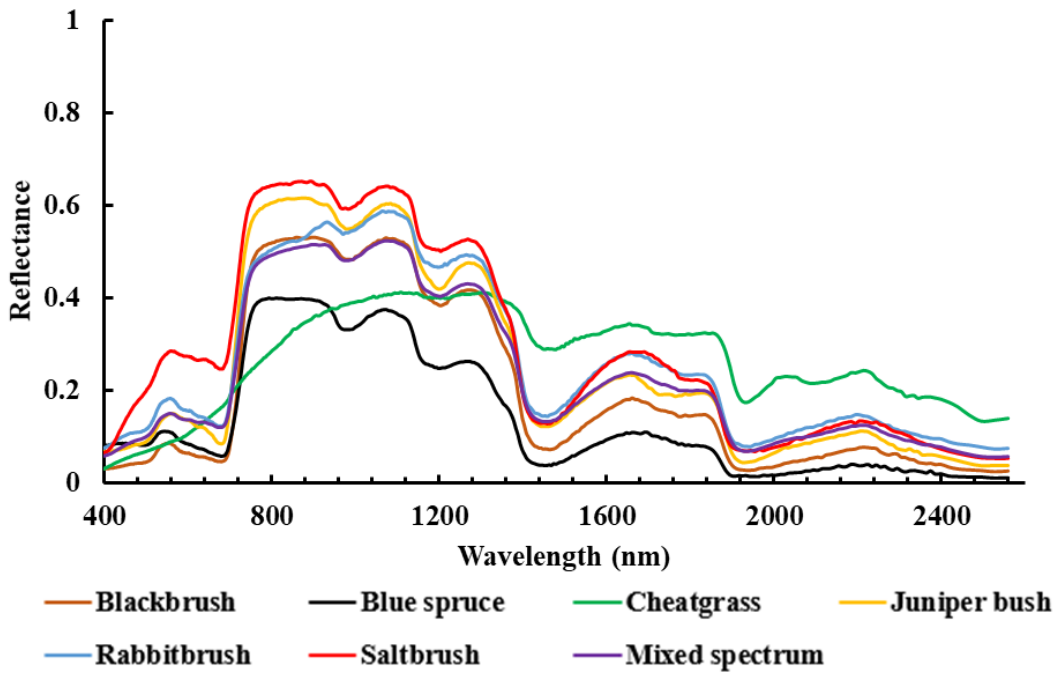


Figure 3.2. The spectral curve of vegetation obtained from the USGS vegetation spectral library.

The hyperspectral data acquired under real atmospheric conditions usually suffer from noise due to sensor calibration, atmospheric attenuations, illumination, and topographic effects (Gao et al., 2009; Goetz, 2009). The noise introduces uncertainty and degrades the spectral characteristics of materials, which significantly affects the performance of SMMs (Van der Meer, 2006). To assess the impact of noise on the discrimination ability of SMMs, random noise of 10% has been added to the minerals (Figure 3.3) and vegetation spectra of the USGS (Figure 3.4).

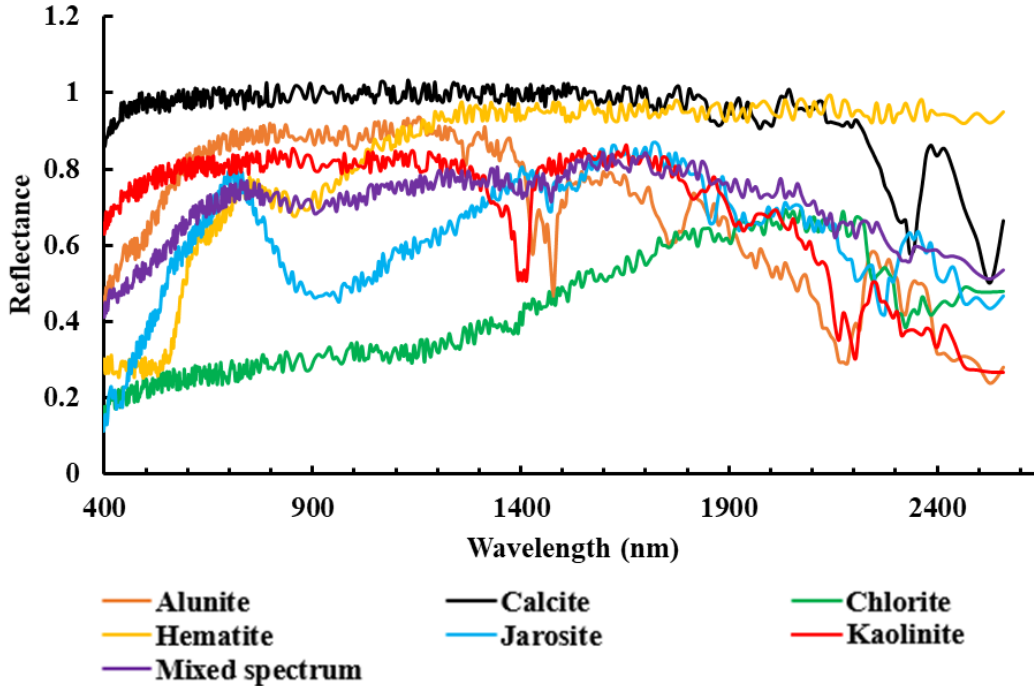


Figure 3.3. The spectral curve of minerals of the USGS mineral spectral library was obtained by adding the 10% of random noise in the spectra.

Spectral mixing is a common phenomenon in remote sensing and causes loss of diagnostics spectral features of material required for successful discrimination and classification using hyperspectral datasets. Spectral mixing causes the mixing of the spectral signature of multiple materials, which results in low intra-class variability of different materials (Clark, 1999). Most of the SMMs and classification techniques suffer due to the intra-pixel mixing of spectral signatures of different targets. We assess the discrimination ability of proposed and existing SMMs with mixed spectra of minerals (Figure 3.5) and mixed vegetation spectra (Figure 3.6). Mixed spectra of minerals obtained by linear mixing of actual mineral class with 50% proportion and other minerals with equal proportion (i.e., 50% of actual class and 8.33% of the remaining six other mineral classes) to avoid complete dilution of the spectral feature of the actual class.

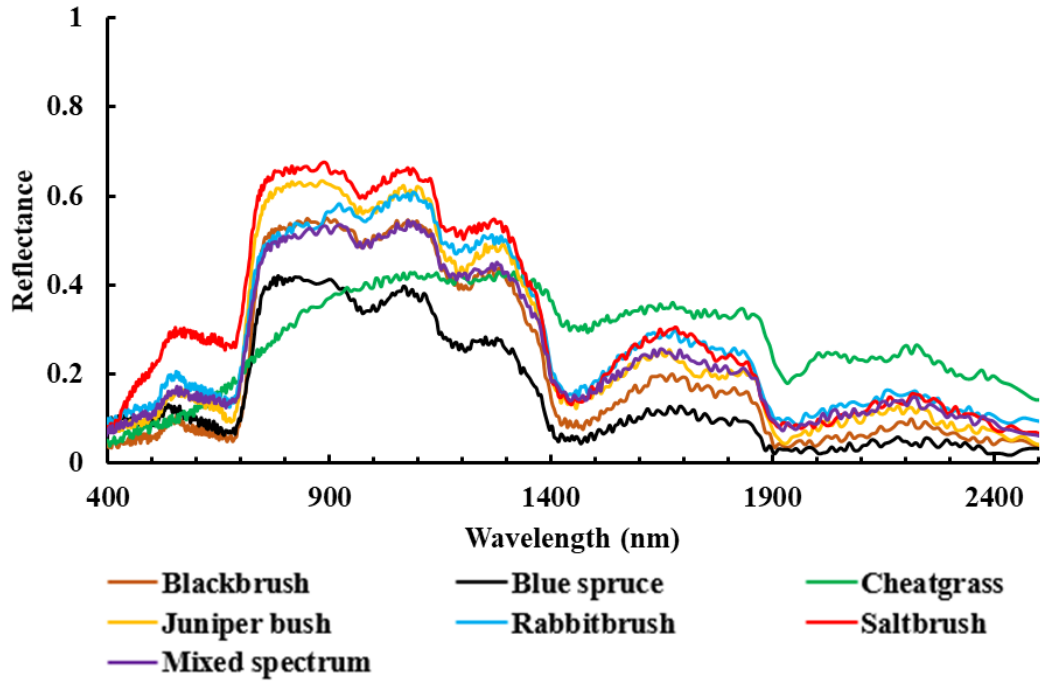


Figure 3.4. The spectral curve of vegetation of USGS vegetation spectral library obtained by adding the 10% of random noise in the spectra.

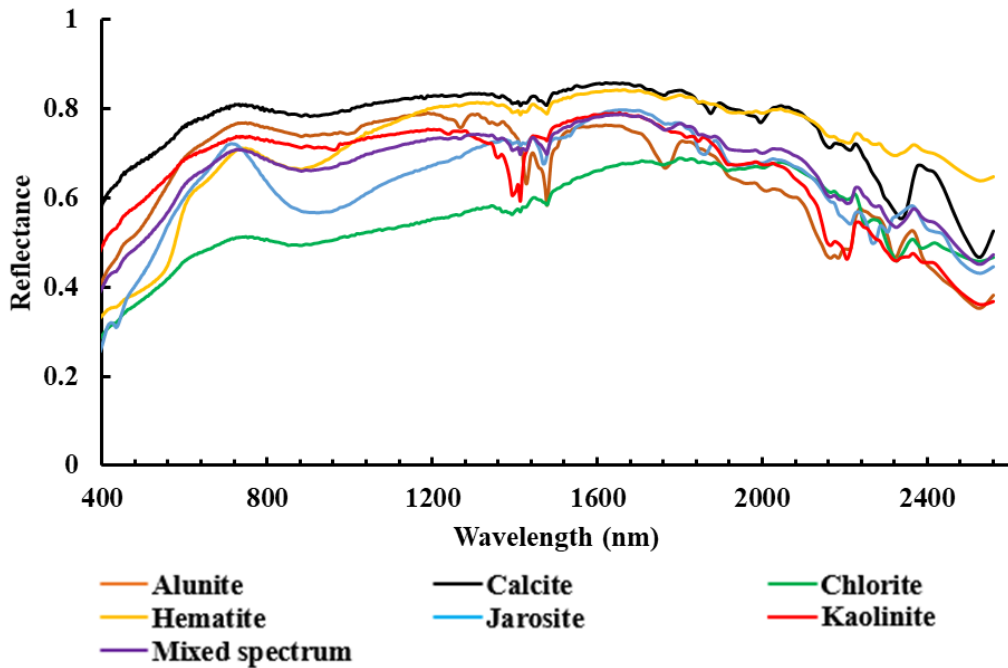


Figure 3.5. Mixed spectra of minerals obtained by linear mixing of actual mineral class and other minerals with equal proportion (i.e., 50% of actual mineral class + 8.33% of other six minerals (i.e., 50%)) using USGS mineral spectral library.

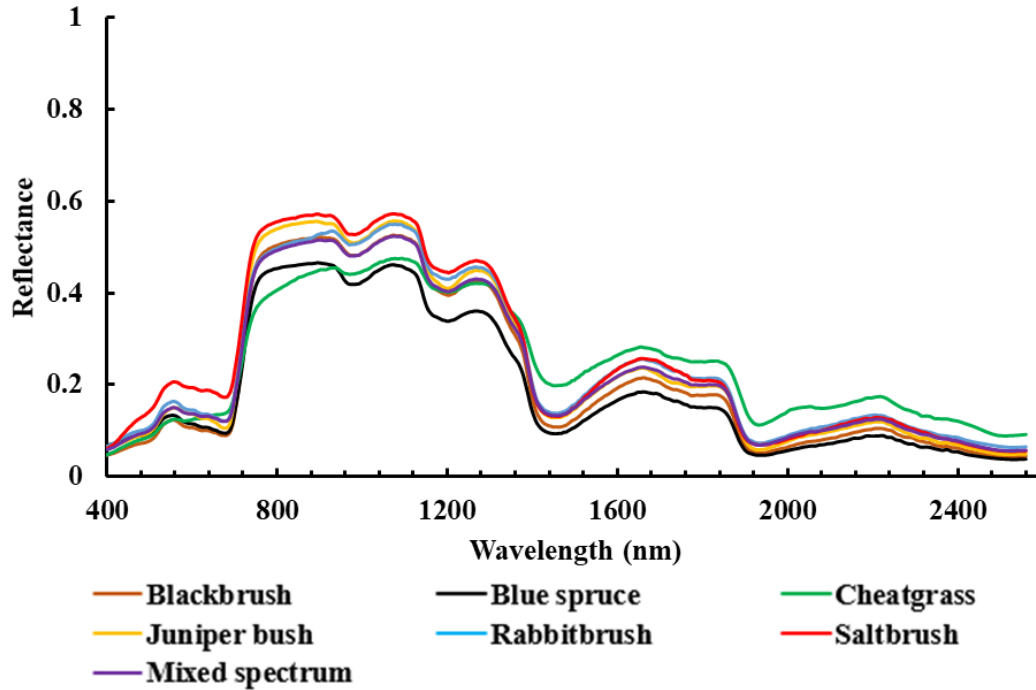


Figure 3.6. Mixed spectra of vegetation obtained by linear mixing of actual vegetation class and other vegetation with equal proportion (i.e., 50% of actual vegetation class + 8.33% of other six vegetation (i.e., 50%)) using USGS vegetation spectral library.

It is necessary to clarify here that the primary purpose of spectral mixing in this study is to induce the spectral characteristics of other classes into the actual class, which decreases the intra-class variability of different spectra and makes spectral discrimination more challenging. The same procedure was used to derive the mixed vegetation spectra.

Furthermore, we used airborne hyperspectral data acquired using a recently developed Airborne Visible/Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) sensor to evaluate the performance of the proposed SMMs. The AVIRIS-NG is an advanced version of the AVIRIS-classic sensor with improved resolution and a higher signal to noise ratio. The data were acquired under a collaborative science mission of the Indian Space Research Organization (ISRO) and National Aeronautics and Space Administration (NASA) in the year 2015/16 and available as level 1 (L1: at sensor radiance) and level 2 (i.e., L2: atmospherically corrected) with geometric correction applied on both the levels. The obtained AVIRIS-NG data of L2 consists of 425 spectral channels in the wavelength range of 380 to 2510 nm at $\sim 5\text{nm}$ and $\sim 7.5\text{m}$ spectral and spatial resolution, respectively. The data were further subjected to removal of uncalibrated bands, where 295 bands were found suitable, and the remaining 130 bands (i.e., 1–15, 146–158, 188–227, 275–328, and 421–425) were excluded due to significant noise. The 295 spectral bands were further subjected to a classical hyperspectral processing system to derive the endmembers. A detailed description of the methodology adopted to derive the hydrothermal alteration minerals (i.e., endmembers) can be referred from our previous work in Kumar et al. (2020a).

The study area is in the part of Hutti-Maski Archean greenstone belt in the eastern Dharwar Craton, south-eastern part of India, which is situated between 16°1'12"–16°24'0" N latitude and 76°36'0"–76°51'0" E longitude (Figure 3.7). The major rock types found in the area are acidic intrusive, amphibolite, granites, metabasalt, and migmatite (Figure 7). The area hosts major gold mineralization and hydrothermal alteration minerals (Hazarika et al., 2015; Kumar et al., 2020a). We used spectra of hydrothermal alteration minerals (i.e., goethite, chlorite, chlorite at the mine site (chlorite mixed with mined materials), kaolinite, muscovite, and mixture (the mixture is derived using linear spectral mixing of all minerals with equal proportion) obtained from our previous work in Kumar et al. (2020a) (Figure 3.8).

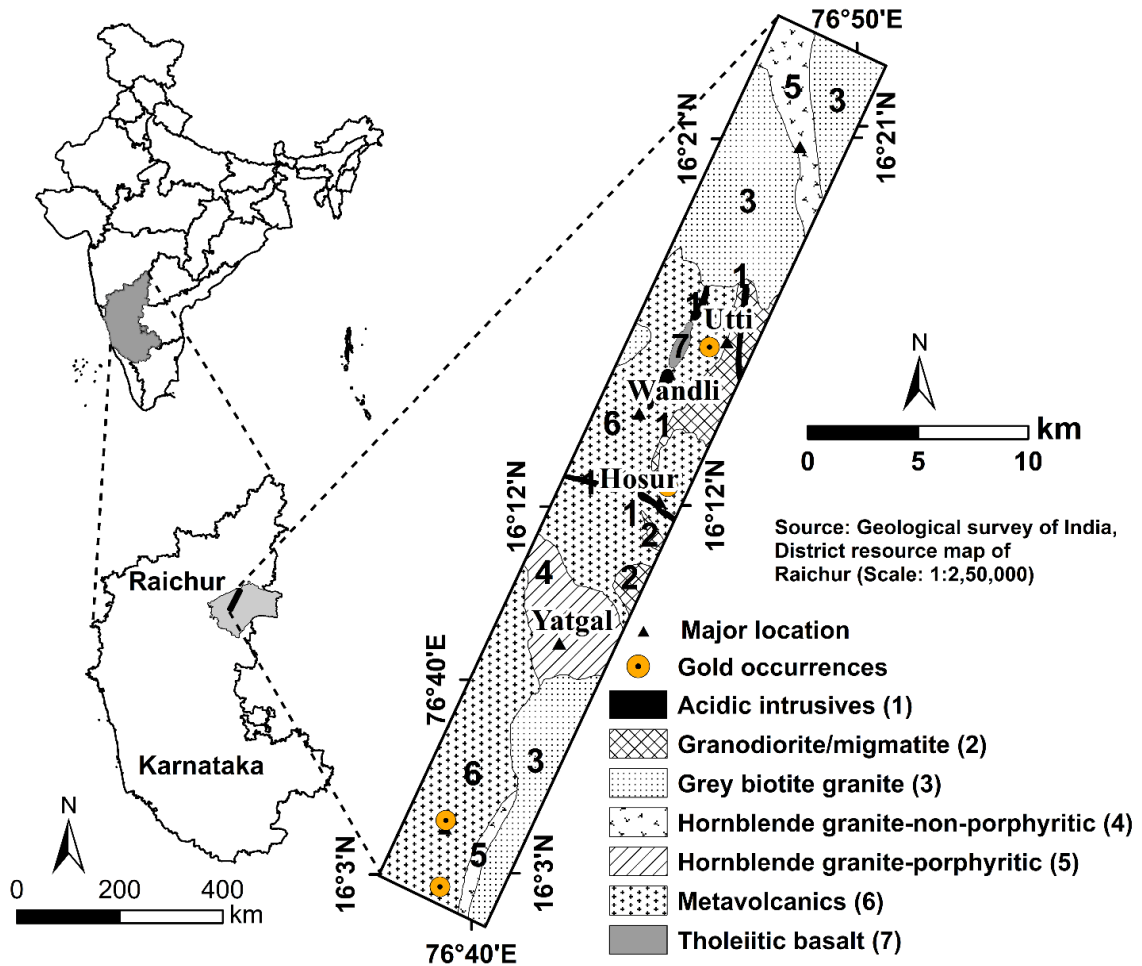


Figure 3.7. Geographical location and geological setting of the area showing major lithological units and gold occurrences reported by the Geological Survey of India.

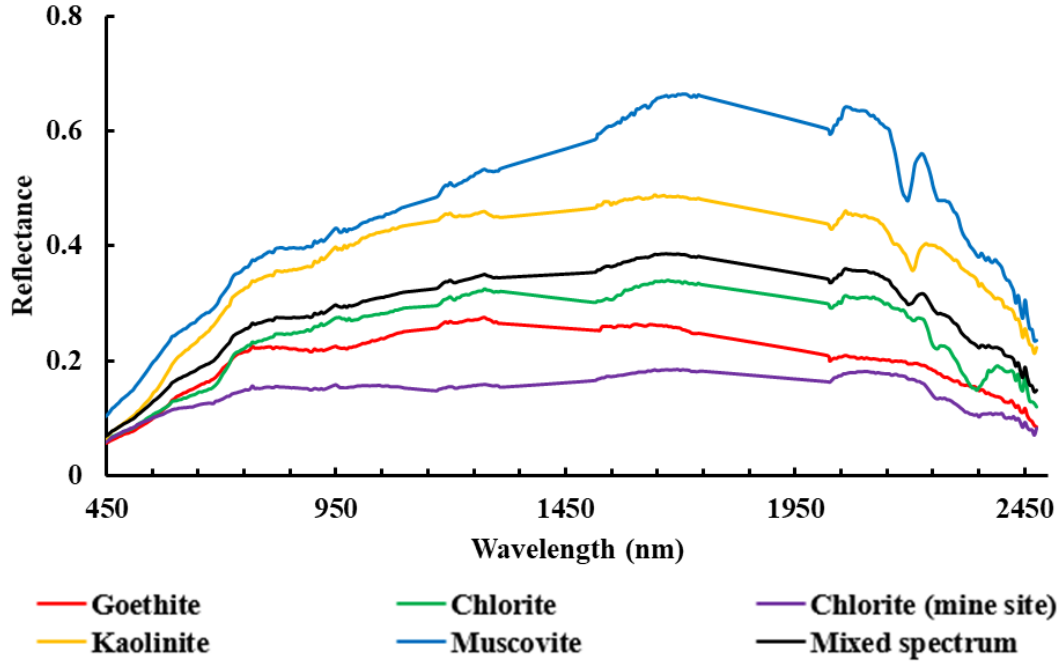


Figure 3.8. The spectra of hydrothermal alteration minerals obtained using AVIRIS-NG airborne hyperspectral data.

3.2.2 Spectral matching measures

A brief mathematical framework of commonly adopted SMMs (such as SAM, SID, and SIDSAM_{tan}), referred hereafter as existing SMMs, and the proposed SMMs (i.e., DSSC, KJSSC, and the hybrid of DSSC and KJSSC, i.e., KJDSSC_{tan}), are discussed in this section.

3.2.2.1 Existing spectral matching measures

- a. Spectral Angle Mapper (SAM): The SAM is a deterministic spectral similarity measure, which computes the spectral angle between known and unknown spectra by treating them as pixel vectors in a space with dimensionality equal to the number of spectral bands (Kruse et al., 1993). The main advantages of SAM are that it suppresses the effects of shadow and illumination effects. The SAM of two spectra $P = (p_1, p_2, \dots, p_L)^T$ and $Q = (q_1, q_2, \dots, q_L)^T$ can be computed using Equation (3.1).

$$SAM(P, Q) = \cos^{-1} \left(\frac{\sum_{l=1}^L p_l q_l}{\left[\sum_{l=1}^L p_l^2 \right]^{\frac{1}{2}} \left[\sum_{l=1}^L q_l^2 \right]^{\frac{1}{2}}} \right) \quad (3.1)$$

where P is reference spectra, Q is unknown spectra, p_l and q_l are reflectance value of spectral band l of spectra P and Q , respectively, L is the number of spectral bands.

- b.** Spectral Information Divergence (SID): The SID calculates the divergence between two spectra (Chang, 1999). The SID is more effective in capturing the subtle spectral variability than SAM (Chang, 2000). The SID of two spectra $P = (p_1, p_2, \dots, p_L)^T$ and $Q = (q_1, q_2, \dots, q_L)^T$ can be computed using Equation (3.2).

$$SID(P, Q) = \sum_{l=1}^L q_l D(p_l \parallel q_l) + \sum_{l=1}^L p_l D(q_l \parallel p_l) \quad (3.2)$$

$$\text{where, } \sum_{l=1}^L q_l D(p_l \parallel q_l) = \sum_{l=1}^L q_l (I_l(q_l) - I_l(p_l)) \quad (3.3)$$

$$\text{and, } \sum_{l=1}^L p_l D(q_l \parallel p_l) = \sum_{l=1}^L p_l (I_l(p_l) - I_l(q_l)) \quad (3.4)$$

$I_l(q_l) = -\log q_l$ and similarly, $I_l(p_l) = -\log p_l$. Measures $I_l(q_l)$ and $I_l(p_l)$ are referred to as the self-information of p_l and q_l for the band l . Equation (3.3) and (3.4) represent the relative entropy of p_l with respect to q_l (indicated with \parallel symbol). In recent years, the SID has gained more appreciation than SAM for spectral discrimination of different materials (van der Meer, 2006; Naresh Kumar et al. 2011; Padma and Sanjeevi, 2014; Kumar et al. 2020a).

- c.** Hybrid of SID and SAM (i.e., $SIDSAM_{tan}$): The cross product of SID and SAM with tangent (tan) trigonometric function provides their hybrid measures (Du et al., 2004). The hybrid approach incorporates the unique characteristics of SAM and SID to improve the discrimination capability. SID can reduce the effects of uncertainty introduced by unknown interferences common in hyperspectral datasets (Du et al., 2004). It can be computed using Equation (3.5).

$$SIDSAM_{tan} = SID \times \tan(SAM) \quad (3.5)$$

3.2.2.2 Proposed spectral similarity measures

We present three effective SMMs, i.e., DSSC, KJSSC, and a hybrid of DSSC and KJSSC, named $KJDSSC_{tan}$, with an enhanced spectral discrimination capability for accurate material discrimination and classification using hyperspectral dataset.

- a.** The DSSC of two spectra can be computed by employing DSC (Dice, 1945). The DSC has been effectively utilized in the field of biological and medical sciences and ecology for image segmentation and measuring the similarity of two molecules and species (Hubálek et al., 1995; McCune et al., 2002; Sampat et al., 2009; Kumar & Zhang, 2018). To the best of our knowledge, DSC's utilization has not been demonstrated for hyperspectral datasets of earth and planetary surface. The DSC computes the inner product between two vectors to measure their similarity (Shanon, 1948; Cha, 2007). It is less sensitive to noise as it gives less weight to outliers and retains the sensitivity in more heterogeneous datasets (McCune et al., 2002). The DSC's similarity value ranges from 0 to 1, where the smaller value indicates a higher match and vice-versa. The DSSC of two spectra $P = (p_1, p_2, \dots, p_L)^T$ and $Q = (q_1, q_2, \dots, q_L)^T$ can be computed using Equation (3.6).

$$DSSC(P, Q) = \frac{2 \sum_{l=1}^L p_l q_l}{\sum_{l=1}^L p_l^2 + \sum_{l=1}^L q_l^2} \quad (3.6)$$

- b.** Kumar-Johnson Spectral Similarity Coefficient (KJSSC): The KJSSC of two spectra can be computed by employing KJSC. The KJSC is a non-parametric symmetric divergence measure of Csiszar's f-divergences and information inequalities (Kumar & Johnson, 2005; Cha, 2007). The KJSC has been successfully used in signal processing for measuring the similarity of the probability distribution (Souza et al., 2017). To the best of our knowledge, the KJSC, as similar to DSC, has not been employed in hyperspectral remote sensing for measuring the similarity between different materials. The KJSC uses arithmetic and geometric mean divergence to calculate the similarity between the given two probability distribution (Kumar & Johnson, 2005). The geometric mean indicates the central tendency computed using their values' products, whereas the arithmetic mean uses their summation. The geometric mean becomes more reliable than the arithmetic mean when the dataset display correlation between variables and contains outliers. In contrast, arithmetic mean provides more accurate results when the datasets are not skewed and show less correlation between variables (Cartwright & Field, 1978). Measuring divergence of two given spectra by incorporating both geometric and arithmetic mean divergence can be an optimal solution to reduce the impact of interference in hyperspectral data resulted due to atmospheric attenuations, sensor calibration and platform disturbances etc. Furthermore, KJSC also incorporates mutual information to reduce the uncertainty of a random variable caused due to the knowledge about other variables (Kumar & Johnson, 2005). The value of KJSSC can range from 0 to infinity, where a value approaches 0 indicates a similar target and greater than 0 indicates a different target. The KJSSC of two spectra $P = (p_1, p_2, \dots, p_L)^T$ and $Q = (q_1, q_2, \dots, q_L)^T$ can be computed using Equation (3.7).

$$KJSSC(P, Q) = \sum_{l=1}^L \left(\frac{(p_l^2 - q_l^2)^2}{2(p_l q_l^{3/2})} \right) \quad (3.7)$$

- c.** Hybrid of KJSSC and DSSC (i.e., KJDSSC_{tan}): An approach of fusing different similarity measures (such as SAM with SID, J-M distance, and others) to develop a hybrid measure for better discrimination and classification of materials have been successfully utilized in the literature (Du et al., 2004; Naresh Kumar et al., 2011; Padma & Sanjeevi, 2014b; Kumar et al., 2020a). We fuse the DSSC and KJSSC to develop a novel hybrid spectral similarity method. The hybrid method uses each measure to improve the discrimination power and reduces the false positive in the identification and classification of spectrally similar materials (Van der Meer, 2006; Naresh Kumar et al., 2011; Kumar et al., 2014; Padma & Sanjeevi, 2014b). We used tangent (*tan*) instead of cosine (*cos*) trigonometric function to combine KJSSC and DSSC. It is important to mention here that *cos* may not be an effective trigonometric function to combine two similarity measures because it computes the projection of one-pixel vector along with the other one, which will reduce or yield less discriminability as compare

with \tan as illustrated in previous studies (Du et al., 2004; Naresh Kumar et al., 2011). The $KJDSSC_{\tan}$ can be obtained using Equation (3.8).

$$KJDSSC_{\tan} = KJSSC \times \tan(DSSC) \quad (3.8)$$

3.3 Computation of SMMs and RSDPW statistics

We computed the spectral similarity matrix for the spectra of minerals, vegetation, mineral spectra with added random noise and vegetation spectra with added random noise, linearly mixed spectra of minerals and linearly mixed spectra of vegetation, and hydrothermal alteration minerals using above discussed SMMs such as SAM (3.1), SID (3.2), SIDSAM_{tan} (3.5), DSSC (3.6), KJSSC (3.7) and $KJDSSC_{\tan}$ (3.8). The computation of these measures was implemented using ‘philtropy’ package available in the R statistical programming language (Drost, 2018).

The spectral similarity values obtained using different SMMs cannot be compared directly, as each method uses a different mathematical framework and measurement units. Chang (2003) introduced a Relative Spectral Discrimination Power (RSDPW), which has been successfully used to compare and evaluate the performance of different SMMs (Du et al., 2004; Van der Meer, 2006; Naresh Kumar et al., 2011; Kumar et al., 2020a). The RSDPW quantifies the ability of SMMs to discriminate and classify a target spectrum (i.e., target pixel vector) from a spectral library or between two spectra with respect to a reference spectrum. Assume $m(.,.)$ is an SMM, $R = (r_1, r_2, \dots, r_L)^T$ is the reference spectrum, $P = (p_1, p_2, \dots, p_L)^T$ and $Q = (q_1, q_2, \dots, q_L)^T$ are the two spectra, the RSDPW of $m(.,.)$, represented by $\Omega^m(P, Q; R)$, can be computed using Equation (3.9).

$$\Omega^m(P, Q; R) = \max \left\{ \frac{m(P,R)}{m(Q,R)}, \frac{m(Q,R)}{m(P,R)} \right\} \quad (3.9)$$

The reference spectrum is usually chosen by considering the spectral similarity of that spectrum with other spectra of different classes. The mixed spectrum can also be obtained by performing a linear spectral mixing using all spectra of different classes (Chang, 2003; Du et al., 2004; Naresh Kumar et al., 2011). We performed a linear spectral mixing using all spectra of different classes with equal proportion to prepare a reference spectrum to be used in the computation of RSDPW statistics to assess the discrimination ability of different SMMs. SMM yielding a higher value of RSDPW statistics indicates a higher capability to accurately discriminate spectra, and vice-versa. The implemented codes and sample data can be available to the scientific community through the GitHub repository (<https://github.com/c-kumar/Spectral-Similarity-Measures>).

3.4 Results and discussion

3.4.1 Spectral similarity analysis

We computed the spectral similarity matrix of various hyperspectral datasets using the six different SMMs. The value of DSSC has been used to illustrate the spectral similarity and intra-class variability of different materials. The spectral similarity matrices of different hyperspectral datasets derived using other SMMs are presented in appendix A. The higher the values of any SMMs, the more the spectra are dissimilar and vice-versa.

The spectra of minerals contain certain similarities with prominent variation, as evident from the absorption position and overall shape of the mineral spectra within the wavelength range of VNIR-SWIR (Figure 3.1). Table 3.1 represents the spectral similarity matrix of minerals using DSSC. The spectral similarity matrix of minerals illustrates that chlorite, hematite, calcite, and jarosite are spectrally dissimilar as they yield a higher value of DSSC than other minerals (i.e., alunite, kaolinite, and muscovite). Similar inferences can be made from the spectral similarity matrix of minerals using other SMMs presented in appendix A. The spectra of vegetation (Figure 3.2) exhibit that most vegetation classes except cheatgrass show absorption position at the same wavelength and similar overall shape. Table 3.2 illustrates that cheatgrass and saltbrush are more dissimilar than other vegetation species or classes.

The noise in the spectral data are common and induce serious challenges for discrimination and classification as they degrade the spectral information embedded in each spectral band. The noise in the spectra can also increase the uncertainty in identifying the absorption position and shape, increasing the chances of false-positive in the identification and classification. The spectra of minerals and vegetation with added random noise are shown in Figures 3.3 and 3.4, respectively. The spectral similarity matrix derived from DSSC for minerals with added noise (Table 3.3) illustrates a slight increment in the similarity of different minerals. The decrement in the similarity value (i.e., DSSC value) indicates the increment in the spectral similarity, which results in low intra-class variability. The spectral similarity value between alunite and chlorite is reduced from 0.3175 to 0.2840 after adding the random noise in the spectra. The spectral similarity value between chlorite and jarosite is reduced from 0.1697 to 0.1490 after adding the noise in the spectra. The spectral similarity between hematite and jarosite is reduced from 0.0514 to 0.0478. Similarly, other minerals have shown a slight reduction in spectral similarity values, which means that the addition of noise in the spectra reduces the intra-class variability. The spectral similarity matrix of vegetation with added random noise (Table 3.4) illustrates that there is a slight increment in the spectral similarity between vegetation classes as displayed by the spectra of minerals. The spectral similarity value between blue spruce and cheatgrass is reduced from 0.1948 to 0.1785 after adding the noise in the spectra. The spectral similarity value between blue spruce and saltbrush is decreased from 0.1967 to 0.1827. The spectral similarity value between the rabbitbrush and saltbrush is decreased from 0.0257 to 0.0248 after adding the noise in the spectral data. Similar effects can be observed with other classes of vegetation (Table 3.4). The addition of random noise causes a decrement in the spectral

similarity value of minerals and vegetation, i.e., the spectra become more similar after inducing the random noise (Tables 3.3 and 3.4).

Material discrimination and classification become more challenging in spectrally mixed materials. We computed the spectral similarity matrix of mixed spectra of minerals (Table 3.5) and mixed spectra of vegetation (Table 3.6). The mixed spectra of minerals and mixed spectra of vegetation are presented in Figures 3.5 and 3.6. Table 3.5 illustrates a significant increment in the spectral similarity of minerals as expected due to linear spectral mixing. For example, the spectral similarity value between alunite and chlorite is significantly reduced from 0.3175 to 0.0477. The spectral similarity value between alunite and kaolinite is decreased from 0.0107 to 0.0021. A similar impact can be observed with other minerals. The spectral similarity matrix of mixed spectra of vegetation (Table 3.6) also exhibits a similar impact in increasing the spectral similarity between different vegetation classes due to linear spectral mixing as observed with mixed spectra of minerals spectra. The spectral similarity value between blue spruce and cheatgrass is significantly decreased from 0.1948 to 0.0275. The spectral similarity value between blue spruce and saltbrush is decreased from 0.1967 to 0.0369. A similar impact can be noticed with other vegetation classes. It is important to notice here that the addition of noise and spectral mixing increases the spectral similarity between different classes and significantly reduces the intra-class variability, and therefore spectral discrimination becomes more challenging.

Furthermore, the spectra of hydrothermal alteration minerals obtained from airborne hyperspectral data (Figure 3.8) show a diverse range of minerals with absorption positions around 950 nm, 2200 nm, and 2350 nm indicating the presence of Fe^{3+} , Al-OH, and Fe/Mg-OH molecules, respectively (Kumar et al., 2020a). These spectra were used to compute the spectral similarity matrix using different SMMs. The spectral similarity matrix (Table 3.7) obtained using DSSC illustrates that chlorite at the mine site (chlorite mixed with mined materials) and muscovite are more dissimilar than other minerals such as goethite, chlorite, and kaolinite.

Table 3.1. Spectral similarity matrix of USGS spectra of minerals using Spectral Angle Mapper (SAM) (in radians). Mixture: linearly mixed spectrum of mineral with equal proportion.

	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.1874	0.5367	0.3946	0.2922	0.1461	0.1878
Calcite	--	0.4280	0.3617	0.2812	0.1224	0.1330
Chlorite		--	0.2652	0.3489	0.5010	0.3564
Hematite			--	0.2384	0.4180	0.2475
Jarosite				--	0.3088	0.1706
Kaolinite					--	0.1796

Table 3.2. Spectral similarity matrix of USGS spectra of vegetation using SAM (in radians). Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rb: rabbit brush, Sb: saltbrush, Ms: linearly mixed spectrum of vegetation with equal proportion.

	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.1742	0.4933	0.1158	0.2008	0.2809	0.1752
Bs	--	0.5585	0.1385	0.2232	0.2090	0.1939
Cg		--	0.4474	0.3496	0.4469	0.3720
Jb			--	0.1285	0.1841	0.0911
Rb				--	0.1601	0.0480
Sb					--	0.1482

Table 3.3. Spectral similarity matrix of USGS mineral spectra after adding 10% of random noise using SAM (in radians).

	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.1794	0.4988	0.3798	0.2799	0.1416	0.1830
Calcite	--	0.3936	0.3452	0.2670	0.1196	0.1296
Chlorite		--	0.2481	0.3255	0.4641	0.3273
Hematite			--	0.2309	0.4006	0.2373
Jarosite				--	0.2942	0.1649
Kaolinite					--	0.1759

Table 3.4. Spectral similarity matrix of USGS vegetation spectra after adding 10% of random noise using SAM.

	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.1788	0.4765	0.1123	0.1928	0.2697	0.1700
Bs	--	0.5321	0.1376	0.2114	0.1930	0.1847
Cg		--	0.4354	0.3386	0.4331	0.3601
Jb			--	0.1277	0.1804	0.0958
Rb				--	0.1583	0.0539
Sb					--	0.1467

Table 3.5. Spectral similarity matrix of mixed spectra of minerals (USGS spectral library) prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e., 50%)) using SAM (units are in radians).

	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0855	0.1777	0.1740	0.1215	0.0641	0.0821
Calcite	--	0.1397	0.1674	0.1232	0.0545	0.0668
Chlorite		--	0.0862	0.1024	0.1642	0.0982
Hematite			--	0.1035	0.1842	0.1096

Jarosite				--	0.1280	0.0664
Kaolinite					--	0.0786

Table 3.6. Spectral similarity matrix of mixed spectra of vegetation (USGS spectral library) prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of different classes (i.e., 50%)) using SAM (in radians).

