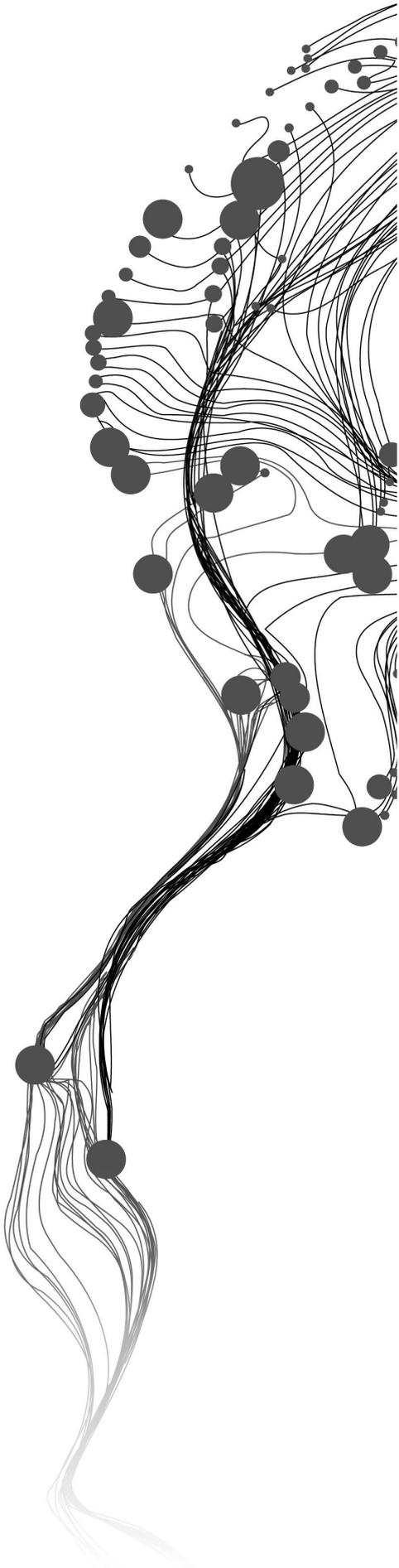


ANALYSIS OF OPENCAST MINING CHARACTERISTICS USING MULTI-TEMPORAL REMOTE SENSING IMAGES AND GOOGLE EARTH ENGINE

JITENDRA SHANKARAI AH
September, 2021

SUPERVISORS:
Shri. Vinay Kumar
Dr. Mariana Belgiu
Shri. Kamal Pandey



ANALYSIS OF OPENCAST MINING CHARACTERISTICS USING MULTI-TEMPORAL REMOTE SENSING IMAGES AND GOOGLE EARTH ENGINE

JITENDRA SHANKARAI AH
Enschede, The Netherlands, September,
2021

Thesis submitted to the Faculty of Geo-information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.
Specialization: Geoinformatics

SUPERVISORS:

Shri. Vinay Kumar
Dr. Mariana Belgiu
Shri. Kamal Pandey

THESIS ASSESSMENT BOARD:

Prof.Dr.Ir. Alfred Stein (Chair)
Dr. Kumar Gaurav (External examiner, IISER, Bhopal)

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Opencast coal mining is the most economical and also the most intense mining practice that drastically alters the land use land cover (LULC) in and around the mining areas. Coal fires are found in coal mines where combustion of coal is extrinsically affected by mining practices. Recent launches of Earth Observation satellites have provided open access to their data thereby giving the opportunity to continuously map opencast coal mines, coal fires and their rapid changes using relatively high spatial and temporal resolution images. This study focuses on using the computational resources of Google Earth Engine (GEE) and its cloud platform for developing web-based application to demonstrate the characteristics of opencast coal mining at Jharia coal field (JCF) in India. Openly accessible multi-temporal data from Sentinel-2 and ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) were used to respectively characterize active coal mining areas, areas affected by coal fires, and their changes between 2019 and 2020.

A Random Forest (RF) classification algorithm was used to classify coal mining areas and other LULC classes with samples generated from random sampling approach, and applying non-heuristic data balancing techniques to address class imbalance problem. The classification achieved overall accuracy of 73.09% and 76.5% for 2019 and 2020 respectively. The change detection results found that the coal mining activity decreased from 16.30 km² in 2019 to 12.51 km² in 2020.

Two dynamic threshold based methods were applied to detect areas affected by coal fires using Land Surface Temperature (LST) images. First method used images with minimum cloud cover and achieved 90% accuracy at detecting coal fires in 2019. A detection ratio was calculated to measure how consistently the pixels were detected as coal fires. The first method achieved 70% accuracy for a detection ratio above 0.5. The LST variability in day and night time images showed that the diurnal images could be independently analysed. The second method used all images to create a 95th percentile composite of day and night images, and achieved 75% accuracy at detecting coal fires. Both methods showed an increase in the extent of the coal fires during the observation period with new fires propagating around the active fires.

The results of this study show effective use of medium-high resolution Sentinel-2 imagery and fine-scale ECOSTRESS LST images to provide consistent information regarding changes in coal mining areas and coal fire areas. The ability to process high volume of data and perform analyses in real-time shows a considerable advance in remote sensing technology towards adopting cloud-based applications in mining operations. Integration of these mining characteristics in an interactive, free and openly accessible application can facilitate stakeholders and policy makers in informed decision making and planning mining operations. The application also provides a framework to facilitate future research through sharing of scripts and ensuring reproducibility.

Keywords

Remote sensing, Image classification, Image segmentation, Random Forest, Opencast coal mine, Jharia, Change detection, Coal fire, LULC change, LST, ECOSTRESS, Sentinel-2, Google Earth Engine

ACKNOWLEDGEMENTS

I would like to thank my supervisors Shri. Vinay Kumar, Dr. Mariana Belgiu and Shri. Kamal Pandey for being the best mentors that I needed throughout the study. I could not have completed the study without their guidance and support. I am grateful to them for taking time to clear my doubts, giving me confidence to study, for phone calls, online meetings, reviewing my work and giving continuous feedback.

I am grateful to Dr. Sameer Saran who played a key role in making our course go smoothly and who knew what had to be done even before we were aware. My gratitude goes towards Dr. Sander Oude Elberink for coordinating the course, Dr. Serkan Girgin for introducing CRIB, and Drs. Petra Budde for data support. I thank everyone at IIRS and ITC, where faculty shared their knowledge, and staff who made our stay memorable, even though it was cut short. And my batchmates from MSc. for being there all the time.

Further acknowledgements go out to Google Earth Engine, R, Python, QGIS, and their developer communities for sharing their knowledge and resources.

I am extremely lucky to have family and friends who have been an integral part of my journey.

I am forever indebted to *Amma, Appaji, Anna* who have been supportive all the time.

And finally, Covid-19, there was an overall loss because of you.

TABLE OF CONTENTS

Abstract	i
Acknowledgements	ii
1 Introduction	3
1.1 Background	3
1.2 Motivation and problem statement	3
1.3 Research identification	5
1.3.1 Research Objectives	5
1.3.2 Research Questions	5
1.3.3 Contribution to science and innovation	5
1.4 Thesis structure	6
2 Literature review	7
2.1 Related work	7
2.2 LULC classification in mining areas	8
2.3 Class imbalance	9
2.4 Random forest classifier	10
2.5 Classification on GEE	10
2.6 Land Surface Temperature	11
2.7 Change detection	11
3 Study area and datasets	13
3.1 Study area	13
3.2 Datasets used	14
3.2.1 Sentinel 2	14
3.2.2 ECOSTRESS	15
4 Methodology	17
4.1 Data preparation	17
4.1.1 Sentinel 2	17
4.1.2 ECOSTRESS	18
4.1.3 LULC classification scheme	18
4.2 Reference data	19
4.3 Class imbalance	20
4.4 RF Classifier	21
4.5 Accuracy assessment and area estimation	22
4.6 Change detection	22
4.7 Detection and change in coal fire areas	23
4.8 Detecting coal fires with percentile approach	23
4.9 Validation of coal fires	24
4.10 Application on GEE	24

5	Results	27
5.1	Effects of balancing methods	27
5.2	Classification results	28
5.2.1	LULC classification and accuracy assessment	28
5.2.2	Area estimation	30
5.3	Change detection	31
5.3.1	Change detection matrix	31
5.3.2	Change in coal mining areas	31
5.4	Coal fire detection	33
5.4.1	Effect of clouds on LST images	33
5.4.2	Threshold of LST images and coal fire detection	34
5.4.3	Percentile approach	34
5.4.4	Validation	35
5.4.5	Relationship between ground temperature and LST	36
5.5	Change in coal fire areas	37
5.6	GEE app	39
6	Discussion	41
6.1	Effect of balancing methods	41
6.2	LULC classification and change detection	41
6.3	Effect of clouds	42
6.4	Coal fire detection	42
6.5	Evaluation of GEE app	44
7	Conclusion and recommendation	45
7.1	Conclusion	45
7.2	Future recommendation	46
	Appendix A	55

LIST OF FIGURES

3.1	False color composite map showing study area of Jharia coalfield, India. Data: Sentinel-2 March, 2020 from https://earthexplorer.usgs.gov/ . Map CRS: WGS84. India map source: Saini, Gupta, and Arora (2016)	13
3.2	Collieries affected by coal fires in JCF. Image source: BCCL (2008)	14
4.1	Overall methodology of the study	17
4.2	Effect of cloud masking on pixel availability of LST images	19
4.3	Variation in indices of samples of 2019 and 2020 along with temporal patterns: Q1 Jan-Mar, Q2 Apr-Jun, Q4 Oct-Dec	21
4.4	Mockup of GEE application showing implementation and interactive elements	24
5.1	LULC classified map	29
a	Year 2019	29
b	Year 2020	29
5.2	Transition between LULC classes during 2019-2020	32
5.3	Net changes of LULC classes during 2019-2020	32
5.4	Change in coal mining areas between 2019 and 2020	33
5.5	Distribution of pixel coverage and number of LST images in 2019 and 2020. Pixel coverage calculated as ratio of number of pixels available after cloud masking and total number of pixels available within colliery boundary	34
5.6	Diurnal spatial variability of LST at coal mine affected by coal fires in JCF and indicated in top left corner of each image (A-C). The color scale in degree Celsius is enhanced for each image. Figure C shows pixels identified as clouds over the coal fire areas. FCC image is acquired from PlanetScope dated December, 2019	35
5.7	Threshold and associated pixel coverage of day and night time images in 2019 and 2020	36
5.8	Number of pixels delineated as fire and their detection ratio for 2019 and 2020	36
5.9	Coal fires detected with detection ratio of pixels above the LST threshold	37
a	Year 2019	37
b	Year 2020	37
5.10	Coal fires detected by percentile approach	38
a	Year 2019	38
b	Year 2020	38
5.11	Scatter plot of correlation between ground temperature and LST values acquired during day and night time acquisitions	39
5.12	Change in coal fire areas between 2019 and 2020	39
a	detection ratio > 0.5	39
b	percentile approach	39
5.13	Image of developed GEE application tool <i>Jharia</i>	40

LIST OF TABLES

3.1	Sentinel 2 and ECOSTRESS images available for the study area without cloud masking.	15
3.2	Spectral bands of Sentinel 2 imagery	15
3.3	8-bit encoding for cloud and water mask of ECOSTRESS L2 Cloud Mask Product (ECO2CLD) SDS	16
4.1	Sentinel 2 and ECOSTRESS images used in the study after cloud masking	18
4.2	Training and validation data used for 2019 classification	20
4.3	Training and validation data used for 2020 classification	20
5.1	Effects of increase and decrease in balancing methods on classification accuracies (in %) for 2019	27
5.2	Effects of increase and decrease in balancing methods on classification accuracies (in %) for 2020	28
5.3	Classification error matrix	30
a	Year 2019	30
b	Year 2020	30
5.4	Mapped areas (in km ²) and post stratified error adjusted estimates (at 95% confidence interval) of LULC classes for 2019 and 2020 classification	31
5.5	Change detection matrix of LULC classes from 2019 to 2020. Mapped areas in km ² for the study area of 447.3 km ²	33
5.6	Validation of detected coal fire pixels with recorded ground observations. Temperatures are in °C. * Locations that were detected within a 70-m from the coal fire pixels and corresponding detection ratio of the nearest pixel is shown. †Locations that did not validate with the percentile approach	38
A.1	Time stamp and corresponding ECOSTRESS granule ID of ECO2LSTE, ECO1BGEO and ECO2CLD data products for the study area. Filenames of the products and processed images can be identified with the granule ID	55
A.2	Land cover mapping scheme used in this study	57
A.3	Citation and references to scripts and datasets used in this study	57

List of abbreviations

LULC	Land Use Land Cover
RF	Random Forests
ECOSTRESS	ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station
GEE	Google Earth Engine
ML	Machine learning
UTC	Coordinated Universal Time
IST	Indian Standard Time
FCC	False Color Composite
ROS	Random oversampling
RUS	Random undersampling
LST	Land Surface Temperature
SDS	Scientific Datasets
RS	Remote sensing
QC	Quality control
WMTS	Web Map Tile Service
OOB	Out of bag
API	Application Programming Interface
OA	Overall accuracy
PA	Producer accuracy
UA	User accuracy
GM	Geometric mean
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
CLAL	Clay Alunite Index
DEM	Digital Elevation Model
MSI	Multispectral Instrument
BOA	Bottom of Atmosphere
TOA	Top of Atmosphere
SR	Surface Reflectance
JCF	Jharia coalfield
BCCL	Bharat Coking Coal Limited
TIR	Thermal infrared
DT	Decision Trees
KNN	k-nearest neighbor
ANN	Artificial neural network
CART	Classification and regression trees
SVM	Support Vector Machines
EVI	Enhanced Vegetation Index
ARD	Analysis Ready Data
MODIS	Moderate Resolution Imaging Spectroradiometer
VIIRS	Visible Infrared Imaging Radiometer Suite
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
NOAA	National Oceanic and Atmospheric Administration
ESA	European Space Agency
NASA	National Aeronautics and Space Administration
LPDAAC	Land Processes Distributed Active Archive Center
UHI	Urban Heat Island

Chapter 1

Introduction

Opencast coal mining is one of the most intensive forms of landscape change moderated by humans (Kuenzer, 2015). Mining significantly changes both natural and anthropogenic systems in the mining and its surrounding areas. The communities living near or over the resources are affected both directly and indirectly by mining operations (Ghosh, 1991; Hina, D., K., & Anupma, 2015). Therefore, it is important to monitor the mining areas which require time series analysis. Currently a lot of satellite based observations are taking place but processing these time series data with conventional approach is not only time consuming but also requires computational resources (Tamiminia et al., 2020). Cloud based platforms are able to provide necessary tools to process such data and can be customized for instance solution (Kumar & Mutanga, 2018). These customized solutions can aid in the visualization, analysis and generation of quick reports from time series satellite data. The derived outputs from such solutions can not only provide decision reference data for the mining industry but also influence environmental protection policies (Amani et al., 2020).

1.1 BACKGROUND

One version of the story of humanity is the story of extraction. Extraction of resources to make tools; generate energy to make more tools; a continuous cycle.

Coal has been a source for energy consumption for a long time. Activities such as opencast mining and underground mining are practiced to extract coal. The effects of the opencast mining such as land cover-land use (LULC) change, coal fires and land subsidence are predominantly found on the land surface. Land subsidence is lowering of the earth surface as a result of mining operations. Distinguishing these effects in spatial and temporal dimensions is a task that is particularly suited to remote sensing (RS) which has advantages of repeated data acquisition, its comprehensive view and data processing. There are also multiple challenges associated with operationalizing remote sensing applications for the mining industry. Remote sensing assessments require skills and knowledge to interpret output of analysis, and derive meaningful results (McKenna, Lechner, Phinn, & Erskine, 2020). On the ground examples of how remote sensing has been integrated into operations are still rare and the scientific literature is dominated by one-off research studies that are limited to a single time.

1.2 MOTIVATION AND PROBLEM STATEMENT

Mining is a major contributor to economy and requires continuous information about attributes such as area, change, reclamation, restoration, social and economic impacts (de Lucia Lobo, Souza-Filho, de Moraes Novo, Carlos, & Barbosa, 2018). Accurate mapping of mining areas can provide such missing information (Chen, Li, He, & Wang, 2017). This lack of information can also be explained due to the scale of the mining activity. Significant changes in the mining activity occur at smaller areas ($\leq 10000\text{m}^2$) which can be detected by satellite images with resolution $\leq 20\text{m}$ (de

Lucia Lobo et al., 2018). Coal fires in the mining areas can be described as extremely weak anomalies when compared to forest fire detection in thermal remote sensing with thermal infrared (TIR) images (Kuenzer, 2015). Coal fires can be characterized as surface and sub-surface fires (Kuenzer, 2015). These fires burn at different depths and at different intensities. Coal fires also occur at coal dump and coal debris areas. Mining operations, fire mitigation measures, and distribution of coal seams, affect temporal intensity and propagation of these fires (Mujawdiya, Chatterjee, & Kumar, 2020; Raju, Gupta, & Prakash, 2013). This has made detection of coal fires in thermal remote sensing very difficult (Kuenzer, 2015; Prakash & Gupta, 1998; Zhang, 2004). Some of these fires have spatial extent smaller than the pixel size of a TIR image and few sub-surface fires produce weak anomaly on the surface which are difficult to detect (Biswal, Raval, & Gorai, 2019; Kuenzer, 2015). Coal fire detection and monitoring with thermal remote sensing have emphasised the importance of images with high temporal and spatial resolution (Chatterjee, 2006; Du et al., 2015; Huo et al., 2015; F. Li, Yang, Liu, Sun, & Liu, 2018; Prakash & Gupta, 1998; Raju et al., 2013; Zhang, 2004). Landsat provides spatial resolution ($>30\text{-m}$) but at low temporal resolution (16 days). Visible Infrared Imaging Radiometer Suite (VIIRS) and Moderate resolution Imaging Spectroradiometer (MODIS) provide high temporal resolution (everyday) but low spatial resolution ($\geq 375\text{-m}$). Since the quality of these images are affected by clouds, the number of images available for analysis is further limited. Until recently, availability of images at both high spatial and temporal resolution were limited. Sentinel-2 constellation of satellites, from 2017 onwards, provide higher spatial ($\leq 20\text{-m}$), temporal (5 days) and spectral resolution data suitable for mapping mining areas (Chen et al., 2017; de Lucia Lobo et al., 2018). The ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS), launched in 2018 is a multispectral thermal infrared radiometer instrument aboard the International Space Station (ISS). It has an irregular orbit, collects data at different times a day, and with a revisit time less than 5 days. The data is made available at 70-m resolution. This provides good spatial and temporal resolution along with diurnal cycle acquisition (Fisher et al., 2020). The primary objective of the mission is to measure evapotranspiration to monitor the health of plants (Fisher et al., 2020). The retrieved data products such as Land Surface Temperature (LST) have been used to monitor forest fires in California (NASA/JPL-Caltech, 2020) and test its efficacy on thermal anomalies over Italian volcanic and geothermal areas (Silvestri et al., 2020). However, there have been no studies using ECOSTRESS for coal fire detection. Thanks to European Space Agency (ESA) and National Aeronautics and Space Administration (NASA), the satellite data is publicly available for use in various studies and applications (Arévalo, Bullock, Woodcock, & Olofsson, 2020; Kumar & Mutanga, 2018).