	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0635	0.1987	0.0451	0.0821	0.1234	0.0705
Bs	--	0.2104	0.0423	0.0769	0.0817	0.0635
Cg		--	0.1833	0.1397	0.1877	0.1495
Jb			--	0.0572	0.0869	0.0406
Rb				--	0.0766	0.0212
Sb					--	0.0715

Table 3.7. Spectral similarity matrix of hydrothermal alteration minerals using SAM (in radians).

	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	0.1154	0.1360	0.1228	0.1933	0.1174
Chlorite	--	0.1334	0.0791	0.1116	0.0544
Chlorite _(mine site)		--	0.1418	0.1603	0.1134
Kaolinite			--	0.1004	0.0463
Muscovite				--	0.0797

3.4.2 Performance evaluation of SMMs using RSDPW statistics

The RSDPW statistics (Eqn. 9) of existing and proposed SMMs were computed to compare and evaluate their discrimination ability using seven different hyperspectral datasets. The SMM yields a higher RSDPW value has more discrimination ability and can produce a more accurate classification of materials. Furthermore, the effects of noise and linear spectral mixing on the performance of SMMs have also been analyzed to evaluate the novelty of proposed SMMs.

3.4.2.1 USGS spectra of minerals

The RSDPW statistics of different SMMs of USGS spectra of minerals have been presented in Table 3.8 and Figure 3.9. The RSDPW of SAM and SID for alunite and calcite are 1.41 and 2.50, respectively, which is lesser than the proposed DSSC and KJSSC. To describe the computation of RSDPW statistics, let us consider the example of two spectra, i.e., alunite (P), calcite (Q), to be discriminated using DSSC with respect to a reference spectrum, i.e., mixture (R). The DSSC value of alunite and mixture and calcite and mixture

are 0.0210 and 0.0645, respectively (Table 3.1). The RSDPW can be calculated as, $\max\left(\frac{DSSC(P \text{ and } R)}{DSSC(Q \text{ and } R)}, \frac{DSSC(Q \text{ and } R)}{DSSC(P \text{ and } R)}\right) = \max\left(\frac{0.0210}{0.0645}, \frac{0.0645}{0.0210}\right) = 3.07$. The RSDPW of DSSC and KJSSC for alunite and calcite are 3.07 and 3.59, which is around 1.5 to 2 times higher than SAM and SID. The hybrid of SIDSAM_{tan} shows similar discrimination power (i.e., 3.56) as DSSC and KJSSC but significantly less than the developed hybrid method, i.e., KJDSSC_{tan}. The RSDPW of KJDSSC_{tan} for discriminating alunite and calcite is 11.02, which is around 3–4 times higher than SIDSAM_{tan}. Similarly, the RSDPW of SAM and SID for discriminating alunite and chlorite are 1.90 and 2.86, respectively. The RSDPW of DSSC and KJSSC for alunite and chlorite are 10.39 and 11.09, which is around 5 times higher than SAM and SID. The developed hybrid measure, i.e., KJDSSC_{tan} shows significantly higher discrimination power (i.e., 117.10) than SIDSAM_{tan} (i.e., 5.60) for discriminating alunite and chlorite. The KJDSSC_{tan} shows around 20 times higher discrimination power to discriminate alunite and chlorite. It can also be noticed that, in some instances, the proposed SMMs, i.e., DSSC and KJSSC, outperform the existing hybrid method SIDSAM_{tan}. Figure 3.9 and Table 3.8 illustrate that the developed SMMs outperform existing SMMs in most of the cases. In a worst-case scenario, the proposed SMMs show similar to slightly less ability to discriminate the spectra of minerals as compared with existing SMMs. The RSDPW of SAM and SID for discrimination between calcite and hematite is 1.86 and 4.30, which is slightly higher than DSSC (i.e., 1.79) and KJSSC (i.e., 1.75). It has also been observed that the discrimination ability of all SMMs appeared similar when making discrimination between alunite and kaolinite.

In the developed SMMs, it has been observed that on a few occasions, KJSSC outperforms DSSC and vice-versa. For example, the RSDPW of KJSSC is slightly higher than DSSC for discriminating alunite and chlorite, whereas the DSSC yields higher RSDPW for discriminating jarosite and chlorite as compared with KJSSC. However, the hybrid of DSSC and KJSSC, i.e., KJDSSC_{tan}, shows significantly higher discrimination ability than the individual measures for discriminating most of the minerals. It has also been observed that the discrimination power of KJDSSC_{tan} is significantly higher than the KJSSC and DSSC, as observed while discriminating between jarosite and chlorite and chlorite and kaolinite. This justifies the requirement of the development of novel hybrid SMM like KJDSSC_{tan} for accurate spectral discrimination using the hyperspectral dataset.

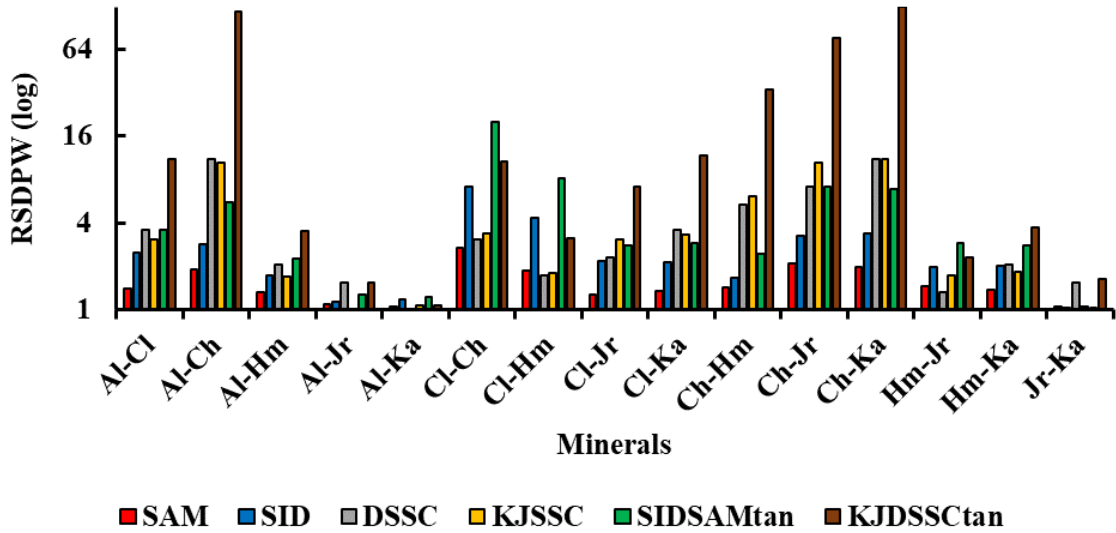


Figure 3.9. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the USGS spectra of minerals.

Table 3.8. RSDPW statistics of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for USGS spectra of minerals.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	1.41	1.90	1.32	1.10	1.05
Calcite	--	2.68	1.86	1.28	1.35
Chlorite		--	1.44	2.09	1.98
Hematite			--	1.45	1.38
Jarosite				--	1.05
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	2.50	2.86	1.72	1.15	1.18
Calcite	--	7.16	4.30	2.18	2.13
Chlorite		--	1.66	3.28	3.36
Hematite			--	1.97	2.02
Jarosite				--	1.02
SIDSAM_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.56	5.60	2.29	1.27	1.23
Calcite	--	19.92	8.13	2.81	2.89
Chlorite		--	2.45	7.10	6.90
Hematite			--	2.89	2.81
Jarosite				--	1.03
DSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.07	10.39	1.71	1.01	1.07

Calcite	--	3.39	1.79	3.10	3.29
Chlorite		--	6.07	10.50	11.15
Hematite			--	1.73	1.84
Jarosite				--	1.06
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.59	11.09	2.05	1.55	1.01
Calcite	--	3.09	1.75	2.32	3.57
Chlorite		--	5.40	7.15	11.02
Hematite			--	1.33	2.04
Jarosite				--	1.54
KJDSSC_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	11.02	117.10	3.52	1.53	1.07
Calcite	--	10.62	3.14	7.19	11.76
Chlorite		--	33.30	76.35	124.86
Hematite			--	2.29	3.75
Jarosite				--	1.64

3.4.2.2 USGS spectra of vegetation

The RSDPW of all SMMs for the USGS spectra of vegetation is presented in Figure 3.10 and Table 3.9. The RSDPW statistics of SAM and SID for discriminating blackbrush and blue spruce are 1.11 and 2.50, whereas the RSDPW of DSSC and KJSSC are 5.09 and 2.52, respectively. The KJDSSC_{tan} yields around 5 times higher RSDPW than SIDSAM_{tan} for discriminating blackbrush and blue spruce. Similarly, the RSDPW of DSSC, KJSSC, and KJDSSC_{tan} are significantly higher than the RSDPW of SAM, SID, and SIDSAM_{tan} for making discrimination between blue spruce and juniper bush. The KJDSSC_{tan} yields around 56 times higher RSDPW than SIDSAM_{tan} for making discrimination between blue spruce and juniper bush. Figure 3.10 and Table 3.9 illustrate that the proposed SMMs outperform previous SMMs to discriminate between most vegetation classes. However, there were a few occasions when existing SMMs show slightly better or similar discrimination ability as compared to proposed SMMs.

It has been observed that the existing SMMs perform reasonable discrimination where the intra-class variability is higher but shows less discrimination ability when dealing with spectrally similar targets. It has also been noticed that all SMMs show a similar ability to discriminate when the intra-class variability is significantly high. For example, the SAM, SID, and SIDSAM_{tan} offer similar discrimination ability than developed SMMs when discriminating spectrally dissimilar classes, such as cheatgrass and blue spruce. However, the discrimination ability of developed SMMs is significantly higher than existing SMMs for discriminating most of the vegetation classes.

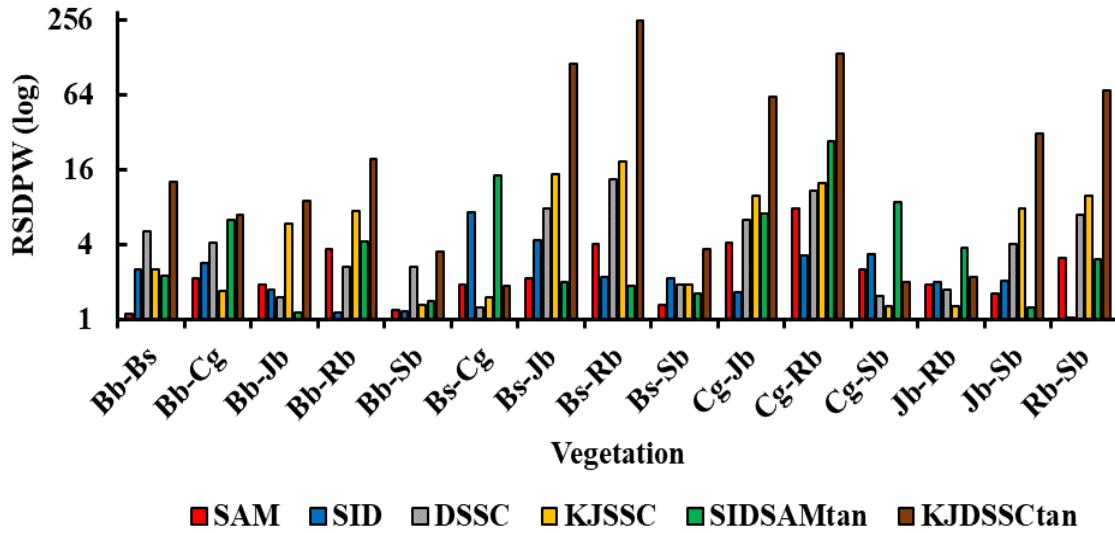


Figure 3.10. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the spectra of vegetation derived from the USGS spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rb: rabbitbrush, and Sb: saltbush.

Table 3.9. RSDPW of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for USGS spectra of vegetation. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rb: rabbitbrush, Sb: saltbush.

SAM	Bs	Cg	Jb	Rb	Sb
Bb	1.11	2.12	1.92	3.65	1.18
Bs	--	1.92	2.13	4.04	1.31
Cg		--	4.08	7.74	2.51
Jb			--	1.90	1.63
Rb				--	3.09
SID	Bs	Cg	Jb	Rb	Sb
Bb	2.50	2.86	1.72	1.15	1.18
Bs	--	7.16	4.30	2.18	2.13
Cg		--	1.66	3.28	3.36
Jb			--	1.97	2.02
Rb				--	1.02
SIDSAM_{tan}	Bs	Cg	Jb	Rb	Sb
Bb	2.26	6.30	1.13	4.23	1.39
Bs	--	14.23	2.00	1.87	1.62
Cg		--	7.10	26.64	8.79
Jb			--	3.75	1.24

Rb				--	3.03
DSSC	Bs	Cg	Jb	Rb	Sb
Bb	5.09	4.12	1.52	2.62	2.66
Bs	--	1.23	7.72	13.32	1.92
Cg		--	6.26	10.80	1.55
Jb			--	1.73	4.03
Rb				--	6.95
KJSSC	Bs	Cg	Jb	Rb	Sb
Bb	2.52	1.68	5.83	7.41	1.32
Bs	--	1.50	14.66	18.63	1.90
Cg		--	9.78	12.43	1.27
Jb			--	1.27	7.70
Rb				--	9.79
KJDSSC_{tan}	Bs	Cg	Jb	Rb	Sb
Bb	12.83	6.93	8.84	19.39	3.51
Bs	--	1.85	113.48	248.85	3.65
Cg		--	61.31	134.44	1.97
Jb			--	2.19	31.05
Rb				--	68.09

3.4.2.3 USGS spectra of minerals and vegetation with added noise

The RSDPW of all SMMs for the spectra of minerals and vegetation classes with added noise are presented in Figure 3.11 and Table 3.10, and Figure 3.12 and Table 3.11, respectively. The addition of random noise increases the uncertainty and causes a decrement in intra-class variability, making spectral discrimination more challenging. The addition of noise components in the spectra of minerals and vegetation classes causes a slight decrement in the RSDPW statistics for all SMMs. The addition of noise results in increasing the spectral similarity, i.e., two similar spectra become more similar. The RSDPW statistics of proposed SMMs are significantly higher than the existing SMMs for both minerals and vegetation spectral data with an added noise component. For example, the discrimination ability of proposed individual SMMs, i.e., DSSC and KJSSC, is around 4 times higher than SAM and SID for discriminating alunite and chlorite after adding the noise (Table 3.10). The proposed hybrid method's discrimination ability, i.e., KJDSSC_{tan}, is around 21 times higher than SIDSAM_{tan} for discriminating alunite and chlorite (Table 3.10).

A similar performance has been noted for the spectra of vegetation with added random noise. The RSDPW statistics of DSSC and KJSSC are around 3 times higher than SAM and SID for discriminating juniper bush and blue spruce (Table 3.11). The discrimination ability of KJDSSC_{tan} is around 8 times higher than SIDSAM_{tan} for discriminating juniper

bush and blue spruce (Table 3.11). It has also been observed that there was hardly an instance when the proposed SMMs fail to distinguish minerals and vegetation spectra due to the addition of noise in the hyperspectral data.

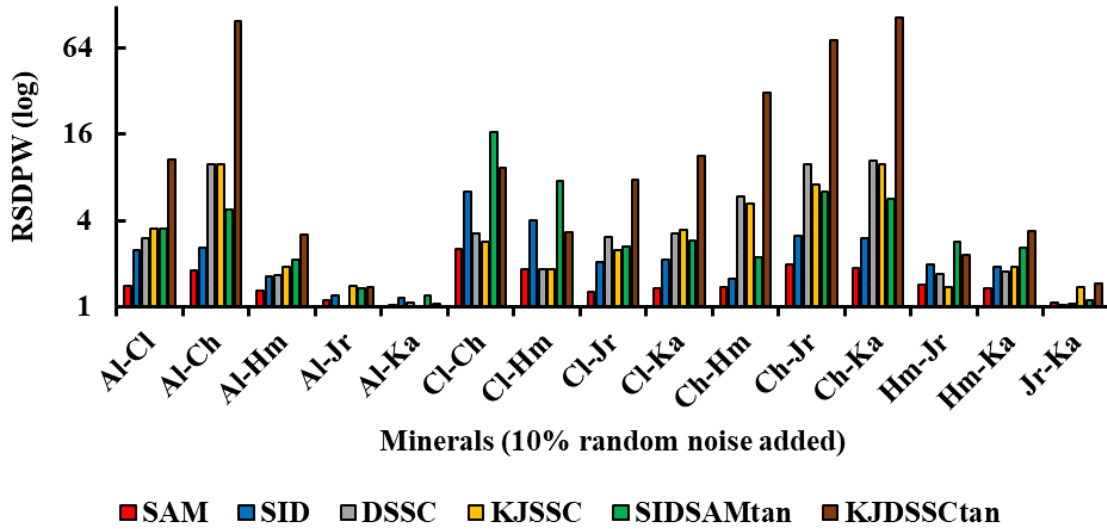


Figure 3.11. Graphical representation of RSDPW statistics of different spectral similarity measures derived using 10% of random noise added in the spectra of minerals derived from the USGS spectral library.

Table 3.10. RSDPW statistics of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for USGS minerals spectra after adding 10% of random noise.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	1.41	1.79	1.30	1.11	1.04
Calcite	--	2.53	1.83	1.27	1.36
Chlorite		--	1.38	1.98	1.86
Hematite			--	1.44	1.35
Jarosite				--	1.07
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	2.47	2.58	1.63	1.21	1.16
Calcite	--	6.37	4.04	2.05	2.13
Chlorite		--	1.58	3.11	2.99
Hematite			--	1.97	1.90
Jarosite				--	1.04
SIDSAM_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.51	4.73	2.14	1.34	1.21

Calcite	--	16.61	7.50	2.62	2.90
Chlorite		--	2.22	6.34	5.72
Hematite			--	2.86	2.58
Jarosite				--	1.11
DSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.04	9.80	1.67	1.01	1.07
Calcite	--	3.23	1.82	3.07	3.24
Chlorite		--	5.87	9.92	10.46
Hematite			--	1.69	1.78
Jarosite				--	1.05
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.50	9.96	1.91	1.40	1.01
Calcite	--	2.84	1.83	2.50	3.48
Chlorite		--	5.21	7.12	9.89
Hematite			--	1.37	1.90
Jarosite				--	1.39
KJDSSC_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	10.64	98.74	3.18	1.38	1.06
Calcite	--	9.28	3.34	7.70	11.28
Chlorite		--	31.00	71.50	104.68
Hematite			--	2.31	3.38
Jarosite				--	1.46

Furthermore, Figures 3.11 and 3.12 present the discrimination power of different SMMs and illustrate that the proposed SMMs outperform the existing SMMs for discriminating most of the minerals and vegetation classes, as displayed for spectra of minerals and vegetation without adding the noise component.

3.4.2.4 USGS spectra of mixed minerals and mixed vegetation

The spectral discrimination of spectrally similar targets commonly occurred due to spectral mixing. We computed the RSDPW of all SMMs using the mixed spectra of mineral (Figure 3.13 and Table 3.12) and mixed spectra of vegetation (Figure 3.14 and Table 3.13) to evaluate the novelty of the proposed SMMs. The proposed SMMs outperform existing SMMs (Figure 3.13 and 3.14).

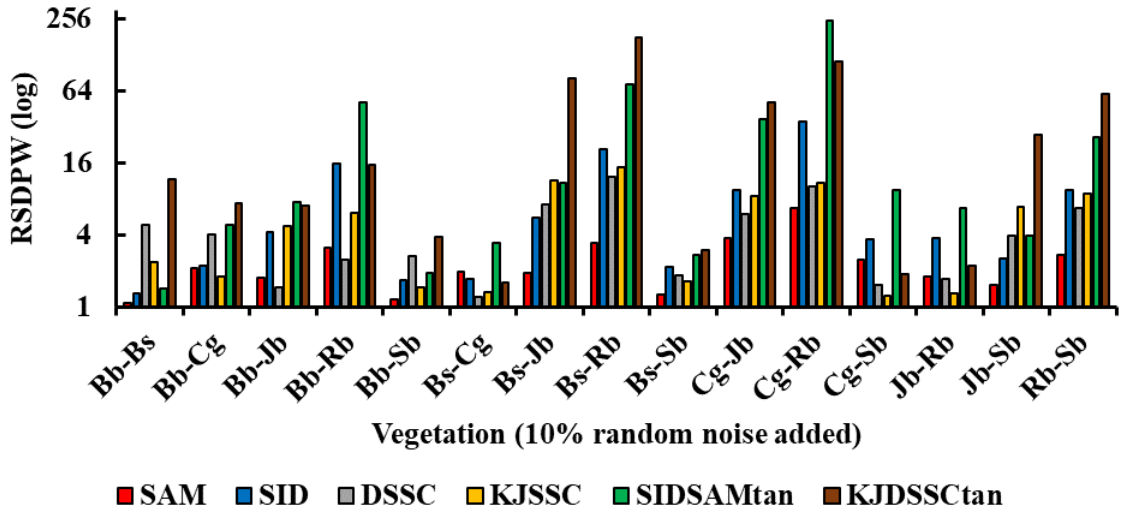


Figure 3.12. Graphical representation of RSDPW statistics of different spectral similarity measures derived using 10% of random noise added in the spectra of vegetation derived from the USGS spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rs: rabbitbrush, and Sb: saltbush.

Table 3.11. RSDPW statistics of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for USGS vegetation spectra after adding 10% of random noise.

SAM	Bs	Cg	Jb	Rb	Sb
Bb	1.09	2.12	1.78	3.15	1.16
Bs	--	1.95	1.93	3.43	1.26
Cg		--	3.76	6.68	2.46
Jb			--	1.78	1.53
Rb				--	2.72
SID	Bs	Cg	Jb	Rb	Sb
Bb	1.30	2.21	4.25	15.88	1.67
Bs	--	1.70	5.54	20.69	2.17
Cg		--	9.39	35.09	3.69
Jb			--	3.74	2.55
Rb				--	9.51
SIDSAM_{tan}	Bs	Cg	Jb	Rb	Sb
Bb	1.42	4.84	7.60	50.54	1.94
Bs	--	3.42	10.77	71.62	2.75
Cg		--	36.82	244.85	9.40
Jb			--	6.65	3.92
Rb				--	26.04

DSSC	Bs	Cg	Jb	Rb	Sb
Bb	4.88	4.05	1.47	2.51	2.66
Bs	--	1.21	7.16	12.24	1.83
Cg		--	5.95	10.15	1.52
Jb			--	1.71	3.91
Rb				--	6.67
KJSSC	Bs	Cg	Jb	Rb	Sb
Bb	2.39	1.80	4.75	6.11	1.46
Bs	--	1.33	11.36	14.63	1.64
Cg		--	8.52	10.98	1.23
Jb			--	1.29	6.93
Rb				--	8.92
KJDSSC_{tan}	Bs	Cg	Jb	Rb	Sb
Bb	11.69	7.28	6.98	15.34	3.88
Bs	--	1.61	81.54	179.34	3.01
Cg		--	50.75	111.63	1.87
Jb			--	2.20	27.10
Rb				--	59.60

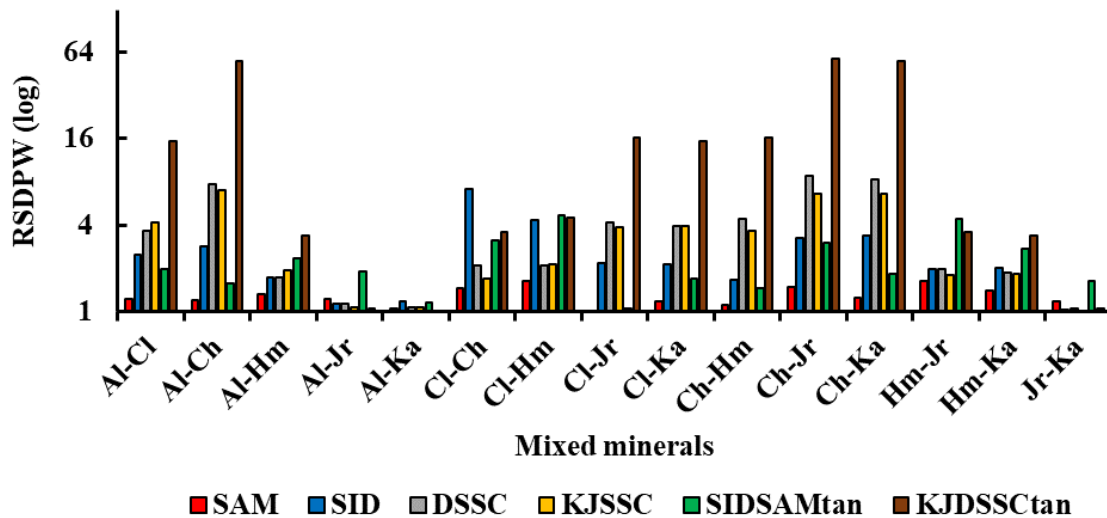


Figure 3.13. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the mixed spectra of minerals prepared by linear mixing of other classes with equal proportion (50% of actual mineral class + 8.33% of six other mineral classes (i.e., 50%)) using USGS mineral spectral library.

Table 3.12. RSDPW of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for mixed spectra of minerals prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e. 50%)) using linear spectral mixing.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	1.23	1.20	1.33	1.24	1.05
Calcite	--	1.47	1.64	1.01	1.18
Chlorite		--	1.12	1.48	1.25
Hematite			--	1.65	1.40
Jarosite				--	1.18
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	2.50	2.86	1.72	1.15	1.18
Calcite	--	7.16	4.30	2.18	2.13
Chlorite		--	1.66	3.28	3.36
Hematite			--	1.97	2.02
Jarosite				--	1.02
SIDSAM_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	1.99	1.59	2.33	1.89	1.16
Calcite	--	3.15	4.64	1.05	1.71
Chlorite		--	1.47	3.00	1.85
Hematite			--	4.41	2.72
Jarosite				--	1.62
DSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	3.66	7.75	1.74	1.13	1.07
Calcite	--	2.11	2.10	4.15	3.93
Chlorite		--	4.45	8.78	8.31
Hematite			--	1.97	1.87
Jarosite				--	1.06
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	4.19	7.06	1.94	1.08	1.07
Calcite	--	1.69	2.16	3.89	3.93
Chlorite		--	3.63	6.56	6.63
Hematite			--	1.81	1.82
Jarosite				--	1.01
KJDSSC_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite
Alunite	15.35	54.71	3.38	1.05	1.01
Calcite	--	3.57	4.53	16.16	15.46
Chlorite		--	16.16	57.63	55.11
Hematite			--	3.57	3.41
Jarosite				--	1.05

There were many instances when the existing SMMs fails to make reliable discriminations between different minerals but proposed SMMs show a better ability to discriminate them with higher confidence (Figure 3.13). The RSDPW of SAM and SID are 1.23 and 2.50, respectively, for distinguishing alunite and calcite, whereas the RSDPW of DSSC and KJSSC are 3.66 and 4.19, respectively, which is around 2–3 times higher than existing SMMs (Table 3.12).

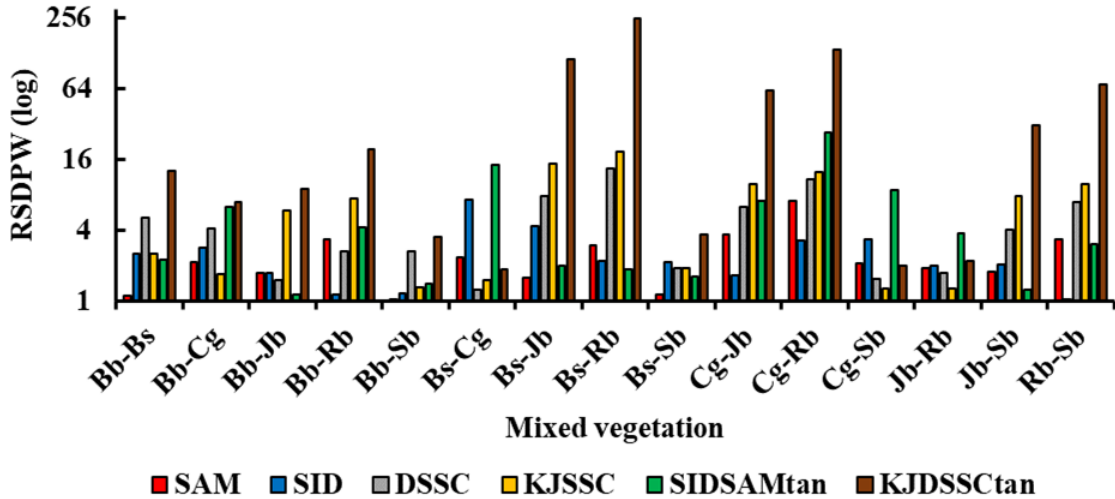


Figure 3.14. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the mixed spectra of vegetation prepared by linear mixing of other classes with equal proportion (50% of actual vegetation class + 8.33% of six other vegetation classes (i.e., 50%)) using USGS vegetation spectral library. Bb: blackbrush, Bs: blue spruce, Cg: cheatgrass, Jb: juniper bush, Rs: rabbitbrush, and Sb: saltbush.

Similarly, the RSDPW statistics of DSSC and KJSSC are around 3–4 times higher than SAM and SID for discriminating alunite and chlorite. The SAM and SID show low ability to distinguish calcite and jarosite, whereas the DSSC and KJSSC show better discrimination ability. The RSDPW of DSSC and KJSSC is around 4 times higher than SAM and SID for discriminating calcite and jarosite. The $KJDSSC_{tan}$ offers a significantly higher ability to discriminate alunite and chlorite, and many other minerals. The RSDPW of $SIDSAM_{tan}$ is 1.85, whereas the $KJDSSC_{tan}$ yields RSDPW statistics of 55.11, which is around 30 times higher. Similarly, the $KJDSSC_{tan}$ shows significantly higher discrimination ability than $SIDSAM_{tan}$ while discriminating against other minerals such as calcite and chlorite, calcite and alunite, chlorite and jarosite, and others.

The proposed SMMs exhibit a similar ability of spectral discrimination of mixed vegetation spectra, as illustrated for mixed mineral spectra. The discrimination of mixed vegetation becomes more challenging due to their similar overall shape and overlapped absorption position. The previous SMMs fail to make satisfactory discrimination of mixed vegetation classes, whereas the proposed SMMs successfully discriminate them with significantly higher RSDPW for most vegetation classes. The RSDPW statistics of SAM and SID for blackbrush and blue spruce are 1.11 and 1.02, which is practically not feasible to

discriminate, whereas the RSDPW statistics of DSSC and KJSSC are 4.40 and 2.14, which is around 2–4 times higher. The $SIDSAM_{tan}$ also shows poor spectral discrimination ability for blackbrush and blue spruce as it yields a low RSDPW, i.e., 1.09, which is significantly less as compared with $KJDSSC_{tan}$ (i.e., 9.42). Similarly, the spectral discrimination ability of $KJDSSC_{tan}$ is higher than $SIDSAM_{tan}$ for discriminating blackbrush and cheatgrass, blackbrush and saltbrush, and blue spruce and rabbitbrush, and others. It has also been observed that on a few occasions, existing SMMs show similar or slightly better spectral discrimination ability as compared with newly proposed SMMs, as seen when discriminating cheatgrass and rabbitbrush. It is also noticed that there has rarely been any instance where the proposed SMMs show lesser discrimination capability than the existing SMMs (as seen while discriminating blue spruce and cheatgrass). The proposed SMMs outperform the existing SMMs for most of the vegetation classes.

Table 3.13. RSDPW statistics of SAM, SID, $SIDSAM_{tan}$, DSSC, KJSSC, and $KJDSSC_{tan}$ for mixed spectra of vegetation (prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e., 50%)) using the linearly mixed spectrum of vegetation as a reference.

SAM	Bs	Cg	Jb	Rb	Sb
Bb	1.11	2.12	1.74	3.32	1.01
Bs	--	2.36	1.56	2.99	1.13
Cg		--	3.68	7.04	2.09
Jb			--	1.91	1.76
Rb				--	3.37
SID	Bs	Cg	Jb	Rb	Sb
Bb	1.02	3.26	3.80	17.46	1.09
Bs	--	3.21	3.87	17.74	1.08
Cg		--	12.42	57.00	2.99
Jb			--	4.59	4.16
Rb				--	19.09
$SIDSAM_{tan}$	Bs	Cg	Jb	Rb	Sb
Bb	1.09	6.92	6.60	57.98	1.11
Bs	--	7.57	6.03	53.02	1.21
Cg		--	45.66	401.24	6.24
Jb			--	8.79	7.31
Rb				--	64.27
DSSC	Bs	Cg	Jb	Rb	Sb
Bb	4.40	4.18	1.37	2.39	3.27
Bs	--	1.05	6.02	10.49	1.35
Cg		--	5.72	9.97	1.28
Jb			--	1.74	4.47

Rb				--	7.79
KJSSC	Bs	Cg	Jb	Rb	Sb
Bb	2.14	2.31	4.10	5.07	2.01
Bs	--	1.08	8.78	10.86	1.07
Cg		--	9.47	11.72	1.15
Jb			--	1.24	8.23
Rb				--	10.18
KJDSSC_{tan}	Bs	Cg	Jb	Rb	Sb
Bb	9.42	9.65	5.61	12.10	6.55
Bs	--	1.02	52.84	113.96	1.44
Cg		--	54.15	116.79	1.47
Jb			--	2.16	36.76
Rb				--	79.29

3.4.2.5 Hydrothermal alteration minerals

We computed and evaluated the performance of SMMs using hydrothermal alteration minerals (i.e., goethite, chlorite, chlorite at the mine site (chlorite mixed with mined materials), kaolinite and muscovite) derived from AVIRIS-NG hyperspectral data (Figure 3.15 and Table 3.14). The proposed SMMs show a better ability to discriminate hydrothermal alteration minerals as compared to existing SMMs. The RSDPW of proposed SMMs is significantly higher than existing SMMs for most hydrothermal alteration minerals (Figure 3.14). The RSDPW statistics of SAM and SID are 2.16 and 3.38, whereas DSSC and KJSSC yield RSDPW of 6 and 5.17, respectively, and these values are 2–3 times higher for discriminating goethite and chlorite (Table 3.14). The RSDPW of SIDSAM_{tan} is 7.31, whereas the RSDPW of KJDSSC_{tan} is 31.08, 4–5 times higher, for discriminating goethite and chlorite. The RSDPW of SAM and SID is 1.46 and 1.44, whereas the DSSC and KJSSC yield the RSDPW of 10.51 and 13.86, which exhibit around 10 times higher ability to discriminate the chlorite and muscovite. The SIDSAM_{tan} also shows the poor ability to discriminate chlorite and muscovite as it yields RSDPW of 2.10, whereas the KJDSSC_{tan} demonstrates significantly higher RSDPW (i.e., 146.18) which exhibits around 70 times higher discrimination power (Table 3.14). Similarly, the discrimination ability of DSSC and KJSSC are significantly better than SAM and SID for several hydrothermal minerals, as illustrated in Figure 3.15 and Table 3.14. It has also been noticed that in several instances (such as when making discrimination between goethite and chlorite (mine site), chlorite and kaolinite, chlorite, and muscovite), the proposed individual SMMs outperform the previous hybrid SIDSAM_{tan}. There were only two instances (i.e., while discriminating goethite and kaolinite, and chlorite (at mine site) and muscovite) when previous SMMs show slightly better discrimination ability than proposed SMMs. Furthermore, the proposed SMMs show a better ability of discrimination for spectrally similar targets, which were not observed with previous SMMs.

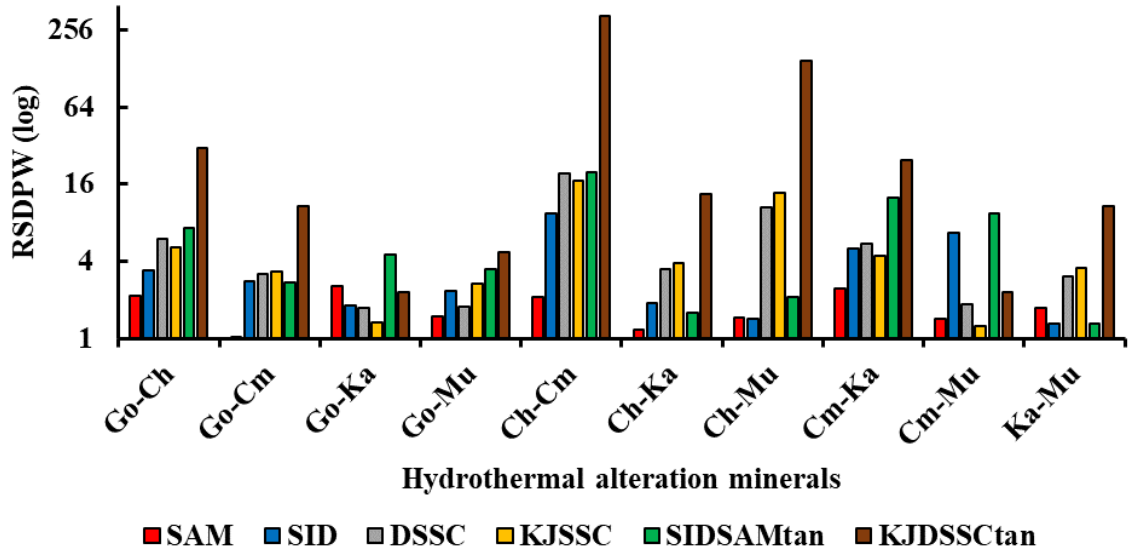


Figure 3.15. Graphical representation of RSDPW statistics of different spectral similarity measures derived using the spectra of hydrothermal alteration minerals of AVIRIS-NG hyperspectral data. Go: goethite, Ch: chlorite, Cm: chlorite at the mine site (a mixture of chlorite and mined materials), Ka: kaolinite, and Mu: muscovite.

Table 3.14. RSDPW statistics of SAM, SID, SIDSAM_{tan}, DSSC, KJSSC, and KJDSSC_{tan} for hydrothermal alteration minerals using the linearly mixed spectrum of all minerals with equal proportion as a reference.