With increase in high resolution and multi-temporal data made available, many at global scale, comes the challenge of implementing RS techniques (Sun & Scanlon, 2019). The challenges of processing and analysing such a high volume of data requires high computational resources along with an architecture to implement RS techniques in safe, reproducible and efficient ways. Cloud computing platforms such as Google Earth Engine (GEE), Amazon Web Services, Microsoft Azure Cloud Services have been addressing this problem by providing architecture and tools to build customised RS based studies and applications (Gomes, Queiroz, & Ferreira, 2020). GEE, launched by Google in 2010 is more popular among the RS researchers with over 400 applications in various domains such as agriculture, land cover, vegetation, hydrology, image processing, urban, and pedosphere, published within a decade (Amani et al., 2020). One of the reasons for emergence of such applications is largely due to availability of pre-built functions which can be easily understood by the user, and can be used to build web-based applications to communicate research output (Gorelick et al., 2017). Pericak et al. (2018) developed a similar application to map yearly extent of surface coal mining over an 83,000 km² Appalachia, using Landsat and GEE.

Jharia coalfield (JCF) as shown in Figure 3.1 is one of the major sources of coking coal in India with

a history of mining for more than 100 years. The active mining area covers 2.96% of the 447 km² area and over the years mining areas have increased by taking up new areas for opencast mining by Bharat Coking Coal Limited (BCCL) (BCCL/CMPDI, 2020). JCF also has a history of coal fires for over 100 years which has now affected around 70% of total coal mines (Pandey, Kumar, Panigrahi, & Singh, 2017). Effects of mining such as land subsidence has also affected land surface around the mining areas as well as human settlements around it (Chatterjee, 2006; Kuenzer, 2015). The scientific literature has been inundated with studies at the Jharia coalfield (Biswal & Gorai, 2020; Gautam, Singh, Mittal, & Sajin, 2008; R K Mishra, Bahuguna, & Singh, 2011; Mujawdiya et al., 2020; Mukherjee, Mukherjee, Chakravarty, & Aikat, 2019). All studies relate to individual characteristics of mining. Application of RS methods which can provide accurate tools has increased the demand for timely information in near real time on the nature and extent of LULC changes in mining areas and coal fire areas (Karan, Samadder, & Maiti, 2016).

1.3 RESEARCH IDENTIFICATION

Integrating different mining characteristics into an RS based application can provide a snapshot of the effects related to mining. Satellite data from the Sentinel-2 and ECOSTRESS missions that are available and openly accessible can be used to map LULC classes and coal fires, monitor change in mining areas and coal fire areas. With the help of resources made available by cloud computing platforms such as GEE, the application can be used as tool by the mining industry and government organisations to monitor the effects of mining which would influence decision making by the stakeholders for planning, mitigation, reclamation, restoration, rehabilitation and closure measures. This would also enable sharing of scripts and data to ensure reproducibility.

1.3.1 Research Objectives

The main objective is to develop a tool to monitor opencast mining characteristics from time series multi-temporal satellite data on GEE platform. This study also aims to attain the following sub-objectives while pursuing the main objective.

1. To assess the change in the opencast(coal) mining area using Sentinel-2 time series multi-temporal data
2. To assess the potential of ECOSTRESS data to identify locations affected by coal fires

1.3.2 Research Questions

1. How has the surface area of the mines changed in the study area?
2. Which ECOSTRESS images are more effective in identifying coal fires?
3. Is there any relationship between LST and the observed temperature at the coal fire locations?

1.3.3 Contribution to science and innovation

Developing the tool using openly accessible data that combines the analysis of opencast mining characteristics on a cloud-based platform such as Google Earth Engine. Until now, no research study has used ECOSTRESS data as a potential data to delineate coal fire locations.

1.4 THESIS STRUCTURE

This document is structured as follows. A literature review and related work is provided in Chapter 2. Chapter 3 describes the study area, and the satellite imagery used. This is followed by Chapter 4 explaining methodology adopted in the study. Results of the LULC classification, coal fire detection, validation, and change detection are shown in Chapter 5. Interpretation of results and its context with previous studies are discussed in Chapter 6. Finally, Chapter 7 summarises the study by answering research questions, and suggesting future recommendations.

Chapter 2

Literature review

2.1 RELATED WORK

Various studies have used supervised and unsupervised methods for land cover-land use classification with time series images (Hina et al., 2015; Müller, Rufin, Griffiths, Siqueira, & Hostert, 2015; Petropoulos, Partinevelos, & Mitraka, 2013). One of the limiting factors that affect choosing images is cloud cover. One method is using a temporal aggregation method, where mean and median are derived from the time series images (Beckschäfer, 2017). Another method is using only cloud free images to make time series data (de Lucia Lobo et al., 2018; Griffiths, van der Linden, Kuemmerle, & Hostert, 2013). do Nascimento Bendini et al. (2019) proposed using a gaussian function as kernel smoother to fill the data gaps in cloud covered pixels to create dense time series. These methods have been used with Landsat, and Sentinel-2 satellite images. Belgiu and Stein (2019) review methods such as spatiotemporal image fusion are used to blend fine spatial resolution images with high temporal resolution images to fill data gaps due to presence of clouds.

Use of Machine Learning (ML) techniques has been effective in mapping coal mining areas and LULC classification and monitoring (de Lucia Lobo et al., 2018; Demirel, Emil, & Duzgun, 2011; Mukherjee et al., 2019). These studies have also emphasised the importance of using high spatial resolution multispectral satellite images for identifying changes in the opencast coal mine areas. Indices such as Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) have been proposed to discriminate land cover classes in mining areas using spectral bands (Karan et al., 2016; Xianju Li, Chen, Cheng, & Wang, 2016; Nascimento et al., 2020). These indices are effective in classifying vegetation areas and water bodies. Mukherjee et al. (2019) proposed a Clay Alunite Index (CLAL), derived from normalized difference of two short wave infrared bands of Landsat-8 to detect opencast coal mines and delineate coal quarry and coal overburden dump regions with average accuracy of 86.24%. This index could be used for classification and monitoring of opencast mining regions.

For LULC classification of mining areas, ML classifiers such as Support Vector Machine (SVM) and Classification And Regression Trees (CART) have been used by de Lucia Lobo et al. (2018), Demirel et al. (2011), and Petropoulos et al. (2013). Validation of classification results in these studies were performed with selection of high resolution satellite images of QuickBird, RapidEye, IKONOS, WorldView-2 and Google Earth. Xianju Li et al. (2016) made comparisons of ML algorithms such as Random Forest (RF), SVM, and Artificial Neural Network (ANN) for classification of opencast mining areas and agricultural landscapes. The study found that RF had greater accuracy followed by SVM and ANN. RF was also less sensitive to feature selection. However, SVM was effective in classifying only opencast mining areas. The study suggested using spectral bands in feature selection along with derived indices. The study also raises the issue of class imbalance of training samples where samples that usually represent classes of interest are under-represented compared to classes that dominate the landscape are over-represented. ML classifiers such as RF and SVM are sensitive to such class imbalance and do not yield high accuracy (Waldner, Chen, Lawes, & Hochman, 2019). Waldner et al. (2019) suggest optimal balancing methods after testing

various undersampling and oversampling methods to balance training data for crop mapping, using RF and SVM algorithms.

Biswal et al. (2019) summarise various studies performed on detection and delineation of coal fires using remote sensed data. One of the methods is to derive Land Surface Temperature (LST) using thermal bands from satellite images such as Landsat, MODIS, VIIRS and ASTER. MODIS offers LST products but they are at 1km resolution. Landsat images have been widely used and algorithms proposed by Ermida, Soares, Mantas, Göttsche, and Trigo (2020) and Wang et al. (2020) use GEE Landsat images to estimate LST. ECOSTRESS is a relatively new mission and offers derived LST from 2018 at 70-m resolution and higher temporal resolution. Such high spatiotemporal resolution data enabled G. Hulley, Shivers, Wetherley, and Cudd (2019) to identify communities in Los Angeles city that were vulnerable to urban heat island effects. Silvestri et al. (2020) used ECOSTRESS for studying thermal anomalies over volcanic and geothermal areas in Italy. The study also compared LST estimations of ASTER and Landsat 8 with ECOSTRESS and found them agreeable. M. C. Anderson et al. (2021) proposed data fusion algorithm to fuse ECOSTRESS, Landsat and MODIS to generate high spatial and temporal resolution time series for mapping evapotranspiration.

Effective method to detect the presence of coal fires is by identifying anomalies in LST with the surrounding areas. One approach is to use field data of coal fire location and temperatures to calibrate the model to delineate fire zones (Pandey et al., 2017) and monitor its trend (Mujawdiya et al., 2020). Another is to use threshold technique to identify temperature range in LST of each acquired image (Rakesh Kumar Mishra, Pandey, Pandey, Kumar, & Roy, 2020; Prakash & Gupta, 1999; Raju et al., 2013). This approach relies on distribution of LST values in the images. The distribution of LST values can vary with seasonal differences, day and night time acquisition and cloud cover (Du et al., 2015; Huo et al., 2015; F. Li et al., 2018; Rakesh Kumar Mishra et al., 2020). Du et al. (2015) proposed a Self-Adaptive Gradient-Based Threshold method for coal fire detection from LST derived from ASTER data. It also uses mathematical morphology thinning method to create boundaries around the coal fire anomalies. This method has been used by F. Li et al. (2018) to identify coal fire areas using night time images with 92% accuracy.

2.2 LULC CLASSIFICATION IN MINING AREAS

LULC classification is a process of assigning land cover classes to pixels in order to categorize them. Extracting LULC data from remote sensing images requires good classification practises. It involves complex processes and requires consideration of many factors such as suitable classification system, choice of remote sensing data, sampling design, image processing, feature extraction, selection of classifier and its parameters, post-processing and accuracy assessment (Lu, Mausel, Brondízio, & Moran, 2004).

Detection of coal mining areas has various applications: LULC mapping, change detection in mining areas, coal fire detection and monitoring, management of mining waste, reclamation and restoration (Mukherjee et al., 2019). Mapping mining areas has been considered challenging in various areas across the world (de Lucia Lobo et al., 2018; Demirel et al., 2011). Mining areas are complex and continuously changing landscapes across different regions of the world. Some surface mining regions are located near towns and cities (Hina et al., 2015; Mukherjee et al., 2019; Prakash & Gupta, 1998). In certain regions, mining areas can be visually distinguished, but detection through algorithms have challenges. Differentiating active, abandoned and closed mines with multispectral data is difficult as they share similar spectral signatures. Numerous studies have used classification techniques such as unsupervised, supervised, parametric, non-parametric, pixel-based, object-based in mapping mining areas (Chen et al., 2017; de Lucia Lobo et al., 2018; Demirel

et al., 2011; Xianju Li et al., 2016; Nascimento et al., 2020; Prakash & Gupta, 1998; Ranjan, Sahoo, & Gorai, 2021). It is important to select an appropriate classification approach based on the nature of the problem (Lu et al., 2004). Use of spectral signatures from multispectral images has been emphasized to distinguish mining areas. Indices such as NDVI, Enhanced Vegetation Index (EVI), NDWI, Normalized Difference Built-up Index (NDBI), and CLAL have been widely used to distinguish vegetation, water bodies, artificial surfaces and mining regions (Mukherjee et al., 2019). The values of these indices vary with seasons and different climatic conditions. The properties of these indices and spectral bands can be captured with multitemporal image composites and in time series to aid supervised classification in applications of LULC mapping, crop monitoring, crop classification and mapping urban areas (Abdi, 2019; Kollert, Bremer, Löw, & Rutzinger, 2021; Pu et al., 2020). Demirel et al. (2011) and Petropoulos et al. (2013) used SVM supervised ML algorithms to map surface coal mines in Turkey and Greece respectively, using multitemporal imagery, and quantified change in mining areas with post classification change analysis.

2.3 CLASS IMBALANCE

Most prediction based learning algorithms assume that the training data they receive for learning is balanced. But in real world data there are always classes represented by a larger number of samples compared to other classes (Batista, Prati, & Monard, 2004). This is called a class imbalance problem. The imbalances can be found in datasets with binary classes where ML classifiers are diagnosing diseases, filtering email spams, fault diagnosis and anomaly detection. This is also evident in datasets for LULC classification with multiple classes (Waldner et al., 2019). Some classes in the study area usually cover smaller regions while others cover larger regions, which makes it difficult to obtain an equal number of samples for all classes, especially in a random sampling design. A multiclass imbalanced dataset is characterized by majority classes that have large instances and minority classes with fewer ones. Most ML classifiers are implicitly or explicitly designed to decrease overall error rate which guides them towards prediction of classes that are over represented in the dataset (Naboureh, Ebrahimy, Azadbakht, Bian, & Amani, 2020).

There are different approaches to solving this problem such as ensemble techniques and algorithm-based approaches, and data resampling (Shariffar, Sarmadian, & Minasny, 2019). Ensemble techniques involve building an ensemble of classifiers rather than one classifier. Here the classifiers such as AdaBoost, RUSBoost are trained to pre-process data before the training stage of the classifier (López, Fernández, García, Palade, & Herrera, 2013). Algorithm-based approach involves modifying existing classifiers by enhancing their discriminatory power to separate classes through methods of objective transformation method, kernel transformation method, clustering algorithm, and task decomposition strategies (Haixiang et al., 2017). Data resampling approach is based on two baseline non heuristic methods, Random undersampling and Random oversampling (Batista et al., 2004). Data resampling is however an easier approach to apply (Waldner et al., 2019). Random undersampling (RUS) balances class distribution by randomly removing samples from majority classes. Random oversampling (ROS) balances class distribution by randomly replicating samples with replacement, from minority classes. Although both methods increase accuracy of classification, particularly for minority classes, there are drawbacks associated with them. RUS can discard useful information from the removed classes and ROS can increase likelihood of overfitting with replicated instances of the same sample (Batista et al., 2004). Several heuristic based approaches have been developed to address this problem. Batista et al. (2004) evaluated performances of these methods and one the results was that ROS was ranked above RUS. Data balancing methods are known to increase overall accuracy of the classification and particularly increase producer's accuracy of the minority classes (Waldner et al., 2019). Two metrics: Overall accuracy (Equation 4.3),

which favors majority classes; Geometric mean of producer's accuracy (Equation 4.6), which is sensitive to minority classes, are proposed to evaluate performance of the classifier providing a summary of accuracy of both majority and minority classes when compared to unbalanced classification (Waldner et al., 2019). The study used full balancing, static and partial balancing methods to generate imbalance datasets to compare different balancing methods on RF and SVM classifiers. The study found that heuristic based methods outperform non-heuristic methods but ROS and RUS methods would still perform better than unbalanced dataset.

2.4 RANDOM FOREST CLASSIFIER

Machine learning (ML) classifiers can be broadly classified into parametric and non parametric. Non-parametric classifiers such as Random Forests (RF), Support Vector Machines (SVM), CART have been used in LULC classification with highly accurate results (Foody & Mathur, 2004; Xianju Li et al., 2016; Petropoulos et al., 2013; Teluguntla et al., 2018).

RF has been a popular classifier due to its high performance and low sensitivity to noise, and is less sensitive to feature selection. Random forests is one of the most popular ML classifiers used for multi-temporal LULC classification (Talukdar et al., 2020). Xianju Li et al. (2016) made comparisons of Random Forests (RF), SVM, and Artificial Neural Network (ANN) ML algorithms for classification of opencast mining areas and agricultural landscapes. The study found that RF had greater accuracy followed by SVM and ANN. Similar findings were reported when comparing ML algorithms for LULC classification (Talukdar et al., 2020). However, the performance is affected when learning from extremely imbalanced datasets which favors prediction accuracy of majority class over minority class (Douzas, Bacao, Fonseca, & Khudinyan, 2019).

2.5 CLASSIFICATION ON GEE

Several studies have emphasized the use of multi-temporal satellite images in remote sensing applications (Tamiminia et al., 2020). Multi-temporal images hosted on GEE have been used in 96% of the published studies between 2011 and 2017, with LULC applications accounting for 10% of them (Kumar & Mutanga, 2018). The classification approach is mostly pixel-based as GEE users do not have access to object based analysis functionalities (Kumar & Mutanga, 2018). GEE provides non parametric algorithms such as DT, CART, KNN, RF, non-linear SVM and ANN classifiers along with integration of TensorFlow capabilities that have been used in various LULC based studies (Gorelick et al., 2017).

One of the important steps in overall classification methodology is pre-processing images. Agencies provide tools to automate the process so that multiple images could be processed in limited time. Some platforms provide pre-processed data at various levels of readiness, also known as Analysis Ready Data (ARD). However, the level of readiness varies depending on the user's needs, which would require additional tools and hardware in processing the images. GEE provides a cloud-based platform where a wide array of such data sets, most at planetary scale, are made available with tools to process the images. The platform hosts remotely sensed data such as Landsat, MODIS, Sentinel - 1, 2, 3, 5-P, ALOS and NOAA AVHRR, spanning over 40 years along with ARD such as EVI, NDVI, LST, DEM at global scale (Gorelick et al., 2017). A number of studies have studied LULC dynamics at different scales. Lee, Cardille, and Coe (2018) fused Landsat data with GlobCover data to improve spatial resolution of data from 300-m to 30-m for Brazil. Hansen et al. (2013), in a landmark study, used computational power of GEE to process 654,178 Landsat images to identify global forest change from 2000 to 2012 at 30-m resolution. According to the study, it took about 100 hours of computation on GEE to generate results. ML classifiers require

high computational power and GEE provides unprecedented opportunities to use this platform for analysis, visualization and applications.

2.6 LAND SURFACE TEMPERATURE

To feel if a surface is hot or cold to touch, we may sometimes only need to put our hand near the surface of the object to feel rather than touching it. Land Surface Temperature (LST) is a similar variable that tells us about the Earth's surface. It is different from the temperature of the air recorded by a weather station on the surface. LST, for bare soil is temperature above its surface, while for a tree, it is the temperature over the canopy (Glynn C Hulley et al., 2019). Radiative energy emitted by Earth's surface can be captured in the thermal infrared (TIR) spectrum (8-12 μ m). For space-borne instruments with different sensors in this spectrum, LST is calculated by estimating surface emitted radiance after accounting for atmospheric correction. The radiation emitted by the surface is different for TIR wavelengths, and is a function of temperature and emissivity. There are deterministic and non-deterministic approaches for LST calculation. F. Li et al. (2018) provide a summary of satellite-derived LST retrieval algorithms. LST retrievals in the TIR have lower uncertainties and higher spatial resolution compared to the Microwave spectrum (3-5 μ m), but are sensitive to presence of clouds. Satellite LST products such as MOD11A1, VIIRS LST are accompanied with quality information, often as bit flags which indicate which LST pixels can be used with sufficient confidence (Hulley and Ghent, 2019). LST has been used in various applications such as crop stress monitoring, Urban Heat Island (UHI) effect, drought monitoring, tracking vector borne diseases and volcanology. MODIS developed a global disturbance index using maximum LST values along with EVI to detect ecosystem differences such as occurrence of wildfires in a forest (Glynn C Hulley et al., 2019). LST values on different land surfaces are affected by the solar illumination. Different surfaces have heat absorption properties and stored heat is dissipated at different rates (Chatterjee, 2006; Glynn C Hulley et al., 2019). So, LST values over the same surface can record different temperatures in images acquired at different times within 24 hours. Image acquisition time plays an important role in studies that characterize thermal anomalies, who prefer images acquired at a particular time of day or night, at different temporal resolutions (Biswal & Gorai, 2020; Chatterjee, 2006; Huo et al., 2014; Mujawdiya et al., 2020; Raju et al., 2013). LST has also been used to detect coal fires in different regions of the world. Coal fires can be categorized into surface and subsurface fires (Prakash & Gupta, 1999). Coal fire anomaly detection, especially of the subsurface fires, are considered difficult to be identified with thermal remote sensing (Kuenzer, 2015). Several methods are employed to detect fires. Most studies used field observations to characterise and delineate areas with coal fires (Gautam et al., 2008). Mujawdiya et al. (2020) studied temporal variations of recorded coal fire and non coal fire locations from 2001 to 2017 using MODIS night time LST products. The intensity of fires varied on an annual basis within and across the fire locations with a maximum annual increase of 0.97K change observed in a growing fire. Methods such as threshold techniques are used to identify and delineate fires (Biswal & Gorai, 2020; Rakesh Kumar Mishra et al., 2020; Prakash & Gupta, 1998). This involves identifying the threshold of LST values in a cloud free image and pixels above the threshold are delineated as coal fires. These studies used images with up to 1% cloud coverage which affects availability of such images at required temporal resolutions.

2.7 CHANGE DETECTION

Change detection is a process of identifying changes in characteristics of a feature from observations recorded at various points in time (Singh, 1989). Timely information on changes occurring

in the features provide important information in understanding the inherent drivers and also assist in decision making. Availability of RS data at different spatial, temporal and spectral resolutions can be attributed to its influence in change detection studies (Ban & Yousif, 2016). For change detection with multi-temporal images, it is necessary for images to be preprocessed to ensure they are spatially and radiometrically comparable (Lu et al., 2004). Preprocessing steps involve geometric corrections such as image registration and orthorectification, and radiometric corrections such as atmospheric corrections. One method of deriving change variables from multi-temporal RS images is by comparing two images. A change image is generated to increase contrast between two images representing changed and unchanged areas. A change map is then generated by thresholding or by classifying the image with unsupervised or supervised techniques (Ban & Yousif, 2016). Lu et al. (2004) provides a comprehensive review of various change detection techniques to compare multi-temporal optical RS images. Prakash and Gupta (1998) used image ratioing and image difference methods to observe changes in coal mine areas in JCF between 1990 and 1994. Image ratioing was found to reduce shadows but produced an overestimated change results. Image differencing between two transformed NDVI images showed changed areas of vegetation and mining areas, but produced random noise in the image. Validation of the change areas was done using field data. Change detection using such thresholding techniques generally have difficulty in identifying true changed areas from detected change areas (Lu et al., 2004). Since, there is inconsistency in availability of multi-temporal images for successive years, change detection based on classification methods is recommended (Ban & Yousif, 2016; Hina et al., 2015; Lu et al., 2004). Of various classification methods, a post-classification comparison logic has been a popular method for change detection analysis (Olofsson, Foody, Stehman, & Woodcock, 2013). The supervised change detection is carried out by classifying each multi-temporal image independently, using the same classification scheme. A pixel-wise comparison of two images generates a transition matrix (Sunar, 1998), also known as change matrix, from which magnitude (areal extent of change) and nature of change (stable, unstable, no change) can be determined (Ban & Yousif, 2016; Hina et al., 2015; Olofsson et al., 2013). Post-classification change detection has been used to characterize changes in coal mining surfaces (Garai & Narayana, 2018; Hina et al., 2015; Pei et al., 2017; Petropoulos et al., 2013).

Chapter 3

Study area and datasets

3.1 STUDY AREA

The study area, Jharia coalfield (JCF) is located in Dhanbad district of Jharkhand state in India between 86.1° to 86.5° E longitude and 23.62° to 23.85° N latitude covering an area of 447 km^2 (Figure 3.1). JCF is the largest producer of coking coal in India consisting of large underground and opencast mines. Dhanbad is the nearest city which is also known as the “Coal capital of India”. Situated in the Damodar river basin, the study area has a history of over a century of diverse opencast as well as underground mining activities. There are human settlements inside the study area that are near and over the mining sites.

In 1971, as a part of nationalisation drive by the Indian government, Bharat Coking Coal Limited

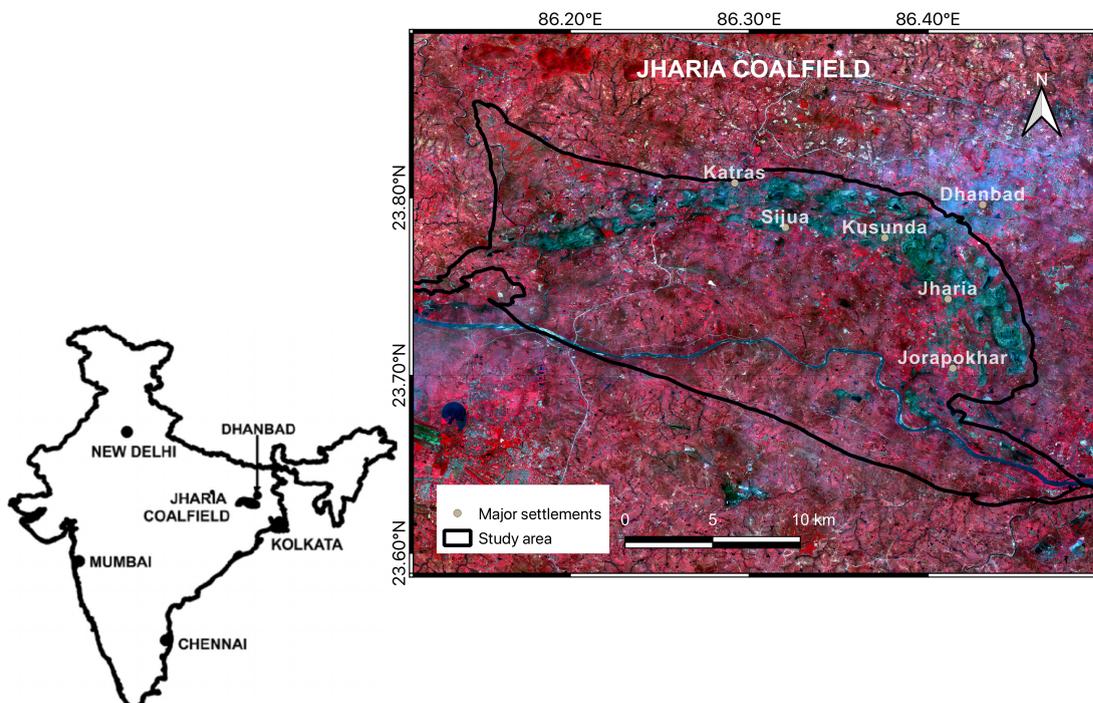


Figure 3.1: False color composite map showing study area of Jharia coalfield, India. Data: Sentinel-2 March, 2020 from <https://earthexplorer.usgs.gov/>. Map CRS: WGS84. India map source: Saini, Gupta, and Arora (2016)

(BCCL) was formed. BCCL inherited about 396 poorly operated collieries from previous private ownership, many of which were affected by coal fire (BCCL, 2008). The coal fires in JCF have been reported since 1916. The fires are characterised as surface and subsurface fires with depths upto 40m (Prakash, Gens, Prasad, Raju, & Gupta, 2013). 67 sites have been affected by coal fires (Figure 3.2), predominantly clustered in eastern part of JCF.

The fires and related land subsidence have resulted in fissure developments in the ground, buildings, roads, bridges and other structures. Field images in Stracher, Prakash, and Sokol (2013) and BCCL (2008), illustrate the extent of damages and accounts of human casualty in JCF due to fire and subsidence effects. Underground fire due to spontaneous combustion of coal has caused further subsidence causing considerable loss to lives and property. Besides causing damage to the coal reserves, the fire damages have extended towards lands, buildings and rail-road networks.

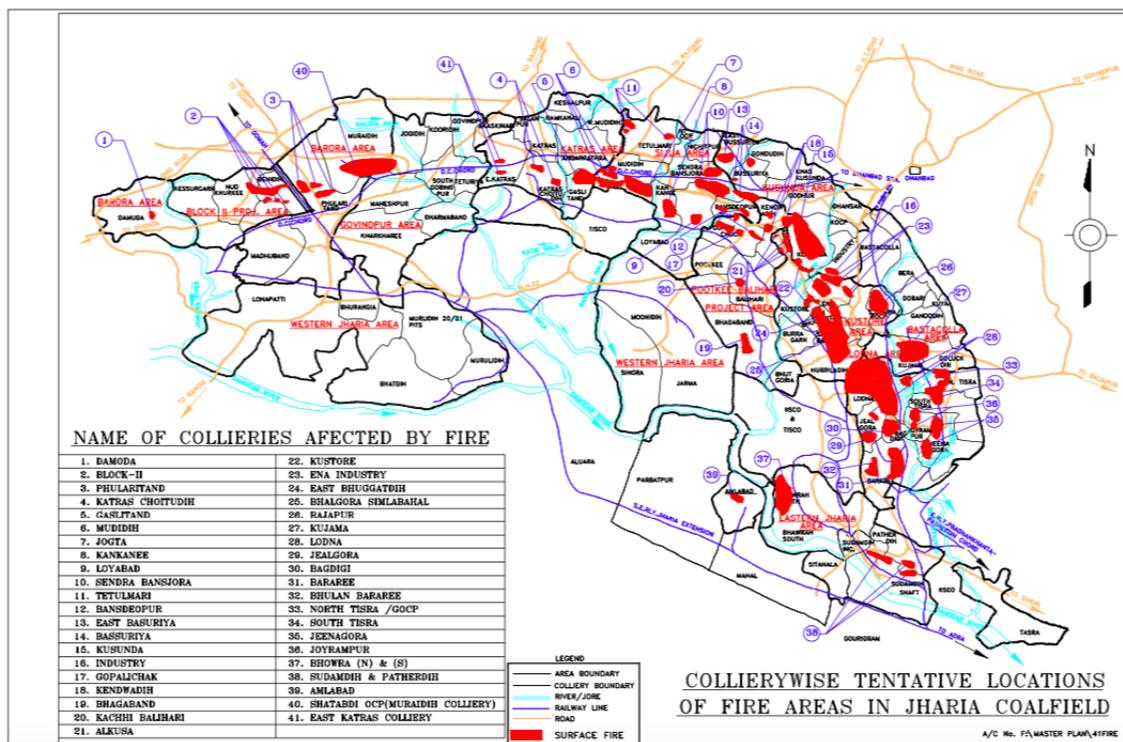


Figure 3.2: Collieries affected by coal fires in JCF. Image source: BCCL (2008)

3.2 DATASETS USED

3.2.1 Sentinel 2

Sentinel 2 is a part of the Earth observation mission from the Copernicus programme. It has launched two satellites Sentinel 2A in 2015 and Sentinel 2B in 2017. The satellites with 5 day revisit time, carry Multispectral Imager (MSI) sensor that deliver 13 spectral bands from 10m to 60m spatial resolution (Table 3.2). This data is openly accessible on ESA and also as public data catalog on the GEE platform. The data is available in two levels: Level 1C for Top of Atmosphere (TOA) and Level 2A for Bottom of Atmosphere (BOA) reflectance images. Unlike BOA images, TOA images are not atmospherically corrected and the correction cannot be carried out on the

GEE environment. So, BOA images (Table 3.1) which are available from January 2019 onwards for the study area were considered in this study.

Table 3.1 Sentinel 2 and ECOSTRESS images available for the study area without cloud masking.

Year	Sentinel 2 images on GEE	Number of ECOSTRESS LST images
2019	146	42
2020	146	42

Table 3.2 Spectral bands of Sentinel 2 imagery

Name	Wavelength (nm)	Resolution (m)	Spectral description
B2	496.6-492.1	10	Blue
B3	560-559	10	Green
B4	664.5-665	10	Red
B5	703.9-703.8	20	Red edge 1
B6	740.2-739.1	20	Red edge 2
B7	782.5-779.7	20	Red edge 3
B8	835.1-833	10	Near Infrared (NIR)
B8A	864.8-864	20	Red edge 4
B9	945-943.2	60	Water vapor
B10	1373.5-1376.9	60	Cirrus
B11	1613.7-1610.4	20	SWIR 1
B12	2202.4-2185.7	20	SWIR 2
QA60	-	60	Cloud mask

3.2.2 ECOSTRESS

ECOSTRESS mission, onboard ISS includes a TIR whiskbroom scanner with five spectral bands from wavelengths 8 to 12.5 μm and an additional band at 1.6 μm for geolocation and cloud detection. The Level 2 product (ECO2LSTE) generated by ECOSTRESS has Land Surface Temperature (LST), five emissivity bands, emissivity error bands and quality control mask. The pixels are aggregated to $\sim 70\text{m} \times 70\text{m}$, which can be visualized as the size of a football field. The 8.29 μm and 9.2 μm bands are not available after May 15, 2019 due to an anomaly detected in the instrument, and respective band values in ECO2LSTE are filled with dummy values (Glynn C. Hulley et al., 2021).

LPDAAC distributes ECOSTRESS swath based data products archived in HD5 format. For every ECO2LSTE data product (Hook & Hulley, 2019b), a corresponding ECO1BGEO (Hook, Smyth, Logan, & Johnson, 2019) and ECO2CLD (Hook & Hulley, 2019a) product is necessary for processing. A python script provided by LPDAAC (Krehbiel, 2021) uses the latitude and longitude arrays in ECO1BGEO to generate projected GeoTIFF images in a geographic coordinate system (EPSG:4326). Each product generates 16 individual GeoTIFF files, which were stacked and cropped to the extent of the study area to create one image for every acquisition. A combination of python and R (R Core Team, 2021) scripts were used to enable batch processing of data. A total of 84 images (Appendix A.1) were acquired over the study area during the period. All images were uploaded into GEE asset as Image Collection.

Table 3.3 8-bit encoding for cloud and water mask of ECOSTRESS L2 Cloud Mask Product (ECO2CLD) SDS

Bit field	Description	Result
0	Cloud Mask Flag	0 = not determined 1 = determined
1	Cloud, if either one of bits 2, 3, or 4 set.	0 = no 1 = yes
2	Thermal Brightness test	0 = no 1 = yes
3	Thermal Difference test 1	0 = no 1 = yes
4	Thermal Difference test 2	0 = no 1 = yes
5	land/water mask	0 = land 1 = water

Indicators of quality of ECO2LSTE product are described in the quality control (QC) Scientific Data Sets (SDS). Description of the 8-bit cloud product, ECO2CLD SDS is provided in Table 3.3. Setting bit 1, masks pixels that were found as clouds by any of the three individual cloud detection tests (Thermal brightness test and two thermal difference tests). Bits 1 and 5 were set to mask pixels detected as cloud and water respectively. Cloud masking was executed in GEE.

Chapter 4

Methodology

The overall methodology (Figure 4.1) adopted to achieve research objectives is described in this chapter. The overall process was executed in GEE. Individual processes adopted are described in each section.

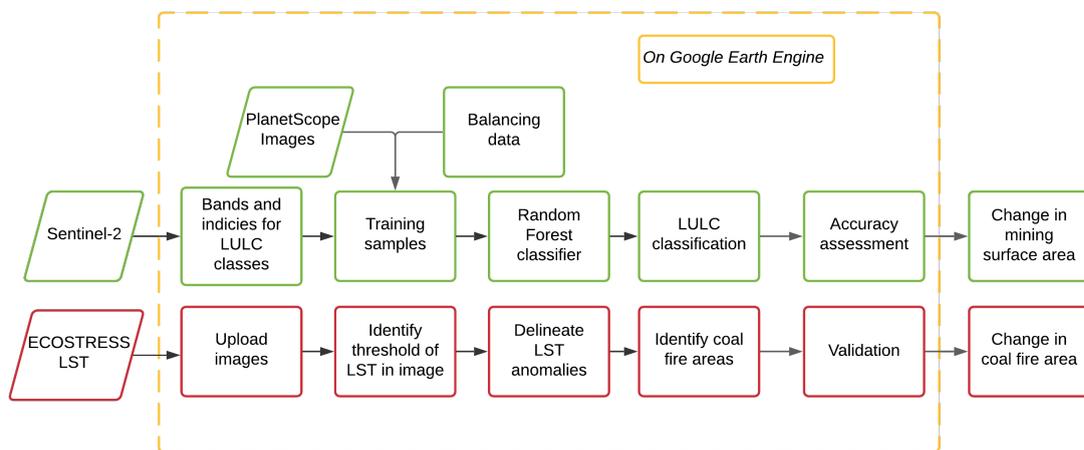


Figure 4.1: Overall methodology of the study

4.1 DATA PREPARATION

4.1.1 Sentinel 2

Sentinel 2 Surface reflectance image collection hosted on GEE platform was used for LULC classification. Images from 2019 and 2020 were filtered from the image collection and multi-temporal images were integrated for the years. The composite dataset was further refined by masking pixels affected by clouds and choosing only images with an overall cloud percentage of less than 15% (Table 4.1). Spatial resolution of 10-m bands were resampled to 20-m before further processing. Seven spectral bands Blue, Green, Red, Red edge-1,2,3 and SWIR 1 were considered. NDVI (Tucker, 1979), NDWI (Gao, 1996), and CLAL (Drury, 1987) indices were calculated (Equations 4.1a, 4.1b and 4.1c) for all pixels and added as bands to the images. These indices have been used to distinguish

vegetation, water bodies, and mining regions (Mukherjee et al., 2019).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (4.1a)$$

$$NDWI = \frac{NIR - SWIR2}{NIR + SWIR2} \quad (4.1b)$$

$$CLAL = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \quad (4.1c)$$

To further capture temporal variation in land use classes, quarterly median, mean and standard deviation composites of all bands were considered. Every quarterly composite is a 3 month group of images starting from January month of each year. Cloud free images could not be obtained for the third quarterly composite for the year 2020, as there were pixels affected by clouds. This resulted in 3 images for each year with each image having 30 bands. All three images were combined into a single image with 90 bands.