SAM	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite
Goethite	2.16	1.03	2.54	1.47
Chlorite	--	2.08	1.18	1.46
Chlorite _(mine site)		--	2.45	1.42
Kaolinite			--	1.72
SID	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite
Goethite	3.38	2.83	1.79	2.35
Chlorite	--	9.54	1.88	1.44
Chlorite _(mine site)		--	5.07	6.64
Kaolinite			--	1.31
SIDSAM_{tan}	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite
Goethite	7.31	2.73	4.57	3.47
Chlorite	--	19.96	1.60	2.10
Chlorite _(mine site)		--	12.47	9.48
Kaolinite			--	1.31
DSSC	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite

Goethite	6.00	3.21	1.72	1.75
Chlorite	--	19.26	3.48	10.51
Chlorite _(mine site)		--	5.53	1.83
Kaolinite			--	3.02
KJSSC	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite
Goethite	5.17	3.31	1.34	2.68
Chlorite	--	17.13	3.87	13.86
Chlorite _(mine site)		--	4.42	1.24
Kaolinite			--	3.58
KJDSSC_{tan}	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite
Goethite	31.08	10.75	2.31	4.70
Chlorite	--	334.08	13.48	146.18
Chlorite _(mine site)		--	24.78	2.29
Kaolinite			--	10.84

3.4.3 Effectiveness of proposed SMMs

The effectiveness of SMMs plays a crucial role in the successful exploitation of hyperspectral dataset for various applications such as minerals and rock type mapping (Bishop et al., 2011; Van der Meer et al., 2012; Zhang & Li, 2014), soil quality mapping and prediction (Govender et al., 2007; Farifteh et al., 2013), vegetation-species discrimination (Govender et al., 2007; Naresh Kumar et al., 2011; Thenkabail & Lyon, 2016), water quality monitoring (Wang et al., 2005; Govender et al., 2007), planetary surface mapping (Chauhan et al., 2015; Kamps et al., 2020) and others (Feng & Sun, 2012; Lu & Fei, 2014). The newly proposed SMMs have illustrated their efficiency in discriminating spectrally similar materials (including a wide range of minerals and vegetation species). It has been evident from the previous discussion that the proposed SMMs also outperform as compared with the existing SMMs when discriminating spectra with random noise and with mixed spectra. The main limitation of the existing SMMs (i.e., SAM and SID) is that they do not consider the intra-class and inter-class variability more efficiently for spectral discrimination of spectrally similar materials (Naresh Kumar et al., 2011; Shanmugam & SrinivasaPerumal, 2014). The robustness of proposed SMMs over the existing SMMs considers the subtle variation of the spectral features and less sensitivity to noise in spectral data (Cartwright & Field, 1978; McCune et al., 2002; Kumar & Johnson, 2005). It is worth highlighting that at numerous occasions (while discriminating minerals (i.e., alunite and chlorite, calcite and kaolinite, chlorite and hematite, chlorite and jarosite, chlorite, and kaolinite (Table 3.8)) and vegetation spectra (i.e., blackbrush and blue spruce, blackbrush and juniper bush, blackbrush and saltbrush, blue spruce and juniper bush, blue spruce and juniper bush, blue spruce and rabbitbrush, blue spruce and saltbrush, cheatgrass and juniper bush, and juniper bush and saltbrush, and rabbitbrush and saltbrush (Table 3.9)), both the DSSC and KJSSC have shown higher spectral discrimination ability

than the existing hybrid measure i.e., $SIDSAM_{tan}$. The higher spectral discrimination ability of DSSC and KJSSC make their hybrid measure i.e., $KJDSSC_{tan}$ more robust for accurate discrimination. Furthermore, the proposed SMMs have shown the least impact of noise and spectral mixing on their spectral discrimination ability due to the fact that DSSC and KJSSC have the ability to efficiently deal with the noise and outliers, and intra-class variability.

It has also been observed that on a few occasions, all SMMs show limited discrimination ability. For example, while discriminating minerals (i.e., alunite and jarosite, alunite and kaolinite, and jarosite and kaolinite (Table 3.8)) and vegetation spectra (i.e., blue spruce and saltbush, and juniper bush and rabbitbrush (Table 3.9)), the values of RSDPW are not high. The spectral wavelength range is a crucial element for accurate discrimination of any materials. Most SMMs use the spectral information available in each spectral band of the hyperspectral data (Shanmugam & SrinivasaPerumal, 2014). It is possible that the consideration of full wavelength range (400–2500 nm) of spectral data may not be an ideal approach for making accurate discrimination of some materials. In such cases, the wavelength range of diagnostic spectral signatures can be considered over the full wavelength range for optimal spectral discrimination, as illustrated in previous studies (Van der Meer, 2006; Naresh Kumar et al., 2011). The prior knowledge of spectral signatures of materials is vital in performing accurate spectral discrimination and identification. However, the automatic selection of suitable spectral bands with a limited number of training samples is an existing hiatus in the field of hyperspectral remote sensing.

3.5 Conclusions

- This study proposed three new effective SMMs, such as DSSC, KJSSC, and $KJDSSC_{tan}$, with higher spectral discrimination ability than existing SMMs (i.e., SAM, SID, and $SIDSAM_{tan}$) of spectrally similar materials.
- A wide range of hyperspectral datasets acquired under laboratory and real atmospheric conditions were utilized to compare and evaluate the spectral discrimination ability of proposed and existing SMMs using RSDPW statistics.
- The effectiveness of SMMs was also evaluated using the spectra with added noise component and mixed spectra to assess the impact of noise and spectral mixing on the spectral discrimination ability.
- The RSDPW statistics of proposed SMMs are significantly higher than existing SMMs for most minerals and vegetation spectra. The proposed SMMs have shown tremendous success over the existing SMMs for accurate discrimination of materials, mainly due to their ability to consider the minute variation of spectral features and less sensitive to noise and outliers.
- The DSSC and KJSSC indicate similar spectral discrimination ability, whereas their hybrid, i.e., $KJDSSC_{tan}$, shows significantly higher discrimination ability of spectrally identical materials. The $KJDSSC_{tan}$ takes unique advantage of both measures for enhanced spectral discrimination.

- The proposed SMMs can be effectively exploited for other essential hyperspectral processing steps, including suitable band selection, intra-class variability analysis, selection of spectrally pure pixels, clustering, and pixel-based image classification for successful hyperspectral data analysis.

4 Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India

(This material has been published in the “*Applied Earth Observation and Geoinformation, Elsevier.*” Please refer to page XIII for the original paper link).

Abstract: In this study, we proposed an automated lithological mapping approach using spectral enhancement techniques and Machine Learning Algorithms (MLAs) using Airborne Visible Infrared Imaging Spectroradiometer-Next Generation (AVIRIS-NG) hyperspectral data in the greenstone belt of the Hutti area, India. We integrated spectral enhancement techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) transformation and different MLAs to accurately map rock types. A conjugate utilization of conventional geological map and spectral enhancement products derived from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data were used to prepare a high-resolution reference lithology map. Feature selection and extraction methods were applied on the AVIRIS-NG data to derive different input dataset such as (a) all spectral bands, (b) shortwave infrared bands, (c) Joint Mutual Information Maximization (JMIM) based optimum bands, and (d) optimum bands using PCA, to choose optimum input dataset for automated lithological mapping. The comparative analysis of different MLAs shows that the Support Vector Machine (SVM) outperforms other Machine Learning (ML) models. The SVM achieved an Overall Accuracy (OA) and Kappa Coefficient (k) of 85.48% and 0.83, respectively, using JMIM based optimum bands. The JMIM based optimum bands were more suitable than other input datasets to classify most of the lithological units within the study area (i.e., metabasalt, amphibolite, granite, acidic intrusive, and migmatite). The sensitivity analysis performed in this study illustrates that the SVM is less sensitive to the number of samples and mislabeling in the model training than other MLAs. The obtained high-resolution classified map with accurate litho-contacts of amphibolite, metabasalt, and granite coupled can be integrated with an alteration map of the area for targeting the potential zone of gold mineralization.

4.1 Introduction

Geological targets like rocks and minerals have different spectral characteristics caused due to electronic and vibrational processes within the spectral domain of 400 to 2500 nm, which leads to wavelength-specific spectral signatures (Clark & Roush, 1984; Cloutis, 1996; Clark, 1999). The spectral signature of rocks and minerals have been extensively used for their identification, discrimination, and mapping (Hunt, 1977; Kruse, 1996; Crosta et al., 1998; Van der Meer, 2004; Vaughan et al., 2005; Bishop et al., 2011; Kumar et al., 2014; Kumar et al., 2015; Guha et al., 2019; Kumar et al., 2020a). The hyperspectral sensor

measures the reflected energy of the surface in many narrow and contiguous spectral bands and provides an opportunity for target detection and discrimination even if targets have their absorption feature closely spaced in the wavelength domain (i.e., few nanometers apart) (Van der Meer et al., 2012). Airborne hyperspectral sensors provide scope to capture a greater number of spectrally pure targets than spaceborne sensors due to higher spatial and spectral resolution and high signal-to-noise ratio (Kruse et al., 2003). Furthermore, the higher resolution of airborne hyperspectral data also provides a better scope to characterize the boundaries of lithological units or rock types.

Both multispectral and hyperspectral datasets have been extensively used in lithological and minerals mapping (Kruse, 1998; Kruse et al., 2003; Ninomiya, 2004; Gomez et al., 2005; Ninomiya et al., 2005; Kruse et al., 2006; Bedini, 2009; Goetz, 2009; Bedini, 2011; Van der Meer et al., 2012; Zhang & Li, 2014; Kumar et al., 2020a). Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Landsat datasets have been well exploited for geological applications (Sabins, 1999; Rowan & Mars, 2003; Ninomiya et al., 2005; Bedini, 2009; Pour & Hashim, 2014a; Pour et al., 2018; Pour et al., 2019). Different spectral enhancement techniques such as Principal Component Analysis (PCA) (Crosta et al., 2003; Amer et al., 2010), Independent Component Analysis (ICA) (Kumar et al., 2015), and band ratios (Rowan et al., 2005; Gad & Kusky, 2007) were extensively used to demarcate different rock types. However, PCA and ICA-based mapping of rocks is usually a manual approach and becomes challenging when defining litho-units' contacts. Researchers also utilized other semi-automated techniques for mapping rock types using different pixel and sub-pixel based spectral mapping techniques using the reference spectra of rocks and minerals (Chen et al., 2010; Murphy et al., 2012; Van der Meer et al., 2012; Asadzadeh & de Souza Filho, 2016; Rao & Guha, 2018; Guha et al., 2019). However, most spectral mapping techniques require a threshold value to obtain the final classification result of target materials, which is subjective and challenging (Shanmugam & SrinivasaPerumal, 2014; Kumar et al., 2020a).

In recent years, Machine Learning Algorithms (MLAs) have shown great success over the conventional classification algorithms for accurate detection and mapping of geological targets (Cracknell & Reading, 2014; Othman & Gloaguen, 2014; Lary et al., 2015; Wei et al., 2016; Othman & Gloaguen, 2017). MLAs use artificial intelligence derived from input data by analyzing its characteristics and pattern with minimal human intervention (Michie et al., 1994). Implementation of MLAs leads to the derivation of semi-automatic tools for mapping the targets in the field of the earth sciences (Mjolsness & DeCoste, 2001; Yu et al., 2012; Cracknell & Reading, 2014; Othman & Gloaguen, 2014; Lary et al., 2015; Othman & Gloaguen, 2017; Bachri et al., 2019). MLAs are widely applied in different geoscientific applications with thrust on mineral mapping and mineral prospectivity mapping (Thompson et al., 2001; Waske et al., 2009; Baykan & Yılmaz, 2010; Abedi et al., 2012; Carranza & Laborte, 2015; Lary et al., 2015; Rodriguez-Galiano et al., 2015; Chen & Wu, 2017). MLAs' main strength lies in their ability to solve complex non-linear mapping, utilize a high volume of information without the assumption of data distribution, and integrate other datasets for better class separability and classification (Zuo, 2017).

MLAs' successful utilization for lithological mapping depends on many factors such as the quality of training dataset, selection of optimum or uncorrelated features or variables, and training parameters (Michie et al., 1994). The quality of training data has a significant impact on the performance of machine learning (ML) models (Cortes et al., 1995; Batista et al., 2004). The typical approach of collecting training datasets for lithological mapping is based on the low-resolution conventional geology map. The poor training data collection strategy from the low-resolution map, especially near the contact boundaries of different litho-units, can significantly degrade the final classification maps (DeFries & Chan, 2000). The feature selection and extraction become more critical when one deals with hyperspectral data due to higher spectral dimensions and inherent spectral redundancy (Archibald & Fann, 2007; Waske et al., 2010). Feature selection method reduces the dimensionality of data by selecting a subset of spectral bands with an intent to maximize the class separability; whereas, the feature extraction method transforms the existing features into a new low-dimensional feature space (Guyon et al., 2004; Hoque et al., 2014; Ren et al., 2014). The training parameters selection for ML models is another critical aspect of successful classification results (Bergstra et al., 2011). The trial and error approach to select the training parameters can over- or under-fit the ML models, which eventually can produce poor classification results (Bergstra et al., 2011). Although some studies have been conducted to address these factors independently, to the best of our knowledge, there is no single study undertaken to address all three factors together to develop an automated lithological mapping model using MLAs.

This study proposes an integrated approach of spectral enhancement techniques and MLAs for automated mapping of rock types using AVIRIS-NG hyperspectral data. The proposed approach will overcome the limitation of the discussed factors for MLAs based lithological mapping. The proposed approach has been implemented in a geological setting consist of spectrally similar rock types such as granite, migmatite, amphibolite, and metabasalt of poorly exposed and weathered surface in subtropical climatic conditions. We performed a comparative analysis of three commonly used MLAs, such as Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine (SVM), using different input datasets derived by feature selection and extraction method using AVIRIS-NG data. Furthermore, we performed a sensitivity analysis to illustrate the impact of reduction and mislabeling or uncertainty in the training dataset on the performance of different MLAs for automated lithological mapping.

4.2 Location and geological setting of the study area

The Hutti greenstone belt is located between 16°1'12" to 16°24'0" N and 76°36'0" to 76°51'0" E in the part of Raichur district, the southern part of India. The geographical location and geological setting of the area are presented in Figure 4.1. The major rock types found in the area include metabasalt, granite, migmatite, amphibolite, and acidic intrusive. The majority of the area is covered by metabasalt, black soil followed by granite, migmatite, and amphibolite. The Hutti greenstone belt hosts three major gold mines, such as Hutti, Utti, and Hirra-Buddini (Mishra et al., 2005). The Hutti Mine is the largest

operation Gold mine in India, has a reserve of >120 t Au, with an average grade of 4.42 g t⁻¹ Au (Hazarika et al., 2015). The Utti mine is an open cast mine and situated in the northeastern part of the area. The area hosts hydrothermal alteration minerals such as chlorite, muscovite, kaolinite, and goethite, indicating the propylitic alteration (Kumar et al., 2020a). Altered metabasalt and amphibolite were keenly studied for understanding the gold formation and associated minerals in the area (Pal & Mishra, 2002; Mishra et al., 2005; Sarma et al., 2008). The area has also been explored for tourmaline minerals using geochemical and petrographic analysis of altered rock samples of sub-surface from the different mine sites as a proxy to gold mineralization (Hazarika et al., 2015). The metabasalt hosts the tourmaline in association with biotite-muscovite foliation, plagioclase, and chlorite minerals (Hazarika et al., 2015).

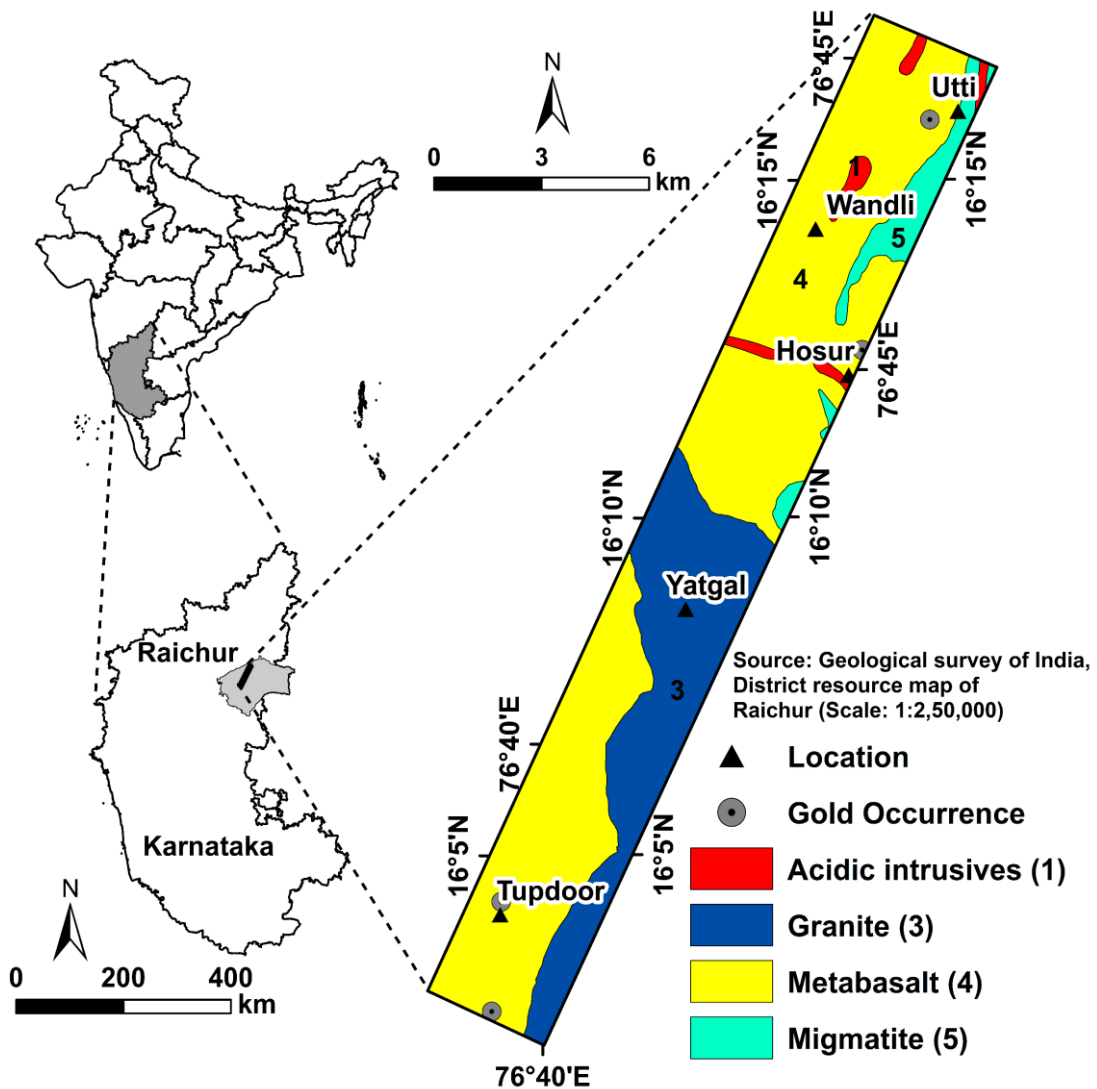


Figure 4.1. Geographical location and geology map of the area showing major lithological units with a few locations and gold occurrences.

4.3 Datasets and methods

The VNIR-SWIR spectral bands of ASTER data were used to derive the spectral enhancement products using PCA and ICA transformation to highlight the lithological units and to refine the conventional lithology map. The AVIRIS-NG data were used to prepare different input datasets to be used in MLAs for automated lithological mapping. The sensor specification and characteristics of ASTER and AVIRIS-NG data have been discussed in the following sections.

4.3.1 Datasets

The ASTER sensor measures the reflected and emitted energy in the wavelength range of VNIR, SWIR, and Thermal Infrared (TIR) at a spatial resolution of 15m, 30m, and 90m, respectively. The detailed sensor specification of ASTER can be found in Rowan and Mars (2003). A cloud-free L1B (level 1: radiometric and geometric corrected) radiance at sensor data (acquired on 12/02/2003) have been obtained through NASA Earth Data Portal (<https://search.earthdata.nasa.gov/search>). The AVIRIS-NG sensor recorded the reflectance energy in 425 spectral channels in the wavelength range of 380 to 2510 nm at 5 nm and 7.5 m spectral and spatial resolution. The data have been acquired and processed (i.e., atmospheric and geometric correction) under a collaborative science mission of the National Aeronautics and Space Administration (NASA) and the Indian Space Research Organization (ISRO) to enhance the geology and mineral information in a different part of India including the study area (Bhattacharya et al., 2019a; Bhattacharya et al., 2019b). The data were acquired on 28/01/2016 and accessed through the NASA-AVIRIS-NG data portal (https://avirisng.jpl.nasa.gov/alt_locator/). The detailed specification of the AVIRIS-NG sensor is presented in Table 4.1.

The software packages and programming languages such as ENVI 5.5 (<https://www.harrisgeospatial.com/Software-Technology/ENVI>), R 3.5 (<https://www.r-project.org/>), and ArcMap 10.6 (<http://desktop.arcgis.com/en/arcmap/>) were used for image processing, automated machine learning implementation, and preparation of spatial layers.

Table 4.1. Specifications of AVIRIS- NG sensor (Thorpe et al., 2016). VNIR: Visible Near Infrared, and SWIR: Shortwave Infrared.

Characteristics	AVIRIS-NG
Spectral range	380 to 2510 nm
Spectral channel	425
Position	5 nm
Calibration	± 0.3 nm
Signal to Noise Ratio (SNR)	>2000 for VNIR and >800 for SWIR
Accuracy	95 % (<5 % uncertainty)
Sample distance	0.3 to 20 m

Radiometric resolution	14 bits
------------------------	---------

4.3.2 Methods

The study utilized ASTER and AVIRIS-NG dataset for automated lithological mapping using MLAs. The ASTER data, along with a geological map and field data, were used to prepare a high-resolution reference map for better training data sampling for ML modeling. A schematic diagram showing the methodology adopted in this study has been presented in Figure 2 and discussed in the following sub-sections.

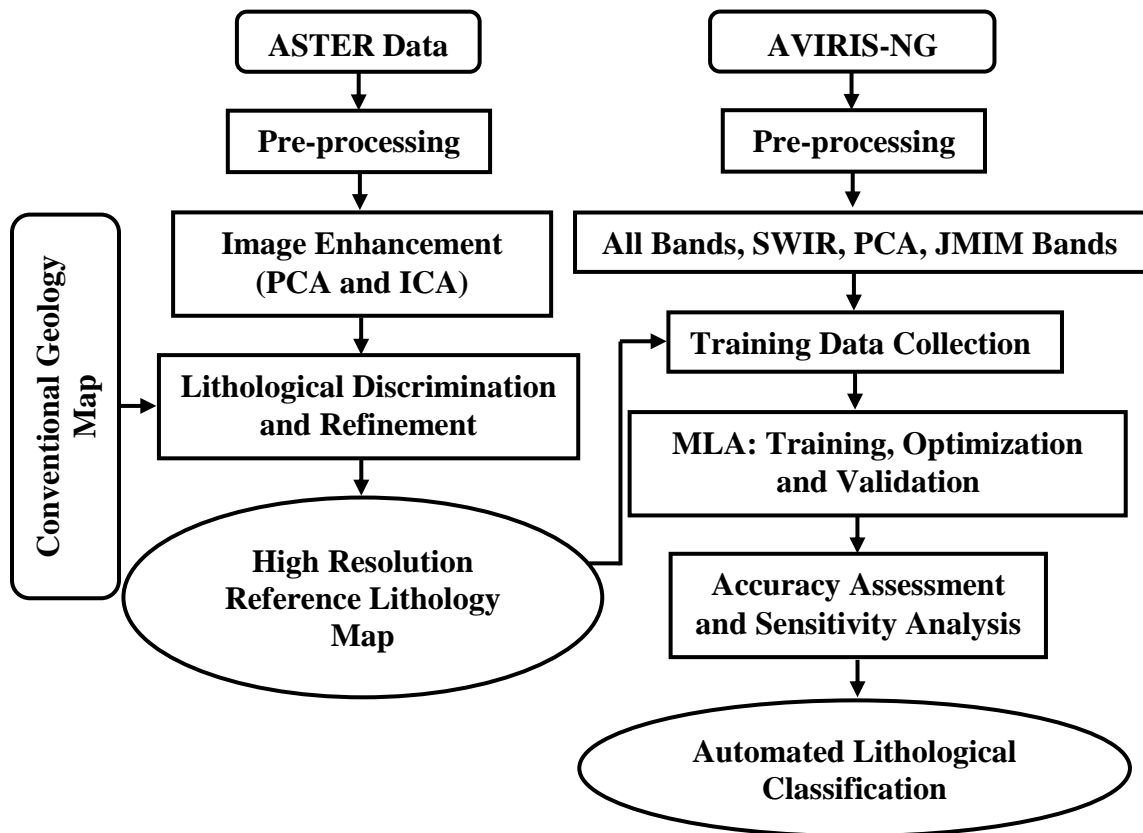


Figure 4.2. The methodology's workflow chart adopted for automated lithological mapping using the integrated spectral enhancement and ML methods.

4.3.2.1 Preparation of high-resolution reference lithology map

A suitable reference map is critical in obtaining good training datasets for the successful mapping of rock types using any MLAs. The most common practice of collecting training data is to use the conventional geological map as a reference map (Cracknell & Reading, 2014; Othman & Gloaguen, 2014, 2017; Bachri et al., 2019). However, it is worth mentioning that the conventional geology map is usually prepared based on the cognitive

utilization of surface and sub-surface information of rock types, including mineralogical composition, texture, formation age, and stratigraphy. In contrast, remote sensing-based rock type mapping is mainly based on the surface spectral characteristics of rocks. Furthermore, the conventional geological map is usually prepared at a large scale (as in this study 1:2,50,000) and may not be a suitable reference map for obtaining a training dataset for classifying high spatial resolution data such as AVIRIS-NG hyperspectral data (in this study).

We adopted an integrated approach of spectral enhancement techniques derived from PCA and ICA transformation using VNIR-SWIR spectral bands of ASTER data along with a geological map and field data at selected traverses to demarcate the lithological boundary and prepare a suitable reference lithology map at the scale of 1:25,000. The band combinations of higher spectral contrast of PCA and ICA transformation have been used in this study. We also used a False Color Composite (FCC) (R:850 nm, G: 650 nm, B: 550 nm) of high-resolution AVIRIS-NG data to update the lithological boundary and to reduce the uncertainty due to discrepancy in spatial resolution of ASTER and AVIRIS-NG datasets. The preference of spectral enhancement products of ASTER data instead of AVIRIS-NG to for refining the reference lithology map was mainly made due to ASTER's large areal coverage and adequate spectral contrast of rock types (Guha et al., 2012; Kumar et al., 2015), which makes an easy and rapid comparison with the conventional geology map for the interpretation of different rock types.

The ASTER data were corrected for the cross-track illumination effect using a cross-track illumination correction tool of ENVI to reduce the impact of energy overspill from band 4 to bands 5 and 9 (Hewson et al., 2005). This was followed by layer stacking of VNIR and SWIR spectral bands with resampling to 30 meters, atmospheric correction using the Fast Line of Sight Atmospheric Analysis of Hypercubes (FLAASH) module (<https://harrisgeospatial.com/docs/FLAASH.html>), and projected to UTM zone 44 True North (referred hereafter processed ASTER data).

4.3.2.2 Preparation of different input datasets using AVIRIS-NG hyperspectral data

The surface reflectance level 2 (L2) AVIRIS-NG data consist of 425 spectral channels were subjected to the removal of noisy bands. A total of 295 spectral bands (referred hereafter as processed hyperspectral data) have been found suitable, whereas remaining bands such as 1–15, 146–158, 188–227, 275–328, and 421–425 were excluded due to significant noise.

The higher dimensionality of hyperspectral data introduces challenges as spectral redundancy and complex model formulation. Obtaining an optimum number of spectral bands can improve the classification accuracy, reduce the model complexity and computational cost by removing the less significant and correlated spectral bands (Archibald & Fann, 2007; Pal & Foody, 2010). The processed hyperspectral data were subjected to feature selection and extraction methods to derive different input datasets from optimizing the input configuration for lithological mapping using MLAs.

In this study, the filter and spectroscopic significance of rocks-based feature selection methods were performed. The filter-based methods evaluate the relevance of variables or spectral bands outside the model and select the variable that passes through certain criteria. A variety of filter-based feature selection methods such as correlation, entropy, mutual information, and chi-squared are available in the literature (Hall & Smith, 1999; Yu & Liu, 2003; Wu et al., 2013). Mutual information-based filters have drawn great attention in recent years. In this study, Joint Mutual Information Maximization (JMIM) (Bennasar et al., 2015) method has been implemented in R using the “praznik” package (<https://github.com/mbq/praznik>), which uses the ‘mutual information’ and the ‘maximum of the minimum’ criterion to select the optimum spectral bands to consist of less band to band correlation. Mutual information is a measure of the mutual dependence between the two spectral bands, whereas the maximum of the minimum approach allows choosing the band, which shows a weaker correlation between bands (Bennasar et al., 2015).

The JMIM was used to calculate the band significance for processed hyperspectral data to select the optimum number of bands for automated lithological mapping. The JMIM produces a score of each band where higher the score, less the band to band correlation and vice-versa. To identify the number of optimum bands, we implemented the first 50 spectral bands of higher JMIM scores in MLAs and marked the accuracy statistics. Subsequently, the next 25 bands of higher JMIM score and so on (such as best 75, 100, 125, 150, 150, 175, 200, 225, 250, 275, and 295) were used in MLAs for lithological mapping. The band numbers yield the maximum OA and k, and the subsequent addition of more bands shows no significant improvement in the accuracy statistics were considered an optimum number of spectral bands (named hereafter JMIM optimum bands). Furthermore, feature selection based on the spectroscopic significance of rocks, as most of the rock-forming minerals show their diagnostic absorption feature in the wavelength range of SWIR, were evaluated and compared with other input datasets for lithological mapping using MLAs.

Moreover, the feature extraction technique, such as PCA, was implemented on the processed hyperspectral data to transform the high dimensional space into a lower dimension. The first 20 principal components of higher eigenvalues (named hereafter PC based optimum bands) were used as input data in MLAs to evaluate its utility for automated lithological classification.

The feature selection and extraction methods applied on processed hyperspectral data produce four different input data configuration such as (a) all spectral bands, (b) SWIR bands (c) JMIM based optimum bands, and (d) PC based optimum bands, have been evaluated and compared for automated lithological classification using MLAs. The regions of other classes, such as vegetation, waterbody, cloud, and shadow regions, were masked using Spectral Angle Mapper (SAM) to avoid misclassification in all input datasets. The SAM is a spectral similarity-based classification method that measures the spectral angle between the image spectra and reference spectra by treating them as a vector in a space with dimensionality equal to the spectral channels (Kruse et al., 1993).

4.3.2.3 Training data collection and spectral separability computation

The reference lithology map was used to collect the training points for automated lithological classification. A collection of good training data samples is crucial in obtaining optimum classification results from any MLAs. We adopted a stratified random sampling to collect the training samples and excluded the data points at the boundary of lithological units to avoid the false labeling of the samples. A proportional number of samples as per the spatial extent of different classes have been collected to avoid the impact of class imbalance on the MLAs (Batista et al., 2004; Oommen et al., 2011). The geo-locations of these samples have been used to extract the reflectance values from all four input datasets and processed further to implement MLAs. The total training samples (i.e., 2170) have been divided into 80% and 20% for training and testing, respectively, for the MLAs.

To understand the intra-class variability of lithological units, a spectral dissimilarity matrix of all four input datasets were computed using Jeffries Matusita (J-M) distance. The J-M distance measures the band wise information to assess the spectral similarity between given spectra (Richards & Richards, 1999) and widely used in remote sensing applications for feature selection, class separability analysis, and spectral matching (Padma & Sanjeevi, 2014b).

4.3.2.4 Implementation of MLAs for automated lithological classification

The training samples derived using different input datasets were used to train the various MLAs, followed by parameter optimization, validation, and accuracy assessment. The implementation of MLAs was done using the “CARET” package available in the “R” programming language (Kuhn et al., 2017). A brief conceptual framework of implemented MLAs has been discussed below.

- a.** Linear Discriminant Analysis (LDA): LDA is a simple and mathematically robust statistical technique for dimension reduction, pattern recognition, and classification. It projects the original high-dimensional data to a lower-dimensional space in a linear combination of variables, which maximizes class separability (Tharwat et al., 2017). LDA assumes that the given data is a Gaussian distribution, and each variable has the same variance. LDA applies three main steps to the given data to perform classification: (i) calculate the separability between different classes (i.e., the distance between the means of different classes); (ii) calculate the distance between mean and samples of each class (i.e., with-in class variability); and (iii) project the original dimensional space to lower dimensions, which maximizes the class separability and minimizes within-class variability (Tharwat et al., 2017). The major limitation of LDA is that it fails to find lower-dimensional space if the dimensions are much higher than the number of samples, which means that the number of samples is crucial while using LDA for image classification. LDA also suffers from the assumption of linearity and Gaussianity; if the classes are non-linearly separable, then it fails to discriminate these classes (Tharwat et al., 2017). The advantage of using LDA is that it does not need any parameter optimization, making it simpler to implement.

- b.** Random Forest (RF): The RF method consists of a large number of Decisions Tree (DT), where each DT is generated using a random vector sampled independently from the input vector, and each DT results in a classification (Pal, 2005; Ge et al., 2018). The majority vote of all DTs is used to assign a final class for each unknown of the input vector (Maxwell et al., 2018). The key to the success of RF is the way it creates the DTs; the tree is planted on the new training set by using random features selection. The new training set is drawn from the bagging method. The bagging method generates a training dataset by randomly drawing with the replacement of the original training set used for each feature/feature combination selected (Pal, 2005). The data not used in training are known as out-of-bag (OOB) data and can be used to provide an independent estimate of OA of RF classification (Maxwell et al., 2018). The implemented RF has one optimization parameter, i.e., the number of trees to grow needs to be optimized for the best results. Some of the most common advantages of RF methods are, it is robust to large and correlated variables and inbuilt feature selection. However, it also suffers from some drawbacks, such as difficult to interpret and computationally costly when it deals with many trees (Kuhn et al., 2017).
- c.** Support Vector Machine (SVM): The SVM is based on statistical learning theory and focuses on the training samples closest in the feature space to produce the optimal separation of classes (Schölkopf et al., 2002; Pal, 2005). The SVM classifier is inherently binary, where classes are linearly separable and identify a single hyperplane (also called a decision boundary) that gives the maximum margin between two classes. The margin is defined as the sum of distances to the hyperplane from two classes' closest points. The binary SVM can be extended as a multi-class classifier by repeatedly applying the classifier to each possible combination of classes (Pal, 2005; Maxwell et al., 2018). The SVM uses kernel functions such as linear, polynomial, sigmoid, and radial basis functions to project the feature spaces to the higher dimensions by assuming that a linear boundary can be obtained in a higher dimensional feature space. The classes are inherently not separable; the hyperplane can be regarded as having a soft-margin, allowing class samples on the wrong side of the boundary (Pal, 2005). The choice between margin and misclassification errors is controlled by a positive user-defined parameter known as cost (C), which needs to be chosen optimally for a better result. Higher C's value produces a more complex hyperplane and less generalized model (Maxwell et al., 2018).

The optimization of training parameters is crucial to obtain accurate classification results from MLAs. Various optimization techniques, such as grid search, simulated annealing, genetic algorithms, have been discussed in the literature (Bergstra & Bengio, 2012; Luo, 2016). However, the grid search algorithm is one of the most commonly adopted methods and has been used in this study. The grid search algorithm was implemented with m-fold cross-validation (m-FCV) for robust parameter estimation. The m-FCV is a statistical resampling technique used to evaluate and select optimum parameters for the model by splitting the training data into an m-distinct subset. Then the model runs m-times, each time withholding one of the subsets, which is used for validation (Kohavi, 1995). The results of each run are validated using withheld data, and the results are averaged across all m runs to produce accuracy statistics such as OA and k (Maxwell et al., 2018). The SVM

model with radial basis function has two parameters: sigma and cost; whereas, the RF contains one parameter, i.e., the number of trees. The LDA does not contain any parameters. The training parameters of SVM and RF were optimized to obtain accurate classification results.

4.4 Results and discussion

4.4.1 Spectral characteristics of rocks

The spectral curves of different lithological units of AVIRIS-NG data were extracted to study their spectral characteristics (Figure 4.3). The spectral curve of rocks exhibits absorption features controlled by their mineral composition, moisture content, surface texture, and illumination effects (Clark, 1999). Most of the rock' spectra show absorption features around 650 and 950 nm caused due to Fe^{3+} and absorption feature around 2200 and around 2300 nm due to Al-OH and Fe/Mg-OH molecules, respectively (Clark, 1999). The spectral curve of acidic intrusive (Figure 4.3 (a)) shows a similar absorption feature around 650 nm and 950 nm due to Fe^{3+} followed by another absorption feature around 2300 nm due to Fe/Mg-OH molecules. The spectral curve of amphibolite (Figure 4.3 (b)), granite (Figure 4.3 (c)), metabasalt (Figure 4.3 (d)), migmatite (Figure 4.3 (e)), and mine site (Figure 4.3 (f)) show similar absorption feature around 650 nm and 950 nm due to Fe^{3+} and absorption feature around 2200 nm and around 2300 nm due to Al-OH and Fe/Mg-OH molecules respectively. The spectral curve of metabasalt (Figure 4.3 (d)) also shows an additional absorption feature around 2260 nm due Fe-OH molecule. Kumar et al. (2020a) identified and mapped hydrothermal alteration minerals such as goethite, muscovite, kaolinite, and chlorite in the study area using AVIRIS-NG data, which show similar absorption features as found in different rock types in this study.

The average spectral curve of different lithological units derived using the training dataset of all spectral bands of the AVIRIS-NG reflectance image is shown in Figure 4.3 ((a). Figure 4.3 (b) presents the boxplot of the average overall reflectance of different lithological units. It can be noticed that the average spectra of different lithological units except the mine site and soil cover show higher spectral similarity, which can also be referred from the spectral dissimilarity matrix of lithological units using all the spectral bands (Table 4.3). The average overall reflectance of metabasalt and migmatite are higher than other rock types, whereas mine site shows the least average overall reflectance. The average overall reflectance of amphibolite and granite are very similar. The average overall reflectance acidic intrusive is less than migmatite and metabasalt but more than other classes. The average overall reflectance of black soil is higher than mine site but less than other classes.

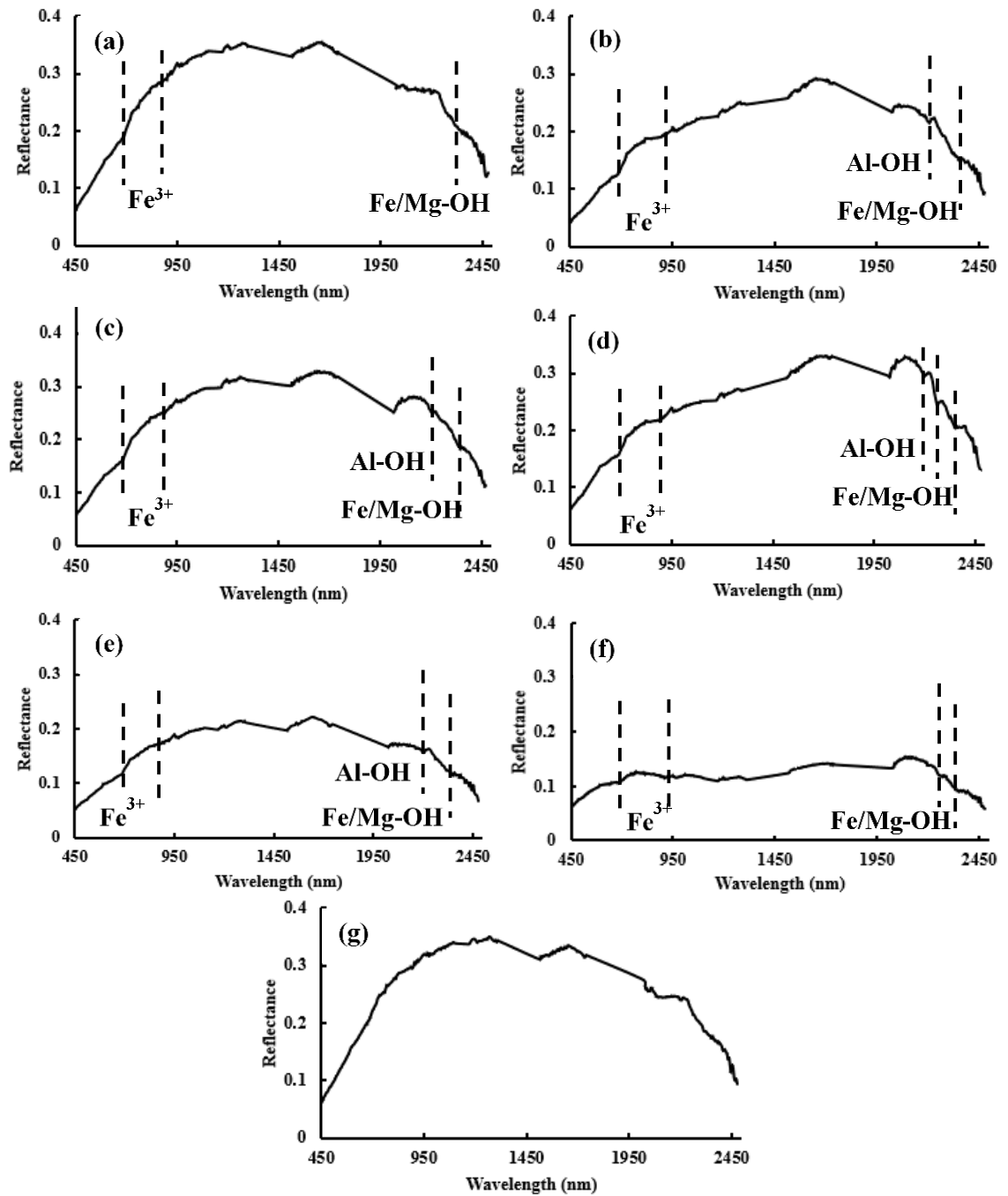


Figure 4.3. Spectral curve of different rock types (a) acidic intrusive, (b) amphibolite, (c) granite, (d) metabasalt, (e) migmatite, (f) mine site, and (g) black soil derived from AVIRIS-NG reflectance image. [Arrow indicates the absorption band position].

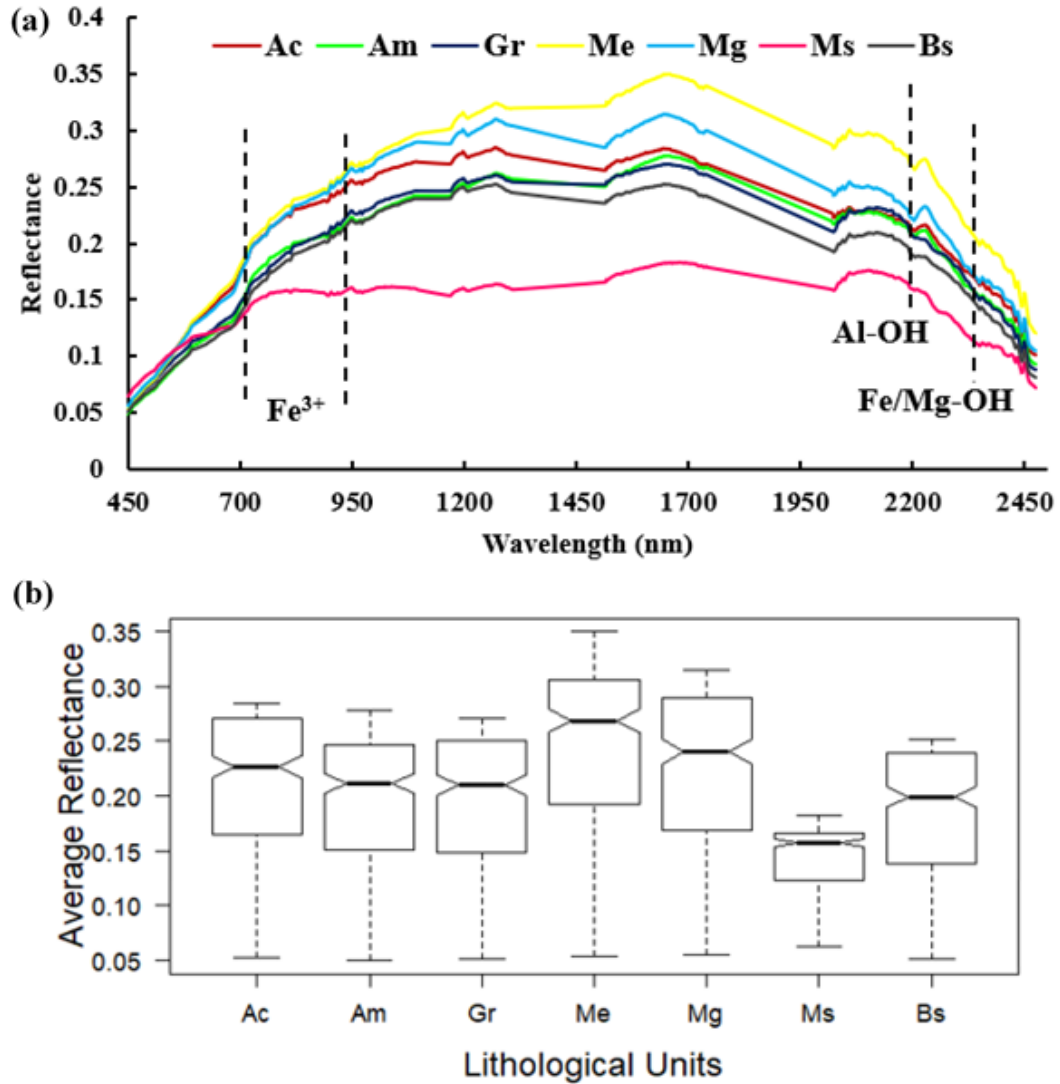


Figure 4.4. The spectral characteristics of different lithological units, including mine site and black soil. (a) the spectral curve of rock types derived from average training pixels of AVIRIS-NG hyperspectral data used in lithological classification, and (b) a box plot of overall average reflectance of different lithological units. [Ac: acidic intrusive, Am: amphibolite, Gr: granite, Me: metabasalt, Mg: migmatite, Ms: Mine site, and Bs: black soil].

4.4.2 High-resolution reference lithology map

The spectral enhancement products derived from PCA and ICA transformation using processed ASTER data were used to prepare the reference lithology map at a scale of 1:25,000. The FCC (R:850 nm, G:650 nm, B: 550 nm) of high resolution processed hyperspectral data (Figure 4.5 (a)), band combination of PCA (R: PC3, G: PC2, B: PC1) (Figure 4.5 (b)) and band combination of ICA (R: IC2, G: IC1, B: IC3) (Figure 4.5 (c))

showed higher spectral contrast and were effectively used in discriminating various lithological units of the area. The interpretation and demarcation of different litho-units were made with the help of a conventional geological map (Figure 4.5 (e)) and prior fieldwork in the study area. The reference lithology map is presented in Figure 4.5 (f). We demarcated a new lithological unit, i.e., amphibolite, in the northern part of the area, as highlighted in the band combinations of PCA and ICA transformation. We have also demarcated a few more units of migmatite, which were not mapped in the conventional geological map due to low-resolution. Furthermore, the demarcation of soil cover from the lithological units can increase the intra-class variability and, therefore, improve classification accuracy. The field validation of the reference lithology map has been discussed in section 4.4.6.

The band combination prepared using PCA and ICA transformation complement each other to discriminate various rock types and soil cover successfully. The band combination prepared using ICA transformation provides better spectral contrast between soil cover and metabasalt, while the band combination of PCA transformation shows better discrimination between amphibolite and soil cover. However, the ICA offers better distinction at litho-contacts' boundaries than PCA for most rock types. The sub-tropical climatic condition, caused by intense weathering and poor rock exposure, makes lithological discrimination challenging.

It is worth mentioning that spectral enhancement products allow preparing a high-resolution reference lithology map at a scale of 1:25,000 by considering the spectral characteristics of rock types. Furthermore, it will enable separating other classes such as soil cover formed by weathered metabasalt and different rock types in the area. The reference lithology map (Figure 4.5 (f)) consist of more units of rocks as compared with a conventional geology map of low resolution (Figure 4.5 (e)). Furthermore, the soil discrimination from the rock types with an accurate lithological boundary allows collecting better training data compared with the conventional geology map (Figure 4.5 (d)). Table 4.2 presents the areal extent of different lithological units, including soil cover and the total number of training and testing samples used in MLAs for lithological classification. The metabasalt covers the largest spatial extent (i.e., 48.90 km²) of the area, followed by amphibolite (11.56 km²), granite (11.03 km²), migmatite (8.95 km²), acidic intrusive (1.55 km²), and mine site (0.32 km²). Black soil (25.48 km²) also covers a majority of the area and is mainly formed as a weathered product of metabasalt and other parent rocks in the area.

4.4.3 Feature selection and extraction-based hyperspectral input datasets

The four input datasets such as all spectral bands, SWIR bands, JMIM, and PCA based optimum bands obtained by implementing feature selection and extraction method using processed AVIRIS-NG hyperspectral data. The utility of these input datasets was evaluated and compared for automated lithological classification using different MLAs. The selection of JMIM based optimum bands was made based on OA and k values.

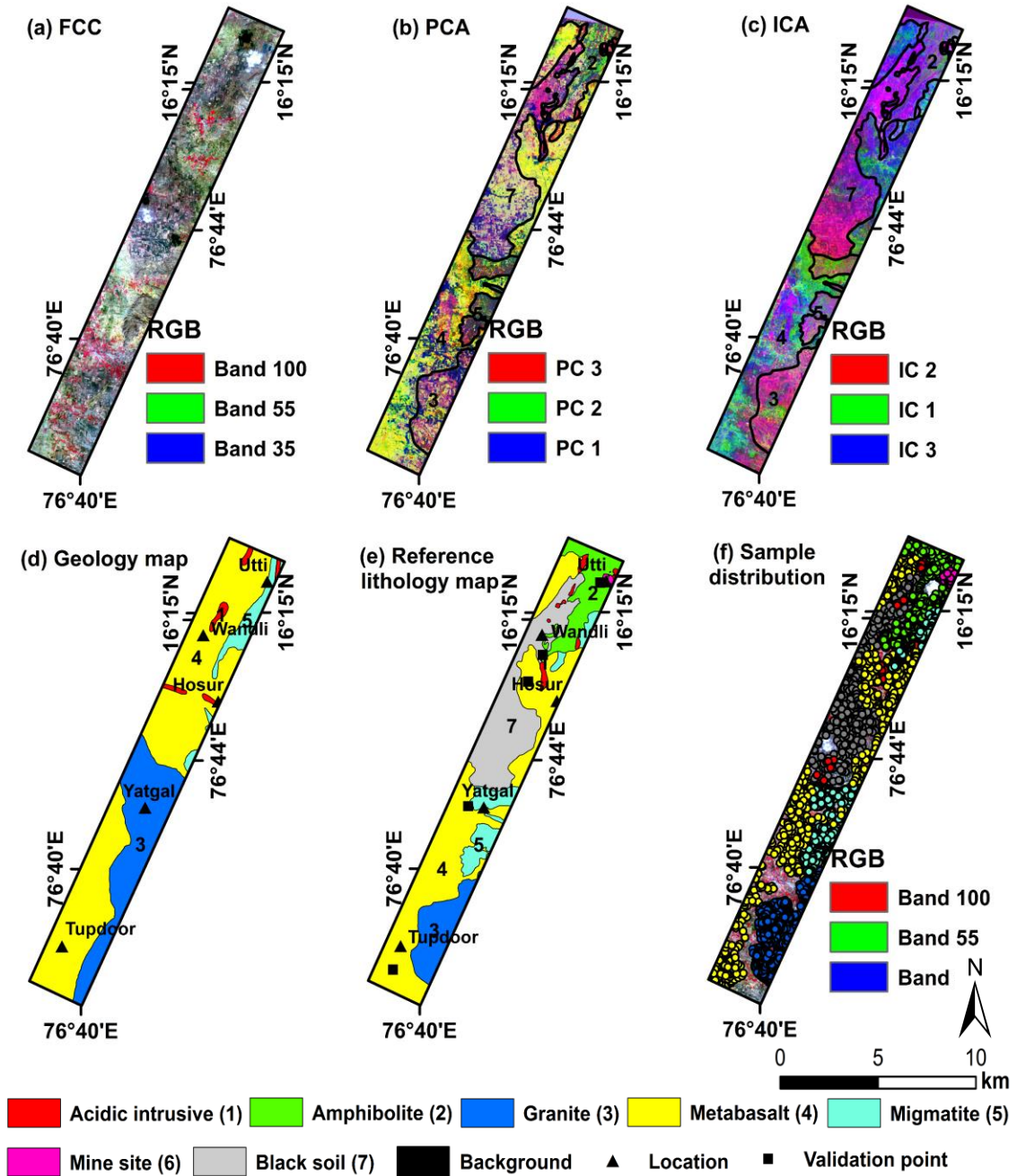


Figure 4.5. The image enhanced products such as (a) FCC (R: 850 nm, G: 650 nm, B: 550 nm) of AVIRIS NG, (b) band combination of PCA (R: PC3, G: PC2, B: PC1) and (c) band combination of ICA (R: IC2, G: IC1, B: IC3) derived from calibrated ASTER data discriminating the litho-units of the area, (d) training data samples overlaid on the FCC of AVIRIS-NG data, (e) conventional geology map prepared from district resource map mapped at the scale of 1:2,50,000 by the geological survey of India, and (f) high-resolution reference lithology map derived from the band combination of PCA and ICA transformation using ASTER data.

Table 4.2. Description of the lithological units, area, number of training and testing samples (Pixels of AVIRIS-NG data) used in the lithological classification using MLAs.

Lithological units	Area (km ²)	Training samples	Testing Samples
Acidic Intrusive	1.55	104 (0.76 km ²)	26 (0.19 km ²)
Amphibolite	11.56	280 (2 km ²)	70 (0.51 km ²)
Granite	11.03	280 (2 km ²)	70 (0.51 km ²)
Metabasalt	48.90	400 (2.9 km ²)	100 (0.73 km ²)
Migmatite	8.95	280 (2 km ²)	70 (0.51 km ²)
Mine Site	0.32	72 (0.52 km ²)	18 (0.13 km ²)
Black Soil	25.48	320 (2.3 km ²)	80 (0.58 km ²)

Figure 4.6 ((a) and (b)) illustrates the accuracy statistics of different spectral bands obtained using JMIM for lithological classification using MLAs. The first 100 JMIM bands achieved maximum or equal accuracy statistics such as OA (Figure 4.6 (a) and k (Figure 4.6 (b)) as compared with total number spectral bands (i.e., 295) to classify the lithological units of the area using MLAs. Figure 4.6 also illustrates that there was a slight decrease in the accuracy statistics when the lithological classification was carried using the first 125 to 225 JMIM bands. Further, MLAs models improve their accuracy statistics when classified with the first 250, first 275, and all spectral bands.

The selection of optimum bands based on the spectroscopic significance of rocks in the wavelength range of SWIR bands was also evaluated based on the achieved accuracy statistics using MLAs (Figure 4.7 and Table 4.4). The selection of an optimum number of PC bands based on higher variance and visual spectral contrast yields higher accuracy statistics than SWIR bands. It is essential to mention here that the feature selection and extraction method need not improve lithological classification accuracy statistics. Still, it reduces the model complexity and computation time by reducing the less important and correlated spectral bands, as illustrated in Figure 4.6.

4.4.4 Spectral separability analysis of hyperspectral input datasets

The spectral dissimilarity matrix of different lithological units, including soil cover and mine site for all input datasets was computed using J-M distance (Table 4.3). The dissimilarity values help in understanding the intra-class variability. A smaller value of J-M distance indicates less intra-class variability and can cause more miss-classification than classes that have larger J-M distance value. The dissimilarity matrix also provides an overview of the suitability of input datasets for lithological classification. Table 4.3 illustrates that the JMIM and PCA based optimum bands yield higher spectral dissimilarity (J-M distance) and intra-class variability. Hence, they would be more suitable datasets than SWIR and all spectral bands for lithological mapping. Most of the spectral bands of JMIM based optimum bands belong to the SWIR and NIR region, as commonly found minerals in the area include chlorite, muscovite, kaolinite, and goethite, have their diagnostic spectral signature in SWIR and NIR (Kumar et al., 2020a). However, the JMIM and PC

based optimum bands yield higher spectral dissimilarity due to the reduction of correlated spectral bands of the hyperspectral data.

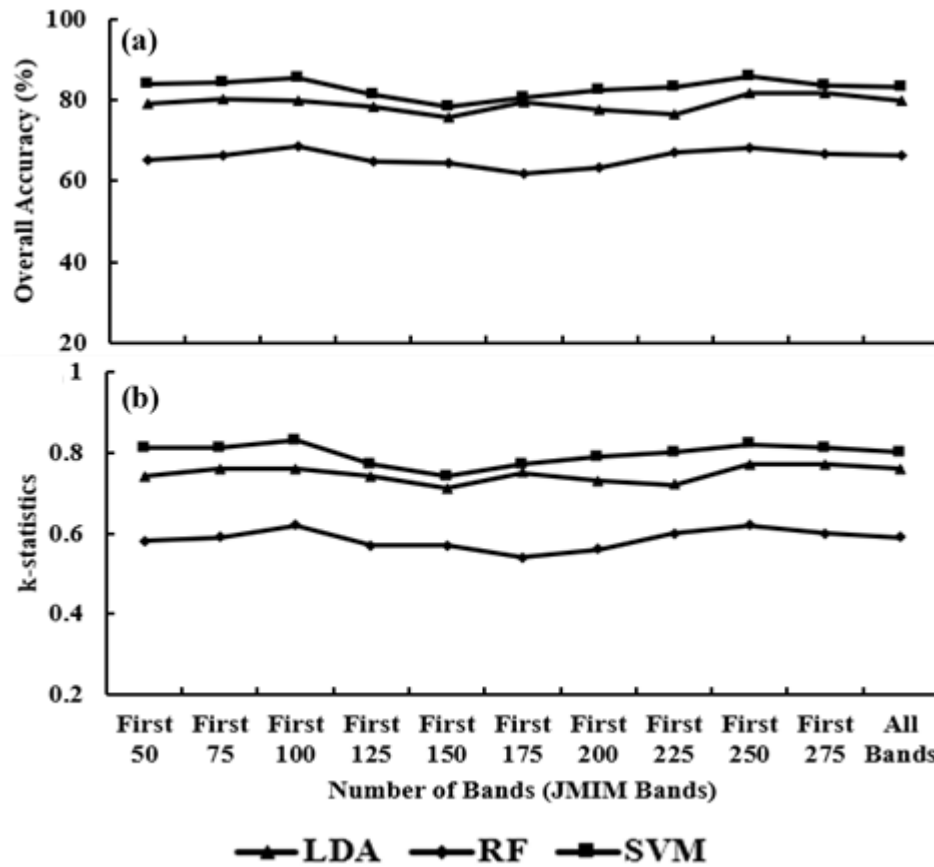


Figure 4.6. The accuracy statistics (a) overall accuracy and (b) kappa coefficient of JMIM bands using different MLAs.

It is interesting to note here that a class that is well separable from other classes in one input dataset becomes less separable in other input datasets (Table 4.3). For example, the amphibolite and acidic intrusive show less spectral dissimilarity in SWIR bands (0.008), JMIM bands (0.009), and all spectral bands (0.0158) as compared to PC bands (0.2795). The input dataset yields higher dissimilarity between classes would produce better classification results. Similarly, PC bands yield more spectral contrast among acidic intrusive, migmatite, and black soil. Amphibolite and granite appear spectrally similar in all input datasets and may yield less individual accuracy in lithological classification. Mine site and black soil appear more spectrally dissimilar as compared with other classes in all input datasets. Amphibolite and granite show less spectral dissimilarity values as compared with other lithological classes and may yield more false positives than other classes. It can be observed that JMIM and PC based optimum bands yield higher dissimilarity values for most of the litho-units and would produce better accuracy statistics as compared with SWIR and all spectral bands in lithological classification. Furthermore, it can also be

noticed that none of the input datasets produce higher spectral dissimilarity among all lithological classes.

Table 4.3. Spectral dissimilarity matrix computed using J-M distance of input datasets such as all spectral bands, SWIR bands, JMIM, and PC based optimum bands. Ai: acidic intrusive, Am: Amphibolite, Gr: Granite, Me: Metabasalt, Mi: Migmatite, Ms: Mine site, Bs: Black soil.

All Bands	Am	Gr	Me	Mg	Ms	Bs
Ac	0.0158	0.0167	0.0797	0.0171	0.5917	0.0533
Am	--	0.0002	0.1455	0.0557	0.4503	0.0115
Gr		--	0.1516	0.0593	0.4474	0.0104
Me			--	0.0239	0.8488	0.2278
Mg				--	0.6970	0.1146
Ms					--	0.3458
SWIR Bands	Am	Gr	Me	Mg	Ms	Bs
Ac	0.0080	0.0147	0.3279	0.0366	0.4960	0.0792
Am	--	0.0014	0.3520	0.0507	0.3909	0.0434
Gr		--	0.3425	0.0515	0.3688	0.0414
Me			--	0.1581	1.0704	0.5724
Mg				--	0.6448	0.1814
Ms					--	0.1953
JMIM Bands	Am	Gr	Me	Mg	Ms	Bs
Ac	0.0090	0.0151	0.1491	0.0356	0.7721	0.0604
Am	--	0.0008	0.2132	0.0710	0.6468	0.0231
Gr		--	0.2366	0.0857	0.6132	0.0153
Me			--	0.0410	1.1538	0.3512
Mg				--	0.9241	0.1655
Ms					--	0.4749
PC Bands	Am	Gr	Me	Mg	Ms	Bs
Ac	0.2795	0.2817	0.1971	0.0379	0.1277	0.9575
Am	--	0.0001	0.6708	0.4394	0.5298	0.4898
Gr		--	0.6741	0.4423	0.5341	0.4879
Me			--	0.0723	0.0567	1.2359
Mg				--	0.0496	1.0818
Ms					--	1.1372

4.4.5 Comparative analysis of MLAs using different input datasets

The four input datasets obtained using processed AVIRIS-NG data were evaluated for automated lithological classification using MLAs. The suitability of the input dataset was evaluated based on the OA and k of testing samples (Table 4.4 and Figure 4.7). The OA is the ratio between total correctly classified pixels and the total number of pixels. The confusion matrices of lithological classification obtained from all input data configuration using MLAs were presented in Table 4.5 to Table 4.8. User accuracy (UA) and Producer Accuracy (PA) were computed to understand individual class accuracy. The UA and PA describe commission and omission errors of individual classes (Congalton, 1991). The k considers the entire confusion matrix of diagonal and represents the measurement of agreement between the classified output and reference data (Brennan & Prediger, 1981).

The lithological classification obtained using JMIM based optimum bands yields better accuracy statistics as compared with other datasets (Figure 4.7). The lithological classification obtained from SVM using JMIM based optimum bands achieved the highest accuracy statistics, i.e., OA and k as 85.48% and 0.83, respectively (Figure 4.8 (b)). The LDA also shows an excellent performance in mapping the lithological units and achieved an OA and k of 79.95% and 0.76 using JMIM bands. The RF shows poor performance in mapping the lithological units using all the datasets (Table 4.4) except PC based optimum bands. It is interesting to note that RF's performance significantly improves with an OA and k of 78.80% and 0.74, respectively, using optimum PC bands. The RF and LDA show poor accuracy in classifying acidic intrusive compared with other classes using SWIR bands (Table 4.4), mainly due to higher spectral similarity with amphibolite and other classes (Table 4.3). However, SVM performs better in classifying acidic intrusive irrespective of higher spectral similarity with other classes. A similar observation can be made from Table 4.5, particularly in the case of RF.

Table 4.6 and Table 4.7 illustrate that most of the litho-units are being classified with adequate accuracy. Intra-class variability is crucial in achieving good classification accuracy. A class that shows higher spectral dissimilarity yields maximum accuracy and vice-versa. Acidic intrusive, amphibolite, and metabasalt show comparatively less accuracy than other classes such as granite, migmatite, black soil, and mine site. The mine site yields maximum accuracy in classification using all input datasets due to higher spectral dissimilarities with other classes.

The obtained high-resolution lithological map using optimum input data and robust MLA (i.e., SVM) with adequate accuracy is important for mineral exploration and other geological applications. The lithological map that maps the litho-contact of metabasalt with granite and amphibolite is of great interest in targeting the gold mineralization in the area (Mishra et al., 2005; Sarma et al., 2008; Guha et al., 2012; Kumar et al., 2020a). Furthermore, the lithological map obtained in this study can be overlaid with the spatial distribution of hydrothermal alteration minerals obtained in Kumar et al. (2020a) to better understand the alteration process and further for gold prospecting in the area.

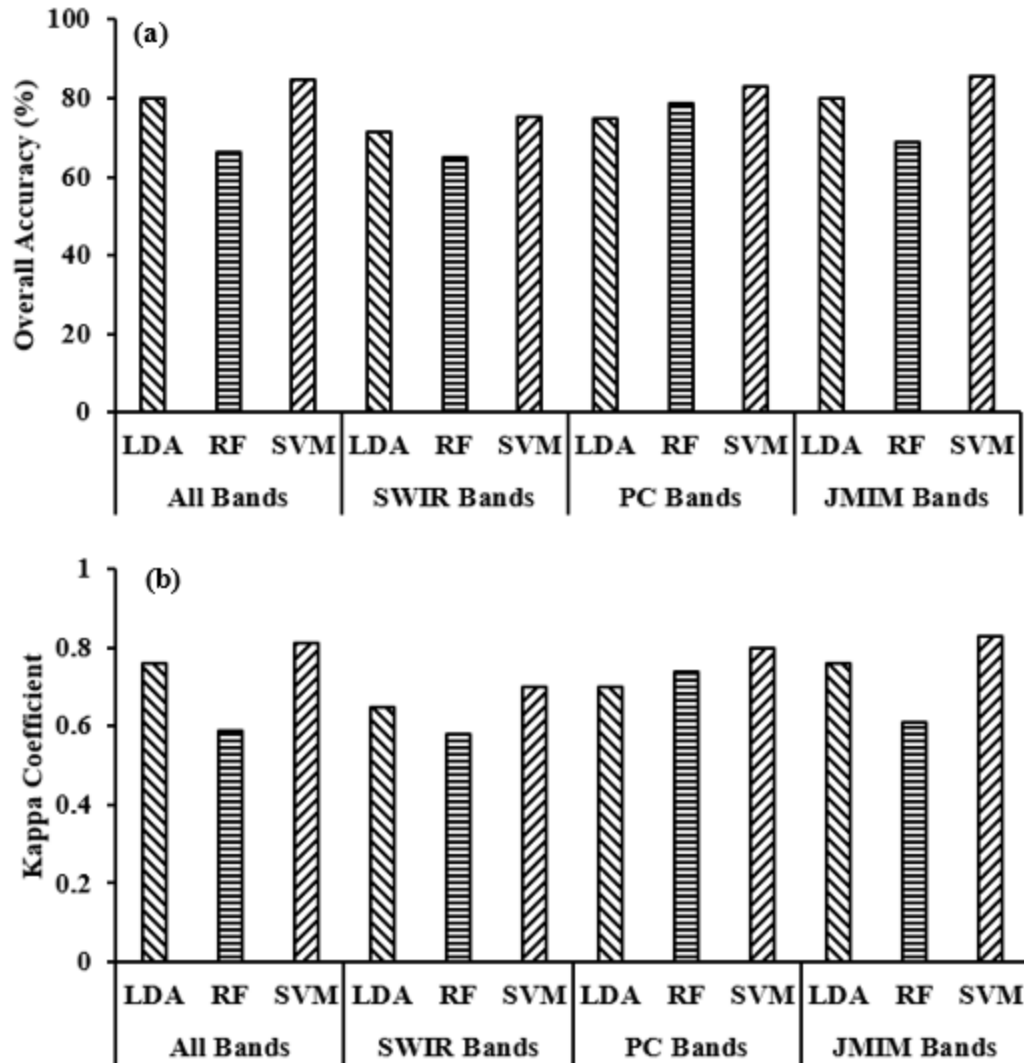


Figure 4.7. Accuracy statistics (a) overall accuracy and (b) kappa coefficient of MLAs using different input datasets such as all spectral bands, SWIR spectral bands, first 20 PC transformed bands of higher variance, and JMIM based optimum bands.

4.4.6 Validation of generated lithological map using field observation

The obtained lithological map from AVIRIS-NG data was validated in the field (Figure 4.8 (b)). Different accessible locations were selected to validate the few of the newly demarcated lithological units in the study area, as shown in Figure 4.8 (a). Figure 4.9 (a) to (f) display various field photographs of rock exposure such as (a) showing the granite found in the Yetgal area, (b) granite and amphibolite contact in Wandli area, (c) highly weathered metabasalt covered by black soil occurring in the pedi-plan, which is flanked by amphibolite in North-West of Hosur area, (d) metabasalt dumped in Utti mine area in the North-East part of the study area, (e) metabasalt, and (f) altered rock in Yetgal area.

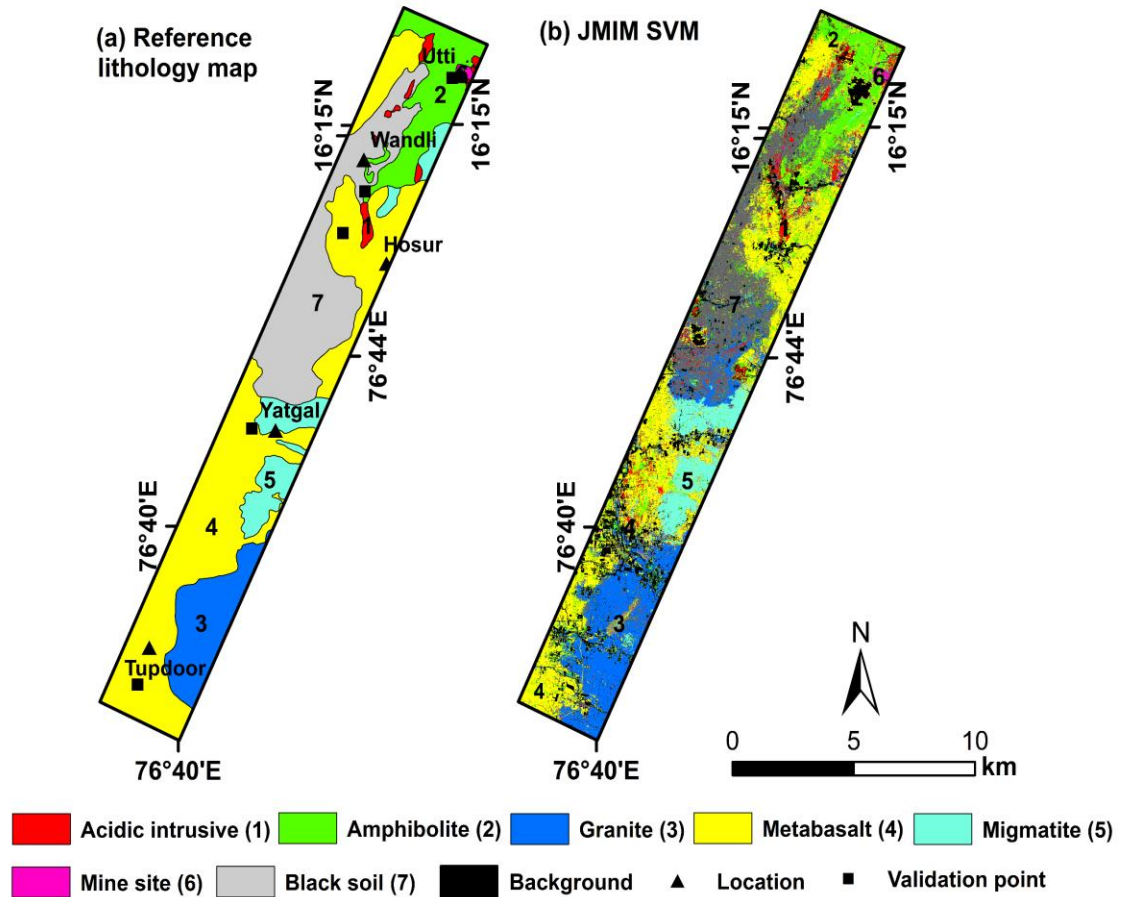


Figure 4.8. (a) Reference lithology map derived from spectral enhancement products using ASTER data; and (b) lithological classification map obtained from SVM using JMIM based optimum bands of AVIRIS-NG hyperspectral data.

The litho-units of the six field locations are nicely agreed with the predicted litho-units in the SVM model's lithological map. It has also been observed that the reference lithology map prepared by conjugative utilization of spectral enhancement products matches with ground existence at six locations. However, it has been noted that the conventional geology map of a large scale shows a considerable discrepancy with ground existence, mainly in the central and northern parts of the study area. The field validation provides excellent confidence to rely on a high-resolution reference map to collect the training data compared with the conventional geology maps for lithological mapping using MLAs.

Table 4.4. Confusion matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using SWIR spectral bands. LDA: linear discriminant analysis, RF: random forest, SVM: support vector machine. Ai: acidic intrusive, Am: Amphibolite, Gr: Granite, Me: Metabasalt, Mi: Migmatite, Ms: Mine site, BS: Black soil. Row: row total, Col: column total, UA: user accuracy, PA: producer accuracy.

LDA	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	15	6	0	11	0	5	4	41	36.59	57.69
Am	0	47	1	12	1	1	1	63	74.60	67.14
Gr	1	0	53	5	0	0	10	69	76.81	75.71
Me	3	6	6	60	6	0	5	86	69.77	60.00
Mi	5	1	2	8	63	0	1	80	78.75	90.00
Ms	0	0	0	0	0	12	0	12	100.00	66.67
BS	2	10	8	4	0	0	59	83	71.08	73.75
Col	26	70	70	100	70	18	80	434		
OA = 71.20%, k = 0.65										
RF	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	9	3	0	7	0	1	3	23	39.13	34.62
Am	6	42	3	11	4	0	5	71	59.15	60.00
Gr	1	2	48	3	0	0	10	64	75.00	68.57
Me	8	10	6	64	13	3	7	111	57.66	64.00
Mi	1	8	4	6	52	2	0	73	71.23	74.29
Ms	1	0	0	0	1	12	0	14	85.71	66.67
BS	0	5	9	9	0	0	55	78	70.51	68.75
Col	26	70	70	100	70	18	80	434		
OA = 64.98%, k = 0.58										
SVM	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	18	1	1	6	0	0	2	28	64.29	69.23
Am	1	52	1	11	0	1	4	70	74.29	74.29
Gr	1	0	54	6	0	0	8	69	78.26	77.14
Me	3	8	5	67	9	3	6	101	66.34	67.00
Mi	1	0	2	4	61	0	0	68	89.71	87.14
Ms	0	0	0	0	0	14	0	14	100.00	77.78
Bs	2	9	7	6	0	0	60	84	71.43	75.00
Col	26	70	70	100	70	18	80	434		
OA = 75.12%, k = 0.70										

Table 4.5. Confusion matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using all spectral bands of VNIR and SWIR.

LDA	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	20	3	0	4	1	0	1	29	68.97	76.92
Am	0	55	1	17	0	2	2	77	71.43	78.57
Gr	2	0	62	5	1	0	4	74	83.78	88.57
Me	2	7	2	59	3	0	3	76	77.63	59.00
Mi	0	0	3	10	65	0	0	78	83.33	92.86
Ms	0	0	0	1	0	16	0	17	94.12	88.89
Bs	2	5	2	4	0	0	70	83	84.34	87.50
Col	26	70	70	100	70	18	80	434		

OA = 79.95%, k = 0.76										
RF	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	11	1	1	4	0	0	3	20	55.00	42.31
Am	1	42	2	10	5	1	5	66	63.64	60.00
Gr	0	0	40	0	0	0	6	46	86.96	57.14
Me	7	14	9	68	6	2	3	109	62.39	68.00
Mi	5	7	1	14	58	0	7	92	63.04	82.86
Ms	0	0	0	1	0	13	0	14	92.86	72.22
Bs	2	6	17	3	1	2	56	87	64.37	70.00
Col	26	70	70	100	70	18	80	434		
OA = 66.36%, k = 0.59										
SVM	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	17	1	0	1	1	0	0	20	85.00	65.38
Am	0	57	2	10	1	1	5	76	75.00	81.43
Gr	0	1	59	2	0	0	2	64	92.19	84.29
Me	7	6	5	79	1	0	2	100	79.00	79.00
Mi	1	0	0	6	67	0	0	74	90.54	95.71
Ms	0	2	0	1	0	17	0	20	85.00	94.44
Bs	1	3	4	1	0	0	71	80	88.75	88.75
Col	26	70	70	100	70	18	80	434		
OA = 84.56%, k = 0.81										

Table 4.6. Confusion Matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using first 20 PC bands.

LDA	Ac	Am	Gr	Me	Mi	Ms	BS	Row	UA (%)	PA (%)
Ac	16	6	0	9	0	0	1	32	50.00	61.54
Am	1	48	2	15	3	1	0	70	68.57	68.57
Gr	1	0	58	1	1	0	7	68	85.29	82.86
Me	7	10	4	57	5	0	2	85	67.06	57.00
Mi	0	0	1	13	61	0	2	77	79.22	87.14
Ms	0	0	0	1	0	16	0	17	94.12	88.89
Bs	1	6	5	4	0	1	68	85	80.00	85.00
Col	26	70	70	100	70	18	80	434		
OA = 74.65%, k = 0.70										
RF	Ac	Am	Gr	Me	Mi	Ms	BS	Row	UA (%)	PA (%)
Ac	20	2	0	0	0	0	0	22	90.91	76.92
Am	0	50	1	10	0	2	5	68	73.53	71.43
Gr	1	0	51	3	0	0	7	62	82.26	72.86
Me	3	11	10	78	8	1	1	112	69.64	78.00
Mi	0	0	2	4	61	0	0	67	91.04	87.14
Ms	0	0	0	0	0	15	0	15	100.00	83.33
Bs	2	7	6	5	1	0	67	88	76.14	83.75

Col	26	70	70	100	70	18	80	434		
OA = 78.80%, k = 0.74										
SVM	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	21	5	0	2	1	0	0	29	72.41	80.77
Am	1	51	1	9	0	1	5	68	75.00	72.86
Gr	1	0	61	2	0	0	2	66	92.42	87.14
Me	2	9	2	77	7	0	1	98	78.57	77.00
Mi	0	0	2	5	62	0	0	69	89.86	88.57
Ms	0	0	0	0	0	17	0	17	100.00	94.44
Bs	1	5	4	5	0	0	72	87	82.76	90.00
Col	26	70	70	100	70	18	80	434		
OA = 83.18%, k = 0.80										

Table 4.7. Confusion Matrix and accuracy statistics of testing datasets of lithological classification obtained by MLAs using JMIM based optimum bands.