4.1.2 ECOSTRESS

84 ECOSTRESS LST products were available for the years 2019 and 2020 respectively. The individual images, each with 16 bands were uploaded to GEE as Image collection asset with image acquisition timestamp in UTC. Some acquisitions were separated by few minutes and partially covered the study area. These images were mosaiced to create one image carrying timestamp of the first acquired image, and only the mosaiced image was considered in the subsequent analyses. Bands considered in this study were 'b1' and 'b14', which correspond to Cloud mask and LST (in Kelvin). The cloud mask band, 'b1' is an 8-bit unsigned bitmask where bits 1 and 5 indicate pixels affected by cloud and water. Pixels masked as water were along the Damodar river within the study area. Other water bodies such as streams, reservoirs, canals, ponds, and water filled quarry were not masked. Masking pixels affected by cloud (Figure 4.2) reduced the number of images to 16 and 20 for 2019 and 2020 respectively (Table 4.1). The LST band 'b14' was converted to Celsius and added as a separate band 'lst' in the image before further processing.

Table 4.1 Sentinel 2 and ECOSTRESS images used in the study after cloud masking

Year	Sentinel 2 images on GEE	Number of ECOSTRESS LST images
2019	65	16
2020	50	20

4.1.3 LULC classification scheme

Land use classification scheme developed by National Remote Sensing Centre (2012), which is primarily based on J. R. Anderson (1976) was adopted to identify classes in the study area on the basis of satellite image data. The major level-1 land cover classes identified in the study were built up, agriculture, forest, barren land, and water bodies. Classes relating to mining activities such as coal overburden dumps and coal mines were identified as separate classes. Water filled quarry are associated with coal mine class. However, it was not possible to spectrally separate this class with water bodies and were not included in the coal mining class. These seven classes were used for subsequent processes such as generating reference data for training and validation, parameters for classifiers, classification and accuracy assessment. Description of classification scheme is shown in Appendix A.2.

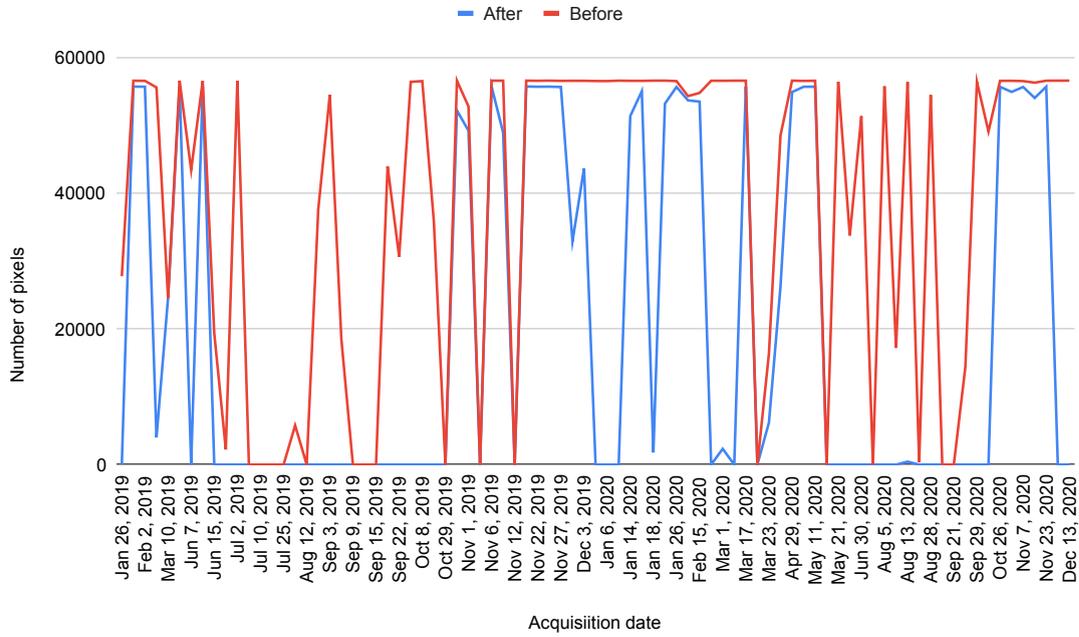


Figure 4.2: Effect of cloud masking on pixel availability of LST images

4.2 REFERENCE DATA

In the absence of reference maps, a simple random sampling design was adopted to generate reference data. Assuming unknown proportions of distribution of classes in the study area, Cochran's formula (J. & Cochran, 1978) for large populations

$$n = \frac{z^2 O(1 - O)}{d^2} \quad (4.2)$$

determines a sample size (n) of 261 for each class for 60% overall accuracy ($O = 0.6$), 90% confidence interval ($z = 1.645$) and with 5% margin of error ($d = 0.05$). With 7 identified classes in the study area the total sample size was estimated to be 1827, each for two years. Also a minimum of 50 samples were required for each class for training data (Lillesand, Kiefer, & Chipman, 2015). Considering the time and effort required to identify the reference data, 3000 samples were randomly generated for each year in GEE for the study area.

To ensure reference data is of higher quality than map data, PlanetScope images at 4.7m spatial resolution, a higher resolution than Sentinel 2 were used along with Google Earth Historical Imagery. PlanetScope images for each year were available in monthly and quarterly mosaic, as Web Map Tile Service (WMTS) on QGIS application (QGIS Development Team, 2021). This minimizes the geolocation variability between map and reference location (Olofsson et al., 2014). To minimize interpreter bias in assigning reference class to the locations, a checklist of classes with its description was referred to as a constant reminder (Appendix A.2).

2431 and 2414 samples were identified representing all the identified classes for 2019 and 2020 respectively. Sample locations with difficulty in identifying class were not chosen. Several locations bordering other classes that were difficult to interpret were also omitted. These locations were mostly in the central eastern part of the study area and areas surrounding the mining areas. This exercise turned out to be time consuming. Minimising interpreter bias is difficult and takes mon-

umental motivation.

Accuracy assessment of classification requires validation data, ideally, ground-based observations which are independent from the sampling design. However, in the absence of such observations, reference data was split for validation and accuracy assessment. Reference data for each class was randomly split into 70% training and 30% validation datasets before addressing class imbalance, and classification. Tables 4.2 and 4.3 show summary of reference data used in this study. The spectral profile of the samples are shown in Figure 4.3.

Table 4.2 Training and validation data used for 2019 classification

	Total samples	Validation	Unbalanced (Training)	RUS fraction	RUS	ROS
Barren land	596	164	432	0.5	216	465
Built up	370	101	269	0.7	188	465
Coal overburden	175	50	125	1	125	465
Coal mine	114	35	79	1	79	465
Agriculture	669	204	465	0.5	233	465
Forest	436	123	313	0.7	219	465
Water bodies	71	18	53	1	53	465
Total	2431	695	1736		1113	3255

Table 4.3 Training and validation data used for 2020 classification

	Total samples	Validation	Unbalanced (Training)	RUS fraction	RUS	ROS
Barren land	565	159	406	0.5	203	455
Built up	362	98	264	0.7	185	455
Coal overburden	197	48	149	1	149	455
Coal mine	91	21	70	1	70	455
Agriculture	664	209	455	0.5	228	455
Forest	460	142	318	0.7	223	455
Water bodies	75	21	54	1	54	455
Total	2414	698	1716		1112	3185

4.3 CLASS IMBALANCE

The problem of class imbalance occurs when there is no balanced ratio in the distribution of classes in the samples which affects classification results. Data driven non-heuristic methods such as ROS and RUS methods were applied to balance majority and minority classes and thus improve classification accuracy. Identifying majority and minority classes, an imbalance ratio can be calculated when there are binary classes. However, it is inappropriate in a multi class scenario, and depends on the sample data (Waldner et al., 2019). Several methods have been proposed to characterize them (Naboureh et al., 2020; Waldner et al., 2019). Ratios from 1:2 to 1:5 were considered as intermediate classes and ratios greater than 1:5 were identified as minority classes. Ratios with greater than 1:10 were considered as extreme minority classes. However, these were not identified in the current data. The classes were divided into three categories: 1. majority classes, which included barren land and agriculture; 2. intermediate classes, including built up and forest, and 3. minority

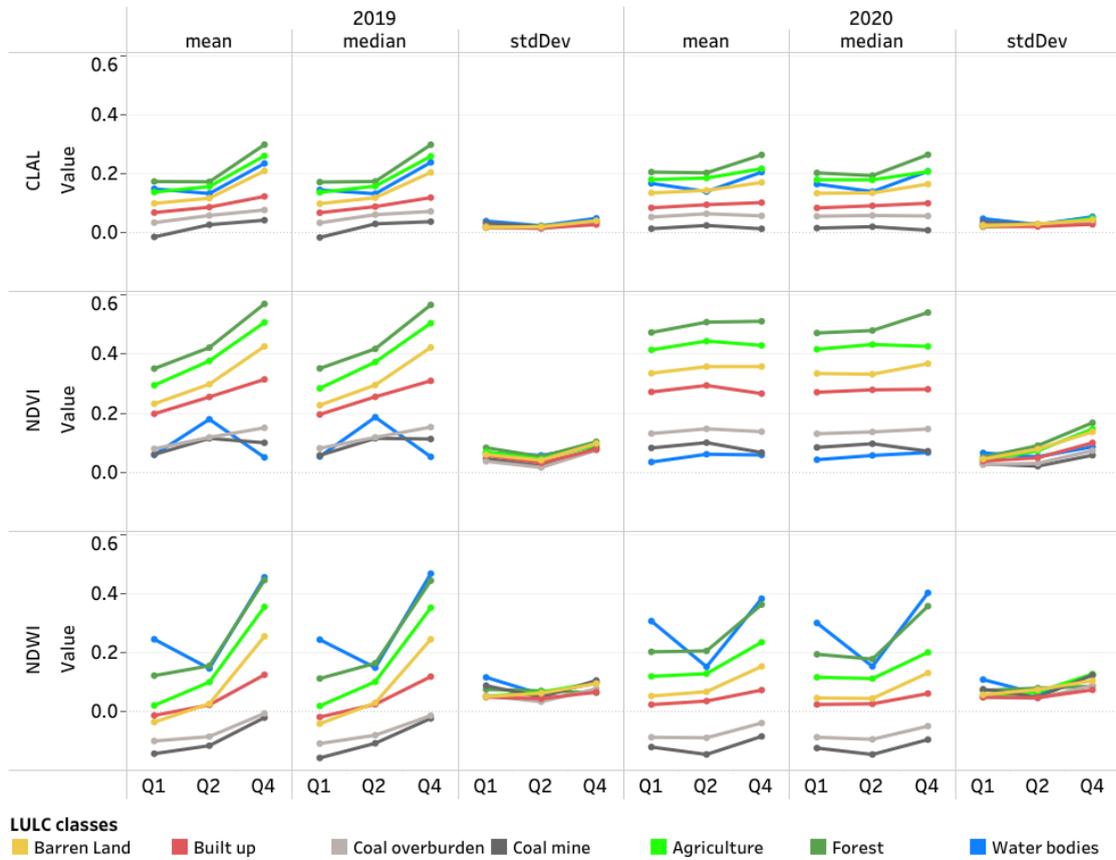


Figure 4.3: Variation in indices of samples of 2019 and 2020 along with temporal patterns: Q1 Jan-Mar, Q2 Apr-Jun, Q4 Oct-Dec

classes, including coal overburden, coal mine and water bodies. Naboureh et al. (2020) identified 50%, 70% and 100% as the fractions that would achieve the best balance for RUS for majority, intermediate and minority classes. In RUS, the classes were randomly undersampled until they reached the fraction of their count of classes. In full balancing ROS, the samples from classes were resampled with replacement until they equalled the majority class with the most number of samples (Tables 4.2 and 4.3). The number of samples for ROS and RUS, calculated for each class were given as `sampling_strategy` in `randomOversampler` and `randomUndersampler` functions of the `imblearn` library, implemented in Python.

4.4 RF CLASSIFIER

Random Forest (RF) (Breiman, 2001) is an algorithm consisting of an ensemble of unpruned decision trees. It uses bootstrap samples from the training data and random feature selection when building trees and reduce correlation among the trees. Prediction of the algorithm for classification is based on majority vote, and for regression is based on averaging. It is widely used as classifier for LULC classification and is known for its robustness and accuracy. Number of trees (`n_tree`) and features at each split (`mtry`) are input parameters for RF. Increasing the number of trees does not considerably increase the accuracy of the classification but increases the computational effort. Belgiu and Drăguț (2016) suggest using upto 500 trees and square root number of features

as parameters for RF. For this study, *ntrree* and *mtry* were specified as 200 and default (square root of the number of features) in `classifier.smileRandomforest`, the RF classifier available in GEE library.

4.5 ACCURACY ASSESSMENT AND AREA ESTIMATION

Classification does not hold value without assessing its accuracy (Congalton, 2001). Accuracy assessment assists in evaluating the performance of the classifier, area estimates of the classes and effect of sampling design. Accuracy assessment was carried out for classified maps of 2019 and 2020 using the ‘confusion matrix’ (Congalton, 1991), also known as ‘Error matrix’. Validation points that were independent of the training data were considered for accuracy assessment. Amongst various available metrics, the traditional metrics of overall accuracy (OA), user accuracy (UA) and producer accuracy (PA) were used to evaluate accuracy assessment of classification. OA (Equation 4.3) provides interpretation of effectiveness in accuracy estimation while UA (Equation 4.4) and PA (Equation 4.5) provide class level evaluation.

$$OA = \frac{\sum_{i=1}^k R_i}{\sum_{i=1}^k N_i} \quad (4.3) \quad UA_i = \frac{R_i}{N_i} \quad (4.4)$$

$$PA_i = \frac{R_i}{M_i} \quad (4.5) \quad GM = \left(\prod_{i=1}^k PA_i \right)^{1/k} \quad (4.6)$$

where k = number of classes, R_i = pixels correctly mapped in class i , N_i = pixels that are mapped to be in class i , M_i = pixels in class i that are in reference data.

These metrics are available on GEE for the RF classifier. Area of mapped classes were calculated by summing the area of each pixel associated with their respective class. Based on the classification errors, areas of mapped classes were adjusted using a post-stratified error estimate (at 95% confidence interval) approach described by Olofsson et al. (2013) with equations 4.7, 4.8, 4.9, 4.10.

$$\hat{p}_{ij} = W_i \cdot \frac{n_{ij}}{n_i} \quad (4.7) \quad \hat{A}_j = A_{total} \cdot \sum_i \hat{p}_{ij} = A_{total} \cdot \hat{p}_{.j} \quad (4.8)$$

$$S(\hat{p}_{.j}) = \sqrt{\sum_{i=1}^k W_i^2 \frac{\frac{n_{ij}}{n_i} \left(1 - \frac{n_{ij}}{n_i}\right)}{n_i - 1}} \quad (4.9) \quad S(\hat{A}_j) = S(\hat{p}_{.j}) \cdot A_{total} \quad (4.10)$$

where \hat{p}_{ij} = proportion of misclassified pixels in mapped class i , W_i = ratio of mapped area of class i and area of all classes (A_{total}), n_{ij} = number of misclassified pixels in reference class, n_i = number of mapped pixels in each class, $S(\hat{p}_{.j})$ = standard error of estimated proportion, $S(\hat{A}_j)$ = standard error of error adjusted area, and at 95% confidence, the error adjusted area is calculated as $\hat{A}_j \pm 1.96 \cdot S(\hat{A}_j)$.

Kappa coefficient, a common metric, was not used as it does not serve a useful role in accuracy assessment or area estimation and is highly correlated to OA (Olofsson et al., 2014).

4.6 CHANGE DETECTION

Change detection here is the process of identifying the transition of LULC classes from multi temporal LULC maps. Here, a post-classification change detection technique was used for analysing changes. This technique implements pixel by pixel comparison of multi temporal classified maps, and minimises impacts of atmospheric and sensor differences between them (Lu et al., 2004). It also provides a transition matrix (Sunar, 1998) which presents important information about spatial distribution of changes generated from classified maps of 2019 and 2020 to assess overall changes in LULC classes.

Changes in coal mine areas were quantified using coal mine classification of 2019 and 2020. Changes were categorized into three states: continue, decrease and increase. Differences between the 2020 and 2019 map of coal mine class provided the conditions for categorizing these states: 1. Coal mine pixels that did not change class were considered as ‘continue’, 2. mines that were present in 2019 and absent in 2020 were categorized as ‘decrease’, and 3. mines that were present in 2020 and absent in 2019 were categorized as ‘increase’.

4.7 DETECTION AND CHANGE IN COAL FIRE AREAS

LST images were acquired over the study area at different periods and at different time of the day. Distribution of LST values in an image vary according to the time of acquisition. Hour of the day was calculated and assigned to each image. This would help in filtering out images acquired during Day and Night. Images acquired from 0300 to 1200hrs UTC (0830 to 1730hrs IST) were classified as Day images and others as Night time images (Glynn C Hulley et al., 2019). Image thresholding technique was adopted to delineate pixels that could potentially be affected by coal fires. Pixels in the image that are within the colliery boundary were considered for thresholding as coal fires occur inside this boundary within the study area (Biswal & Gorai, 2020). Each acquired image affected by clouds, would reduce the total number of pixels available for thresholding. A pixel coverage ratio was calculated as ratio of number of pixels available for thresholding after cloud masking and total number of 70-m pixels that cover the colliery boundary (56567 pixels). Images with high pixel coverage were considered for thresholding. Mean and standard deviation of each image was calculated and threshold values (Equation 4.11) were set for each image (Biswal & Gorai, 2020; Huo et al., 2014).

$$\nu_{LST} = mean_{LST} + 2\sigma_{LST} \quad (4.11)$$

The LST pixel values in an image were used to determine the threshold value. Pixels above this threshold (ν_{LST}) were identified as coal fires in each image. To identify the potential of a coal fire pixel to be a coal fire pixel in a year, a ratio (referred as detection ratio) of the number of coal fire pixels and number of images considered for the year was calculated.