LDA	Ac	Am	Gr	Me	Mi	Ms	BS	Row	UA (%)	PA (%)
Ac	16	4	0	3	0	0	2	25	64.00	61.54
Am	2	54	1	15	0	1	1	74	72.97	77.14
Gr	0	1	61	5	1	0	6	74	82.43	87.14
Me	5	3	2	61	2	0	0	73	83.56	61.00
Mi	2	0	1	7	67	0	0	77	87.01	95.71
Ms	0	0	0	1	0	17	0	18	94.44	94.44
Bs	1	8	5	8	0	0	71	93	76.34	88.75
Col	26	70	70	100	70	18	80	434		
OA = 79.95%, k = 0.76										
RF	Ac	Am	Gr	Me	Mi	Ms	BS	Row	UA (%)	PA (%)
Ac	14	2	0	1	0	0	1	18	77.78	53.85
Am	1	39	1	14	3	0	4	62	62.90	55.71
Gr	1	4	40	4	0	0	10	59	67.80	57.14
Me	2	14	9	71	10	0	5	111	63.96	71.00
Mi	5	8	1	5	55	1	0	75	73.33	78.57
Ms	0	0	0	0	0	17	1	18	94.44	94.44
Bs	3	3	19	5	2	0	59	91	64.84	73.75
Col	26	70	70	100	70	18	80	434		
OA = 67.97%, k = 0.61										
SVM	Ac	Am	Gr	Me	Mi	Ms	Bs	Row	UA (%)	PA (%)
Ac	19	2	0	1	0	0	0	22	86.36	73.08
Am	2	57	2	9	1	1	4	76	75.00	81.43
Gr	0	1	66	4	3	0	6	80	82.50	94.29
Me	2	4	1	77	0	0	1	85	90.59	77.00

Mi	2	0	1	3	66	0	0	72	91.67	94.29
Ms	0	0	0	1	0	17	0	18	94.44	94.44
Bs	1	6	0	5	0	0	66	78	84.62	85.71
Col	26	70	70	100	70	18	77	434		
OA = 85.48%, k = 0.83										

4.4.7 Impact of number and quality of training samples on MLAs

The number of training samples is crucial for obtaining an optimum classification result. A suitable number of training samples are always required by most of the MLAs (Maxwell et al., 2018). It is worth considering how different MLA performs while reducing the training datasets. To assess the impact of the number of samples on MLAs' performance, a reduction of 15%, 30%, and 45% of total training samples (training and testing datasets) have been used (Table 4.8). All the models were optimized after reduction for both training and testing datasets. The OA (Figure 4.10 (a)) and k statistic (Figure 4.10 (b)) of each MLA have been computed for training and testing datasets. Figure 4.10 and Table 4.8 highlight a slight reduction in all the models' accuracy statistics when training datasets are reduced by 15 to 45% of total datasets of each class (2170 for all the classes). The LDA, RF, and SVM have decreased the OA by ~5%, ~2%, and ~1 %. The model's sensitivity to the number of samples is crucial in selecting the MLA for geological application. It has been found that SVM is less sensitive to the size of training datasets as compared to RF and LDA.

To test the effects of mislabeling on different MLAs, an intentional mislabeling of 15%, 30%, and 45% of total datasets of each class has been prepared to test the model's sensitivity due to mislabeling (Figure 4.11 and Table 4.9). All the models were optimized after each mislabeling for both training and testing datasets. The OA (Figure 4.11 (a)) and kappa coefficient (Figure 4.11 (b)) of each MLA have been computed for training and testing datasets. A drastic reduction in the performance of all MLA has been observed when training datasets are mislabeled by 15 to 45%. The OA and k of training datasets of LDA, RF, and SVM decrease from 69.60 to 30% (OA) and 0.61 to 0.16 (k), 55.44 to 22.52%, and 0.46 to 0.07, 74.09 to 44.06% and 0.69 to 0.32 respectively (Table 4.9).

The testing datasets also show a similar drastic reduction in the performance of all the MLAs. The OA and k of testing datasets of LDA, RF, and SVM decrease from 69.75 to 35.25% (OA) and 0.64 to 0.22 (k), 55.43 to 22.58%, and 0.46 to 0.07, 74.36 to 46.31% and 0.69 to 0.35 respectively (Table 4.9). The RF and LDA completely fail in discriminating the lithological classes and have the highest impact of mislabeling and yield very poor coefficient statistic; whereas, the SVM shows the least impact of mislabeling.



Figure 4.9. Field photographs of different rock types were acquired during fieldwork in the study area at the validation points shown in Figure 4 (a). (a) exposure of granite in North-East of the Tupdoor area, (b) granite and amphibolite contact was taken from Wandli area, and (c) highly weathered metabasalt covered by black soil occurring in pedi-plain, which is flanked by amphibolite in North-West of Hosur area, (d) metabasalt dumped in Utti mine, (e) metabasalt, and (f) altered rock in Yetgal area.

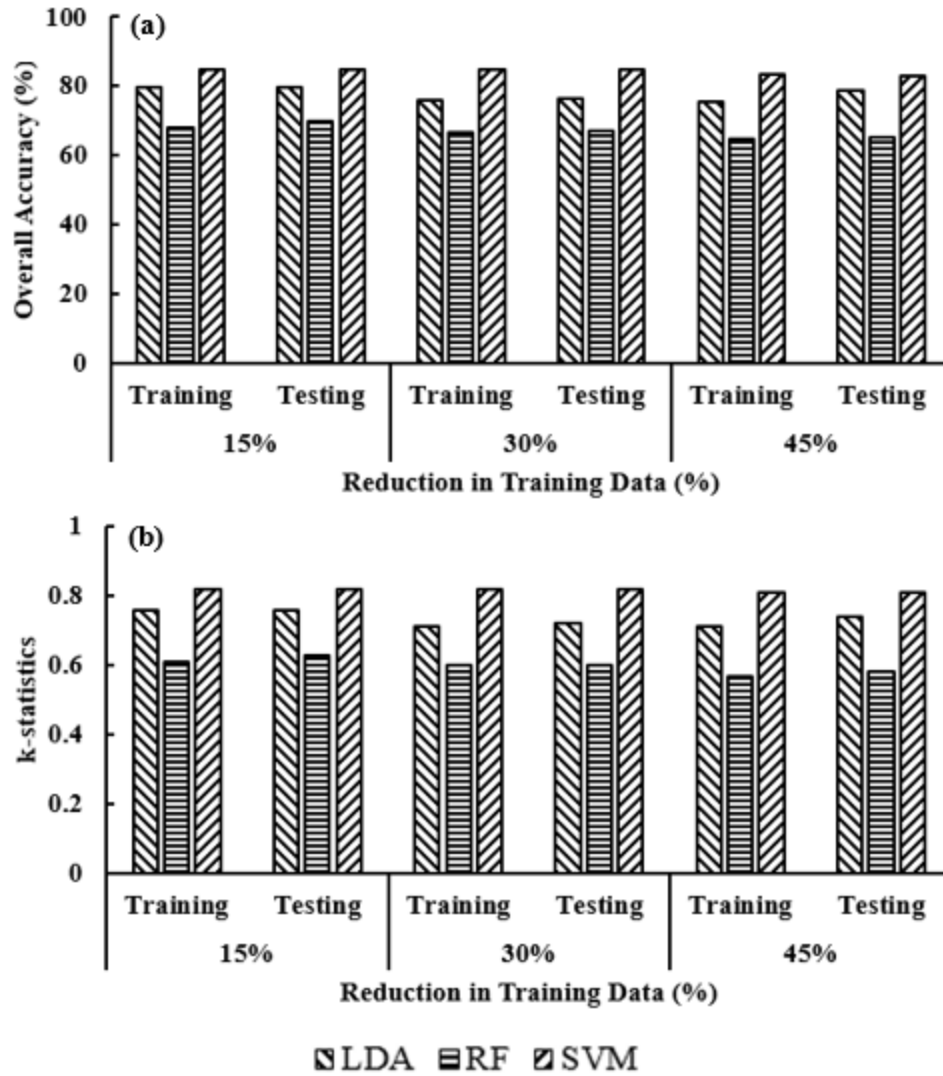


Figure 4.10. Reduction in sample datasets by 15%, 30%, and 45% and their corresponding accuracy statistics of training and testing samples (a) OA (%) and (b) k of different MLAs.

Table 4.8. Reduction in training datasets by 15%, 30%, and 45%, and their impact on the performance of different MLA. OA: overall accuracy and k: kappa coefficient.

	15%				30%			
	Training		Testing		Training		Testing	
	OA (%)	k	OA (%)	k	OA (%)	k	OA (%)	k
LDA	79.73	0.76	79.83	0.76	76.82	0.72	76.57	0.72
RF	67.86	0.61	69.89	0.63	66.70	0.60	67.33	0.60
SVM	84.89	0.81	84.74	0.81	84.70	0.81	84.82	0.81
45%								
	Training		Testing					

	OA (%)	k	OA (%)	k				
LDA	75.61	0.71	74.81	0.70				
RF	64.62	0.57	65.25	0.58				
SVM	83.47	0.80	83.16	0.80				

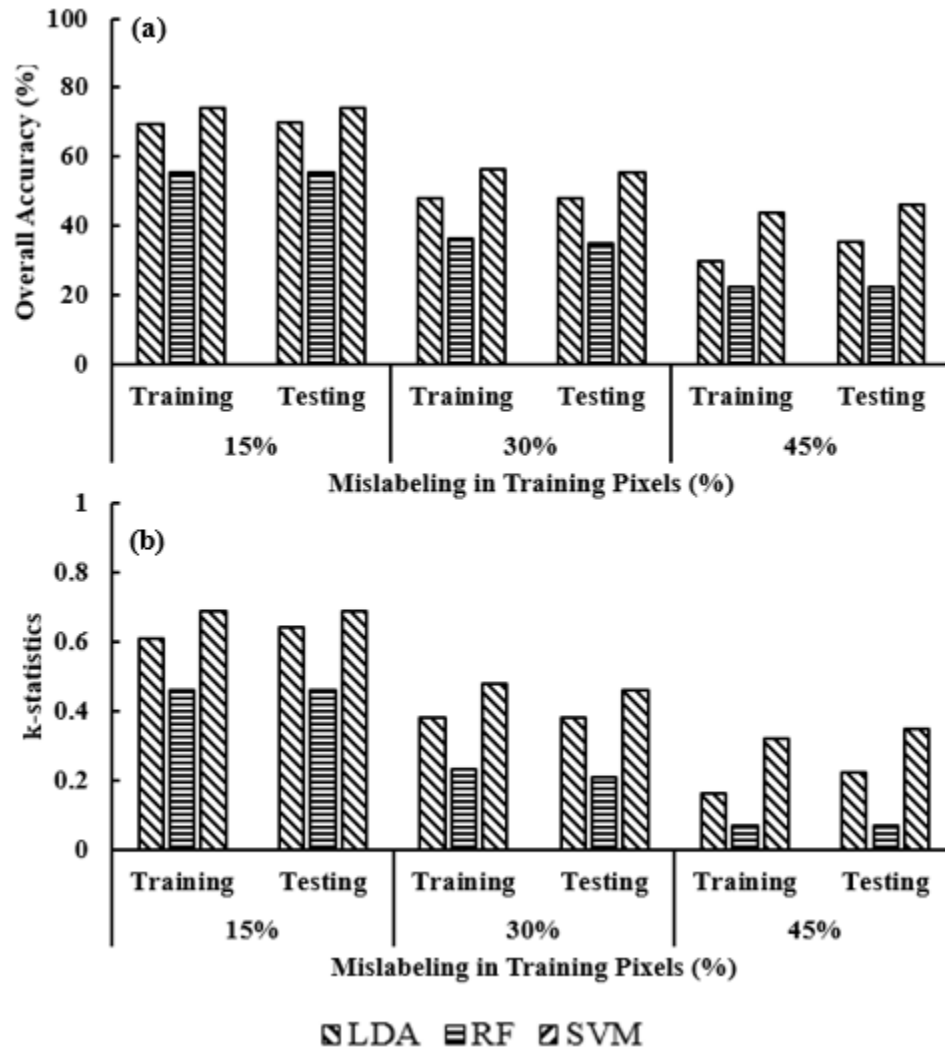


Figure 4.11. Mislabeling in sample datasets by 15%, 30%, and 45% and their corresponding accuracy statistics of training and testing samples. (a) Overall accuracy (%) and (b) kappa coefficient of different MLAs.

Table 4.9. Mislabeling in training datasets by 15%, 30%, and 45% and their impact on the performance of different MLA. OA: overall accuracy (%) and k: kappa coefficient.

	15%				30%			
	Training		Testing		Training		Testing	
	OA (%)	k	OA (%)	k	OA (%)	k	OA (%)	k
LDA	69.60	0.61	69.75	0.64	47.98	0.38	47.93	0.38
RF	55.44	0.46	55.43	0.46	36.52	0.23	35.02	0.21
SVM	74.09	0.69	74.36	0.69	56.34	0.48	55.3	0.46
	45%							
	Training		Testing					
	OA (%)	k	OA (%)	k				
LDA	30.00	0.16	35.25	0.22				
RF	22.52	0.07	22.58	0.07				
SVM	44.06	0.32	46.31	0.35				

4.5 Conclusions

- The study developed a novel approach of integrating spectral enhancement techniques and MLAs for an automated lithological mapping using AVIRIS-NG hyperspectral data. The high spectral and spatial resolution of AVIRIS-NG data successfully captured the minute variation of spectrally similar rock types and successfully classified in the study area.
- The conjugate utilization of conventional geology maps coupled with spectral enhancement products derived from PCA and ICA transformation using ASTER data provided a suitable reference lithology map for obtaining better training sample collection for lithological mapping using MLAs.
- The JMIM based optimum bands are more suitable than other input datasets (i.e. SWIR bands and all spectral bands) for obtaining an accurate lithological map with higher classification accuracy.
- The SVM outperforms other methods irrespective of the input dataset used in the classification. SVM appeared to be less sensitive to the number of samples and mislabeling or uncertainty in training datasets as compared to other ML models.
- The lithological map obtained using JMIM based optimum bands using SVM accurately classify most of the rock types, including other classes such as soil cover and mine site in the area. The high-resolution lithological map with accurate and distinct litho-contacts of amphibolite, metabasalt, and granite can be integrated with an alteration map of the area for targeting the potential zone of gold mineralization.

5 Multi-sensor datasets-based optimal integration of spectral, textural, and morphological characteristics of rocks for lithological classification using machine learning models

(This material has been submitted for possible publication in the “*Geocarto International, Taylor and Francis.*” Please refer to page XIV for the original paper link).

Abstract: This study presents a multi-sensor datasets integration of spectral, textural, and morphological characteristics of rock types for accurate lithological classification using Machine Learning (ML) models. We utilized multi-sensor datasets such as Advanced Spaceborne Thermal and Reflection Radiometer (ASTER), Advanced Land Observing (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR), Sentinel-1 (SAR), and Digital Elevation Model (DEM) of optical and microwave regions to derive different characteristics of rock surface. The derived characteristics have been used to prepare various input features such as spectral, spectral, and transformed spectral (principal and independent component transformed bands), morphological (SAR indices), textural, and optimum hybrid features to evaluate their ability to classify different rock types accurately (i.e., limestone, sandstone, shale, chert breccia, metabasalt, granite-gneiss, ultramafic rocks, and dike) found in the area. A normalized joint mutual information maximization filter-based feature selection method was used to select important variables for each input feature to improve the intra-rock variability and optimal classification results. We compared and evaluated the novelty of different input features for lithological classification using optimized Random Forest (RF) and Support Vector Machine (SVM) models. The classified lithological maps were assessed for their accuracy using the confusion matrix and field data. The comparative analysis illustrated that the optimal hybrid input features consist of important variables derived from different surface characteristics of rock types have shown great success over other input features to accurately classify different rock types found in the study area. The maximum accuracy statistics achieved in lithological classification by SVM (Overall Accuracy (OA): 77.78%; Kappa Coefficient (k): 0.74) and RF (OA: 76.77%; k: 0.73) using optimal hybrid input features, which is around 15% higher than as obtained using spectral input features. The SVM slightly outperforms across different input features to accurately classify different rock types. The higher accuracy of optimal hybrid input features illustrated the novelty of multi-sensor datasets-based integration over the commonly used spectral data alone approach for successful lithological classification using ML models.

5.1 Introduction

The spectral characteristics of rocks have been extensively utilized in lithological discrimination and mapping. Various multispectral and hyperspectral sensors have shown their ability to map the geological targets like rocks and minerals in different geological settings and climatic conditions (Cloutis, 1996; Kruse, 1998; Kruse et al., 2006; Gad &

Kusky, 2007; Tan et al., 2009; Amer et al., 2010; Bishop et al., 2011; Van der Meer et al., 2012; Pour & Hashim, 2014a; Zhang & Li, 2014; Kumar et al., 2015; Adep et al., 2016; Carrino et al., 2018; Feng et al., 2018; Ge et al., 2018; Bhattacharya et al., 2019b; Guha et al., 2020; Kumar et al., 2020a; Kumar et al., 2020b). However, lithological or rock-type classification becomes more challenging where the geological site consists of several rock types of similar mineral composition and weathered surfaces, which diminishes the intra-rock variability (Jones et al., 2004). A few recent studies demonstrated the advantages of integrating physical properties such as morphological and textural information with spectral properties of surface for improving the intra-rock variability and classification accuracy (Othman & Gloaguen, 2014; Wei et al., 2016; Masoumi et al., 2017; Othman & Gloaguen, 2017). The integration of multiple datasets has been successfully employed in a wide range of remote sensing applications, including land-use/landcover classification (De Martino et al., 2003; Jin et al., 2008), forest/vegetation mapping (Franklin et al., 2000; Dian et al., 2015; Seifi Majdar & Ghassemian, 2017), urban mapping (De Martino et al., 2003)(Dian et al. 2015), and lithological mapping (Wei et al., 2016; Masoumi et al., 2017; Othman & Gloaguen, 2017).

Different sensors operate in various parts of the electromagnetic spectrum, including optical and microwave regions can be effectively used to derive various physical properties of the surface for better material discrimination, identification, and classification (Lillesand et al., 2015). Microwave remote sensors operate at different frequencies, and polarization were successfully employed to derive the physical and dielectric properties of surface materials (Schmullius & Evans, 1997). Synthetic Aperture Radar (SAR) measures the backscattered energy, representing surface roughness and dielectric constant of surface (Hajnsek et al., 2003). The amount of surface roughness is mainly controlled by grain size and general topography. In contrast, the dielectric constant is highly influenced by volumetric moisture content and the degree of heterogeneity of sediments (Hajnsek et al., 2003). Imaging SAR has tremendous potential in surface discrimination (such as soil) based on surface roughness, grain size, volumetric moisture content, and heterogeneity (van Zyl et al., 1990).

The application of imaging SAR datasets is well established in the geological structural mapping (Pal et al., 2007; Pour & Hashim, 2014b, 2015; Guha et al., 2020) but rarely highlighted in lithological classification (Tan et al., 2009; Wang et al., 2018). Dual and full polarization SAR can extract surface physical properties such as surface roughness, grain size, and morphology (Wang et al., 2018; LIU et al., 2019). Furthermore, the topography of the surface can be a vital characteristic to be used in lithological mapping as different rock types may show different impacts of weathering and erosional processes, results in variability in morphology and textural characteristics (Othman & Gloaguen, 2014; Masoumi et al., 2017; Othman & Gloaguen, 2017). Morphological and textural characteristics can be used to infer the landforms and other physical properties of the surface due to physical weathering and erosional processes (Othman & Gloaguen, 2017).

The commonly exploited methods in lithological mapping include image enhancement techniques, pixel, and sub-pixel spectral mapping methods. The Principal Component

Analysis (PCA) (Gomez et al., 2005; Amer et al., 2010), Independent Component Analysis (ICA) (Kumar et al., 2015; Pour et al., 2019; Kumar et al., 2020b) and band ratios (Gad & Kusky, 2007) are most common spectral enhancement techniques, which have been extensively used to enhance the spectral contrast of different rock types. Several pixels and sub-pixel spectral mapping methods such as spectral angle mapper (Kruse et al., 1993), spectral information divergence (Chang, 1999), mixture tuned matched filtering (Boardman, 1998), and others have been employed on multispectral and hyperspectral datasets to classify different rock types (Harris et al., 2005; Goodarzi Mehr et al., 2013; Hadigheh & Ranjbar, 2013; Zhang & Li, 2014). The common limitation of spectral enhancement techniques is that they do not produce a labeled classification output and require a human intervention to demarcate the lithological boundary manually. This become more challenging particularly in a vegetated and weathered surface areas (Kumar et al., 2020b). Nevertheless, the main limitation of the spectral mapping methods is that they require a threshold value to obtain a final classification output, which is usually a trial and error approach, and become a more challenging task where there is a lack of comprehensive field data and visual interpretation of land classes are difficult (Kumar et al., 2020a; Kumar et al., 2020b).

In recent decades, Machine Learning (ML) algorithms have demonstrated remarkable success over the traditional classification methods in classification and prediction of geological targets and others (Thompson et al., 2001; Waske et al., 2009; Baykan & Yılmaz, 2010; Waske et al., 2010; Abedi et al., 2012; Wu et al., 2013; Cracknell & Reading, 2014; Othman & Gloaguen, 2014; Carranza & Laborte, 2015; Lary et al., 2015; Rodriguez-Galiano et al., 2015; Othman & Gloaguen, 2017). The ML algorithms utilize artificial intelligence derived from input features by analyzing their characteristics and pattern with minimal human intervention for decision-making (Michie et al., 1994; Mjolsness & DeCoste, 2001; Mohri et al., 2018). ML models' main advantages are that it can learn complex non-linear relationships and structure from the data itself and efficient in dealing with a higher number of variables without the assumption of data distribution (Zuo, 2017; Mohri et al., 2018). Furthermore, it can integrate many features derived from multi-sensor datasets, which can increase class separability and improve classification accuracy (Zuo, 2017). Understanding the variable significance and its contribution to increasing the intra-class variability is an important step toward optimal classification (Dash & Liu, 1997; Guyon et al., 2004). ML models' successful utilization highly depends on various factors, including quality and sufficient quantity of training datasets, selection of optimum and less correlated features, and optimal training parameters (Batista et al., 2004; Oommen et al., 2011; Kumar et al., 2020b). The feature selection is a crucial task mainly when one deals with many spectral bands or variables to assess their significance in improving the intra-class variability and classification accuracy (Yu & Liu, 2003; Guyon et al., 2004; Waske et al., 2010).

The lithological classification becomes a more challenging task in a complex geological setting and poor rock exposure (Jones et al., 2004). Furthermore, homogeneity in major mineral composition of different rock types and weathered surface cover makes their spectral characteristics less distinctive and illustrative to be used as a primary input features

in the lithological classification. To overcome this limitation, we propose an optimal integration of spectral, morphological, and textural characteristics of rock type surface derived from different sensors such as ASTER, PALSAR, Sentinel-1, and DEM for accurate lithological classification using SVM and RF models. We compare and evaluate the novelty of different input features derived from various datasets. It is worth to mention here that to the best of our knowledge, the integration of different SAR (i.e., L- band PALSAR and C-band Sentinel-1) with spectral and DEM datasets has not been explored in lithological classification using ML models. Furthermore, the unique contribution of the proposed research is that we are developing a multi-sensor data integration with optimum feature selection within the ML framework for lithological classification.

5.2 Location and geological setting

The study area is situated in the south-west part of the Chattarpur district in Madhya Pradesh, India (Figure 5.1 (a)). It is located between 24°38'0" to 24°40'0"N latitude and 79°42'0" to 80°0'0"E longitude. The False Color Composite (FCC) (Figure 5.1 (b)) and elevation map (Figure 5.1 (c)) prepared using ASTER spectral and DEM data exhibit overall landcover and topographical variation of the area. The major rock types found in the area are limestone, sandstone, Fe-bearing sandstone, shale, chert breccia, metabasalt, and granite-gneiss, as displayed in the geology map of the study area (Figure 5.1 (d)). The other major landcover found in the study area includes water bodies and sparse to moderate vegetation cover. The elevation of the study site ranges from 193m to 473m from the mean sea level.

5.3 Datasets

Four primary datasets such as (a) ASTER multispectral, (b) PALSAR, (c) Sentinel-1, and (d) Digital Elevation Model (DEM) of ALOS PALSAR were used to derive spectral, morphological, and textural features of rocks. A brief sensor specification of ASTER, PALSAR, Sentinel-1, and DEM is presented in Table 5.1.

The ASTER sensor records the reflected and emitted energy in the electromagnetic spectrum of VNIR, SWIR, and Thermal Infrared (TIR) at a ground resolution of 15m, 30m, and 90m, respectively. A cloud-free L1B (radiometric and geometric corrected) radiance at the sensor image of the VNIR and SWIR region (acquired on 15/03/2005) was accessed through NASA Earth Data Portal (<https://search.earthdata.nasa.gov/search>) and used to derive the spectral characteristics of rocks. The dual-polarization (horizontal transmitting and horizontal receiving (HH) and horizontal transmitting and vertical receiving (HV)) PALSAR level 1.5 (radiometrically terrain corrected) and DEM were obtained at the spatial resolution of 12.5m from the Alaska Satellite Facility (<https://asf.alaska.edu/>).

Table 5.1. A brief sensor specification of different datasets used in this study.

Sensor	ASTER multispectral	ALOS PALSAR (level 1.5)	Sentinel-1 GRD (level 1)	PALSAR DEM
features	Spectral	Morphological	Morphological	Textural
Wavelength /frequency	VNIR and SWIR	L band (1.2 GHz)	C band (5.405 GHz)	--
Spectral bands	3 (VNIR) and 6 (SWIR)	Dual polarization: HV and HH	Dual polarization: VV and VH	--
Resolution	VNIR: 15m and SWIR: 30m	12.5 m	10 m	12.5 m

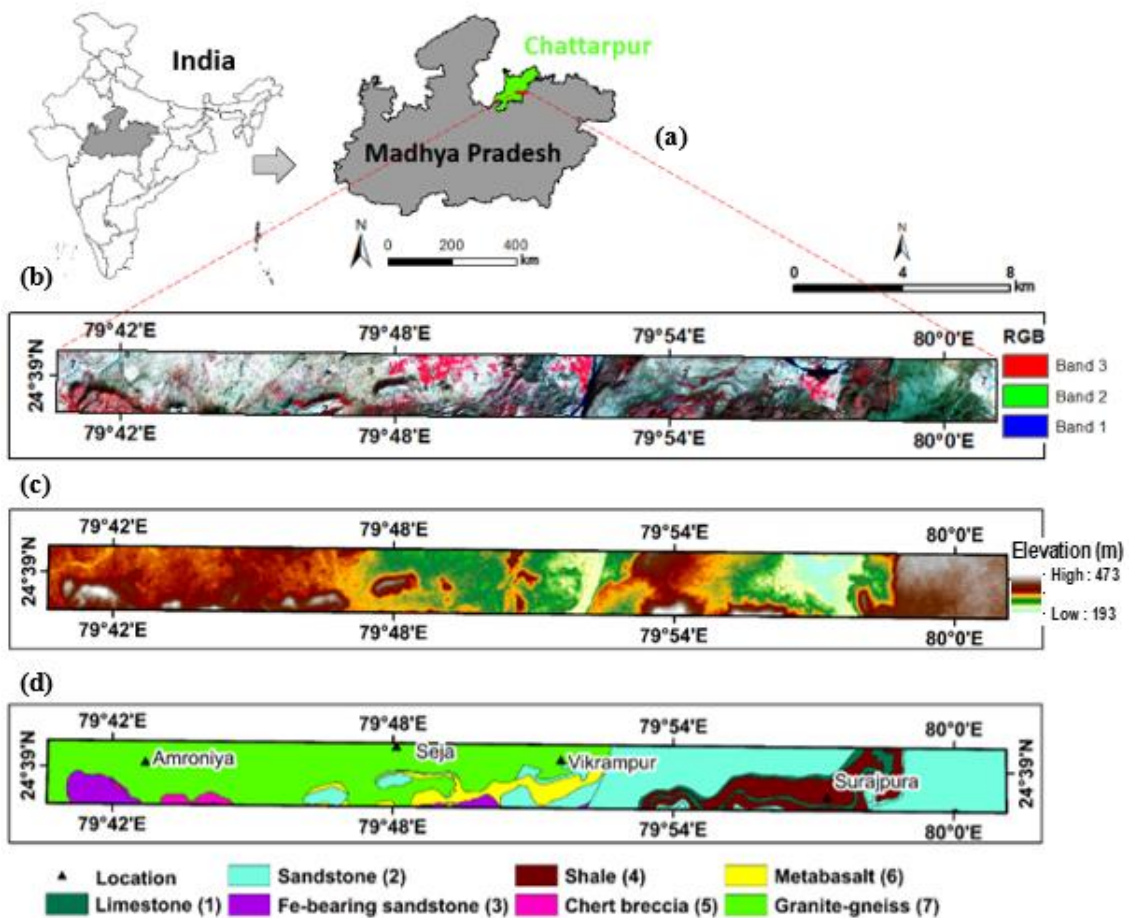


Figure 5.1. (a) Location of the study area, (b) a false-color composite map prepared from ASTER data showing overall landcovers, (c) DEM showing the overall topography, and (d) geology map exhibits major rock types found in the study area.

The dual-polarization (i.e., vertical transmitting and vertical receiving (VV) and vertical transmitting and horizontal receiving (VH)) Sentinel-1 level 1 Ground Range Detected (GRD) of full resolution (i.e., 10m) data were accessed from Sentinel Online Portal of the European Space Agency (<https://scihub.copernicus.eu/dhus/#/home>). The L1 GRD Sentinel products consist of focused SAR data that has been detected, multi-looked, and projected to ground range using an Earth ellipsoid model where the phase information is lost. The obtained product has approximately a square spatial resolution and square pixel spacing with reduced speckle (<https://sentinel.esa.int/web/sentinel/missions/sentinel-1/data-products>).

It is important to mention here that The PALSAR measures the backscattered energy at lower frequency (i.e., 1.2 GHz) as compared with Sentinel-1 (i.e., 5.4 GHz). Both the sensors measure the backscattered energy in different polarization. The multi-polarization and multi-frequency SAR sensors have better potential to extract the surface properties (Gstaiger et al., 2012; LIU et al., 2019) and, therefore, can provide useful information to increase the intra-class variability of different rock types. To the best of our knowledge there is no study making a concrete discussion on the utility of these two sensors for their suitability in the lithological classification. Furthermore, the conjugate utilization of different SAR datasets to discriminate different rock types remains unexplored in literature. The PALSAR, Sentinel-1, and DEM datasets were used to derive various morphological and textural characteristics from improving the intra-class variability of different rock types for accurate lithological mapping using ML models. All downloaded datasets are cloud-free and retained the same projection system.

The software packages and programming languages such as ENVI 5.5 (<https://www.harrisgeospatial.com/Software-Technology/ENVI>), SNAP 7 (<https://step.esa.int/main/toolboxes/snap/>), R 3.5 (<https://www.r-project.org/>), and ArcMap 10.7 (<http://desktop.arcgis.com/en/arcmap/>) were used for image processing and analysis of multispectral, DEM, SAR datasets, ML models' development and implementation, and preparation of GIS spatial layers.

The software packages and programming languages such as ENVI 5.5 (<https://www.harrisgeospatial.com/Software-Technology/ENVI>), SNAP 7 (<https://step.esa.int/main/toolboxes/snap/>), R 3.5 (<https://www.r-project.org/>), and ArcMap 10.7 (<http://desktop.arcgis.com/en/arcmap/>) were used for image processing and analysis of multispectral, DEM, SAR datasets (i.e., PALSAR and Sentinel-1), ML models' development and implementation, and preparation of spatial layers.

5.4 Methods

The methodology adopted to integrate the spectral, textural, and morphological characteristics of rocks derived from multi-sensor datasets for improved lithological mapping using SVM and RF models is presented in Figure 5.2 and discussed in the following sections.

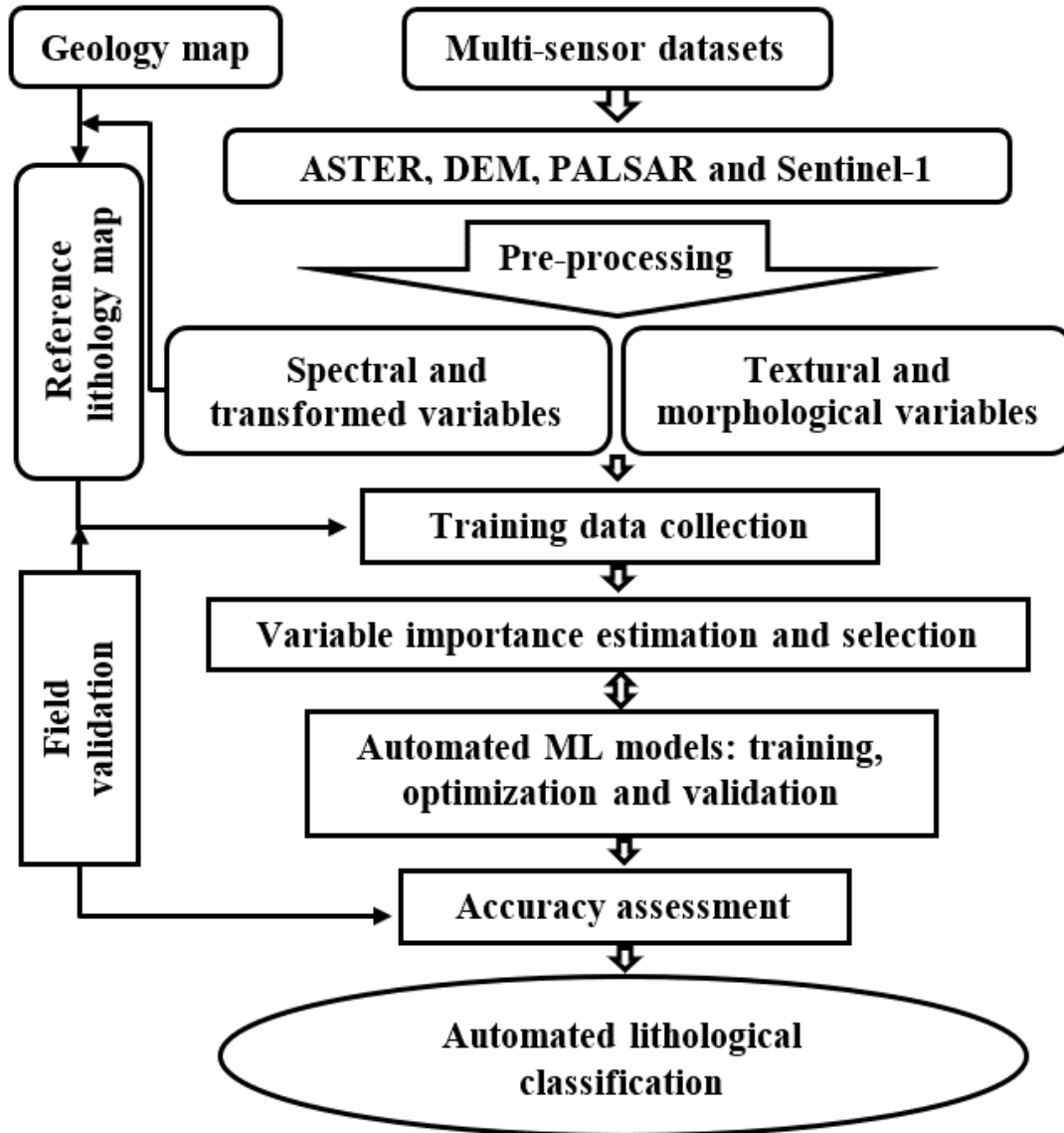


Figure 5.2. The workflow chart of the overall methodology adopted for improved lithological mapping.

5.4.1 Data pre-processing

The four datasets, such as ASTER, PALSAR, Sentinel-1, and DEM, were pre-processed to remove the preliminary noise for optimal results. The ASTER data were corrected for the cross-track illumination effect using the cross-track illumination correction tool of ENVI to reduce the impact of energy overspill from spectral bands 4 to 5 and 9 (Hewson et al., 2005). This was followed by layer stacking of VNIR and SWIR spectral bands and resampled their spatial resolution to 30 m ground resolution, radiometric correction, and conversion to reflectance using ENVI's radiometric calibration tool

(<https://www.harrisgeospatial.com/docs/RadiometricCalibration.html>). Subsequently, the log residuals correction method (<https://www.harrisgeospatial.com/docs/TDPerformAtmosphericCorrection.html>) was applied to enhance the spectral absorption features of rock-forming minerals (named hereafter processed ASTER data).

The PALSAR data were pre-processed for border noise and inherent speckles associated with SAR datasets using border noise correction and speckle filtering followed by conversion to the unit less backscattered value of the surface of the SNAP tool (Zuhlke et al., 2015). The Sentinel-1 GRD data were pre-processed by employing the following operations such as thermal and border noise correction, calibration, speckle filtering, range-Doppler terrain correction followed by conversion to the unit less backscattered value of surface using SNAP tool (Zuhlke et al., 2015). The DEM data were corrected for the sink using the Fill tool of spatial analyst of ArcMap to remove the minor imperfection in the elevation data (<https://desktop.arcgis.com/en/arcmap/10.7/tools/spatial-analyst-toolbox/fill.htm>). All the datasets were resampled at the spatial resolution of processed ASTER data (i.e., 30m) using the nearest neighborhood resampling method for integration and information extraction.

5.4.2 Preparation of reference lithology map

A suitable reference map is essential in selecting good training datasets for accurate lithological classification using any classification methods. The utilization of commonly used geology map as a reference map may not be an ideal source due to their large scale (i.e., 1:2,50,000 in this study) where the boundary of lithological units is usually an approximation (Grebby et al., 2011; Cracknell & Reading, 2014; Othman & Gloaguen, 2014; Bachri et al., 2019; Kumar et al., 2020b). Furthermore, the geological maps prepare based on the mineralogical composition of rock types and consider the stratigraphic succession and formation age. In contrast, remote sensing-based rock-type classification is usually based on rocks' surface properties (Kumar et al., 2020b). This warrants the modification in lithological units' boundary for good training data collection for remotely sensed data-based lithological classification. Kumar et al. (2020b) demonstrated the utility of PCA and ICA transformation for enhancing different rock types for preparing a suitable reference map and their subsequent utilization in training datasets collection. We effectively used the standard False Color Composite (FCC), and band combinations of PCA and ICA transformation derived using the processed ASTER data to accurately demarcate different litho-contacts to prepare a suitable reference map at a lower scale (i.e., 1:25,000). It is important to mention here that the interpretation of different band combinations of PCA and ICA transformation was carried out using the study site's geological map and prior field knowledge. Random stratified sampling was adopted to collect the geo-location of training datasets. A proportional and well-distributed samples as per the spatial extent of each lithological unit was collected to reduce the influence of class-imbalance on the performance of ML models (Oommen et al., 2011). The geo-location of samples was further used to derive the training datasets using different data sources.

5.4.3 Derivation of different input features

The derivation of suitable input features for lithological classification is crucial in obtaining optimal classification results. We have extracted various spectral, morphological, and textural features of rocks from processed ASTER, PALSAR, Sentinel-1, and DEM datasets to evaluate their ability to accurately classify lithological units found in the study site using RF and SVM models.

5.4.3.1 Spectral and transformed spectral input features

The spectral signatures of rock-forming minerals obtained in the wavelength range of VNIR-SWIR of the electromagnetic spectrum have been used as a primary input dataset in lithological mapping. The original spectral bands may not achieve great success in discriminating rock types of similar mineral composition and weathered surface, causing a reduction in rock types' spectroscopic characteristics. The utility of spectral enhancements such as PCA and ICA transformation was successfully applied on remotely sensed datasets for enhancing the spectral contrast of different rock types in various geological settings (Kumar et al., 2015; Pour et al., 2019; Kumar et al., 2020b). The advantage of spectral enhancement techniques is that it can increase the intra-class variability of lithological classes by removing the correlated spectral band and reducing the noise from the remotely sensed data. We applied PCA and ICA transformation on the processed ASTER data to derive transformed features (i.e., PCA and ICA transformed variables).

5.4.3.2 Morphological features

Spectral characteristics of rock types alone may not be an optimal dataset to discriminate the lithological units of similar mineral composition and in a weathered surface. However, different rock types show a different response to weathering, erosional process, and tectonics, which leads to diversified morphology, landforms, and surface textures (Othman & Gloaguen, 2014; Masoumi et al., 2017; Othman & Gloaguen, 2017). We derived several SAR indices using dual-polarization bands, i.e., HH and HV of PALSAR and VV and VH of Sentinel-1 data, to derive the morphological characteristics of the surface from being used in the lithological mapping. We derived several SAR indices to extract the morphological characteristics of the surface. However, we are not aiming to quantify and classify different morphology features rather interested in capturing the surface morphology of different rock types in SAR images of different polarization and frequency for improving lithological discrimination. The SAR indices have been successfully used to enhance the discrimination of different earth surface features, including vegetation, forest type, and landcover, which play an essential role in improving the classification accuracy (Schmullius & Evans, 1997; Walker et al., 2010; Dong et al., 2012; Qin et al., 2015; De Alban et al., 2018). The summary of SAR indices derived in this study using PALSAR and Sentinel-1 datasets is presented in Table 5.2.

5.4.3.3 Textural features

The utilization of textures as additional variables to improve rock types' intra-class variability has been recommended in recent studies (Masoumi et al., 2017). The texture of the rock surface can be useful in deriving the rocks' surface's physical properties, which are characterized by weathering and erosional processes. The texture represents the tonal variance in the spatial domain of an image (Chica-Olmo & Abarca-Hernandez, 2000). Haralick et al. (1973) proposed a grey-level co-occurrence matrix method to derive the texture and is commonly used in literature to derive the surface from various remotely sensed datasets. We computed six commonly used textural variables: mean, variance, homogeneity, entropy, second moment, and correlation (Haralick et al., 1973; Li et al., 2011). The commonly used dataset to derive the textural features is the spectral bands of the remotely sensed data of higher standard deviation (Masoumi et al., 2017). In this study, we derive the texture using ASTER spectral bands of higher standard deviation, i.e., band 6 and DEM, to evaluate their suitability for lithological mapping. To the best of our knowledge, the DEM has not been tested to extract the textural features and their subsequent utilization in lithological classification. We computed and compared the variables importance of textures derived from both the datasets using the Normalized Joint Mutual Information Maximization (NJMIM) feature selection method. A detailed description of NJMIM can be referred to in section 5.4.3.5. The selection of window size is crucial in obtaining the information from texture images (Masoumi et al., 2017). A variety of window sizes ranging from 3 by 3 to 65 by 65 was computed to find the optimal window size of different textural indices to characterize the surface of rock types for better discrimination and classification. We calculated the NJMIM score of textural images of the mentioned window sizes and selected the optimum based on the higher NJMIM score. The texture image yields a higher NJMIM score indicates their higher contribution in increasing the intra-class variability of different rock types for accurate lithological classification.

Table 5.2. Description of indices prepared using PALSAR (P: P₁₋₁₂) and Sentinel-1 (S: S₁₋₁₂) data to derive the surface's morphological features. HH: horizontal transmitting and horizontal receiving, HV: horizontal transmitting and vertical receiving, VV: vertical transmitting and vertical receiving, and VH: vertical transmitting and horizontal receiving.

SAR datasets	
PALSAR (P)	Sentinel-1 (S)
<i>HH</i> (P ₁)	<i>VH</i> (S ₁)
<i>HV</i> (P ₂)	<i>VV</i> (S ₂)
<i>HH/HV</i> (P ₃)	<i>VV/VH</i> (S ₃)
<i>HV/HH</i> (P ₄)	<i>VH/VV</i> (S ₄)
$(HH + HV)/2$ (P ₅)	$(VV + VH)/2$ (S ₅)
$HH - HV$ (P ₆)	$VV - VH$ (S ₆)
$(HH - HV)/(HH + HV)$ (P ₇)	$(VV - VH)/(VV + VH)$ (S ₇)
$(HH \times HV)/(HH + HV)$ (P ₈)	$(VV \times VH)/(VV + VH)$ (S ₈)
$(HH \times HV)/(HH - HV)$ (P ₉)	$(VV \times VH)/(VH - VH)$ (S ₁₉)

$HV - HH$ (P_{10})	$VH - VV$ (S_{10})
$(HH + HV)/(HH \times HV)$ (P_{11})	$(VV + VH)/(VV \times VH)$ (S_{11})
$(HH - HV)/(HH \times HV)$ (P_{12})	$(VV - VH)/(VV \times VH)$ (S_{12})

5.4.3.4 Training data collection and class separability computation

The quality and number of training datasets are crucial in achieving optimal results from any classification algorithm (Batista et al., 2004). The stratified random sampling was used to collect the geo-location of labeled samples using the prepared reference lithology map. A proportional number of training datasets by considering the spatial extent of different lithological units have been collected to reduce the influence of class imbalance on the performance of ML models (Batista et al., 2004; Oommen et al., 2011). A total of 2890 labeled training datasets were selected and subsequently divided into the proportion of training (75%, i.e., 2168) and testing samples (25%, i.e., 722) used in the development of SVM and RF models for lithological classification (Table 5.3). The geo-locations of labeled training samples were used to extract the pixel values of the aforementioned variables derived in this study.

The intra-class variability of different classes is crucial to understand the complexity of a classification problem and to find an optimal solution. To understand the intra-rock variability of spectral data, a dissimilarity matrix of spectral training datasets was prepared using Jeffries Matusita (J-M) distance statistics. The J-M distance values range from 0 to 2, where higher values indicate the optimal separability and vice-versa. The J-M distance method has been widely used in variable selection, class separability, and spectral discrimination and classification (Padma & Sanjeevi, 2014b).

Table 5.3. A summary of lithological units, areal extent, number of training, and testing datasets used in lithological classification using ML models.

Lithological units	Area (km²)	Training samples	Testing samples
Limestone	12.44	300	100
Sandstone	22.22	375	125
Fe-bearing sandstone	14.94	338	113
Shale	5.91	191	64
Chert breccia	0.77	150	50
Metabasalt	2.52	169	56
Granite-gneiss	42.06	420	140
Ultramafic	0.022	113	38
Dike	0.11	113	38

5.4.3.5 Variable importance measure

The feature selection is an essential task mainly when dealing with a large number of variables derived from various data sources to remove correlated and less suitable variables

to improve the performance of ML models (Archibald & Fann, 2007; Pal & Foody, 2010). The feature selection method also reduces the model complexity, computation cost and improves the model generalization. In this study, the Normalized Joint Mutual Information Maximization (NJMIM), a filter-based feature selection method, was used to measure variable importance. It uses ‘mutual information’ to derive the dependency of variables and the ‘maximum of the minimum’ criterion to select the optimum variables (Bennasar et al., 2015). The NJMIM technique was executed in “R” using the “praznik” package (<https://github.com/mbq/praznik>) to measure the variable importance of different input features derived from multi-sensor datasets. The NJMIM method provides the importance value (i.e., score) of each variable where a higher value indicates a more significant variable and vice-versa. The filter-based feature selection method's advantage is that it measures the variable importance outside the model or classification method and computational efficient.

To assess the novelty of different input features obtained from multi-sensor datasets, we applied the NJMIM method to measure the importance of different variables. Subsequently, to find the optimum number of variables for each input feature, we implemented SVM and RF with tuned parameters and marked their accuracy statistics. A set of variables (i.e., 5, 10, 15, and so on) provide the highest accuracy statistics and show no significant improvement in accuracy statistics of classification results. It has been considered an optimum number of variables for that particular input features. It is worth emphasizing that it is crucial to select the important and uncorrelated variables and remove the correlated variables from the input features to achieve optimal results and assess their novelty for lithological classification. The feature selection-based variables selection of different input features is important because few variables could decrease the intra-class variability of rocks and reduce the accuracy statistics and mislead in assessing their novelty for lithological classification.

It is worth mentioning here that although the feature selection method does a great job in selecting the important variable and reducing the correlated and less significant variables, but lacks in providing a descriptive view on the contribution of a variable to improve the specific class-separability. To overcome this limitation, we computed the boxplot of different variables that yield a higher NJMIM score for each input feature. This helps in understanding how other variables of each input feature contribute to improving the intra-class variability. A boxplot is a practical approach of graphically exhibits the structure of numerical data through their quartiles. The other advantage of using the boxplot is that it displays the variation in training datasets without making any assumptions of the inherent statistical distribution (Williamson et al., 1989).

5.4.3.6 Machine learning algorithms

The ML models were implemented using the “CARET” package available in the “R” programming language (Kuhn et al., 2017). The training datasets derived from different features were utilized to train, optimize, and validate the RF and SVM models. A brief framework of RF and SVM algorithms has been discussed below.

- a. RF: The RF algorithm is an ensemble method introduced by Breiman (2001) that employs a large number of decision trees to classify an image (Masoumi et al., 2017). It can solve both regression and classification problems with adequate accuracy (Grinand et al., 2013). The success of RF highly depends on the way it creates its decision trees. The tree is planted on the new training dataset by using random training data selection. The new training dataset is obtained using the bagging method, which generates the training dataset by randomly obtaining with the substitution of the original training dataset used for each feature/variable or feature set selected (Pal, 2005). The training dataset not utilized in training the RF model is known as out-of-bag data and used to compute an independent assessment of the overall accuracy of the RF model (Maxwell et al., 2018).
- b. SVM: The SVM algorithm was proposed by Vapnik (1999) developed based on statistical learning concepts. The SVM focuses on the training dataset nearest in the feature space to provide an optimum decision boundary (also known as hyperplane) to separate different classes (Schölkopf et al., 2002; Pal, 2008). The binary SVM develops a single separable hyperplane that maximizes the boundary between two classes. The multi-class SVM is an extension of binary SVM, which is prepared by repeatedly applying different binary SVM to each possible combination of classes (Pal, 2008; Seifi Majdar & Ghassemian, 2017; Maxwell et al., 2018). The SVM algorithm provides the flexibility of employing different kernel functions such as linear, polynomial, sigmoid, and radial basis for projecting the feature spaces to the higher dimensions by assuming that a linear boundary can be achieved in higher dimensional feature space (Pal, 2008). The choice between hyperplane and misclassification errors is contained by cost, which is a positive user-defined parameter and needs to be selected carefully for an optimum classification. A higher value of cost produces a more complex hyperplane and less generalized model and vice-versa (Maxwell et al., 2018).

Both RF and SVM have been successfully used in the lithological mapping (Yu et al., 2012; Cracknell & Reading, 2014; Lary et al., 2015; Othman & Gloaguen, 2017; Maxwell et al., 2018; Bachri et al., 2019; Kumar et al., 2020b). The selection of optimum training parameters is crucial to achieving optimal classification results from any ML models. Several optimization methods, such as grid search, simulated annealing, and genetic algorithms, are found in the literature (Bergstra et al., 2011; Bergstra & Bengio, 2012; Luo, 2016). A grid search optimization method is commonly employed and implemented with n-fold cross-validation (n-FCV) (Kohavi, 1995) for optimal parameter selection in this study. The SVM-radial basis function consists of two optimization parameters, such as cost and sigma. In contrast, the RF algorithm consists of one optimization parameter, i.e., the number of trees to grow (Kuhn et al., 2017). The SVM-radial basis function and RF were optimized using the 10-fold cross-validation for optimal results.

5.5 Results and discussion

5.5.1 Reference lithology map

The geology map (Figure 5.1 (d)) prepared at a coarser scale (i.e., 1:2,50,000 scale) displays a generalized boundary of litho-contact and may not be an ideal source to derive the geo-location of the training datasets. The PCA and ICA transformation applied to processed ASTER data to derive the enhanced spectral products. The FCC (R:850nm, G:650nm, and B: 550nm) of ASTER data (Figure 5.2 (b)), the band combinations of PCA (R: PC4, G: PC3, B: PC2) (Figure 5.3 (a)), and ICA (R: IC4, G: IC3, B: IC1) (Figure 5.3 (b)) transformation in conjunction with geology map and prior field knowledge successfully used to enhance various lithological units and interpretation, which was further used to prepare a suitable reference lithology map (Figure 5.3 (c)). Both PCA and ICA transformation complement each other in the accurate demarcation of different rock types found in the area.

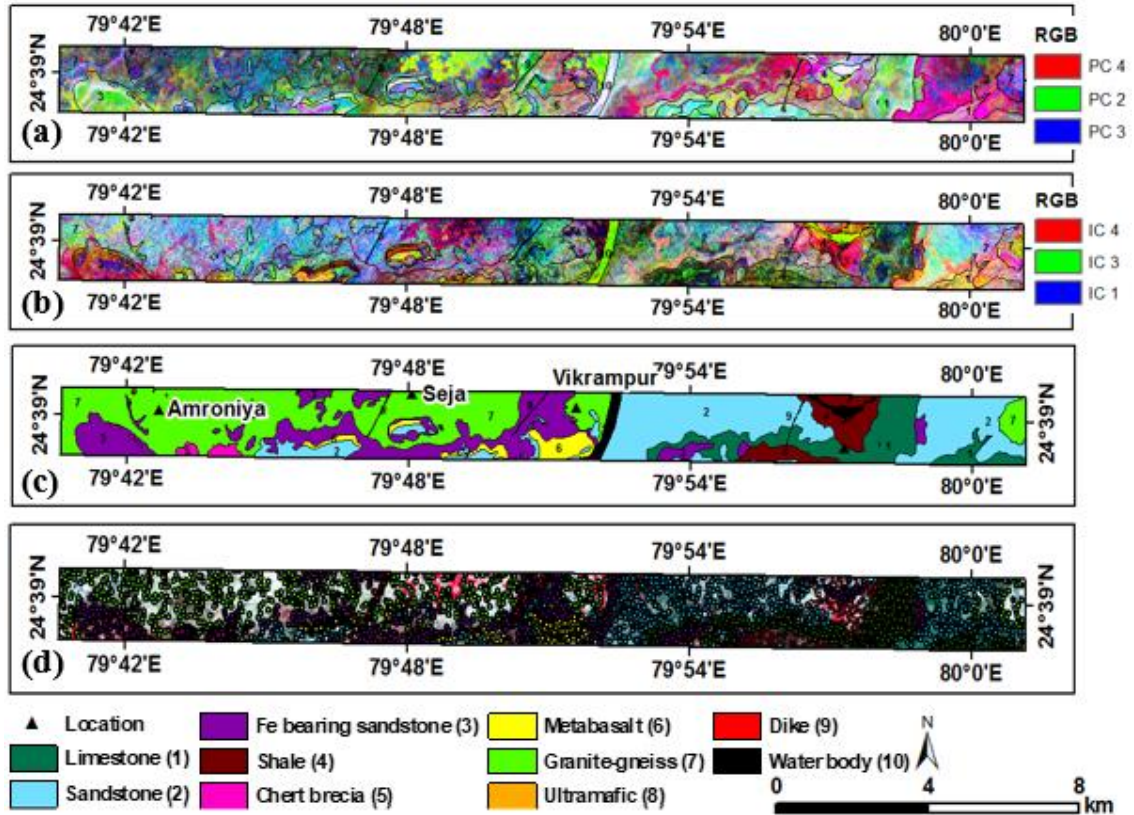


Figure 5.3. The enhanced spectral products are discriminating different rock types found in the study area. (a) band combination of PCA transformation (R: PC4, G: PC2, B: PC1), (b) band combination of ICA (R: IC4, G: IC3, B: IC1) derived from processed ASTER data, (c) reference lithology map, and (d) the geo-location of training samples overlaid on the false-color composite of ASTER data.

However, PCA provides better spectral contrast between granite-gneiss and sandstone, whereas ICA shows a better spectral contrast between chert breccia and granite-gneiss. The derived reference lithology map prepared at the scale of 1:25,000 was further used in collecting the geo-location of training samples (Figure 5.3 (d)). We successfully demarcated a few rock types, such as dike and ultramafic, which were confirmed in the field but not located in the area's geology map. A zoomed view of the band combination of PCA and ICA transformation highlighting the spectral contrast of different rock types presented in Figure 5.4. Figure 5.5 displays the field photographs of a few rock types such as (a) sandstone of the Kaimpur group in the Vikrampur area, (b) granite of Bundelkhand granitoid complex captured in Amroniya area, and (c) shale of Semri group.

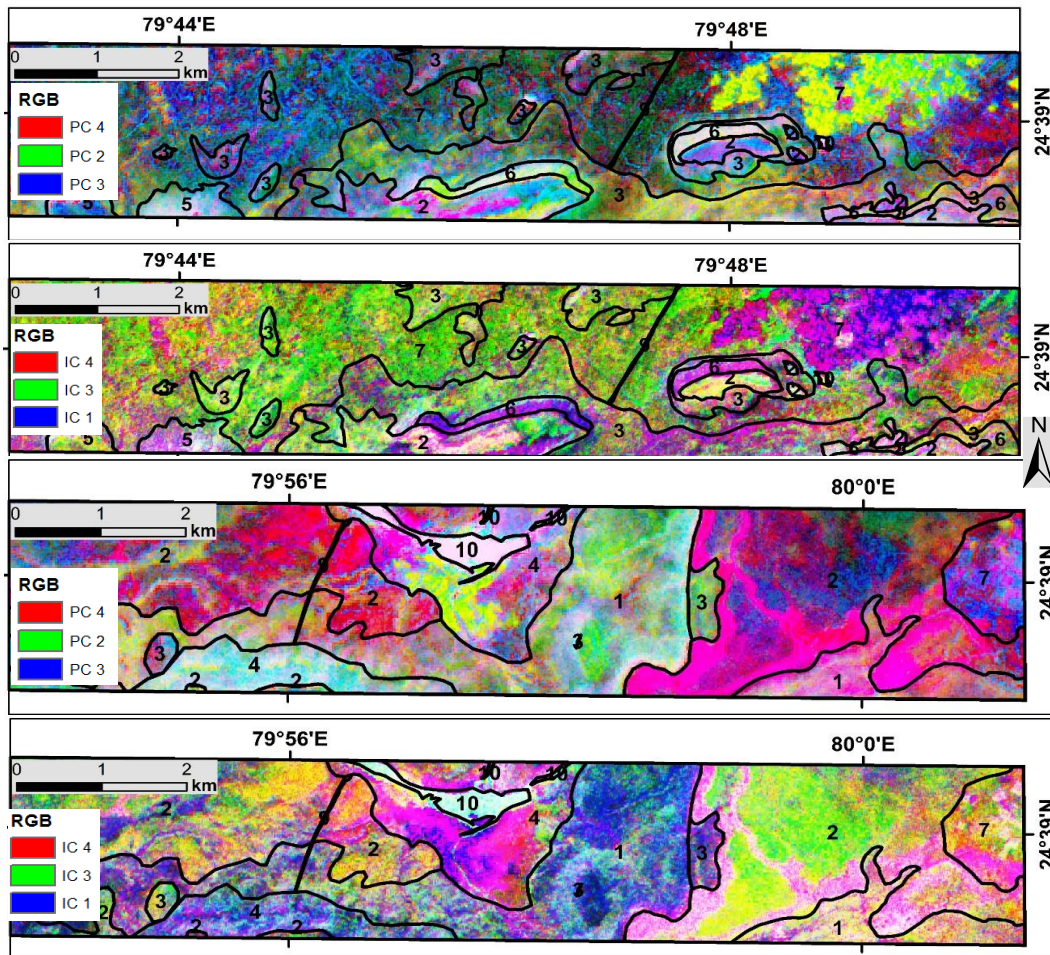


Figure 5.4. The zoomed view of PCA and ICA transformation enhanced spectral products discriminating different rock types found in the study area. 1: limestone, 2: sandstone, 3: Fe-bearing sandstone, 4: shale, 5: chert breccia, 6: metabasalt, 7: granite-gneiss, 8: ultramafic, 9: dike, and 10: water body.

5.5.2 Spectral characteristics of rocks

The spectral signatures of different rock types are important in assessing the major rock-forming minerals. The diagnostic spectral signatures caused due to electronic and vibrational processes in the VNIR-SWIR were captured with some limitation in ASTER spectral data due to their coarser spectral and spatial resolution. Figures 5.6 and 5.7 present the reflectance spectra and boxplots of different rock types derived using average training datasets of processed ASTER data. The majority of rock types found in the area show absorption feature around 0.95 μm , 2.2 μm , and 2.30-2.35 μm caused due to Fe^{3+} metals, AL-OH, and Fe/Mg-OH molecules (Clark, 1999) respectively. The spectral curve of ultramafic, granite-gneiss, and dike exhibits a distinguishable absorption feature around 2.2 μm and 2.3 μm due to Al-OH and Fe/Mg-OH molecule; whereas, limestone, sandstone, Fe-bearing sandstone, shale, chert breccia, and metabasalt don't exhibit any dominant absorption feature around 2.2 μm . Three rock types, i.e., ultramafic, granite-gneiss, and dike, do not show a distinct absorption feature around 2.3 to 2.35 μm .



Figure 5.5. Field photographs of a few rock types at the selected traverse captured during the field visit in the study area. (a) the sandstone of the Kaimpur group in the Vikrampur area, (b) granite of Bundelkhand granitoid complex captured in the Amroniya area, (c) shale of Semri group.

The average reflectance of different rock types showing considerable variability, which enables their discrimination and classification. Figure 5.7 presents the boxplot of different rock types found in the study area. The boxplot of different rock types illustrates that there is considerable overlap in most of the rock types' quartiles, which indicates their similarities. The metabasalt and limestone show the least average reflectance, whereas ultramafic and granite-gneiss show maximum overall average reflectance.

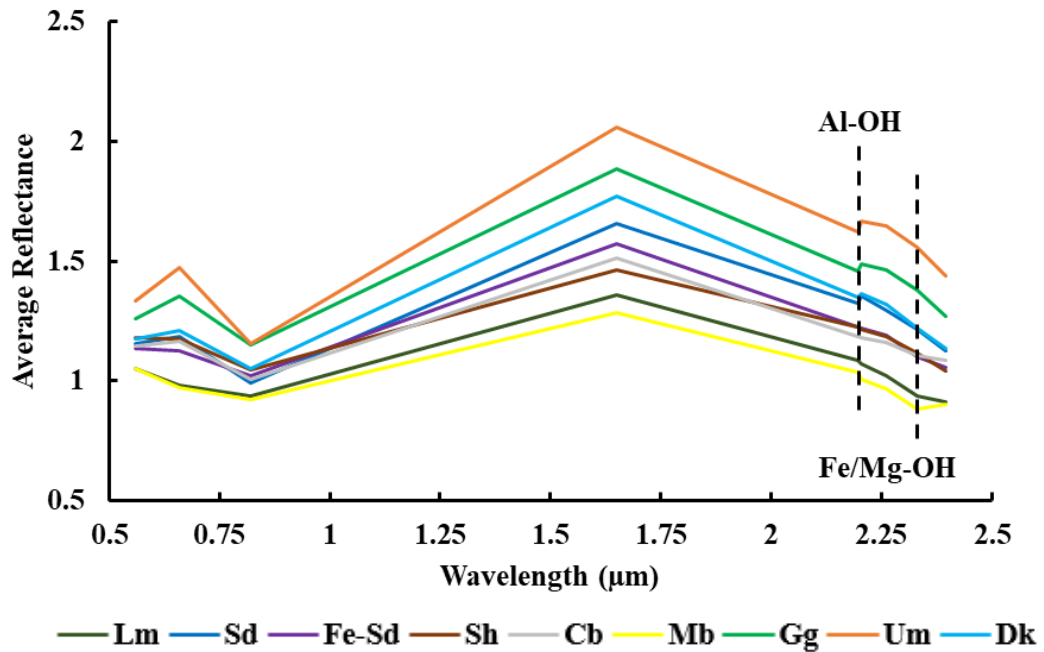


Figure 5.6. The overall average spectra of different rock types derived from processed ASTER data. [Dashed line indicates the diagnostic absorption position]. Lm: limestone, Sd: sandstone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Cb: chert breccia, Mb: metabasalt, Gg: granite-gneiss, um: ultramafic rock, Dk: dike.

The overall average reflectance of chert breccia, Fe-bearing sandstone, and shale is higher than metabasalt and limestone but less than other rock types, including dike, granite-gneiss, sandstone, and ultramafic rocks. The overall average reflectance of granite-gneiss is less than ultramafic but higher than other rocks found in the study area. It is important to note here that chert breccia, Fe-bearing sandstone, and shale show a similar median and shows higher overlaps of quartiles of overall average reflectance and exhibit less intra-class variability. Similarly, the dike and sandstone yield a similar median and shows higher similarity among them. The median and quartiles of ultramafic and granite-gneiss exhibit higher dissimilarities in median and quartiles. The boxplot of rock indicates the least similarity, and overlap in the median and quartiles, respectively, of other rocks, indicates higher dissimilarities and can cause less-false positive and vice-versa.

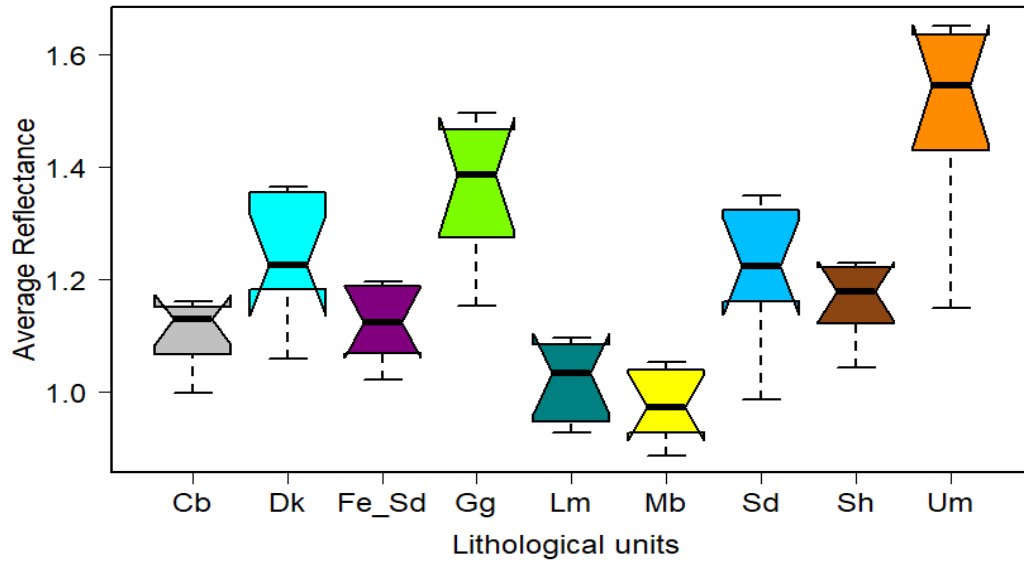


Figure 5.7. A boxplot of overall average spectra of different rock types derived from the training dataset of processed ASTER data.

5.5.3 Spectral-intra-class variability

The intra-class variability of different rock types derived from the spectral input features derived from the training datasets of processed ASTER data using spectral dissimilarity matrix computed using the J-M distance method (Table 5.4).

Table 5.4. Spectral dissimilarity matrix of overall average reflectance of training datasets derived from processed ASTER data using the J-M distance measure. Lm: limestone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Mb: metabasalt, Gg: granite-gneiss, Um: ultramafic rock, and Dk: dike.

	Sd	Fe-Sd	Sh	Cb	Mb	Gg	Um	Dk
Lm	0.4409	0.2343	0.2862	0.2198	0.0263	0.9081	1.1769	0.5162
Sd	--	0.0554	0.1295	0.1077	0.6225	0.1521	0.4232	0.0120
Fe-Sd		--	0.032	0.0122	0.3960	0.3699	0.6895	0.1100
Sh			--	0.0078	0.4718	0.4946	0.8315	0.2083
Cb				--	0.3855	0.4718	0.8031	0.1800
Mb					--	1.0905	1.3227	0.6920
Gg						--	0.1008	0.0850
Um							--	0.3083

The accuracy of a classification result highly depends on the intra-class variability of different classes. The spectral dissimilarity matrix can be easily interpreted by considering each row of Table 5.4. For example, while considering the row of limestone, the least dissimilarity value yields by metabasalt (i.e., 0.0263), which illustrates that there is a

higher probability while classifying limestone that metabasalt can yield false-positive as compared with other classes. Similarly, while considering the row of sandstone, the least dissimilarity value yield by the dike (i.e., 0.0120), Fe-bearing sandstone (0.0554) and chert breccia (0.1022), which means that while classifying sandstone there is a higher probability of mixed-up with dike, Fe-bearing sandstone, and chert breccia, and can cause false positive. Similarly, the Fe-bearing sandstone yields less dissimilarity value with chert-breccia and shale as compared with other classes. Shale shows higher similarity with chert breccia (i.e., 0.0078). It can be noticed that ultramafic and granite-gneiss show higher intra-class variability among the rock types found in the study area and have less probability of miss-classification than other classes.

5.5.4 Variable importance and selection

The selection of optimal variables is crucial in improving the intra-class variability of rocks for accurate lithological classification using ML models. The variables importance of spectral input feature derived from processed ASTER data is shown in Figure 5.8. The spectral band 7 yields the highest NJMIM score, which indicates its maximum contribution in discriminating various rock types. The other eight spectral bands of ASTER data produce similar NJMIM scores and show approximately equal contribution in discriminating different rock types found in the area.

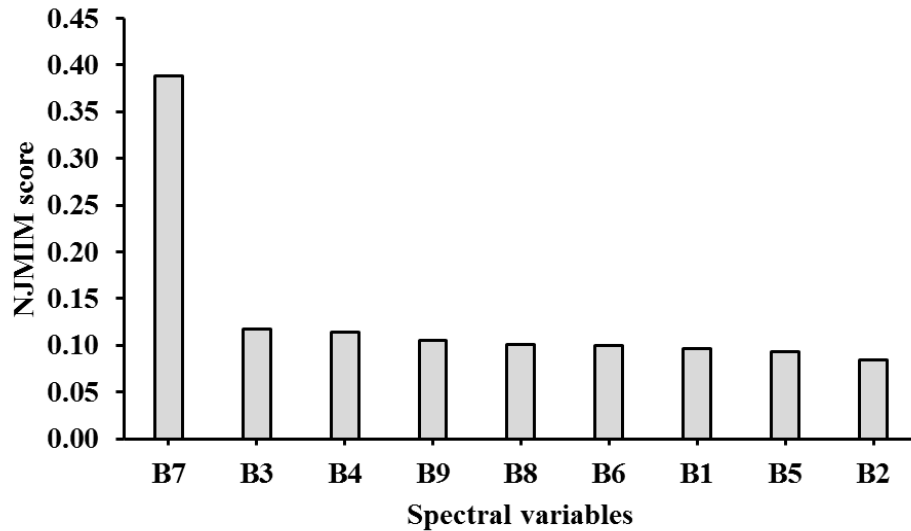


Figure 5.8. Variable importance of spectral input features derived from processed ASTER data. B: spectral bands of VNIR and SWIR.

The variable importance of transformed spectral features derived from PCA and ICA transformation using processed ASTER data is shown in Figure 5.9. The first five important variables of transformed spectral features include PC1, PC6, IC2, PC5, and PC2. The first few PCA and ICA transformation components consist of more information content than the last components. It is interesting to note here that the few lower-order

components of PCA and ICA transformation yield higher NJMIM scores than higher-order components. For example, PC6 and PC5 yield a higher NJMIM score than PC2 and PC4. In other words, the contribution of PC6 and PC5 are more in increasing the intra-class variability than PC2 and PC4. Similar observations can be noticed with the component of ICA transformation. The IC2 yields higher NJMIM score than IC5, IC1 and other IC components.

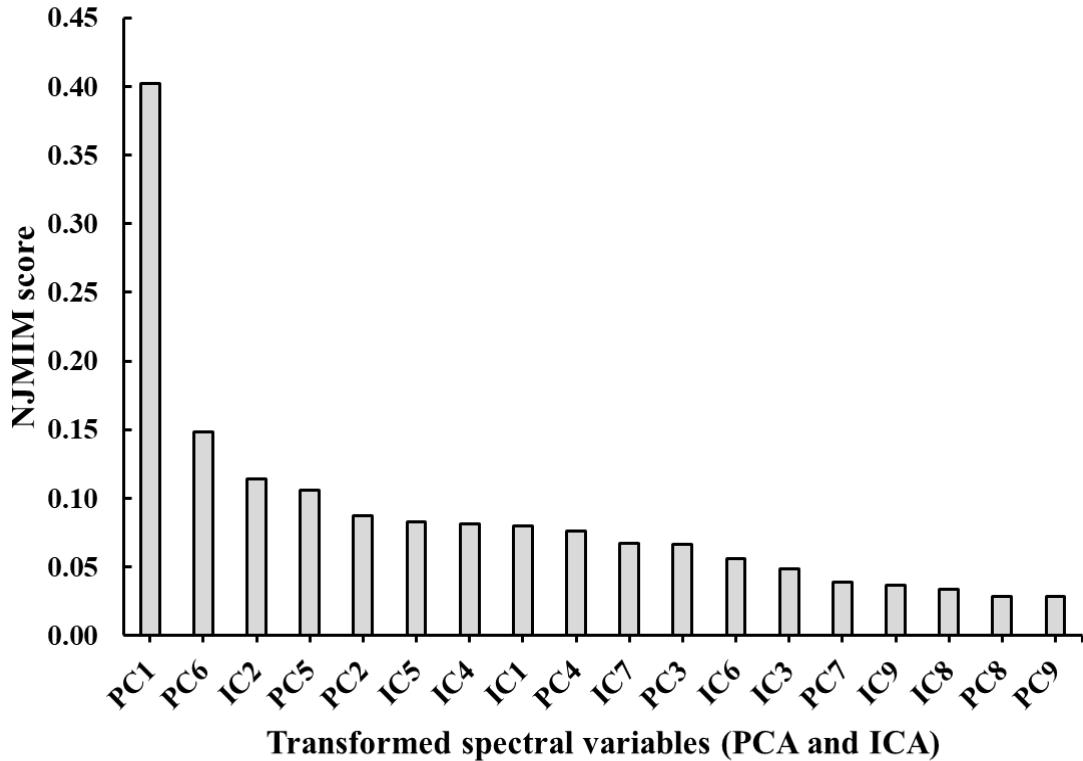


Figure 5.9. Variable importance of transformed spectral input features derived from processed ASTER spectral data using PCA and ICA transformation.

The variable importance of morphological features extracted from various indices (Table 5.2) derived from dual-polarization PALSAR and Sentinel-1 SAR datasets are displayed in Figure 5.10. Most of the index derived from PALSAR data yields a higher NJMIM score than indices derived from Sentinel-1 data. Four out of the first five important variables are derived from PALSAR data. Nevertheless, the VH band of sentinel-1 data marked as the second most important variable. The lower score of indices derived from the Sentinel-1 data indicates their less contribution in discriminating different rock types found in the area. One possible reason for the less effectiveness of Sentinel-1 data in discriminating different rock types is their lower wavelength compared with PALSAR data (Table 5.1). The higher wavelength of PALSAR data with higher penetration ability effectively extracts the surface's diversified morphological characteristics. It hence contributes to increasing the intra-class variability of different rock types. The SAR indices P₈ of PALSAR (Figure 5.11 (a)) and S₁ of Sentinel-1 ((Figure 5.11 (b)) data are the two best indices yields higher

NJMIM score among all the SAR indices indicates their suitability in lithological discrimination. The qualitative comparison of Figure 5.11 demonstrates that index P₈ has a better ability to discriminate different rock types as compared with S₁. The spectral band 6 of processed ASTER data and DEM were used to extract the textural features in this study. The DEM outperforms the spectral band 6 of ASTER data in extracting the textural features to increase the intra-class variability of different rock types.

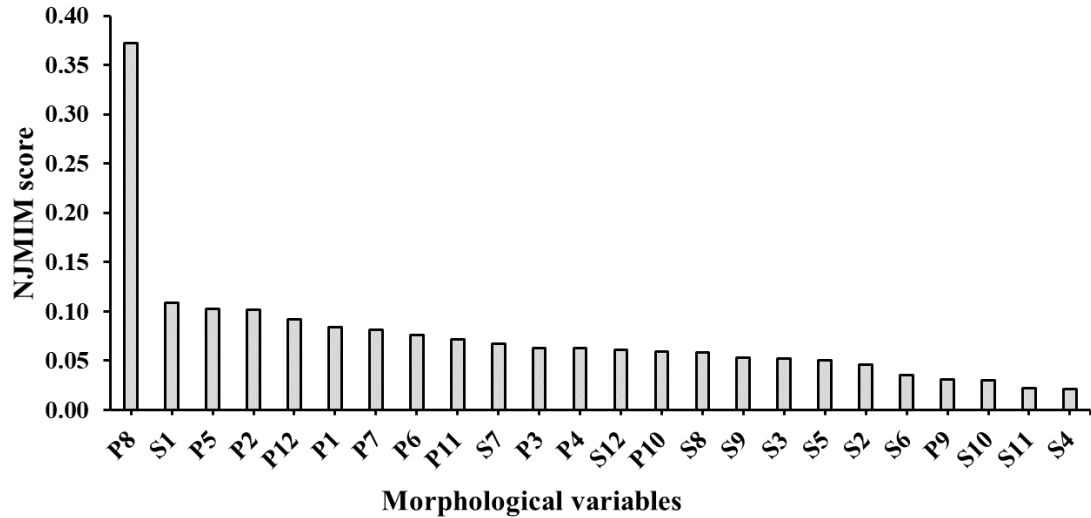


Figure 5.10. Variable importance of different morphological input features derived from PALSAR (P) and Sentinel-1 (S) data.