Coal fires are dynamic and propagate with time (Kuenzer, 2015; Mujawdiya et al., 2020). With time, coal fires continue to burn while few become dormant, and new fires are propagated. To identify these changes between years, we need pixels that have been consistently detected as coal fires. Detection ratio can be considered as how often a pixel is detected as coal fire. Ratio above 0.5 indicate that pixels were detected in more than 50% of the images in the year. Pixels below 0.5 detection ratio will contribute to coal fires for the year and not for changes between years. To identify changes two coal fire zone images from 2019 and 2020 were subtracted, and pixels were categorized into ‘active’, ‘dormant’, ‘increasing’.

4.8 DETECTING COAL FIRES WITH PERCENTILE APPROACH

The Day and Night time images were considered for this method. For each year, 95th percentile composite of LST values were generated separately for Day and Night time images to capture the magnitude of change in temperature. So, each year would have one Day time and one Night time composite. Threshold of LST values for each composite were calculated using Equation 4.11. Pixels above the threshold value were selected from the composites. Two images were mosaiced to create one image to generate coal fire pixels for the corresponding year. Changes in status of coal mine areas between 2019 and 2020 were followed as explained in Section 4.7.



Figure 4.4: Mockup of GEE application showing implementation and interactive elements

4.9 VALIDATION OF COAL FIRES

Accuracy of coal fire detection was validated with field data published by Biswal and Gorai (2020). Validation data was available for 2019 and not available for 2020. There were 20 locations identified in the study area where temperature measurements were collected using thermal imaging camera. The observations were recorded on April 2, 2019. Validation was performed in two steps: 1. Ground observations that were within the coal fire pixels, 2: Ground observations that were within 70-m from the nearest coal fire pixels.

To identify the relationship between the observed temperature and ECOSTRESS product, LST values were extracted from the images acquired on May 27, 2019 at 10:18 UTC and October 30, 2019 at 20:05 UTC, which were the nearest Day and Night time images available that were not affected entirely by clouds.

4.10 APPLICATION ON GEE

”A picture can speak a thousand words”. But an interactive application can make pictures tell a story. The web-based user interface provided by GEE allows building interactive web application powered by cloud computing architecture. GEE provides client libraries, Javascript API, and Python API for data storage, analysis, and building custom applications.

In this study the developed application is based on Javascript, a web-based interface which is provided at code.earthengine.google.com and the features can be used with a user signup. The platform

offers data stored in repositories. Before being made available, the data is pre-processed to enable efficient access and thereby addressing data management problems (Gorelick et al., 2017). GEE allows limited user data storage as a private repository which can be used to integrate with the data catalog on GEE datasets. GEE also allows free publishing and hosting of the applications on their platform. The application can be accessed by user in a web-browser which requires *client-side* user interface functions, also known as widgets. These widgets are made available through the `ui` package. The `ui` package allows access to in-built functions to design graphical interface and building an application. The application can have simple widgets such as buttons (`ui.Button()`), labels (`ui.Label()`) and sliders (`ui.Slider()`), more complex ones such as charts (`ui.Chart()`), maps (`ui.Map()`), panels (`ui.Panel()`) for application layout, and event handlers (`onClick()` and `onChange()`) for user and widget interactions. Appropriate parameters need to be specified while using the widgets. These parameters govern the behavior of widgets in the application. Mockup of the application as shown in Figure 4.4 can provide the parameter specification. Split panel is one of the user interfaces available on GEE that allow comparison of two images. The user interface consists of left and right panels that overlay two images. Both panels are separated by a divider, which when moved resizes the map area displayed in the panels. After selecting the 'start' and the 'end' year, the user can select two map layers simultaneously in the split panel. The legend of the map layers are shown accordingly. Summary of the map layers are shown as bar charts in the panel. The user is able to select location on the map to generate timeseries charts of features such as NDVI, CLAL and LST.

Chapter 5

Results

5.1 EFFECTS OF BALANCING METHODS

Training samples were generated for 2019 and 2020 based on balancing methods RUS and ROS. Classification for 2019 and 2020 was performed using RF classifier and validated with independent dataset. UA, PA, OA and GM were calculated for each classification. To evaluate the effect of balancing methods in addressing class imbalance, differences in OA (ΔOA) and GM (ΔGM) with unbalanced data were computed. Regarding different LULC classes, minority classes showed stronger response to balancing methods by improvement in accuracy.

For 2019 (Table 5.1), OA and GM ranged between 72.66%-73.09% and 71.73%-74.47% respectively. Between all the methods, RUS had the least OA and GM and ROS the highest. Compared to unbalanced, RUS showed increased PA for minority class coal overburden and decreased PA for water bodies. There were no significant changes in the UA and PA in other minority classes. ΔOA and ΔGM indicate negative gain in OA and GM. ROS showed increased PA for all minority classes and increased UA for coal mine. However, there was a significant decrease in UA value (7.64%) for coal overburden. There was no gain in OA ($\Delta OA = 0$), but positive gain in GM ($\Delta GM = 1.69$).

For 2020 (Table 5.2), OA and GM ranged between 72.9%-76.5% and 72.12%-77.16% respectively.

Table 5.1 Effects of increase and decrease in balancing methods on classification accuracies (in %) for 2019

2019	Unbalanced		RUS		ROS	
	UA	PA	UA	PA	UA	PA
Barren land	65.10	59.15	+2.35	+0.61	+2.73	0
Built up	77.65	65.35	0	0	+1.66	+2.97
Coal overburden	71.43	70.00	-0.57	-2.00	-7.64	+4.00
Coal mine	69.70	65.71	-1.27	+2.85	+0.89	+2.86
Agriculture	72.65	83.33	+0.62	0	+1.11	-3.43
Forest	78.91	82.11	-1.74	+0.81	-3.54	0
Water bodies	94.12	88.89	+0.37	+5.56	+0.32	+5.55
OA	73.09		72.66		73.09	
GM	72.78		71.73		74.47	
ΔOA			-0.43		0.00	
ΔGM			-1.04		1.69	

Minimum OA and GM were observed in ROS and unbalanced, while RUS had the highest. Compared to unbalanced, RUS showed an increase in UA for all minority classes and significant increase in PA value for coal overburden (8.33%) and coal mine (19.05%). There was no change in PA for water bodies. RUS had positive gain in OA and GM with $\Delta OA = 5.04$ and $\Delta GM = 2.00$ respectively.

The results indicate that ROS and RUS methods performed better at addressing class imbalance

for 2019 and 2020 respectively.

Table 5.2 Effects of increase and decrease in balancing methods on classification accuracies (in %) for 2020

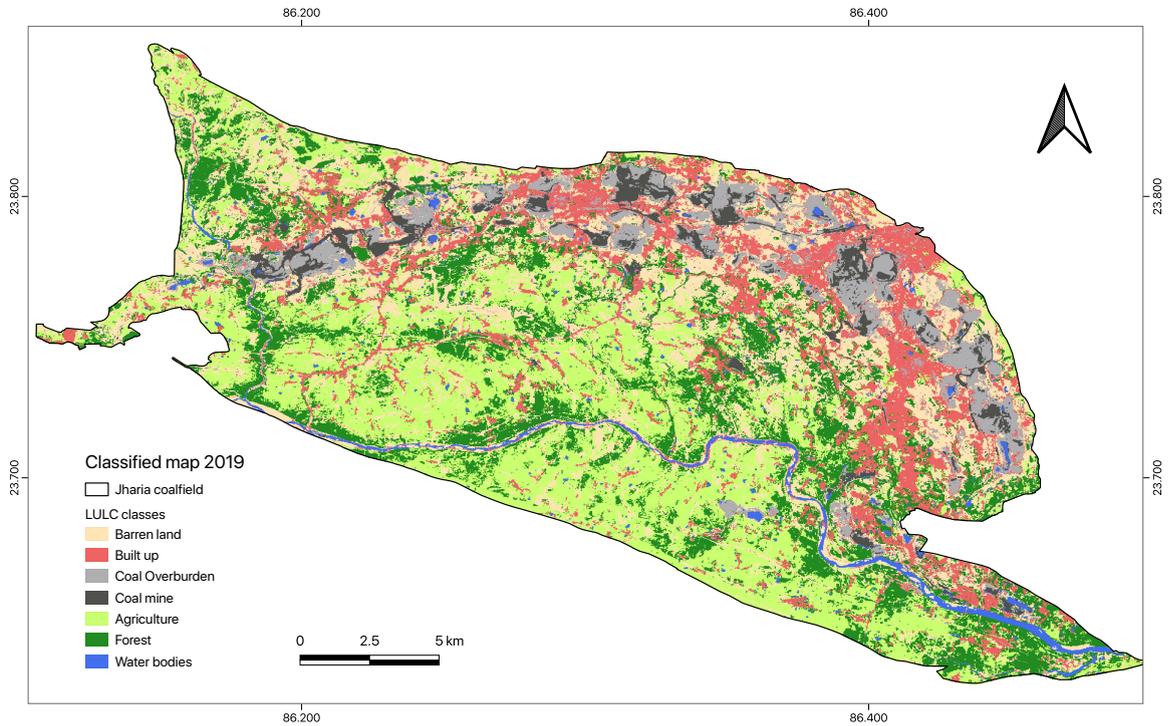
2020	Unbalanced		RUS		ROS	
	UA	PA	UA	PA	UA	PA
Barren land	63.52	63.52	+3.36	+1.26	-0.36	-3.14
Built up	76.92	71.43	+1.73	0	-0.78	-3.06
Coal overburden	70.00	72.92	+5.00	+8.33	-5.59	+6.25
Coal mine	73.33	52.38	+5.62	+19.05	+5.24	0
Agriculture	78.64	82.78	+1.08	0	0	-5.27
Forest	79.86	78.17	-0.55	+2.82	-5.18	+2.82
Water bodies	79.17	90.48	+7.19	0	+0.83	+4.76
OA	74.50		76.50		72.92	
GM	72.12		77.16		72.21	
Δ OA			2.00		-1.58	
Δ GM			5.05		0.09	

5.2 CLASSIFICATION RESULTS

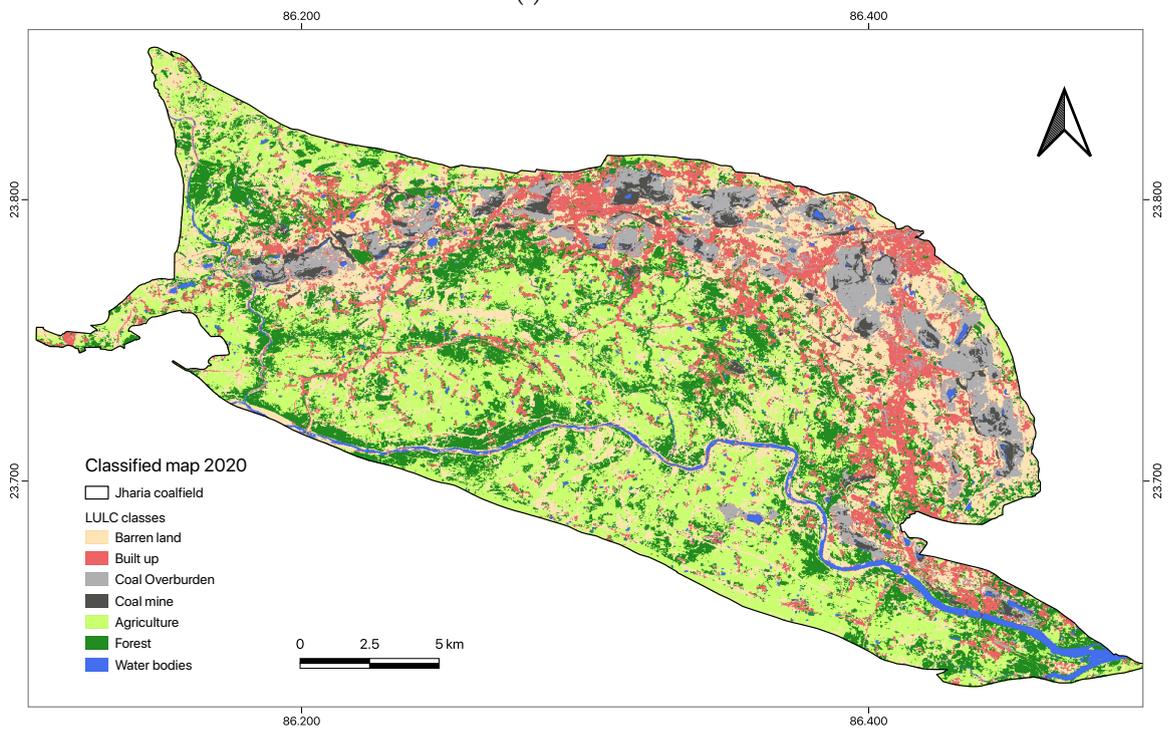
5.2.1 LULC classification and accuracy assessment

A pixel-based classification was applied in the study area for 2019 and 2020 using training data that addressed the class imbalance problem. The corresponding classified maps are shown in Figures 5.1a and 5.1b. Accuracy assessment of individual classified maps was performed using independent validation data that was initially separated from the reference data (Tables 4.2 and 4.3). Accuracy assessment results are summarised in Tables 5.3a and 5.3b.

The results indicate that classification accuracies were similar for both years. The OA level of 73.09% and 76.5% was obtained for 2019 and 2020 respectively. The UA and PA were also consistent within the individual classes for each year. In 2019, UA and PA ranged between 63.79%-94.44% and 59.15%-94.44% respectively. In 2020, UA and PA ranged between 66.88%-86.36% and 64.78%-90.48% respectively. Water bodies class achieved the highest UA and PA for both years, while coal overburden and barren land had the respective lowest UA and PA for 2019, and barren land had the lowest UA and PA for 2020. Misclassification within the classes were observed in both classified maps and error matrices. In 2019, Barren land had confusion with all of the classes but mostly with agriculture, built up and forest which affected its UA and PA values. Few of the sandy areas along the river bed, which belong to barren land class, were classified as built up. Agriculture and forest had confusion within themselves which contributed to decrease in their respective UA and PA and thus affecting OA. Coal overburden had confusion with barren land, built up and coal mine classes that lowered its UA and PA. These patterns were observed in both 2019 and 2020. In 2019, coal mine class had confusion with barren land, built up and coal overburden that affected its UA (70.59%) and PA (68.57%). In 2020, coal mine class had slight confusion with coal overburden, forest and water bodies that affected its UA (78.95%) and PA (71.43%). The results show that the coal mine class had confusion mostly with the coal overburden class.



(a) Year 2019



(b) Year 2020

Figure 5.1: LULC classified map

Table 5.3 Classification error matrix

		Reference							UA(%)	PA(%)
		1	2	3	4	5	6	7		
Classified	1. Barren land	97	14	2	3	18	9	0	67.83	59.15
	2. Built up	9	69	4	0	3	1	1	79.31	68.32
	3. Coal overburden	8	5	37	8	0	0	0	63.79	74
	4. Coal mine	1	2	7	24	0	0	0	70.59	68.57
	5. Agriculture	37	9	0	0	163	12	0	73.76	79.9
	6. Forest	12	2	0	0	19	101	0	75.37	82.11
	7. Water bodies	0	0	0	0	1	0	17	94.44	94.44
									OA= 73.09%	

(a) Year 2019

		Reference							UA(%)	PA(%)
		1	2	3	4	5	6	7		
Classified	1. Barren land	103	15	3	0	23	9	1	66.88	64.78
	2. Built up	10	70	3	0	0	6	0	78.65	71.43
	3. Coal overburden	4	4	39	5	0	0	0	75	81.25
	4. Coal mine	0	0	2	15	0	1	1	78.95	71.43
	5. Agriculture	31	2	0	0	173	11	0	79.72	82.78
	6. Forest	8	7	1	1	13	115	0	79.31	80.99
	7. Water bodies	3	0	0	0	0	0	19	86.36	90.48
									OA= 76.5%	

(b) Year 2020

5.2.2 Area estimation

Post classification, areas of the mapped classes were calculated by multiplying the number of pixels per class with the area of each pixel. Accounting for magnitude of classification errors, areas of mapped classes were estimated at 95% confidence level, using the approach described in Olofsson et al. (2013). These post stratified error adjusted estimates for 2019 and 2020 are shown in Table (5.4).

In 2019, mapped majority classes barren land and agriculture occupied 22.92% and 29.28% of the study area, followed by intermediate classes built up and forest with 14.82% and 19.65%, and minority classes coal overburden, coal mine and water bodies covering 7.33%, 3.64% and 2.37% respectively. Similar values were observed in 2020 for the mapped classes barren land, agriculture, built up, forest, coal overburden, coal mine and water bodies respectively occupying 24.09%, 28.3%, 14.6%, 20.4%, 7.01%, 3.08% and 2.54% of the study area.

In 2019, compared to mapped areas, error adjusted coal mine area was overestimated by 11.47% from 16.3 to 18.37 km². Overestimates were also observed in barren land by 8.45%, and in built up by 10.18%. Coal overburden, agriculture and forest classes were underestimated by 12.29%, 4.68% and 8.32% respectively. Similar to 2019, in 2020 the coal mine area was overestimated by 9.99% from 12.51 to 13.76 km², and built up by 10.88%. Barren land, coal overburden, agriculture and forest were underestimated by 2.08%, 6.91%, 0.88% and 1.82% respectively. There was no significant difference between mapped and adjusted areas for water bodies in 2019 and 2020 which had achieved good classification accuracy.

Table 5.4 Mapped areas (in km²) and post stratified error adjusted estimates (at 95% confidence interval) of LULC classes for 2019 and 2020 classification

	2019			2020		
	mapped Area	error adjusted Area	SE(±)	mapped Area	error adjusted Area	SE(±)
Barren land	102.54	111.21	12.24	110.05	107.76	11.79
Built up	66.29	73.04	8.93	58.88	65.29	8.45
Coal overburden	32.78	28.75	5.86	33.69	31.36	5.61
Coal mine	16.30	18.17	4.57	12.51	13.76	3.82
Agriculture	130.99	124.85	11.15	127.72	126.59	10.21
Forest	87.89	80.57	8.71	92.95	91.25	8.93
Water bodies	10.58	10.76	1.89	11.55	11.35	2.55

5.3 CHANGE DETECTION

5.3.1 Change detection matrix

After quantifying accuracy of classification, change detection analysis was carried out by post-classification technique as described in Section 4.6. The results are summarised in a change detection matrix shown in Table 5.5.