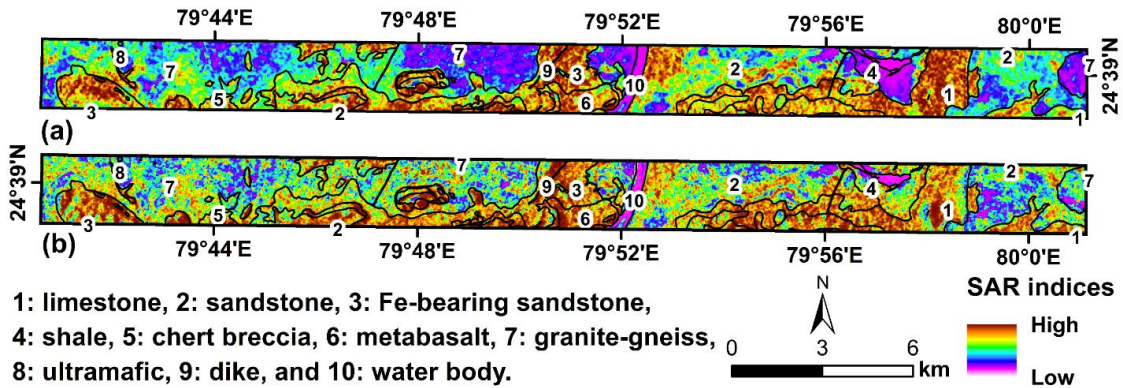


Figure 5.11. SAR indices (a) P₈ and (b) S₁ derived from PALSAR and Sentinel-1 data respectively displaying the backscattered energy of different rock types found in the area.

The variable importance of textural features such as correlation, mean, entropy, variance, second moment, and homogeneity derived from DEM with optimum window size is displayed in Figure 5.12. Table 5.5 presents different textural features and their corresponding derived optimum window size. The DEM has been considered as a textural variable and included in the computation of the NJMIM score. It is essential to highlight here that most of the textural features, including DEM, yield a higher NJMIM score than

other input features (such as transformed spectral and morphological input features). The DEM and correlation yield a higher NJMIM score than other textural features indicating their higher contribution in increasing the intra-class variability of different rock types found in the area. Figure 5.13 presents the three important textural variables, such as DEM (Figure 5.13 (a)), correlation (Figure 5.13 (b)), and entropy (Figure 5.13 (c)), illustrating their ability to discriminate different rock types found in the area. Other rock types show considerable variability with the elevation. The mean elevation of chert breccia and sandstone is higher, whereas shale shows the least mean elevation than other rock types found in the study area. The Fe-bearing sandstone, granite-gneiss, and ultramafic rocks show similar mean elevation slightly higher than the dike, limestone and metabasalt.

Table 5.5. Textural features derived from optimum window size using DEM.

Textural variables	Optimum window	Textural variables	Optimum window
Correlation	17×17	Variance	17×17
Mean	11×11	Second moment	21×21
Entropy	13×13	Homogeneity	29×29

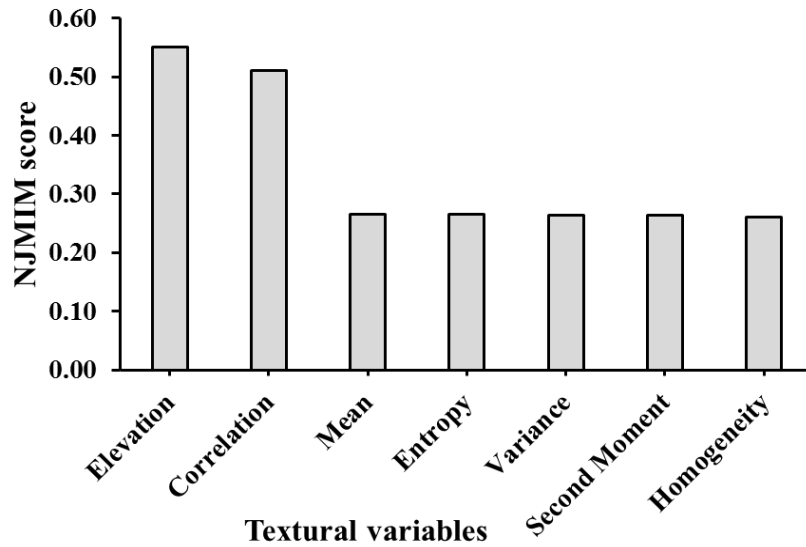


Figure 5.12. Variable importance of textural input features derived from DEM with an optimum window.

The optimum hybrid input features (i.e., 30) consist of various important variables of high NJMIM score derived from all input features to increase the intra-class variability of different rock types for accurate lithological classification using ML models. It is most likely that optimal input features may yield better accuracy statistics in classifying most of the rock types found in the area. Furthermore, it is worth considering that the number of variables is about 50% less than the actual number of variables (i.e., 58), which gives the possible best accurate lithological classification results and makes the model less complex

and computationally efficient. The optimum number of variables of different input features extracted from multi-sensor datasets is presented in Table 5.6.

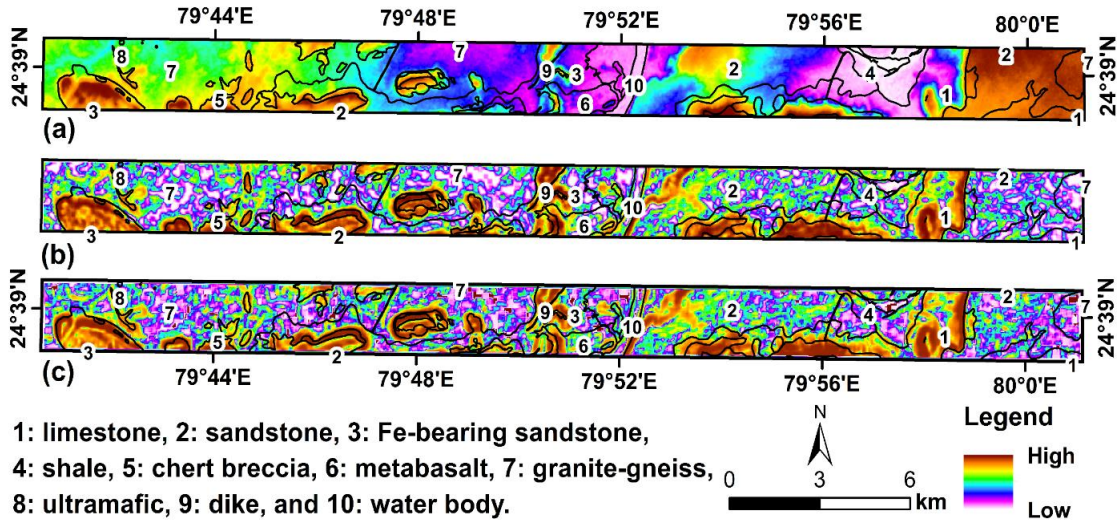


Figure 5.13. Textural variables such as (a) elevation, (b) correlation, and (c) entropy derived from the optimum window size using DEM illustrating their ability to discriminate different rock types found in the area.

Table 5.6. Summary of different input features extracted from multi-sensor datasets and the optimum number of variables used in lithological classification.

Input features	Number of variables	Optimum number of variables	Sensor/dataset
Spectral	9	9	ASTER
Spectral and transformed spectral	27	5	ASTER
Spectral and morphological	33	5	PALSAR and Sentinel-1
Spectral and textural	14	7	PALSAR DEM
Optimal hybrid	58	30	Multi-sensor datasets

5.5.5 Comparison of ML models using different input datasets

Different input features such as original spectral, spectral and transformed, spectral and morphological, spectral and textural, and optimal hybrid input features derived from multi-sensor datasets were evaluated for lithological mapping using SVM and RF models. To achieve optimal result from each input dataset and make a fair comparison, the score of NJMIM was used to find the important and suitable variable. To find an optimal number of variables in each input feature (except original spectral input features), a different number of variables were employed to derive the accuracy statistics for lithological

mapping using RF and SVM models. A number gives higher accuracy statistics and shows no significant improvement by adding a further variable in that input features were considered as an optimal number of variables for that input features. The exploited input features show the potential to classify different rock types found in the study area. Several accuracy statistics such as Overall Accuracy (OA %), Kappa Coefficient (k), User Accuracy (UA %), Producer Accuracy (PA %) were computed for each input feature using SVM and RF models. The accuracy statistics of different input features to accurately classify different rock types found in the study area using RF and SVM models have been presented in Figure 5.14.

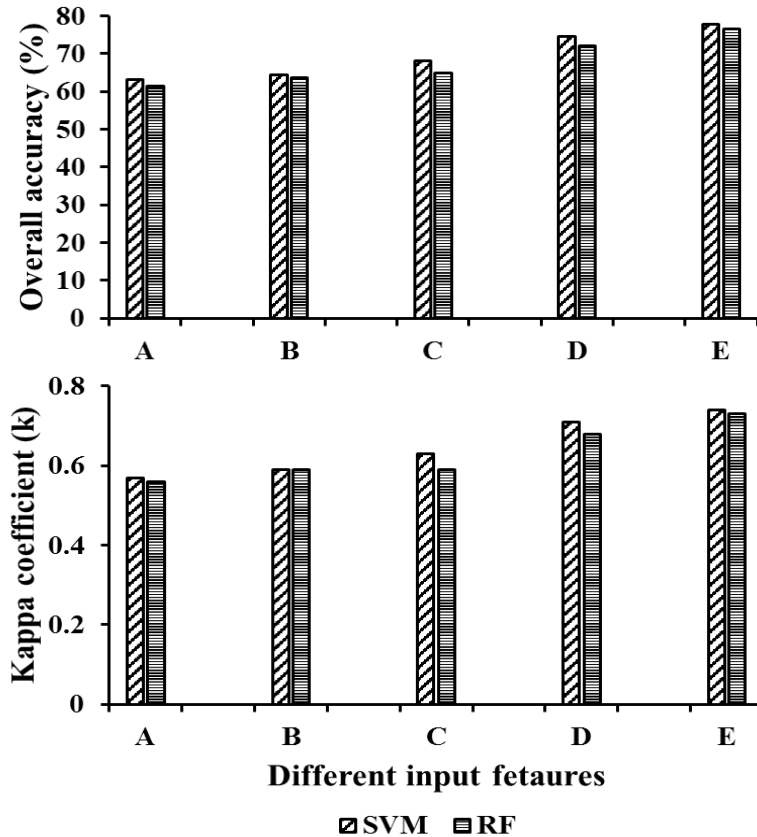


Figure 5.14. The accuracy statistics of lithological classification obtained from different input features using SVM and RF models. A: spectral input features, B: spectral and transformed input features, C: spectral and morphological input features, D: spectral and textural input features, and E: optimal hybrid input features.

The spectral input features yield the least accuracy statistics than other exploited input features (Table 5.7). The SVM and RF produce 63.06% and 61.39% of OA and 0.57 and 0.56 of k (Figure 5.14). The spectral input data show a poor ability to map the Fe-bearing sandstone and metabasalt. SVM's UA and PA for classifying the Fe-bearing sandstone and metabasalt is 36.11% and 34.82%, 59.52%, and 44.64%, respectively. The RF also shows similar performance as the UA and PA of RF for Fe-bearing sandstone (37% and 33%) and

metabasalt (52.94% and 32.14%). It is worth mentioning here that the Fe-bearing sandstone and metabasalt have shown considerable spectral similarity with other rock types, as illustrated in the spectral dissimilarity matrix (Table 5.4). Other rock types have shown slightly better accuracy statistics in lithological classification using RF and SVM. It has also been noticed that both RF and SVM show somewhat different abilities to classify different rock types correctly. The rock types such as limestone, sandstone, shale, and chert breccia were better classified by SVM, whereas RF better classified granite-gneiss, ultramafic, and dike. Few rock types, including Fe-bearing sandstone and metabasalt, were classified by RF and SVM with similar accuracy statistics. Furthermore, the ultramafic was classified with higher accuracy statistics (i.e., SVM: 82.93% (UA) and 91.89% (PA), and RF: 85.37% (UA) and 94.59% (PA)) as compared with other rock types mainly due to their higher spectral dissimilarity as illustrated in Table 5.2.

The integrated spectral and transformed spectral input features slightly improve the lithological mapping by increasing the accuracy statistics around 2% in OA and k (Table 5.7). The SVM yields 64.31% (OA) and 0.59 (k) whereas RF yields 63.75% (OA) and 0.59 (k) (Figure 5.14). The improvement in classification accuracy statistics of spectral and transformed input features over spectral data alone can be counted towards the addition of important variables derived from PCA and ICA transformation. The addition of important PCA and ICA transformed bands may improve the intra-class variability of most of the rock types. It has been observed that there is a slight decrement (i.e., 1-2%) in the individual accuracy statistics (i.e., UA and PA) of a few rock types such as limestone, sandstone, and shale, whereas other rock types, including Fe-bearing sandstone, chert breccia, metabasalt, granite-gneiss, ultramafic and dike show a significant increment (i.e., 4-14%) in individual accuracy statistics. The Fe-bearing sandstone and metabasalt were classified with higher accuracy statistics as classified using spectral input data alone. The UA of SVM and RF have increased by around 10%. The PA of SVM has not shown any improvement, whereas the PA of RF is slightly improved, i.e., ~4%. Like spectral input data, the SVM and RF show a slightly different ability to correctly classify different rock types. Sandstone, shale, and granite-gneiss better classified by SVM, whereas RF does a better job in classifying chert breccia, ultramafic, and dike (Table 5.7).

The integrated spectral and morphological input features show better performance than spectral input data and integrated spectral and transformed datasets for mapping various rock types using SVM and RF (Table 5.8). The SVM yields 68.19% (OA) and 0.63 (k) whereas RF yields 64.86% (OA) and 0.59 (k) (Figure 5.14). The improvement in classification accuracy statistics can be accounted for by integrating important variables derived from SAR datasets, which leads to an increase in the intra-class variability of different rock types. The accuracy statistics of SVM and RF have increased by ~5% and 3%, respectively, as obtained by spectral input data alone. Furthermore, it shows better performance than integrated spectral and transformed input features, particularly by SVM (i.e., ~4%), whereas the RF model does not show much improvement (i.e., ~1%). It is observed that most of the rock types except limestone and sandstone have been efficiently mapped with higher individual accuracy as compared with spectral data and integrated spectral and transformed input datasets. As observed with the previous two input datasets,

the SVM and RF show a slightly diverse capability to accurately classify different rock types. The SVM shows a better ability to classify shale, chert breccia, metabasalt, and granite-gneiss, whereas the RF does a better classification of ultramafic and dike. However, few rock types such as limestone, sandstone, and Fe-bearing sandstone were classified with similar accuracy statistics.

Table 5.7. The accuracy statistics of lithological classification are derived from spectral input data and integrated spectral and transformed spectral input datasets using SVM and RF models. Lm: limestone, Fe-Sd: Fe-bearing sandstone, Sh: shale, Mb: metabasalt, Gg: granite-gneiss, Um: ultramafic rock, and Dk: dike. OA: overall accuracy (%), k: kappa coefficient, UA: user accuracy (%), PA: producer accuracy (%).

	Spectral input features				Spectral and transformed spectral input features			
	OA (%)		k		OA (%)		k	
SVM	63.06		0.57		64.31		0.59	
RF	61.39		0.56		63.75		0.59	
	SVM		RF		SVM		RF	
Rocks	UA	PA	UA	PA	UA	PA	UA	PA
Lm	61.94	83.00	57.35	78.00	58.27	81.00	58.27	81.00
Sd	68.66	73.60	63.38	72.00	67.72	68.80	65.85	64.80
Fe-Sd	36.11	34.82	37.00	33.04	45.88	34.82	45.65	37.50
Sh	70.18	63.49	61.40	55.56	67.31	55.56	62.96	53.97
Cb	70.45	62.00	71.79	56.00	67.80	80.00	69.23	72.00
Mb	59.52	44.64	52.94	32.14	61.70	51.79	56.25	32.14
Gg	69.92	66.43	68.49	71.43	69.54	75.00	66.05	76.43
Um	82.93	91.89	85.37	94.59	88.57	83.78	94.59	94.59
Dk	62.96	45.95	84.00	56.76	68.00	45.95	86.21	67.57

The integrated spectral and textural input features show a great success over spectral data, integrated spectral and transformed data, and integrated spectral and morphological datasets for lithological mapping using SVM and RF (Table 5.8). The SVM yields 74.58% (OA) and 0.71 (k) whereas RF yield 72.08% (OA) and 0.68 (k) (Figure 5.14). The accuracy statistics of both SVM and RF models increased by 10–11% compared with previously employed input datasets. The integration of textural variables with the spectral input data helps increase the intra-class variability of different rocks found in the area. It is interesting to note here that most of the rock types have significantly improved their accuracy statistics. Both SVM and RF does a satisfactory classification for most of the rock types (Table 5.8).

Table 5.8. The accuracy statistics of lithological classification derived from integrated spectral and morphological input dataset, and integrated spectral and textural input dataset using SVM and RF models.

	Spectral and morphological input feature				Spectral and textural input features			
	OA (%)		k		OA (%)		k	
SVM	68.19		0.63		74.58		0.71	
RF	64.86		0.59		72.08		0.68	
	SVM		RF		SVM		RF	
Rocks	UA	PA	UA	PA	UA	PA	UA	PA
Lm	60.77	79.00	60.61	80.00	62.40	78.00	61.03	83.00
Sd	68.84	76.00	71.43	68.00	78.63	82.40	85.44	70.40
Fe-Sd	47.31	39.29	43.90	48.21	57.00	50.89	52.58	45.54
Sh	75.44	68.25	64.58	49.21	88.89	76.19	83.05	77.78
Cb	81.40	70.00	70.45	62.00	77.19	88.00	74.07	80.00
Mb	60.00	53.57	52.38	39.29	75.61	55.36	59.09	46.43
Gg	75.71	75.71	70.47	75.00	81.62	79.29	76.10	86.43
Um	92.11	94.59	94.59	94.59	92.50	100.00	94.59	94.59
Dk	77.42	64.86	92.31	64.86	77.78	75.68	83.87	70.27

The optimum hybrid input features consist of different suitable variables derived from spectral, morphological, and textural characteristics of rocks to increase the intra-class variability for accurate lithological classification. The optimal input features have shown great success over other input features for lithological mapping using SVM and RF models (Table 5.9). The SVM yields 77.78% (OA) and 0.74 (k), whereas RF yields 76.67% (OA) and 0.73 (k), which is highest among all input features exploited in lithological mapping using SVM and RF models (Figure 5.14). It shows around 10–15% of improvement over spectral input features.

Table 5.9. The accuracy statistics of lithological classification are derived from the optimal hybrid input dataset using SVM and RF models.

	Optimal hybrid input features			
	OA (%)		k	
SVM	77.78		0.74	
RF	76.67		0.73	
	SVM		RF	
Rocks	UA	PA	UA	PA
Lm	71.43	85.00	65.91	87.00
Sd	81.97	80.00	84.62	79.20
Fe-Sd	60.00	53.57	56.44	50.89
Sh	87.10	85.71	82.81	84.13
Cb	90.91	80.00	88.00	88.00
Mb	64.29	64.29	57.14	50.00
Gg	82.19	85.71	85.11	85.71
Um	97.37	100.00	100.00	97.30

Dk	84.85	75.68	93.33	75.68
-----------	-------	-------	-------	-------

Furthermore, most of the rock types' accuracy statistics have significantly improved compared with other input datasets. As noticed with other input datasets, the SVM slightly outperforms RF to classify different rock types found in the study area accurately. The individual accuracy statistics of Fe-bearing sandstone obtained using spectral data is 36.11% (UA) and 34.82% (PA), which improved to 60% (UA) and 53.57% (PA) using optimal hybrid data. The individual accuracy of limestone achieved using spectral data is 61.94% (UA) and 83% (PA), which increased to 71.43% (UA) and 85% (PA) using optimal hybrid data. Similar improvements have been achieved for other rock types found in the area.

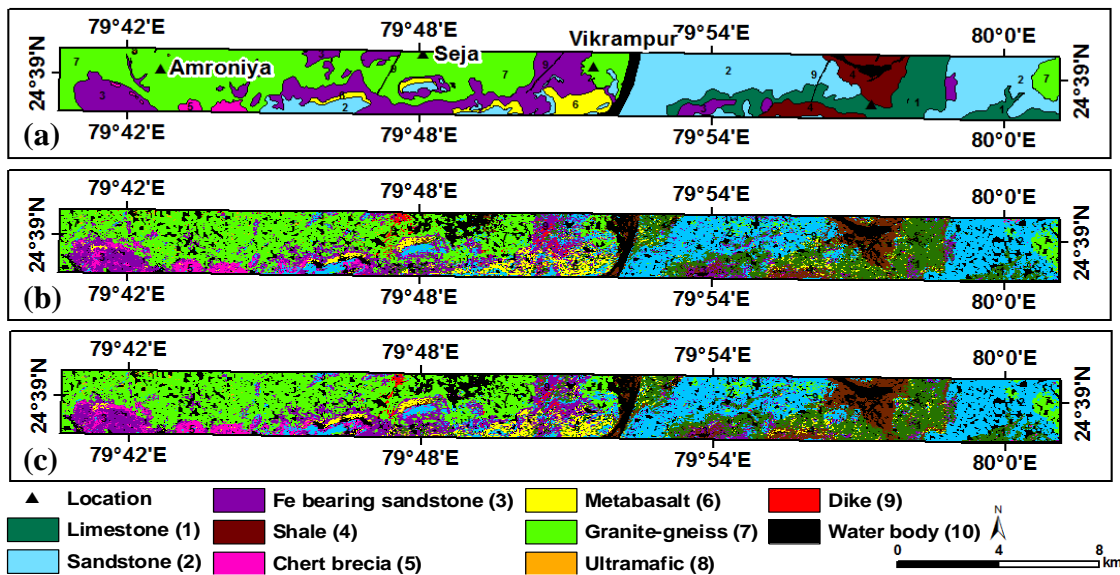


Figure 5.15. The lithological classification was obtained from the optimal hybrid input data using RF and SVM. (a) Reference lithology map, (b) lithological classification map using RF, and (c) lithological classification map using SVM.

The lithological map obtained from the optimal input features derived using RF (Figure 5.15 (b)) and SVM model (Figure 5.15 (c)) shows high similarity among them as observed statistically using their accuracy statistics (Figure 5.14 and Table 5.9). We performed a qualitative validation of the derived lithological classification map using a conjugative reference lithology map (Figure 5.15 (a)) and with the field data. The classified maps well agreed with the field data and high-resolution reference lithology map, which provides great confidence to rely on the proposed multi-sensor approach of integrating spectral, morphological, and textural characteristics of rocks for accurate lithological classification using SVM and RF models.

5.6 Conclusions

- We utilized multi-sensor datasets such as ASTER, PALSAR, Sentinel-1, and DEM to derive surface characteristics of rocks from preparing various input features such as (a) spectral, (b) spectral and transformed spectral, (c) spectral and morphological, (d) spectral and textural, and (e) optimal hybrid input features, to evaluate their ability in lithological classification using ML models.
- The NJMIM feature selection method was used to select the important and uncorrelated variables of each input feature to achieve optimal results and assess their novelty for accurate lithological classification using ML models.
- The comparative evaluation illustrated that the optimal hybrid input features outperform other input features to classify different rock types found in the study area accurately. The poor performance of SVM and RF obtained using the spectral data alone indicate that spectral data may not be an ideal input data for accurate lithological mapping in a geological setting where different rocks exhibit spectral similarity and weathered surface.
- The textural variables (i.e., elevation, correlation, entropy, mean, variance, homogeneity, and second moment) show a better ability to discriminate different rock types than other derived variables such as transformed spectral and morphological (SAR indices). However, the index derived from PALSAR data shows a higher ability to discriminate different rock types than the Sentinel-1 data.
- The optimal hybrid input features consist of several vital variables derived from multi-sensor datasets that yield a higher NJMIM score than other input features. The optimal hybrid input features produce maximum individual accuracy statistics for all the lithological units found in the area. The maximum accuracy statistics obtained in lithological classification by SVM and RF models using optimal hybrid input features are 77.78% and 76.77% of OA and 0.74 and 0.73 of k, respectively, which is ~15% higher as obtained using spectral data.
- The obtained lithological maps well agreed with the field data and reference lithology map, which provides great confidence to rely on the proposed multi-sensor datasets-based integration of spectral, morphological, and textural characteristics of rocks for accurate lithological classification using ML models.

6 Conclusions and future work

This research aims to develop innovative methods for discrimination and classification of geological targets, including minerals and rock types/lithological units. The developed methods have been evaluated at two study sites, the Hutti-Maski greenstone belt and the Chattarpur area located in India. The Hutti-Maski site is explored for different minerals and rock types classification using hyperspectral data, whereas the Chattarpur site is studied for lithological classification using multi-sensor datasets. Chapter 2 evaluates the utility of high-resolution hyperspectral data and develops an optimum threshold selection method using spectral similarity matrix of the target and non-target classes, which addressed the first objective (i.e., to evaluate the utility of AVIRIS-NG data for hydrothermal alteration mineral mapping and to develop an optimum threshold selection method for SMMs). The spatial distribution of hydrothermal alteration minerals such as chlorite, goethite, muscovite, and kaolinite have been mapped, which primarily belongs to the propylitic type of hydrothermal alteration in the Hutti-Maski area. The hybrid SMM, i.e., SIDSAM_{tan} (OA: 94.40; k: 0.93), outperforms the SAM and SID in mineral classification. The advantage of the proposed threshold selection method is that it does not require extensive field data to validate the selected threshold and classification results' reliability. Chapter 3 addressed the third objective of the study (i.e., to develop effective SMMs for accurate discrimination of spectrally similar minerals) by proposing three new effective SMMs such as DSSC, KJSSC, and KJDSSC_{tan} for accurate discrimination of spectrally similar minerals. The proposed SMMs outperform exiting SMMs (i.e., SAM, SID, and SIDSAM_{tan}) across different hyperspectral datasets of synthesized mixed and noisy spectra acquired under laboratory and real atmospheric conditions. The developed KJDSSC_{tan} shows a significantly higher ability to discriminate spectrally similar materials as compared with other SMMs. The developed SMMs can also be used in optimum band selection, class separability analysis, endmember selection, clustering, and image classification for successful hyperspectral data analysis. Chapter 4 presents a novel approach of automated lithological classification using ML models, which addressed the fourth objective (i.e., to evaluate the utility of AVIRIS-NG hyperspectral data for automated lithological classification using ML models) of the study. The developed approach integrates the spectral enhancement techniques and MLAs for an automated lithological classification using high-resolution AVIRIS-NG hyperspectral data. The study highlighted the effectiveness of conjugate utilization of geology maps coupled with spectral enhancement products derived from PCA and ICA transformation using ASTER multispectral data to prepare a reference lithology map for obtaining suitable training datasets collection. The JMIM feature selection based optimum bands is suitable than other input datasets (i.e., SWIR bands and all spectral bands) to obtain accurate lithological maps using ML models. The SVM (OA: 85.48%; k: 0.83) outperforms RF and LDA across different input datasets used in the classification. SVM's performance shows the least sensitivity to the quantity and quality of training datasets compared with RF and LDA. Chapter 5 addressed the fifth objective (i.e., to develop a multi-sensor datasets-based optimal integration of spectral, morphological, and textural characteristics of rocks for lithological classification using ML models) of this study. A multi-sensor datasets of

different wavelength regions, including optical and microwave, have accurately classified different rock types using SVM and RF models. Several input features such as (a) spectral, (b) spectral and transformed spectral, (c) spectral and morphological, (d) spectral and textural, and (e) optimal hybrid input features were evaluated for their ability in accurate lithological classification. The optimal hybrid input features outperform other input features to classify various rock types found in the Chattarpur area accurately. The spectral data alone with their poor performance of SVM and RF models, indicate that they may not be an ideal input features for accurate rock types classification in a geological setting of weathered and similar spectral characteristics. The textural variables, including elevation, correlation, entropy, mean, variance, homogeneity, and second moment show a better ability to discriminate various rock types than other derived variables such as transformed spectral and morphological (SAR indices). The PALSAR data show a higher ability than the Sentinel-1 dataset to discriminate different rock types found in the area. The optimal hybrid input features yield maximum accuracy statistics in lithological classification using SVM and RF models, which is around 15% higher as obtained using spectral data alone. The SVM slightly outperforms the RF model. The developed methods in this study are not only applicable to discrimination and classification of geological targets. They can be effectively utilized in other applications of remote sensing such as vegetation/forest mapping, water quality monitoring, and soil classification.

The research presented in this dissertation provides the foundation for the following future work to develop robust geological remote sensing classification methods:

- Future work should focus on developing automated optimum threshold selection for accurate classification and hyperspectral data analysis. The addition of spectral feature parameters can be used as a constrained parameter in solving the threshold selection problem.
- Future work should consider developing an integrated spectral-machine learning method for accurate classification. The successful development of this tool would be an innovative breakthrough in the field of hyperspectral remote sensing.
- The study should also explore the advantage of fusing spectral data and fully polarimetric SAR of different frequencies to evaluate their ability for accurate lithological classification. The products of polarimetric decomposition and different SAR indices (developed in this study) can be assessed in lithological classification.
- Future work should also focus on developing the ensemble and hybrid feature selection method for suitable band selection for accurate lithological classification using ML models.

References

- Abedi, M., Norouzi, G.-H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, *46*, 272-283.
- Adep, R. N., Vijayan, A. P., Shetty, A., & Ramesh, H. (2016). Performance evaluation of hyperspectral classification algorithms on AVIRIS mineral data. *Perspectives in Science*, *8*, 722-726.
- Amer, R., Kusky, T., & Ghulam, A. (2010). Lithological mapping in the Central Eastern Desert of Egypt using ASTER data. *Journal of African Earth Sciences*, *56*(2-3), 75-82.
- Archibald, R., & Fann, G. (2007). Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and remote sensing letters*, *4*(4), 674-677.
- Ardjmand, M., Mirzajanzadeh, M., & Zabihi, F. (2014). Measurement and correlation of solid drugs solubility in supercritical systems. *Chinese Journal of Chemical Engineering*, *22*(5), 549-558.
- Asadzadeh, S., & de Souza Filho, C. R. (2016). A review on spectral processing methods for geological remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *47*, 69-90.
- Awad, M. E., Amer, R., López-Galindo, A., El-Rahmany, M. M., del Moral, L. F. G., & Viseras, C. (2018). Hyperspectral remote sensing for mapping and detection of Egyptian kaolin quality. *Applied Clay Science*, *160*, 249-262.
- Ayeldeen, H., Shaker, O., Hegazy, O., & Hassanien, A. E. (2015). Distance similarity as a CBR technique for early detection of breast cancer: An Egyptian case study. In *Information Systems Design and Intelligent Applications* (pp. 449-456). Springer.
- Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A. C., & Benbouziane, A. (2019). Machine learning algorithms for automatic lithological mapping using remote sensing data: a case study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *ISPRS International Journal of Geo-Information*, *8*(6), 248.
- Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, *6*(1), 20-29.
- Baum, K., Schreyer, E., Totterman, S., Farber, J., Tamez-Peña, J., & González, P. (2010). Application of the Dice Similarity Coefficient (DSC) for failure detection of a fully-automated atlas based knee mri segmentation method. ISMRM Annual Meeting.
- Baykan, N. A., & Yılmaz, N. (2010). Mineral identification using color spaces and artificial neural networks. *Computers & Geosciences*, *36*(1), 91-97.
- Bedini, E. (2009). Mapping lithology of the Sarfartoq carbonatite complex, southern West Greenland, using HyMap imaging spectrometer data. *Remote sensing of Environment*, *113*(6), 1208-1219.
- Bedini, E. (2011). Mineral mapping in the Kap Simpson complex, central East Greenland, using HyMap and ASTER remote sensing data. *Advances in Space Research*, *47*(1), 60-73.

- Ben-Dor, E., Patkin, K., Banin, A., & Karnieli, A. (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—a case study over clayey soils in Israel. *International Journal of Remote Sensing*, 23(6), 1043-1062.
- Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using joint mutual information maximisation. *Expert Systems with Applications*, 42(22), 8520-8532.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1), 281-305.
- Bergstra, J. S., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*.
- Bhattacharya, B. K., Green, R. O., Rao, S., Saxena, M., Sharma, S., Kumar, K. A., Srinivasulu, P., Sharma, S., Dhar, D., & Bandyopadhyay, S. (2019a). An overview of AVIRIS-NG airborne hyperspectral science campaign over India. *Current Science*, 116, 1082-1088.
- Bhattacharya, S., Kumar, H., Guha, A., Dagar, A. K., Pathak, S., Rani, K., Mondal, S., Kumar, K. V., Farrand, W., & Chatterjee, S. (2019b). Potential of airborne hyperspectral data for geo-exploration over parts of different geological/metallogenic provinces in India based on AVIRIS-NG observations. *Curr. Sci.*, 116(7), 1143-1156.
- Bierwirth, P., Huston, D., & Blewett, R. (2002). Hyperspectral mapping of mineral assemblages associated with gold mineralization in the Central Pilbara, Western Australia. *Economic Geology*, 97(4), 819-826.
- Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2), 6-36.
- Bishop, C. A., Liu, J. G., & Mason, P. J. (2011). Hyperspectral remote sensing for mineral exploration in Pulang, Yunnan Province, China. *International Journal of Remote Sensing*, 32(9), 2409-2426.
- Boardman, J. W. (1998). Leveraging the high dimensionality of AVIRIS data for improved sub-pixel target unmixing and rejection of false positives: mixture tuned matched filtering. *Summaries of the seventh JPL Airborne Geoscience Workshop*, JPL Publication, 1998,
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement*, 41(3), 687-699.
- Carranza, E. J. M., & Laborte, A. G. (2015). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences*, 74, 60-70.
- Carrino, T. A., Crósta, A. P., Toledo, C. L. B., & Silva, A. M. (2018). Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect. *International Journal of Applied Earth Observation and Geoinformation*, 64, 287-300.
- Cartwright, D., & Field, M. (1978). A refinement of the arithmetic mean-geometric mean inequality. *Proceedings of the American Mathematical Society*, 36-38.
- Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1(2), 1.

- Chang, C.-I. (1999). Spectral information divergence for hyperspectral image analysis. IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293),
- Chang, C.-I. (2000). An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on information theory*, 46(5), 1927-1932.
- Chang, C.-I. (2003). *Hyperspectral imaging: techniques for spectral detection and classification* (Vol. 1). Springer Science & Business Media.
- Chauhan, P., Kaur, P., Srivastava, N., Sinha, R. K., Jain, N., & Murty, S. (2015). Hyperspectral remote sensing of planetary surfaces: an insight into composition of inner planets and small bodies in the solar system. *Current Science*, 915-924.
- Chen, X., Warner, T. A., & Campagna, D. J. (2010). Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada: a rule-based system. *International Journal of Remote Sensing*, 31(7), 1733-1752.
- Chen, Y., & Wu, W. (2017). Mapping mineral prospectivity using an extreme learning machine regression. *Ore geology reviews*, 80, 200-213.
- Chica-Olmo, M., & Abarca-Hernandez, F. (2000). Computing geostatistical image texture for remotely sensed data classification. *Computers & Geosciences*, 26(4), 373-383.
- Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of remote sensing*, 3(3-58), 2-2.
- Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*, 89(B7), 6329-6340.
- Clark, R. N., Swayze, G. A., & Gallagher, A. (1993). Mapping minerals with imaging spectroscopy. *US Geological Survey, Office of Mineral Resources Bulletin*, 2039(141-150), 69.
- Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., & Gent, C. A. (2003). Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research: Planets*, 108(E12).
- Clark, R. N., Swayze, G. A., Wise, R. A., Livo, K. E., Hoefen, T. M., Kokaly, R. F., & Sutley, S. J. (2007). *USGS digital spectral library splib06a (2327-638X)*.
- Cloutis, E. A. (1996). Review article hyperspectral geological remote sensing: evaluation of analytical techniques. *International Journal of Remote Sensing*, 17(12), 2215-2242.
- Colarusso, P., Kidder, L. H., Levin, I. W., Fraser, J. C., Arens, J. F., & Lewis, E. N. (1998). Infrared spectroscopic imaging: from planetary to cellular systems. *Applied Spectroscopy*, 52(3), 106A-120A.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of Environment*, 37(1), 35-46.
- Cortes, C., Jackel, L. D., & Chiang, W.-P. (1995). Limits on learning machine accuracy imposed by data quality. *Advances in neural information processing systems*,
- Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in

- the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 22-33.
- Crosta, A., De Souza Filho, C., Azevedo, F., & Brodie, C. (2003). Targeting key alteration minerals in epithermal deposits in Patagonia, Argentina, using ASTER imagery and principal component analysis. *International Journal of Remote Sensing*, 24(21), 4233-4240.
- Crosta, A. P., Sabine, C., & Taranik, J. V. (1998). Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. *Remote sensing of Environment*, 65(3), 309-319.
- Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(3), 131-156.
- De Alban, J. D. T., Connette, G. M., Oswald, P., & Webb, E. L. (2018). Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing*, 10(2), 306.
- De Carvalho, O. A., & Meneses, P. R. (2000). Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18,
- De Martino, M., Causa, F., & Serpico, S. B. (2003). Classification of optical high resolution images in urban environment using spectral and textural information. IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477),
- DeFries, R., & Chan, J. C.-W. (2000). Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote sensing of Environment*, 74(3), 503-515.
- Dian, Y., Li, Z., & Pang, Y. (2015). Spectral and texture features combined for forest tree species classification with airborne hyperspectral imagery. *Journal of the Indian Society of Remote Sensing*, 43(1), 101-107.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
- Ding, J., Li, X., & Huang, L. (2015). A Novel Method for Spectral Similarity Measure by Fusing Shape and Amplitude Features. *Journal of Engineering Science & Technology Review*, 8(5).
- Dong, J., Xiao, X., Sheldon, S., Biradar, C., & Xie, G. (2012). Mapping tropical forests and rubber plantations in complex landscapes by integrating PALSAR and MODIS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74, 20-33.
- Drost, H.-G. (2018). Philentropy: information theory and distance quantification with R. *Journal of Open Source Software*, 3(26), 765.
- Du, Y., Chang, C.-I., Ren, H., Chang, C.-C., Jensen, J. O., & D'Amico, F. M. (2004). New hyperspectral discrimination measure for spectral characterization. *Optical Engineering*, 43(8), 1777-1787.
- Farifteh, J., Nieuwenhuis, W., & García-Meléndez, E. (2013). Mapping spatial variations of iron oxide by-product minerals from EO-1 Hyperion. *International Journal of Remote Sensing*, 34(2), 682-699.

- Farooq, S., & Govil, H. (2014). Mapping Regolith and Gossan for Mineral Exploration in the Eastern Kumaon Himalaya, India using hyperion data and object oriented image classification. *Advances in Space Research*, 53(12), 1676-1685.
- Feng, J., Rogge, D., & Rivard, B. (2018). Comparison of lithological mapping results from airborne hyperspectral VNIR-SWIR, LWIR and combined data. *International Journal of Applied Earth Observation and Geoinformation*, 64, 340-353.
- Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 52(11), 1039-1058.
- Franklin, S., Hall, R., Moskal, L., Maudie, A., & Lavigne, M. (2000). Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, 21(1), 61-79.
- Friedrich, G., Helmert, M., & Wotawa, F. (2016). *KI 2016: Advances in Artificial Intelligence: 39th Annual German Conference on AI, Klagenfurt, Austria, September 26-30, 2016, Proceedings* (Vol. 9904). Springer.
- Gad, S., & Kusky, T. (2007). ASTER spectral ratioing for lithological mapping in the Arabian–Nubian shield, the Neoproterozoic Wadi Kid area, Sinai, Egypt. *Gondwana research*, 11(3), 326-335.
- Gao, B.-C., Montes, M. J., Davis, C. O., & Goetz, A. F. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote sensing of Environment*, 113, S17-S24.
- Ge, W., Cheng, Q., Tang, Y., Jing, L., & Gao, C. (2018). Lithological classification using sentinel-2A data in the Shibanjing ophiolite complex in inner Mongolia, China. *Remote Sensing*, 10(4), 638.
- Goetz, A. F. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote sensing of Environment*, 113, S5-S16.
- Goetz, A. F., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for earth remote sensing. *science*, 228(4704), 1147-1153.
- Gomez, C., Delacourt, C., Allemand, P., Ledru, P., & Wackerle, R. (2005). Using ASTER remote sensing data set for geological mapping, in Namibia. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(1-3), 97-108.
- Goodarzi Mehr, S., Ahadnejad, V., Abbaspour, R. A., & Hamzeh, M. (2013). Using the mixture-tuned matched filtering method for lithological mapping with Landsat TM5 images. *International Journal of Remote Sensing*, 34(24), 8803-8816.
- Govender, M., Chetty, K., & Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, 33(2).
- Gower, J. C. (1985). Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra and its Applications*, 67, 81-97.
- Grebby, S., Naden, J., Cunningham, D., & Tansey, K. (2011). Integrating airborne multispectral imagery and airborne LiDAR data for enhanced lithological mapping in vegetated terrain. *Remote sensing of Environment*, 115(1), 214-226.
- Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65-74.

- Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., & Vieilledent, G. (2013). Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. *Remote sensing of Environment*, 139, 68-80.
- Gstaiger, V., Huth, J., Gebhardt, S., Wehrmann, T., & Kuenzer, C. (2012). Multi-sensoral and automated derivation of inundated areas using TerraSAR-X and ENVISAT ASAR data. *International Journal of Remote Sensing*, 33(22), 7291-7304.
- Guha, A., Chakraborty, D., Ekka, A., Pramanik, K., Kumar, K. V., Chatterjee, S., Subramaniam, S., & Rao, D. A. (2012). Spectroscopic study of rocks of Hutti-Maski schist belt, Karnataka. *Journal of the Geological Society of India*, 79(4), 335-344.
- Guha, A., Chatterjee, S., Oommen, T., Kumar, K. V., & Roy, S. K. (2020). Synergistic use of ASTER, L-band ALOS PALSAR, and Hyperspectral AVIRIS-NG data for exploration of lode type gold deposit—a study in Hutti Maski Schist Belt, India. *Ore geology reviews*, 103818.
- Guha, A., Kumar, K. V., Porwal, A., Rani, K., Sahoo, K., Kumar, S. A., Singaraju, V., Singh, R., Khandelwal, M., & Raju, P. (2019). Reflectance spectroscopy and ASTER based mapping of rock-phosphate in parts of Paleoproterozoic sequences of Aravalli group of rocks, Rajasthan, India. *Ore geology reviews*, 108, 73-87.
- Guyon, I., Gunn, S., Ben-Hur, A., & Dror, G. (2004). Result analysis of the NIPS 2003 feature selection challenge. *Advances in neural information processing systems*, 17, 545-552.
- Hadigheh, S. M. H., & Ranjbar, H. (2013). Lithological mapping in the eastern part of the central Iranian volcanic belt using combined ASTER and IRS data. *Journal of the Indian Society of Remote Sensing*, 41(4), 921-931.
- Hajnsek, I., Pottier, E., & Cloude, S. R. (2003). Inversion of surface parameters from polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4), 727-744.
- Hall, M. A., & Smith, L. A. (1999). Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. FLAIRS conference,
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*(6), 610-621.
- Harris, J., Rogge, D., Hitchcock, R., Ijewliw, O., & Wright, D. (2005). Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering. *Canadian Journal of Earth Sciences*, 42(12), 2173-2193.
- Harsanyi, J. C., & Chang, C.-I. (1994). Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), 779-785.
- Hazarika, P., Mishra, B., & Pruseth, K. L. (2015). Diverse tourmaline compositions from orogenic gold deposits in the Hutti-Maski greenstone belt, India: implications for sources of ore-forming fluids. *Economic Geology*, 110(2), 337-353.
- Hecker, C., Van der Meijde, M., van der Werff, H., & van der Meer, F. D. (2008). Assessing the influence of reference spectra on synthetic SAM classification results. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12), 4162-4172.

- Hewson, R., Cudahy, T., Mizuhiko, S., Ueda, K., & Mauger, A. (2005). Seamless geological map generation using ASTER in the Broken Hill-Curnamona province of Australia. *Remote sensing of Environment*, 99(1-2), 159-172.
- Homayouni, S., & Roux, M. (2004). Hyperspectral image analysis for material mapping using spectral matching. ISPRS Congress Proceedings,
- Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2014). MIFS-ND: A mutual information-based feature selection method. *Expert Systems with Applications*, 41(14), 6371-6385.
- Hubálek, Z., Pow, I., Reid, H., & Hussain, M. (1995). Antigenic similarity of central European encephalitis and louping-ill viruses. *Acta Virologica*, 39(5-6), 251-256.
- Hunt, G. R. (1977). Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, 42(3), 501-513.
- Jin, H., Li, P., & Fan, W. (2008). Land cover classification using multitemporal CHRIS/PROBA images and multitemporal texture. IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium,
- Jones, R. R., McCaffrey, K. J., Wilson, R. W., & Holdsworth, R. E. (2004). Digital field data acquisition: towards increased quantification of uncertainty during geological mapping. *Geological Society, London, Special Publications*, 239(1), 43-56.
- Kamps, O., Hewson, R., van Ruitenbeek, F., & van der Meer, F. (2020). Defining surface types of Mars using global CRISM summary product maps. *Journal of Geophysical Research: Planets*, 125(8), e2019JE006337.
- Keshava, N. (2004). Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1552-1565.
- King, T. V., & Clark, R. N. (2000). Verification of remotely sensed data. In *Remote sensing for site characterization* (pp. 59-62). Springer.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai,
- Kokaly, R., Clark, R., Swayze, G., Livo, K., Hoefen, T., Pearson, N., Wise, R., Benzel, W., Lowers, H., & Driscoll, R. (2017). Usgs spectral library version 7 data: Us geological survey data release. *United States Geological Survey (USGS): Reston, VA, USA*.
- Kruse, F. (1996). Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra. *International Journal of Remote Sensing*, 17(9), 1623-1632.
- Kruse, F. A. (1988). Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern grapevine mountains, Nevada, and California. *Remote sensing of Environment*, 24(1), 31-51.
- Kruse, F. A. (1998). Advances in hyperspectral remote sensing for geologic mapping and exploration. Proceedings 9th Australasian remote sensing conference,
- Kruse, F. A. (2012). Mapping surface mineralogy using imaging spectrometry. *Geomorphology*, 137(1), 41-56.
- Kruse, F. A., Boardman, J. W., & Huntington, J. F. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1388-1400.

- Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., & Goetz, A. (1993). The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. AIP Conference Proceedings,
- Kruse, F. A., Perry, S. L., & Caballero, A. (2006). District-level mineral survey using airborne hyperspectral data, Los Menucos, Argentina. *Annals of Geophysics*, 49(1).
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., & Benesty, M. (2017). Caret: classification and regression training. 2016. *R package version*, 4.
- Kumar, A., & Zhang, K. Y. (2018). Advances in the development of shape similarity methods and their application in drug discovery. *Frontiers in chemistry*, 6, 315.
- Kumar, C., Chatterjee, S., & Oommen, T. (2020a). Mapping hydrothermal alteration minerals using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski gold deposit area, India. *International Journal of Remote Sensing*, 41(2), 794-812.
- Kumar, C., Chatterjee, S., Oommen, T., & Guha, A. (2020b). Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *International Journal of Applied Earth Observation and Geoinformation*, 86, 102006.
- Kumar, C., Shetty, A., Raval, S., Champatiray, P. K., & Sharma, R. (2014). Sub-pixel mineral mapping using EO-1 hyperion hyperspectral data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(8), 455.
- Kumar, C., Shetty, A., Raval, S., Sharma, R., & Ray, P. C. (2015). Lithological discrimination and mapping using ASTER SWIR Data in the Udaipur area of Rajasthan, India. *Procedia Earth and Planetary Science*, 11, 180-188.
- Kumar, P., & Johnson, A. (2005). On a symmetric divergence measure and information inequalities. *Journal of Inequalities in pure and applied Mathematics*, 6(3).
- LaPara, T. M., Nakatsu, C. H., Pantea, L. M., & Alleman, J. E. (2002). Stability of the bacterial communities supported by a seven-stage biological process treating pharmaceutical wastewater as revealed by PCR-DGGE. *Water Research*, 36(3), 638-646.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2015). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 30, 1e9.
- Lehnert, L., Meyer, H., & Bendix, J. (2016). Hsdar: manage, analyse and simulate hyperspectral data in R. *R Package Version 0.4*, 1.
- Li, N., Frei, M., & Altermann, W. (2011). Textural and knowledge-based lithological classification of remote sensing data in Southwestern Prieska sub-basin, Transvaal Supergroup, South Africa. *Journal of African Earth Sciences*, 60(4), 237-246.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.
- LIU, C.-a., CHEN, Z.-x., Yun, S., CHEN, J.-s., Hasi, T., & PAN, H.-z. (2019). Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture*, 18(3), 506-525.

- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1), 010901.
- Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 5(1), 18.
- Magendran, T., & Sanjeevi, S. (2014). Hyperion image analysis and linear spectral unmixing to evaluate the grades of iron ores in parts of Noamundi, Eastern India. *International Journal of Applied Earth Observation and Geoinformation*, 26, 413-426.
- Masoumi, F., Eslamkish, T., Abkar, A. A., Honarmand, M., & Harris, J. R. (2017). Integration of spectral, thermal, and textural features of ASTER data using Random Forests classification for lithological mapping. *Journal of African Earth Sciences*, 129, 445-457.
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784-2817.
- McCune, B., Grace, J. B., & Urban, D. L. (2002). *Analysis of ecological communities* (Vol. 28). MjM software design Gleneden Beach, OR.
- Meyer, D., & Buchta, C. (2009). proxy: Distance and similarity measures. *R package version 0.4-3*, URL <http://CRAN.R-project.org/package=proxy>.
- Meyer, D., C. Buchta, and M. D. Meyer. (2018). Package 'proxy'. <https://cran.r-project.org/web/packages/proxy/index.html>
- Michie, D., Spiegelhalter, D. J., & Taylor, C. (1994). Machine learning. *Neural and Statistical Classification*, 13(1994), 1-298.
- Mishra, B., & Pal, N. (2008). Metamorphism, fluid flux, and fluid evolution relative to gold mineralization in the Hutti-Maski Greenstone Belt, Eastern Dharwar Craton, India. *Economic Geology*, 103(4), 801-827.
- Mishra, B., Pal, N., & Sarbadhikari, A. B. (2005). Fluid inclusion characteristics of the Uti gold deposit, Hutti-Maski greenstone belt, southern India. *Ore geology reviews*, 26(1-2), 1-16.
- Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: state of the art and future prospects. *science*, 293(5537), 2051-2055.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Molan, Y. E., Refahi, D., & Tarashti, A. H. (2014). Mineral mapping in the Maherabad area, eastern Iran, using the HyMap remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 27, 117-127.
- Murguía, M., & Villaseñor, J. L. (2003). Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications. *Annales Botanici Fennici*,
- Murphy, R. J., Monteiro, S. T., & Schneider, S. (2012). Evaluating classification techniques for mapping vertical geology using field-based hyperspectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8), 3066-3080.
- Naresh Kumar, M., Seshasai, M., Vara Prasad, K., Kamala, V., Ramana, K., Dwivedi, R., & Roy, P. (2011). A new hybrid spectral similarity measure for discrimination

- among Vigna species. *International Journal of Remote Sensing*, 32(14), 4041-4053.
- Nidamanuri, R. R., & Zbell, B. (2011). Use of field reflectance data for crop mapping using airborne hyperspectral image. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5), 683-691.
- Ninomiya, Y. (2004). Lithologic mapping with multispectral ASTER TIR and SWIR data. *Sensors, Systems, and Next-Generation Satellites VII*,
- Ninomiya, Y., Fu, B., & Cudahy, T. J. (2005). Detecting lithology with Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) multispectral thermal infrared "radiance-at-sensor" data. *Remote sensing of Environment*, 99(1-2), 127-139.
- Oommen, T., Baise, L. G., & Vogel, R. M. (2011). Sampling bias and class imbalance in maximum-likelihood logistic regression. *Mathematical Geosciences*, 43(1), 99-120.
- Othman, A. A., & Gloaguen, R. (2014). Improving lithological mapping by SVM classification of spectral and morphological features: The discovery of a new chromite body in the Mawat ophiolite complex (Kurdistan, NE Iraq). *Remote Sensing*, 6(8), 6867-6896.
- Othman, A. A., & Gloaguen, R. (2017). Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq. *Journal of Asian Earth Sciences*, 146, 90-102.
- Padma, S., & Sanjeevi, S. (2014a). Jeffries Matusita-Spectral Angle Mapper (JM-SAM) spectral matching for species level mapping at Bhitarkanika, Muthupet and Pichavaram mangroves. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(8), 1403.
- Padma, S., & Sanjeevi, S. (2014b). Jeffries Matusita based mixed-measure for improved spectral matching in hyperspectral image analysis. *International Journal of Applied Earth Observation and Geoinformation*, 32, 138-151.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.
- Pal, M. (2008). Multiclass approaches for support vector machine based land cover classification. *arXiv preprint arXiv:0802.2411*.
- Pal, M., & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2297-2307.
- Pal, N., & Mishra, B. (2002). Alteration geochemistry and fluid inclusion characteristics of the greenstone-hosted gold deposit of Hutti, Eastern Dharwar Craton, India. *Mineralium Deposita*, 37(8), 722-736.
- Pal, S., Majumdar, T., & Bhattacharya, A. (2007). Usage of ERS SAR data over the Singhbhum shear zone, India for structural mapping and tectonic studies. *Geocarto International*, 22(4), 285-295.
- Pirajno, F. (2009). Orogenic, Amagmatic and Hydrothermal Mineral Systems of Uncertain Origin. In *Hydrothermal Processes and Mineral Systems* (pp. 885-1023). Springer.

- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., & Gualtieri, A. (2009). Recent advances in techniques for hyperspectral image processing. *Remote sensing of Environment*, *113*, S110-S122.
- Plaza, A., & Chang, C.-I. (2005). Fast implementation of pixel purity index algorithm. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*,
- Pontual, S., Merry, N., & Gamson, P. (2008). Spectral interpretation-field manual. GMEX. Spectral analysis guides for mineral exploration. *Victoria, AusSpec International Pty.*
- Pour, A. B., & Hashim, M. (2011). Identification of hydrothermal alteration minerals for exploring of porphyry copper deposit using ASTER data, SE Iran. *Journal of Asian Earth Sciences*, *42*(6), 1309-1323.
- Pour, A. B., & Hashim, M. (2014a). ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration. *SpringerPlus*, *3*(1), 130.
- Pour, A. B., & Hashim, M. (2014b). Structural geology mapping using PALSAR data in the Bau gold mining district, Sarawak, Malaysia. *Advances in Space Research*, *54*(4), 644-654.
- Pour, A. B., & Hashim, M. (2015). Structural mapping using PALSAR data in the Central Gold Belt, Peninsular Malaysia. *Ore geology reviews*, *64*, 13-22.
- Pour, A. B., Hashim, M., Park, Y., & Hong, J. K. (2018). Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. *Geocarto International*, *33*(12), 1281-1306.
- Pour, A. B., Park, Y., Park, T.-Y. S., Hong, J. K., Hashim, M., Woo, J., & Ayoobi, I. (2019). Evaluation of ICA and CEM algorithms with Landsat-8/ASTER data for geological mapping in inaccessible regions. *Geocarto International*, *34*(7), 785-816.
- Qin, Y., Xiao, X., Dong, J., Zhang, G., Shimada, M., Liu, J., Li, C., Kou, W., & Moore III, B. (2015). Forest cover maps of China in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI. *ISPRS Journal of Photogrammetry and Remote Sensing*, *109*, 1-16.
- Ramirez-Lopez, L., & Stevens, A. (2016). Resemble: regression and similarity evaluation for memory-based learning in spectral chemometrics. *R package version*, *1*(2), 2.
- Rao, D. A., & Guha, A. (2018). Potential utility of spectral angle mapper and spectral information divergence methods for mapping lower vindhyan rocks and their accuracy assessment with respect to conventional lithological map in jharkhand, india. *Journal of the Indian Society of Remote Sensing*, *46*(5), 737-747.
- Ren, H., & Chang, C.-I. (2000). Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images. *Optical Engineering*, *39*(12), 3138-3145.
- Ren, J., Zabalza, J., Marshall, S., & Zheng, J. (2014). Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]. *IEEE Signal Processing Magazine*, *31*(4), 149-154.
- Richards, J. A., & Richards, J. (1999). *Remote sensing digital image analysis* (Vol. 3). Springer.

- Robila, S. A., & Gershman, A. (2005). Spectral matching accuracy in processing hyperspectral data. *International Symposium on Signals, Circuits and Systems*, 2005. ISSCS 2005.,
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore geology reviews*, 71, 804-818.
- Rogers, A. J., Kolb, J., Meyer, F. M., & Armstrong, R. (2007). Tectono-magmatic evolution of the Hutti-Maski Greenstone Belt, India: Constrained using geochemical and geochronological data. *Journal of Asian Earth Sciences*, 31(1), 55-70.
- Rowan, L. C., & Mars, J. C. (2003). Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote sensing of Environment*, 84(3), 350-366.
- Rowan, L. C., Mars, J. C., & Simpson, C. J. (2005). Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). *Remote sensing of Environment*, 99(1-2), 105-126.
- Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore geology reviews*, 14(3-4), 157-183.
- Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., & Markey, M. K. (2009). Complex wavelet structural similarity: A new image similarity index. *IEEE transactions on image processing*, 18(11), 2385-2401.
- Sarma, D. S., Mcnaughton, N. J., Fletcher, I. R., Groves, D. I., Mohan, M. R., & Balaram, V. (2008). Timing of gold mineralization in the Hutti gold deposit, Dharwar Craton, South India. *Economic Geology*, 103(8), 1715-1727.
- Schmullius, C., & Evans, D. (1997). Review article Synthetic aperture radar (SAR) frequency and polarization requirements for applications in ecology, geology, hydrology, and oceanography: A tabular status quo after SIR-C/X-SAR. *International Journal of Remote Sensing*, 18(13), 2713-2722.
- Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Schwarz, J., & Staenz, K. (2001). Adaptive threshold for spectral matching of hyperspectral data. *Canadian Journal of Remote Sensing*, 27(3), 216-224.
- Seifi Majdar, R., & Ghassemian, H. (2017). A probabilistic SVM approach for hyperspectral image classification using spectral and texture features. *International Journal of Remote Sensing*, 38(15), 4265-4284.
- Shanmugam, S., & SrinivasaPerumal, P. (2014). Spectral matching approaches in hyperspectral image processing. *International Journal of Remote Sensing*, 35(24), 8217-8251.
- Shanon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-623.
- Souza, D. M., Costa, I. A., & Nóbrega, R. A. (2017). A study of distance/similarity measurements in the context of signal processing (density estimation). 2017 2nd

- International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT),
- Tan, Q., Gao, J., & Li, X. (2009). Application of sar remote sensing data to lithological mapping: a case study in railway geological survey. 2009 IEEE International Geoscience and Remote Sensing Symposium,
- Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190.
- Thenkabail, P. S., & Lyon, J. G. (2016). *Hyperspectral remote sensing of vegetation*. CRC press.
- Thompson, S., Fueten, F., & Bockus, D. (2001). Mineral identification using artificial neural networks and the rotating polarizer stage. *Computers & Geosciences*, 27(9), 1081-1089.
- Thorpe, A., Frankenberg, C., Aubrey, A., Roberts, D., Nottrott, A., Rahn, T., Sauer, J., Dubey, M., Costigan, K., & Arata, C. (2016). Mapping methane concentrations from a controlled release experiment using the next generation airborne visible/infrared imaging spectrometer (AVIRIS-NG). *Remote sensing of Environment*, 179, 104-115.
- Van der Meer, F. (2004). Analysis of spectral absorption features in hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 5(1), 55-68.
- Van der Meer, F. (2006). The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 3-17.
- van der Meer, F., Kopačková, V., Koucká, L., van der Werff, H. M., van Ruitenbeek, F. J., & Bakker, W. H. (2018). Wavelength feature mapping as a proxy to mineral chemistry for investigating geologic systems: An example from the Rodalquilar epithermal system. *International Journal of Applied Earth Observation and Geoinformation*, 64, 237-248.
- Van der Meer, F. D., & De Jong, S. M. (2011). *Imaging spectrometry: basic principles and prospective applications* (Vol. 4). Springer Science & Business Media.
- Van der Meer, F. D., Van der Werff, H. M., Van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., Van Der Meijde, M., Carranza, E. J. M., De Smeth, J. B., & Woldai, T. (2012). Multi-and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 14(1), 112-128.
- van Zyl, J. J., Zebker, H. A., & Elachi, C. (1990). Polarimetric SAR applications. *Radar Polarimetry for Geoscience Applications*, 315, 360.
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999.
- Vaughan, R. G., Hook, S. J., Calvin, W. M., & Taranik, J. V. (2005). Surface mineral mapping at Steamboat Springs, Nevada, USA, with multi-wavelength thermal infrared images. *Remote sensing of Environment*, 99(1-2), 140-158.
- Vishnu, S., Nidamanuri, R. R., & Bremananth, R. (2013). Spectral material mapping using hyperspectral imagery: a review of spectral matching and library search methods. *Geocarto International*, 28(2), 171-190.

- Walker, W. S., Stickler, C. M., Kellndorfer, J. M., Kirsch, K. M., & Nepstad, D. C. (2010). Large-area classification and mapping of forest and land cover in the Brazilian Amazon: A comparative analysis of ALOS/PALSAR and Landsat data sources. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(4), 594-604.
- Wang, S., Yan, F., Zhou, Y., Zhu, L., Wang, L., & Jiao, Y. (2005). Water quality monitoring using hyperspectral remote sensing data in Taihu Lake China. Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.,
- Wang, W., Ren, X., Zhang, Y., & Li, M. (2018). Deep Learning Based Lithology Classification Using Dual-Frequency Pol-SAR Data. *Applied Sciences*, 8(9), 1513.
- Waske, B., Benediktsson, J. A., Árnason, K., & Sveinsson, J. R. (2009). Mapping of hyperspectral AVIRIS data using machine-learning algorithms. *Canadian Journal of Remote Sensing*, 35(sup1), S106-S116.
- Waske, B., van der Linden, S., Benediktsson, J. A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7), 2880-2889.
- Wei, J., Liu, X., & Liu, J. (2016). Integrating textural and spectral features to classify silicate-bearing rocks using Landsat 8 data. *Applied Sciences*, 6(10), 283.
- Williamson, D. F., Parker, R. A., & Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. *Annals of internal medicine*, 110(11), 916-921.
- Wu, H., Zhang, X., Xie, H., Kuang, Y., & Ouyang, G. (2013). Classification of solder joint using feature selection based on Bayes and support vector machine. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 3(3), 516-522.
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. Proceedings of the 20th international conference on machine learning (ICML-03),
- Yu, L., Porwal, A., Holden, E.-J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45, 229-239.
- Zadeh, M. H., Tangestani, M. H., Roldan, F. V., & Yusta, I. (2014). Spectral characteristics of minerals in alteration zones associated with porphyry copper deposits in the middle part of Kerman copper belt, SE Iran. *Ore geology reviews*, 62, 191-198.
- Zhang, L., Zhang, L., Tao, D., Huang, X., & Du, B. (2013). Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8), 4955-4965.
- Zhang, X., & Li, P. (2014). Lithological mapping from hyperspectral data by improved use of spectral angle mapper. *International Journal of Applied Earth Observation and Geoinformation*, 31, 95-109.
- Zuhlke, M., Fomferra, N., Brockmann, C., Peters, M., Veci, L., Malik, J., & Regner, P. (2015). SNAP (sentinel application platform) and the ESA sentinel 3 toolbox. *ESASP*, 734, 21.

Zuo, R. (2017). Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural Resources Research*, 26(4), 457-464.

Appendix A

Table A1. Spectral similarity matrix of USGS spectra of minerals using Spectral Angle Mapper (SAM), Spectral Information Divergence (SID), $SIDSAM_{tan}$, Kumar-Johnson Spectral Similarity Coefficient (KJSSC), and $KJDSSC_{tan}$.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.1874	0.5367	0.3946	0.2922	0.1461	0.1878
Calcite	--	0.4280	0.3617	0.2812	0.1224	0.1330
Chlorite		--	0.2652	0.3489	0.5010	0.3564
Hematite			--	0.2384	0.4180	0.2475
Jarosite				--	0.3088	0.1706
Kaolinite					--	0.1796
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0204	0.1574	0.0975	0.0547	0.0121	0.0223
Calcite	--	0.0960	0.0785	0.0489	0.0100	0.0089
Chlorite		--	0.0367	0.0617	0.1348	0.0638
Hematite			--	0.0323	0.1061	0.0384
Jarosite				--	0.0596	0.0194
Kaolinite					--	0.0190
$SIDSAM_{tan}$	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0039	0.0937	0.0406	0.0165	0.0018	0.0042
Calcite	--	0.0438	0.0297	0.0141	0.0012	0.0012
Chlorite		--	0.0100	0.0225	0.0738	0.0238
Hematite			--	0.0079	0.0471	0.0097
Jarosite				--	0.0190	0.0033
Kaolinite					--	0.0034
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	95.36	540.95	152.30	104.41	15.02	28.55
Calcite	--	1080.64	277.23	372.42	70.93	102.52
Chlorite		--	340.59	199.96	546.60	316.64
Hematite			--	66.13	187.24	58.64
Jarosite				--	152.84	44.25
Kaolinite					--	28.73
$KJDSSC_{tan}$	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	4.74	177.77	11.77	6.35	0.16	0.60
Calcite	--	449.83	25.08	49.23	2.82	6.62
Chlorite		--	84.93	34.26	171.32	70.33
Hematite			--	3.40	16.20	2.11
Jarosite				--	10.05	0.92
Kaolinite					--	0.56

Table A2. Spectral similarity matrix of USGS spectra of vegetation using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}. Bb: blackbrush, Bs: blue spruce, Cg: cheat grass, Jb: juniper bush, Rb: rabbit brush, Sb: saltbrush, Ms: linearly mixed spectrum of vegetation with equal proportion.

SAM	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.1742	0.4933	0.1158	0.2008	0.2809	0.1752
Bs	--	0.5585	0.1385	0.2232	0.2090	0.1939
Cg		--	0.4474	0.3496	0.4469	0.3720
Jb			--	0.1285	0.1841	0.0911
Rb				--	0.1601	0.0480
Sb					--	0.1482
SID	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0204	0.1574	0.0975	0.0547	0.0121	0.0223
Bs	--	0.0960	0.0785	0.0489	0.0100	0.0089
Cg		--	0.0367	0.0617	0.1348	0.0638
Jb			--	0.0323	0.1061	0.0384
Rb				--	0.0596	0.0194
Sb					--	0.0190
SIDSAM_{tan}	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0036	0.0846	0.0113	0.0111	0.0035	0.0039
Bs	--	0.0600	0.0109	0.0111	0.0021	0.0017
Cg		--	0.0176	0.0225	0.0646	0.0249
Jb			--	0.0042	0.0198	0.0035
Rb				--	0.0096	0.0009
Sb					--	0.0028
KJSSC	Bs	Cg	Jb	Rb	Sb	Ms
Bb	27.46	115.34	21.32	43.91	144.72	24.77
Bs	--	302.74	64.84	99.52	164.40	62.30
Cg		--	70.73	43.13	113.25	41.56
Jb			--	6.18	33.29	4.25
Rb				--	20.52	3.34
Sb					--	32.73
KJDSSC_{tan}	Bs	Cg	Jb	Rg	Sb	Ms
Bb	1.78	13.82	0.47	1.44	12.70	0.42
Bs	--	59.72	7.96	12.96	32.76	5.39
Cg		--	7.98	3.13	16.37	2.91
Jb			--	0.05	0.91	0.05
Rb				--	0.53	0.02
Sb					--	1.47

Table A3. Spectral similarity matrix of USGS mineral spectra after adding 10% of random noise using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.1794	0.4988	0.3798	0.2799	0.1416	0.1830
Calcite	--	0.3936	0.3452	0.2670	0.1196	0.1296
Chlorite		--	0.2481	0.3255	0.4641	0.3273
Hematite			--	0.2309	0.4006	0.2373
Jarosite				--	0.2942	0.1649
Kaolinite					--	0.1759
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0184	0.1348	0.0878	0.0488	0.0112	0.0208
Calcite	--	0.0806	0.0694	0.0425	0.0092	0.0084
Chlorite		--	0.0324	0.0534	0.1145	0.0535
Hematite			--	0.0294	0.0948	0.0339
Jarosite				--	0.0523	0.0172
Kaolinite					--	0.0179
SIDSAM_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0033	0.0734	0.0350	0.0140	0.0016	0.0038
Calcite	--	0.0335	0.0250	0.0116	0.0011	0.0011
Chlorite		--	0.0082	0.0180	0.0573	0.0182
Hematite			--	0.0069	0.0401	0.0082
Jarosite				--	0.0158	0.0029
Kaolinite					--	0.0032
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	90.59	474.16	141.25	94.47	14.68	28.09
Calcite	--	930.05	243.63	316.70	68.78	98.36
Chlorite		--	307.70	178.43	472.45	279.66
Hematite			--	63.25	169.97	53.64
Jarosite				--	130.11	39.27
Kaolinite					--	28.29
KJDSSC_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	4.19	138.38	10.12	5.28	0.15	0.55
Calcite	--	351.03	20.30	38.76	2.58	5.91
Chlorite		--	68.69	26.78	131.23	54.80
Hematite			--	3.03	13.52	1.77
Jarosite				--	7.80	0.77
Kaolinite					--	0.52

Table A4. Spectral similarity matrix of USGS vegetation spectra after adding 10% of random noise using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}.

SAM	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.1788	0.4765	0.1123	0.1928	0.2697	0.1700
Bs	--	0.5321	0.1376	0.2114	0.1930	0.1847
Cg		--	0.4354	0.3386	0.4331	0.3601
Jb			--	0.1277	0.1804	0.0958
Rb				--	0.1583	0.0539
Sb					--	0.1467
SID	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0832	0.1625	0.0206	0.0457	0.0906	0.0397
Bs	--	0.2543	0.0411	0.0512	0.0485	0.05169
Cg		--	0.1302	0.0847	0.1539	0.08768
Jb			--	0.0125	0.0360	0.00934
Rb				--	0.0255	0.0025
Sb					--	0.02377
SIDSAM_{tan}	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0150	0.0839	0.0023	0.0089	0.0250	0.0068
Bs	--	0.1497	0.0057	0.0110	0.0095	0.0097
Cg		--	0.0606	0.0298	0.0712	0.0330
Jb			--	0.0016	0.0066	0.0009
Rb				--	0.0041	0.0001
Sb					--	0.0035
KJSSC	Bs	Cg	Jb	Rb	Sb	Ms
Bb	25.73	96.42	19.40	38.04	123.46	21.60
Bs	--	211.33	57.21	82.44	142.96	51.6755
Cg		--	65.38	39.86	103.63	38.7844
Jb			--	6.06	31.42	4.54984
Rb				--	19.94	3.5329
Sb					--	31.526
KJDSSC_{tan}	Bs	Cg	Jb	Rb	Sb	Ms
Bb	1.5443	10.7867	0.4092	1.1775	10.2834	0.3493
Bs	--	38.1270	6.4886	9.8655	26.4088	4.0839
Cg		--	6.9394	2.7033	14.0514	2.5418
Jb			--	0.0499	0.8327	0.0501
Rb				--	0.4950	0.0228
Sb					--	1.3571

Table A5. Spectral similarity matrix of mixed spectra of minerals prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e., 50%)) using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}.

SAM	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0855	0.1777	0.1740	0.1215	0.0641	0.0821
Calcite	--	0.1397	0.1674	0.1232	0.0545	0.0668
Chlorite		--	0.0862	0.1024	0.1642	0.0982
Hematite			--	0.1035	0.1842	0.1096
Jarosite				--	0.1280	0.0664
Kaolinite					--	0.0786
SID	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0038	0.0169	0.0168	0.0082	0.0021	0.0038
Calcite	--	0.0105	0.0152	0.0084	0.0017	0.0023
Chlorite		--	0.0039	0.0053	0.0144	0.0050
Hematite			--	0.0055	0.0185	0.0066
Jarosite				--	0.0091	0.0025
Kaolinite					--	0.0034
SIDSAM_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.0003	0.0030	0.0029	0.0010	0.0001	0.0003
Calcite	--	0.0015	0.0026	0.0010	0.0001	0.0002
Chlorite		--	0.0003	0.0005	0.0024	0.0005
Hematite			--	0.0006	0.0035	0.0007
Jarosite				--	0.0012	0.0002
Kaolinite					--	0.0003
KJSSC	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	15.99	53.00	20.94	13.59	2.59	4.58
Calcite	--	102.38	35.41	42.10	12.10	19.20
Chlorite		--	38.41	20.45	53.19	32.37
Hematite			--	10.39	25.02	8.91
Jarosite				--	16.73	4.93
Kaolinite					--	4.89
KJDSSC_{tan}	Calcite	Chlorite	Hematite	Jarosite	Kaolinite	Mixture
Alunite	0.1799	2.5328	0.3167	0.1440	0.0053	0.0177
Calcite	--	8.0007	0.7325	1.1788	0.1091	0.2718
Chlorite		--	1.4532	0.4247	2.4319	0.9691
Hematite			--	0.0951	0.4235	0.0600
Jarosite				--	0.1918	0.0168

Kaolinite					--	0.0176
-----------	--	--	--	--	----	--------

Table A6. Spectral similarity matrix of mixed spectra of vegetation prepared by linear mixing of other classes with equal proportion (50% of actual class + 8.33% of other classes (i.e. 50%)) using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}.

SAM	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0635	0.1987	0.0451	0.0821	0.1234	0.0705
Bs	--	0.2104	0.0423	0.0769	0.0817	0.0635
Cg		--	0.1833	0.1397	0.1877	0.1495
Jb			--	0.0572	0.0869	0.0406
Rb				--	0.0766	0.0212
Sb					--	0.0715
SID	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0105	0.0320	0.0026	0.0075	0.0184	0.0059
Bs	--	0.0440	0.0045	0.0058	0.0084	0.0060
Cg		--	0.0280	0.0190	0.0365	0.0192
Jb			--	0.0025	0.0093	0.0015
Rb				--	0.0069	0.0003
Sb					--	0.0064
SIDSAM_{tan}	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0007	0.0064	0.0001	0.0006	0.0023	0.0004
Bs	--	0.0093	0.0002	0.0004	0.0007	0.0004
Cg		--	0.0051	0.0027	0.0069	0.0029
Jb			--	0.0001	0.0008	0.0001
Rb				--	0.0005	0.0000
Sb					--	0.0005
KJSSC	Bs	Cg	Jb	Rb	Sb	Ms
Bb	3.47	14.11	2.86	5.62	16.50	3.06
Bs	--	20.95	7.95	10.63	20.73	6.55
Cg		--	11.03	7.20	17.67	7.07
Jb			--	1.07	5.92	0.75
Rb				--	4.05	0.60
Sb					--	6.14
KJDSSC_{tan}	Bs	Cg	Jb	Rb	Sb	Ms
Bb	0.0313	0.2787	0.0115	0.0332	0.2946	0.0088
Bs	--	0.5765	0.1581	0.2205	0.7646	0.0825
Cg		--	0.2291	0.0955	0.5231	0.0845
Jb			--	0.0018	0.0354	0.0016
Rb				--	0.0226	0.0007

Sb					--	0.0574
----	--	--	--	--	----	--------

Table A7. Spectral similarity matrix of hydrothermal alteration minerals using SAM, SID, SIDSAM_{tan}, KJSSC, and KJDSSC_{tan}.

SAM	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	0.1154	0.1360	0.1228	0.1933	0.1174
Chlorite	--	0.1334	0.0791	0.1116	0.0544
Chlorite _(mine site)		--	0.1418	0.1603	0.1134
Kaolinite			--	0.1004	0.0463
Muscovite				--	0.0797
SID	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	0.0085	0.0160	0.0136	0.0181	0.0069
Chlorite	--	0.0249	0.0060	0.0060	0.0021
Chlorite _(mine site)		--	0.0383	0.0287	0.0196
Kaolinite			--	0.0059	0.0039
Muscovite				--	0.0029
SIDSAM_{tan}	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	0.0010	0.0022	0.0017	0.0035	0.0008
Chlorite	--	0.0033	0.0005	0.0007	0.0001
Chlorite _(mine site)		--	0.0055	0.0046	0.0022
Kaolinite			--	0.0006	0.0002
Muscovite				--	0.0002
KJSSC	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	7.25	13.03	69.80	157.07	18.84
Chlorite	--	36.02	32.29	84.97	3.64
Chlorite _(mine site)		--	161.59	302.11	62.36
Kaolinite			--	12.49	14.10
Muscovite				--	50.45
KJDSSC_{tan}	Chlorite	Chlorite _(mine site)	Kaolinite	Muscovite	Mixture
Goethite	0.20	0.82	11.61	44.16	1.15
Chlorite	--	5.02	2.51	14.27	0.04
Chlorite _(mine site)		--	55.20	139.56	12.39
Kaolinite			--	0.34	0.50
Muscovite				--	5.42