Barren land saw a net increase of 7.5 km² in 2020, mainly contributed by built up area (11.6 km²). Similar extent of agriculture, coal overburden and forest transitioned and also contributed to the barren land class. Built up saw a net decrease of 11.7%, when 17.5% of built up area was converted to barren land. 2.5 km² of built up area transformed into coal overburden in 2020. Built up also had no significant transitions and conversions among other classes. Coal mine of 9.3 km² was retained between 2019 and 2020 and there was a net decrease of 3.8 km² during the period. The decrease can be attributed to 31.28% of coal mine area transitioning to coal overburden from 2019 to 2020, whereas 2.3 km² (18.4%) of coal overburden added to coal mine area. Of 16.3 km², coal mine area saw transition of 4.9% to barren land, 3% to built up and water bodies, and 0.61% to forest. Agriculture, with the highest mapped area saw a less significant net decrease of 2.49%, where 12.9% and 9.5% were transformed to barren land and forest, respectively. Also barren land and forest respectively added 13.5% and 6.9% to agriculture. Forest had an overall increase of 5.77% mainly contributed by 12.5km² of agriculture, which was converted to forest, while 8.9 km² of forest was lost to agriculture. With a little bit of surprise, 6 km² of barren land was converted to forest and equally 6.1 km² of forest was lost to barren land. Water bodies of 8.7 km² extent were retained from 2019 to 2020, while the net increase of water bodies is mainly attributed to barren land (1.4 km²). Transitions of water bodies to other classes were not significant.

The transition of changes among the LULC classes can be visualized with a chord diagram shown in Figure 5.2. The net decrease and increase in extent of the changes are shown in the Figure 5.3.

5.3.2 Change in coal mining areas

The coverage of coal mining area was found to be 16.3 and 12.5 km² for the years 2019 and 2020, respectively. This indicated a net decrease in coal mining area by 23.2%. 9.26 km² of coal mining areas continued activity during 2019 and 2020. An extent of 3.25 km² of coal mining areas were added during 2020, whereas 7.03 km² of coal mining areas decreased after 2019. The figure 5.4 shows the change in coal mining areas.

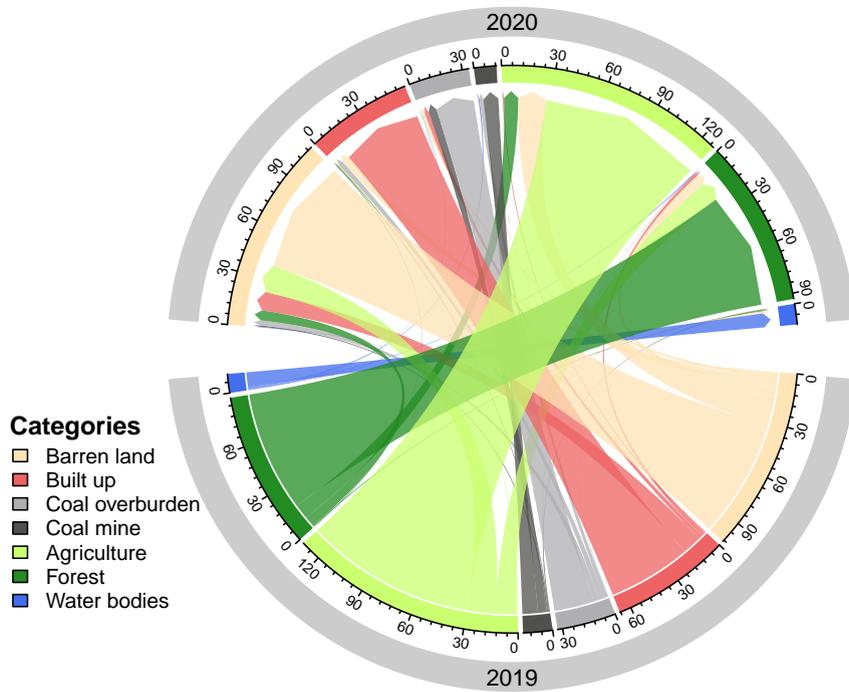


Figure 5.2: Transition between LULC classes during 2019-2020

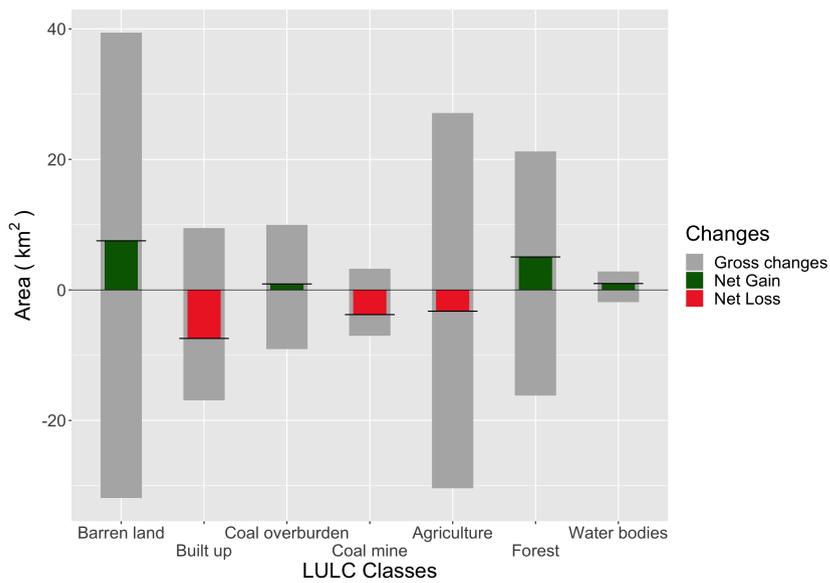


Figure 5.3: Net changes of LULC classes during 2019-2020

Table 5.5 Change detection matrix of LULC classes from 2019 to 2020. Mapped areas in km² for the study area of 447.3 km²

2019	2020							2019 Total
	1	2	3	4	5	6	7	
1. Barren land	70.8	4.7	2.1	0.2	17.3	6.1	1.4	102.5
2. Built up	11.6	49.4	2.5	0.2	0.7	1.9	0	66.3
3. Coal overburden	3.2	3.2	23.7	2.3	0	0.1	0.2	32.8
4. Coal mine	0.8	0.5	5.1	9.3	0	0.1	0.5	16.3
5. Agriculture	17.0	0.6	0	0	100.7	12.5	0.2	131.0
6. Forest	6.3	0.4	0	0	8.9	71.8	0.5	87.9
7. Water bodies	0.4	0.1	0.3	0.6	0.1	0.5	8.7	10.6
2020 Total	110.0	58.9	33.7	12.5	127.7	92.9	11.5	
Change (%)	7.32	-11.17	2.78	-23.20	-2.49	5.77	9.13	

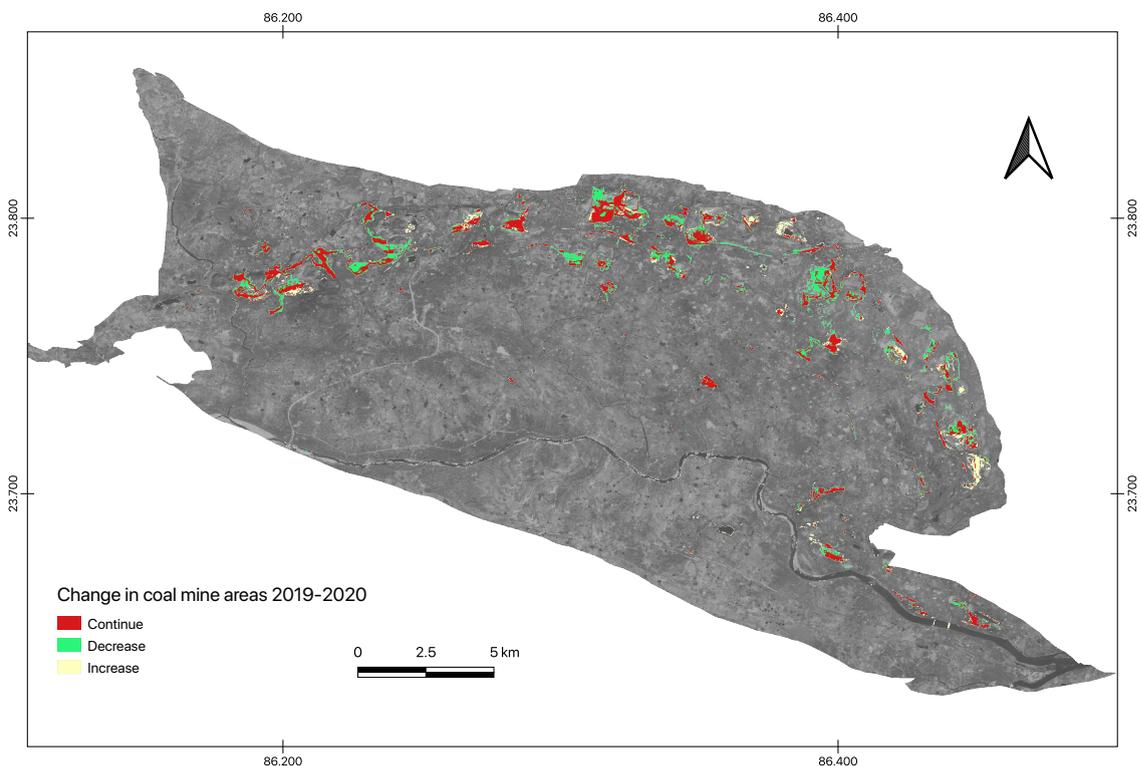


Figure 5.4: Change in coal mining areas between 2019 and 2020

5.4 COAL FIRE DETECTION

5.4.1 Effect of clouds on LST images

To consider images that were least affected by clouds, pixel coverage was calculated for each image and the Figure 5.5 shows the distribution of pixel coverage for every image acquired in 2019 and 2020. Out of 36 images, 9 images had pixel coverage above 0.95 in 2019 and 11 images in 2020. These images were considered for thresholding. However, all 36 images were considered for the percentile approach. Some pixels were detected as clouds by cloud detection tests in few of the

night images over the coal mines where fires were recorded by ground observations, as shown at one location in Figure 5.6-C.

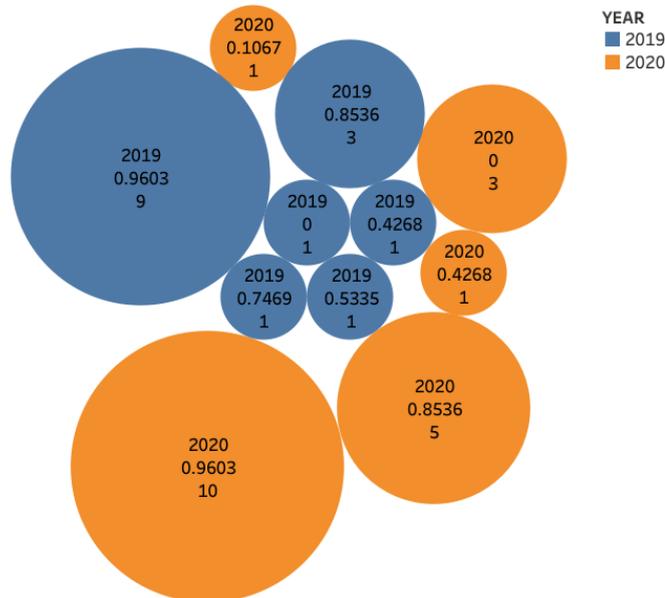


Figure 5.5: Distribution of pixel coverage and number of LST images in 2019 and 2020. Pixel coverage calculated as ratio of number of pixels available after cloud masking and total number of pixels available within colliery boundary

5.4.2 Threshold of LST images and coal fire detection

Threshold of LST in each image with pixel coverage greater than 0.95 was calculated using Equation 4.11. The threshold values of 20 images for 2019-2020 are shown in Figure 5.7. Out of these images, 8 images were acquired during ‘day’ time with an average threshold value of 36.23°C. Similarly 12 images were acquired during ‘night’ time with an average threshold value of 23.6°C. To detect pixels as coal fire pixels, all pixels in each image that were above the threshold were assigned as coal fire pixels. Detection ratios calculated for pixels acquired during 2019 and 2020, are shown in Figure 5.8.

A total of 4795 pixels (21.40 km²) and 7304 pixels (32.6 km²) were above threshold values and were detected as coal fires. The detection ratio ranged between 0 and 1, and the ratios were binned into 4 categories: 0-0.25, 0.25-0.5, 0.5-0.75, 0.75-1. In 2019, the area of coal fire pixels for these categories were 8.47, 3.59, 3.25, 5.02 and 1.08 km². Similarly, in 2020, the areas were 16.71, 5.54, 5.32, 4.53 and 0.51 km². The areas affected by coal fires for both years are shown in Figures 5.9a and 5.9b. The maps also show false positive fire areas, especially along the river that were detected as coal fires.

5.4.3 Percentile approach

Out of 36 images available for analysis for both years, 2019 had 16 images, out of which 9 images belonged to day time acquisition and 7 for night time acquisition. Similarly for the year 2020, there were 20 images available, but there were only 3 day time images and 17 night time images.

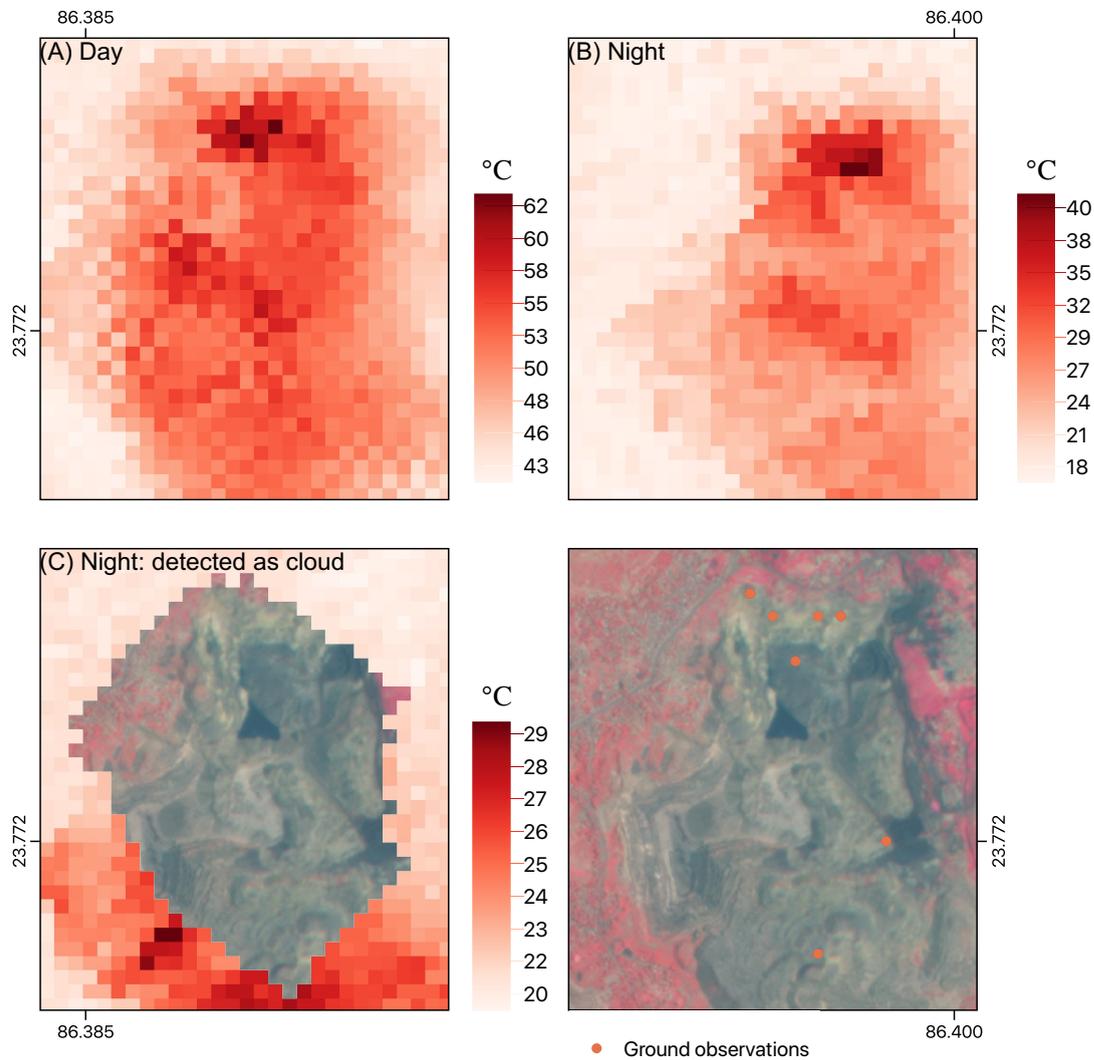


Figure 5.6: Diurnal spatial variability of LST at coal mine affected by coal fires in JCF and indicated in top left corner of each image (A-C). The color scale in degree Celsius is enhanced for each image. Figure C shows pixels identified as clouds over the coal fire areas. FCC image is acquired from PlanetScope dated December, 2019

Threshold (Equation 4.11) was applied to 95 percentile composites of day and night LST images before mosaicing them to create a coal fire map for each year. The extent of the delineated coal fires were calculated at 13.04 km² and 14.39 km², for 2019 and 2020. The delineated coal fire maps are shown in Figures 5.10a and 5.10b.

5.4.4 Validation

Validation of coal fire detection was carried out for the year 2019 with details listed in Table 5.6. Out of 20 locations, 18 locations were within the detected coal fire pixel achieving an accuracy of 90%. Two locations ID: 9 and 11, which did not validate, were found to be within 70-m of the

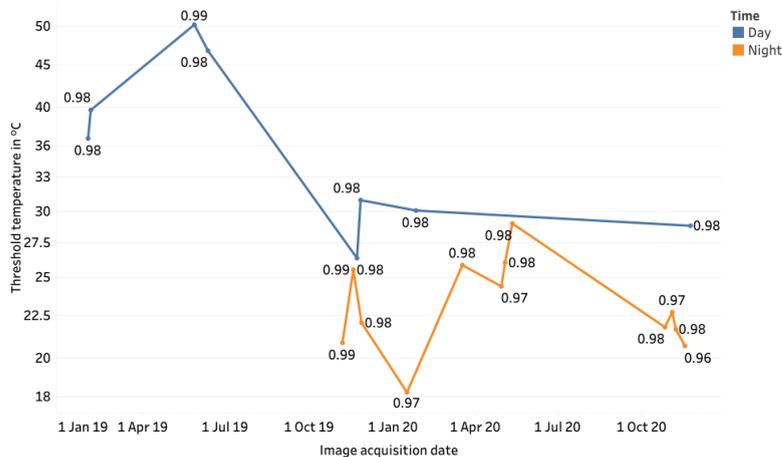


Figure 5.7: Threshold and associated pixel coverage of day and night time images in 2019 and 2020

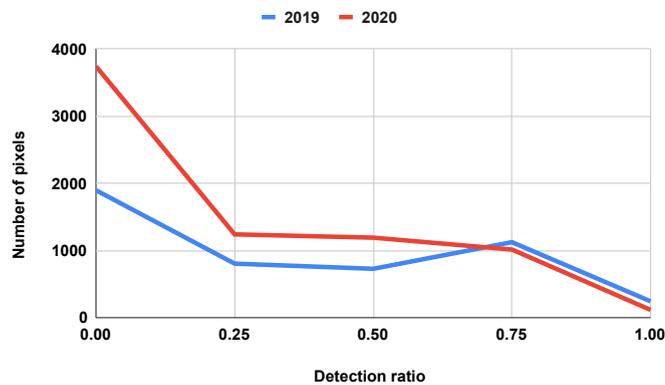
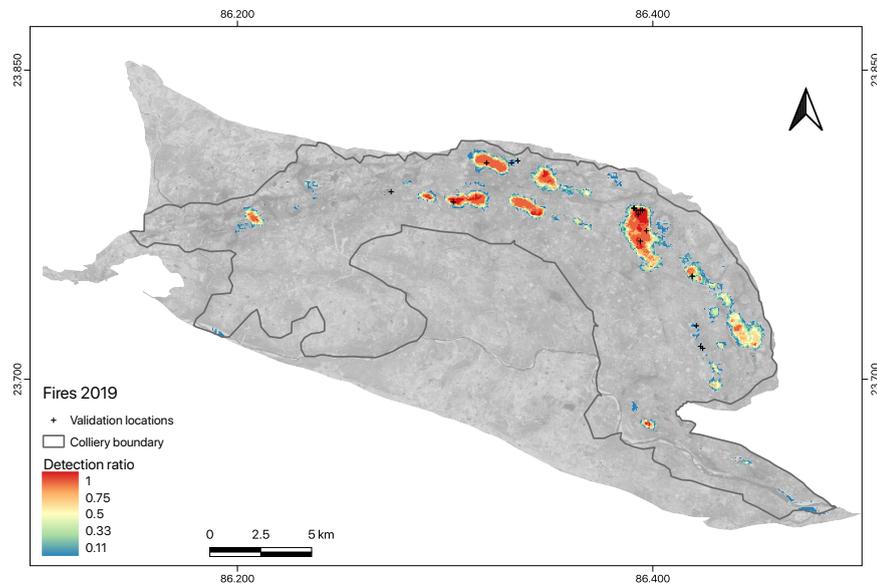


Figure 5.8: Number of pixels delineated as fire and their detection ratio for 2019 and 2020

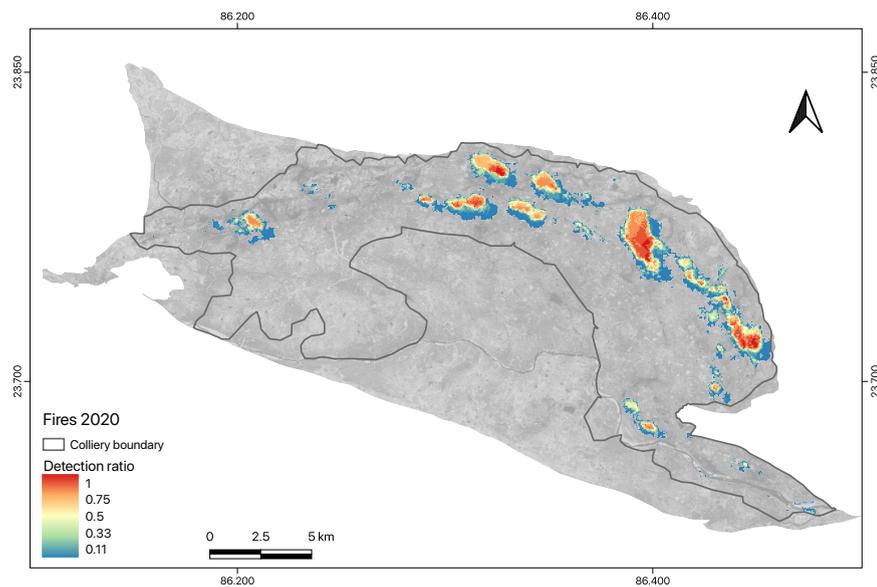
nearest coal fire pixel. These nearest pixel to these locations had detection ratio of 0.11. Similarly, validation was carried out for the percentile approach, which achieved an accuracy of 75% with 15 ground locations within coal fire pixels. Five locations ID: 1, 9, 11, 17 and 20, which did not validate, 1 and 9 were found to be within 70-m from the nearest coal fire pixel. Locations 9, 11, 17 and 20 had detection ratio of 0.11 and location 1 had detection ratio of 0.22. The validation locations are shown in the Figures 5.9a and 5.10a.

5.4.5 Relationship between ground temperature and LST

The mean threshold values of day and night acquisitions were different. The ECOSTRESS LST values for the 20 locations were extracted from the nearest available day and night acquisitions to see if there was any relationship between the ground temperatures and LST (Table 5.6). The Pearson's correlation coefficient (R) between the ground temperature and LST were determined as shown in Figure 5.11. It was observed that ground temperature was highly correlated with Day time LST with R of 0.78, and night time was weakly correlated with R of 0.48.



(a) Year 2019

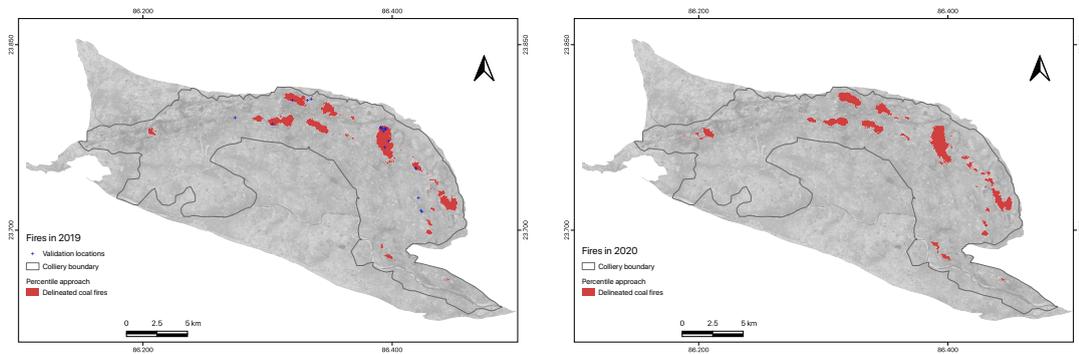


(b) Year 2020

Figure 5.9: Coal fires detected with detection ratio of pixels above the LST threshold

5.5 CHANGE IN COAL FIRE AREAS

Coal fire pixels that had detection ratio above 0.5 were considered for analysing change in coal fire areas for the years 2019 and 2020. However, this decreased the validation accuracy from 90% to 70%. The extent of the coal fire areas were 9.35 km² and 10.35 km², for the year 2019 and 2020, a change of 10.69%. The changes in the status of coal fire areas between the years were identified as 'active', 'dormant' and 'increasing' (Figure 5.12a and 5.12b). In coal fire areas above detection ratio



(a) Year 2019

(b) Year 2020

Figure 5.10: Coal fires detected by percentile approach

of 0.5, 7.99 km² were found ‘active’, 1.35 km² were ‘dormant’, and 2.35 km² were ‘increasing’. For the percentile approach, the extent of the coal fires were respectively 13.04 km² and 14.39 km² for 2019 and 2020, a change of 10.35%. The status of fires changed where 10.46 km² were found

Table 5.6 Validation of detected coal fire pixels with recorded ground observations. Temperatures are in °C. * Locations that were detected within a 70-m from the coal fire pixels and corresponding detection ratio of the nearest pixel is shown. † Locations that did not validate with the percentile approach

ID	lat	lon	Name	ground °C	LST day	LST night	detection ratio
1	23.806	86.335	Chandaur	39.02	46.81	19.35	0.22†
2	23.805	86.32	Ram Kanali	45.9	54.07	27.51	0.89
3	23.786	86.304	Kumarjuri	50.33	55.75	27.25	1.00
4	23.78	86.393	Godhar	49.14	60.05	26.17	1.00
5	23.772	86.397	Keshka	46.05	51.11	28.81	0.78
6	23.767	86.394	Alkusha	44.13	54.57	26.01	0.89
7	23.783	86.391	Godhar	41.03	48.49	20.83	0.56
8	23.782	86.395	Godhar	44.33	52.45	21.05	1.00
9	23.805	86.332	Chandaur	40.55	49.09	21.87	0.11*†
10	23.75	86.419	Rajapur	43.78	49.17	30.13	0.56
11	23.791	86.274	Tentulia	42.04	48.49	19.61	0.11*†
12	23.783	86.391	Godhar	43.56	48.49	20.83	0.56
13	23.783	86.391	Godhar	43.97	48.49	20.83	0.56
14	23.782	86.394	Godhar	43.72	56.07	21.77	1.00
15	23.75	86.419	Rajapur	41.22	49.17	30.13	0.56
16	23.716	86.423	Tisra	42.16	46.57	20.57	0.22
17	23.715	86.424	Tisra	42.02	46.25	21.99	0.11†
18	23.782	86.395	Godhar	44.21	52.45	21.05	1.00
19	23.782	86.392	Godhar	43.96	50.53	23.21	0.89
20	23.726	86.421	Bhaghtdih	42.72	47.25	21.71	0.11†

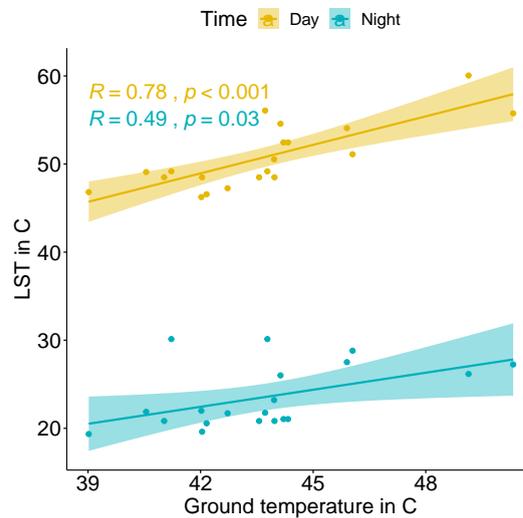
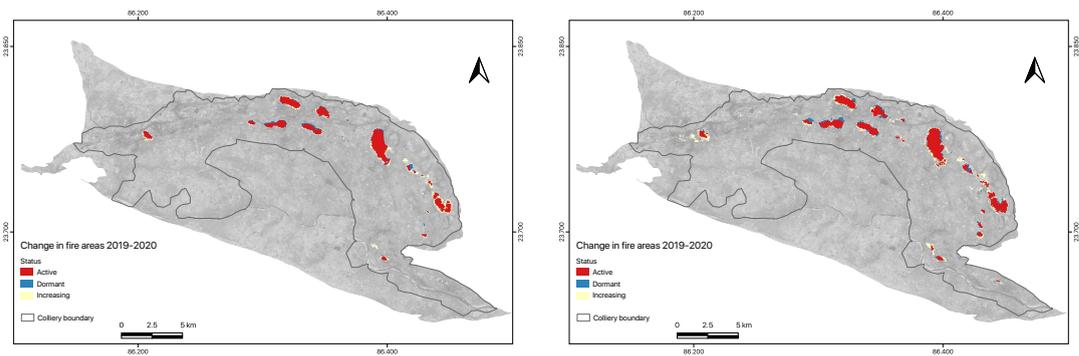


Figure 5.11: Scatter plot of correlation between ground temperature and LST values acquired during day and night time acquisitions



(a) detection ratio > 0.5

(b) percentile approach

Figure 5.12: Change in coal fire areas between 2019 and 2020

‘active’, 2.58 km² were ‘dormant’, and 3.92 km² were ‘increasing’.

5.6 GEE APP

The application was developed around the initial design (Figure 4.4) and published on the GEE platform which can be accessed at [shankaraiah.users.earthengine.app/view/fjharria](https://earthengine.google.com/view/fjharria). Features such as split panel, opacity slider, summary tables, legend, color palettes, map selector, map style selector, and chart generator were implemented to enable visualization, interactivity of the outputs of the analyses. Dictionaries and lists were used to assimilate information into the workflow. The application enables user to select one of the three basemaps. An animation allows user to visualize distribution of LST values for different acquisitions in the study area. A user can select six map layers for ‘start’ and ‘end’ year. For 2019, the layers are: 1. LULC map-2019, 2. Detected coal fire zones-2019, 3. Coal fire zones-2019, 4. Sentinel-2 FCC-2019, 5. NDVI-2019, and 6. CLAL-2019. Additional two layers, Change in mining areas and Change in fires, show the change detection

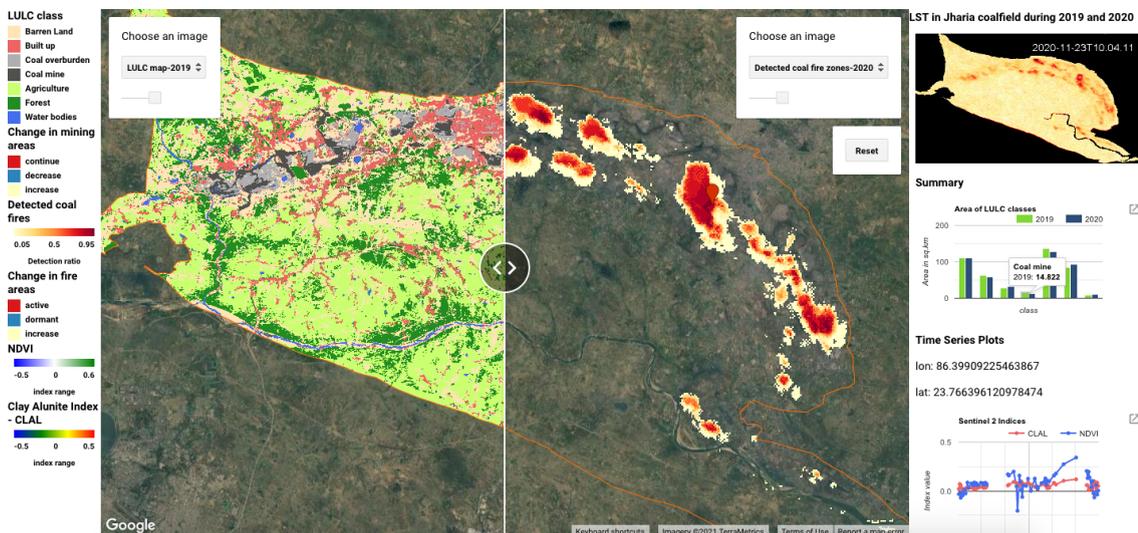


Figure 5.13: Image of developed GEE application tool *Jharia*

results. The legends corresponding to the layers are displayed in a panel. A summary chart is generated showing areas of LULC classes for both years. The user can choose location on the map to generate two time series charts, one showing NDVI and CLAL indices, and other showing LST values of the location. All of the charts are interactive and highlight information when hovered over the chart areas. These charts can also be downloaded by user in both image and comma separated value formats.

The structure of the application has 3 javascripts; 1. `landuse.js` contains functions which perform classification for each year, 2. `ecostress.js` contains functions for coal fire detection, 3. `appScript.js` which is a controller script to create layout, interaction, visualisation, chart generation, handle user inputs and interact with functions from `ecostress.js` and `landuse.js` scripts for change detection. This structure was implemented to accommodate analysis for multiple years. Outputs of detection ratio based approach for coal fire delineation were implemented in the application. One way to improve the speed of the application is to use analysis outputs as data assets instead of performing analysis such as classification and reprojection which are time consuming. This was not implemented in this application in order to enable reproducibility and to extend analysis in future. An image the published application is shown in Figure 5.13.

Chapter 6

Discussion

6.1 EFFECT OF BALANCING METHODS

There was evidence of class imbalance among the LULC classes which was addressed by ROS and RUS methods. In 2019, among the balancing methods, there was a gain of 2.00 in ΔGM without any loss in OA for ROS. In 2020, RUS had gains in both ΔGM and ΔOA . In both 2019 and 2020, ROS decreased the UA value of the coal overburden class by around 7%, even though there was improvement in its PA. Barren land and built up classes, which had already affected UA of coal overburden in unbalanced data contributed to its further decrease when oversampling replicated more samples that misclassified barren land and built up as coal overburden. It was noted that the randomness involved in non-heuristic approach could affect the accuracies when a different random seed was used to resampling training data. Balancing methods with heuristic approaches would definitely address this (Batista et al., 2004; Haixiang et al., 2017). Overall, the balancing methods did not show significant gains but still performed better than unbalanced data, which was also noted by Douzas et al. (2019) and Waldner et al. (2019). The results show effectiveness of including balancing methods in the classification procedure to improve both overall classification accuracy and the individual class accuracies, especially those of minority classes.

6.2 LULC CLASSIFICATION AND CHANGE DETECTION

Random sampling approach was adopted to generate samples which could be subject to introducing interpreter error, and the spectral profiles show some evidence of it. This affected the classification accuracy. Coal mining areas are active mining areas whereas coal overburden are inactive mining areas. So, changes in active mining areas cannot be correlated to changes in inactive mining areas as they both have different operations and management. So, coal overburden, which generally belongs to coal mining areas reported by Garai and Narayana (2018) and Ranjan et al. (2021), was not part of the coal mine class in this study. BCCL, in their BCCL/CMPDI (2020) reported active coal mine area of 11.64 km² for 2019. This was an underestimate compared to 16.3 km² mapped in this study. The road networks which are used to transport coal are mostly covered with coal dust. These roads were classified as coal mine areas. This could contribute to the difference in the coal mine areas. There were no studies to compare with 2020 mapped coal mine area of 12.51 km². CLAL index was chosen to differentiate coal mine from the coal overburden class. The index was able to discriminate not only coal mine and coal overburden but also other classes in all quarterly composites, proving a useful feature in distinguishing all the classes (Figure 4.3). Barren land, built up and coal overburden shared similar spectral signatures in visible spectrum, also mentioned by other studies (Isidro, McIntyre, Lechner, & Callow, 2017; Mukherjee et al., 2019; Ranjan et al., 2021), while built up and forest shared overlapping signatures in SWIR1 spectrum. These problems could be addressed using microwave satellite data that can discriminate features based on their dielectric property and backscattering intensities (Forkuor, Ullmann, & Griesbeck, 2020).

The change detection analysis was carried out using a post-classification method with the classified maps. Overall, coal mine areas saw a decrease in activity during 2019-2020. Most of the coal mine areas were converted to coal overburden dumps indicating continued coal mining activity. Coal overburden contributed to coal mine areas in 2020 and there was no significant contribution from other classes. The road networks and coal dump areas that were classified as coal mines in 2019 had reduced in 2020. This also shows a decrease in overall mining activity. This decrease could also have been affected by the impact of lockdowns due to the pandemic (Jain, 2021). However, these impacts specifically at JCF have not been quantified. Other noticeable results were the contribution of barren land to the forest in 2020. This could be because of most of the barren land shown as forest in the map was indeed forest in the reference data. Extent of water bodies increased in 2020 which can be because of rains during the year. There was also conversion of barren lands along the Damodar river into water bodies. An independent accuracy assessment is necessary to evaluate the change detection map to assess if the changes did really happen or the misclassifications of the mapped classes propagated into the changes (Olofsson et al., 2013). The accuracy of the change detection map was not measured in this study. The users are recommended to independently perform the accuracy assessment of change detection map following practices mentioned in Olofsson et al. (2014). These practices should include probability based sampling design; a response design with temporally consistent reference data which is of higher quality than map data; Use unbiased estimators of area and accuracies to report changes.

6.3 EFFECT OF CLOUDS

57% of the 84 images acquired over the study period could not be used as they were entirely affected by clouds. The cloud masking further reduced the LST pixels in the remaining 36 images. These were quantified using pixel coverage within the study area. Pixel coverage provided a metric to choose images for coal fire detection. Clouds were also detected for LST pixels over coal fire areas which were noticeable in night time acquisitions (Figure 5.6-C). The emissions from the burning coal fires that contribute to atmospheric sulfate aerosols are known to affect cloud detection tests and also LST retrieval algorithms by lowering solar radiance on the earth's surface (Glynn C Hulley et al., 2019; Stracher & Taylor, 2004). A thorough understanding of effects of aerosols and how cloud detection tests and LST retrieval algorithms account for aerosols is necessary.

6.4 COAL FIRE DETECTION

55% of the 36 images had pixel coverage above 0.95 which were used for coal fire detection. This shows the ability of ECOSTRESS to provide cloud-free data with high temporal resolution. The pixel coverage would be high when the image covered the entire study area. So, images that partially covered the study area were omitted. Threshold for each image was calculated for both day and night images. Detection ratio was calculated as a measure to see how often the LST pixels were detected above threshold. For 2019, validation of coal fire pixels with ground observations achieved 90% accuracy. Ground observations that were within 70-m from nearest coal fire pixel had detection ratio within 0-0.25, indicating possible weaker anomalies at these locations. There is also a possibility that the spatial extent of these fires is less than the 70-m pixel size which would produce weak anomalies.

Coal fires are known to burn for long periods of time. BCCL in 2008 reported an estimate of 17.32 km² at JCF being affected by coal fires based on data collected from ground observations (BCCL, 2008). Studies at JCF by Gautam et al. (2008), Rakesh Kumar Mishra et al. (2020), Mujawdiya et al. (2020), and Pandey et al. (2017) have characterised temporal changes in nature of coal fires

since 1998. Spatial extent of fires have been reported to be between 8 and 12 km² for various years. Changes in spatial extent of fires have been between 4 to 5 year intervals with change not more than 2 km² between intervals. One way to interpret the area estimates of these studies is that they used one image to determine the extent of coal fires representing the study interval. This could be an underestimate of the actual extent of coal fires.

Coal fires increased by 11.2 km² in 2020 from 21.4 km² in 2019, which can be mainly attributed to pixels within detection ratio 0-0.25, which was 51% of the coal fire area in 2020. This could be due to LST variability in night time images (9 in 2020 compared to 3 in 2019) that contributed to it. So, the change in extent of coal fires in 2020 when compared to 2019 may not be realistic. Pixels detected as coal fires could be refined with field knowledge and land cover data to filter out false positive pixels. Water bodies also contribute to false positive pixels because of their high thermal inertia compared to land surfaces which takes them longer to lose heat. These are noticeable in night time images. One approach to filter out false positive pixels was to choose fires that were consistently above threshold value. So coal fires above detection ratio 0.5 were chosen to see change in status of fires between the years. The extent of fires in 2020 increased from 2019. Although not significant, spatial extent of the 'increasing' fires were higher than 'dormant' fires, indicating spreading of fires. Similar pattern was also observed in the percentile approach. Comparing both the approaches, there were some noticeable pixels, especially on the periphery of coal fire zones which shared different fire status. These could be attributed to environmental factors such as wind that can affect LST values, and also variations in LST of neighboring pixels. Validation of these changes with in situ data can evaluate the accuracy of these changes. This also raises a point that is one year a good window to see changes in coal fires status, given that these fires burn for longer periods. Some fires that lay dormant due to fire mitigation measures can reignite. A one year window might be able to capture success and failure of these measures.

Percentile approach on the other hand used all the images that were available to create composite images with day and night images. It achieved higher accuracy of 75% when compared to coal fires with detection ratio >0.5, showing a better approach to see change in the status of fires. Detection of coal fires by the percentile approach however did not look like it was affected by false positive pixels but would need further investigation to confirm the results. The study would be more interesting if there were day and night acquisitions for summer and winter periods since LST values are also influenced by the seasons. Finally, the collocated LST values and ground observations showed high correlation for day images but lower for night images, which could be due to sub-pixel variations in LST over the surface.

Thresholding is one of the approaches for delineating coal fires from LST images and equation 4.11 is not a universal equation. Much of information was discarded based on this approach where spatial coverage was required for the entire study area. This could be optimised by using in situ data and reducing the spatial extent from study area to individual collieries or coal fire affected areas. Use of thresholds on LST images were able to identify coal fire areas with high accuracy, indicating the ability of ECOSTRESS sensor to capture coal fires at 70-m spatial resolution. One method to improve coal fire detection is to downscale spatial resolution from 70-m to 30-m, and then spatially sharpened to 30-m Landsat SR bands using data mining approach as discussed in M. C. Anderson et al. (2021) and Xue et al. (2020). This process enhances the sharpness of the LST pixel boundaries. Another method is to use LST along with in situ data such as elevation, land use and subsidence to differentiate coal fire pixels from non coal fire pixels (Liu et al., 2021).

The potential of ECOSTRESS spatial and temporal resolution have been evaluated by various studies dedicated to identifying communities vulnerable to UHI (G. Hulley et al., 2019); evaluating spatiotemporal variation in water use in wildfire affected areas (Poulos, Barton, Koch, Kolb, & Thode, 2021); increased monitoring of thermal anomalies in volcanoes and geothermal areas (Sil-

vestri et al., 2020); predicting amount of carbon absorbed by plants over diurnal cycle (Xing Li, Xiao, Fisher, & Baldocchi, 2021); generate daily evapotranspiration datacubes at 30-m resolution (M. C. Anderson et al., 2021).

6.5 EVALUATION OF GEE APP

Following the requirement and intended use, the application was designed and developed utilizing storage and analysis capabilities of GEE. The results of the LULC classification, coal fire detection, change in mining areas, change in coal fire areas, and indices allows users to interact and visualize map layers, charts and time series plots. There are however limitations with the application. Most notably, the application is restricted to one interface. A tool can have multiple interfaces integrated into the application and enable users a different perspective of the analysis results. Further, the application is restricted to functions provided by GEE which inhibits creating user defined functions for visualisation. This can be solved by developing the front end of the application independently while using storage and computing resources of GEE to run analysis. This however, involves hosting application on a web server and deployment on application engine for computation, which requires additional resources that are not freely available. Additional functionalities such as downloading the map layers from the application interface can enable users to use them for independent validation.

The application is intended for stakeholders and organisations to facilitate decision making. It is essential to evaluate if the application is serving the intended purpose. It is also important to assess the usability of the application. Usability is also a key component in the design of any interactive application. Feedback from testing, users and stakeholders help in evaluating and re-design of the application to make it more 'user friendly'. It also evaluates if the application serves the intended use and requirement of the end-user. Due to time constraints, usability evaluation of the developed application could not be carried out in this study.

Chapter 7

Conclusion and recommendation

7.1 CONCLUSION

The main objective of the study was to build an application on the GEE platform to enable users to access results of the analysis of opencast coal mining characteristics for the years 2019 and 2020. Active coal mining areas were identified with pixel-based classification of LULC classes at JCF using Sentinel-2 imagery. Random sampling technique was adopted to generate training samples which produced unbalanced datasets under-representing coal mine as minority class. Balancing methods of ROS and RUS addressed the class imbalance problem by improving the accuracy of minority classes and overall accuracy of the classification. RF classifier was able to use a high dimensional dataset with 90 features and contributed to the accuracy of the LULC maps. However, it is emphasized that the results of classification can only be good as the training data that is utilized. CLAL proved to be an useful index in separating classes, particularly coal mine and coal overburden, and capturing their temporal patterns. Post-classification change detection provided insights into change in LULC classes and coal mine areas between the years.

Coal fires were detected with thresholding methods using LST images. ECOSTRESS was able to provide multi-temporal cloud free images. Both day and night time images contributed to coal fire detection with good accuracy. Pixel coverage and detection ratio, respectively, were able to quantify the number of pixels available for analysis and potential of each pixel being a coal fire. There was also the presence of false positive pixels that contributed to an increase in the extent of detected coal fires. The percentile based approach used all of the LST images by analysing day and night images to identify coal fires. This approach was able to remove false positive pixels and provide a reliable extent of coal fire areas. Changes in the status of coal fires could be assessed within the coal fire areas. The ECOSTRESS LST images are highly recommended for characterizing coal fires.

GEE proved to be extremely resourceful in performing the study, given the volume of data and processing intensity. Implementation of results of LULC classification, change detection and coal fire detection on a cloud-based interactive application has the potential to be integrated into mining operations and decision making. The application is also capable of integrating results of subsequent years. All scripts developed during the thesis are released in a public repository licensed under CC by 4.0. Datasets can also be accessed from the repository links provided in the Appendix A.3. This will ensure reproducibility and also encourage future research and development with inter-comparable results.

Answers to research questions

- How has the surface area of mines changed in the study area?

Although overall coal mining activity reduced in 2020 when compared to 2019, coal mining continued operations by acquiring coal overburden areas.

- Which ECOSTRESS images are more effective in identifying coal fires?

Two approaches were used to detect coal fires. The first method used ECOSTRESS LST images with pixel coverage greater than 0.95 were able to identify coal fires with 90% accuracy, while the second method used all images to detect fires with 75% accuracy. However, the first method also detected false positive fire pixels. Presence of clouds played an important role in the availability of pixels for analysis. Results of cloud detection tests over coal fires areas also reduced pixel availability.

- Is there any relationship between LST and observed temperatures at the coal fire locations?
There was a high correlation between day time temperature and observed temperatures. But the night time temperatures had low correlation indicating LST variability with the background and within the pixels

7.2 FUTURE RECOMMENDATION

The following recommendations are suggested for future research

Analysis of coal mining characteristics

- Use of reference maps, in situ data, and adopting sampling techniques such as stratified sampling for generating reference data.
- Implement heuristic-based balancing methods to address class imbalance issues.
- Explore methods such as *boruta* (Kursa & Rudnicki, 2010) and *varSelRF* (Díaz-Uriarte & Alvarez de Andrés, 2006) for feature selection that can reduce computational intensity of the classifier.
- Explore Object based approach to LULC classification which unlike pixel-based classification, includes not only spectral properties but also spatial and texture characteristics of the classes.
- Conduct thorough analysis of cloud detection tests in ECOSTRESS LST images.
- Assess potential of ECOSTRESS LST to distinguish surface and subsurface fires at both pixel and sub-pixels levels.
- Study influence of coal mining activity with temporal variations of LST at coal fires locations.

GEE App

- Application can take input from the user such as field data to perform validation of classification and change detection maps.
- Implementing balancing methods on the GEE platform.
- Integration of multiple tools under one framework to provide users with different perspectives of the analysis.
- Develop application where user-interface is developed from scratch and uses GEE in the back-end for processing, data storage and deployment.

Bibliography

- Abdi, A. M. (2019). GIScience & Remote Sensing Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. doi:10.1080/15481603.2019.1650447
- Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., ... Brisco, B. (2020). Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13. doi:10.1109/JSTARS.2020.3021052
- Anderson, J. R. (1976). A land use and land cover classification system for use with remote sensor data. *US Government Printing Office*, 964.
- Anderson, M. C., Yang, Y., Xue, J., Knipper, K. R., Yang, Y., Gao, F., ... Rey-Sanchez, C. (2021). Interoperability of ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales. *Remote Sensing of Environment*, 252. doi:10.1016/j.rse.2020.112189
- Arévalo, P., Bullock, E. L., Woodcock, C. E., & Olofsson, P. (2020). A Suite of Tools for Continuous Land Change Monitoring in Google Earth Engine. *Frontiers in Climate*, 2. doi:10.3389/fclim.2020.576740
- Ban, Y., & Yousif, O. (2016). Change Detection Techniques: A Review. (Chap. 2). doi:10.1007/978-3-319-47037-5{_}2
- Batista, G. E. A. P. A., 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). doi:10.1145/1007730.1007735
- BCCL. (2008). Report on master plan for dealing with fire, subsidence and rehabilitation in lease holds of BCCL. Central Mine Planning and Design Institute Limited, Regional Institute.
- BCCL/CMPDI. (2020). Report on Land Use / Vegetation Cover Mapping of Jharia Coalfield based on Satellite Data of the Year 2019. Remote Sensing Cell Geomatics Division, Central Mine Planning & Design Institute Ltd.
- Beckschäfer, P. (2017). Obtaining rubber plantation age information from very dense Landsat TM & ETM + time series data and pixel-based image compositing. *Remote Sensing of Environment*, 196. doi:10.1016/j.rse.2017.04.003
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114. doi:10.1016/j.isprsjprs.2016.01.011
- Belgiu, M., & Stein, A. (2019). Spatiotemporal Image Fusion in Remote Sensing. *remote sensing*, 11(818). doi:10.3390/rs11070818
- Biswal, S. S., & Gorai, A. K. (2020). Change detection analysis in coverage area of coal fire from 2009 to 2019 in Jharia Coalfield using remote sensing data. *International Journal of Remote Sensing*, 41(24). doi:10.1080/01431161.2020.1800128
- Biswal, S. S., Raval, S., & Gorai, A. K. (2019). Delineation and mapping of coal mine fire using remote sensing data – a review. *International Journal of Remote Sensing*, 40(17). doi:10.1080/01431161.2018.1547455
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1). doi:10.1023/A:1010933404324

- Chatterjee, R. S. (2006). Coal fire mapping from satellite thermal IR data – A case example in Jharia Coalfield, Jharkhand, India. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2). doi:10.1016/j.isprsjprs.2005.12.002
- Chen, W., Li, X., He, H., & Wang, L. (2017). A Review of Fine-Scale Land Use and Land Cover Classification in Open-Pit Mining Areas by Remote Sensing Techniques. *Remote Sensing*, 10(2). doi:10.3390/rs10010015
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1). doi:10.1016/0034-4257(91)90048-B
- Congalton, R. G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 10(4). doi:10.1071/WF01031
- de Lucia Lobo, F., Souza-Filho, P. W. M., de Moraes Novo, E. M. L., Carlos, F. M., & Barbosa, C. C. F. (2018). Mapping Mining Areas in the Brazilian Amazon Using MSI/Sentinel-2 Imagery (2017). *Remote Sensing*, 10(8). doi:10.3390/rs10081178
- Demirel, N., Emil, M. K., & Duzgun, H. S. (2011). Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. *International Journal of Coal Geology*, 86(1). doi:10.1016/j.coal.2010.11.010
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1). doi:10.1186/1471-2105-7-3
- do Nascimento Bendini, H., Fonseca, L. M. G., Schwieder, M., Körting, T. S., Rufin, P., Sanches, I. D. A., ... Hostert, P. (2019). Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. *International Journal of Applied Earth Observation and Geoinformation*, 82. doi:10.1016/j.jag.2019.05.005
- Douzas, G., Bacao, F., Fonseca, J., & Khudinyan, M. (2019). Imbalanced Learning in Land Cover Classification: Improving Minority Classes' Prediction Accuracy Using the Geometric SMOTE Algorithm. *remote sensing*, 11(3040). doi:10.3390/rs11243040
- Drury, S. A. (1987). Image interpretation in geology. *Geocarto International*, 2(2). doi:10.1080/10106048709354098
- Du, X., Cao, D., Mishra, D., Bernardes, S., Jordan, T., & Madden, M. (2015). Self-Adaptive Gradient-Based Thresholding Method for Coal Fire Detection Using ASTER Thermal Infrared Data, Part I: Methodology and Decadal Change Detection. *Remote Sensing*, 7(6). doi:10.3390/rs70606576
- Ermida, S. L., Soares, P., Mantas, V., Göttsche, F.-M., & Trigo, I. F. (2020). Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series. *Remote Sensing*, 12(9). doi:10.3390/rs12091471
- Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., ... Hook, S. (2020). ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. *Water Resources Research*, 56(4). doi:10.1029/2019WR026058
- Foody, G., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6). doi:10.1109/TGRS.2004.827257
- Forkuor, G., Ullmann, T., & Griesbeck, M. (2020). Mapping and Monitoring Small-Scale Mining Activities in Ghana using Sentinel-1 Time Series (2015–2019). *Remote Sensing*, 12(6). doi:10.3390/rs12060911
- Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. doi:10.1016/S0034-4257(96)00067-3

- Garai, D., & Narayana, A. C. (2018). Land use/land cover changes in the mining area of Godavari coal fields of southern India. *The Egyptian Journal of Remote Sensing and Space Science*, 21(3), 375–381. doi:10.1016/J.EJRS.2018.01.002
- Gautam, R. S., Singh, D., Mittal, A., & Sajin, P. (2008). Application of SVM on satellite images to detect hotspots in Jharia coal field region of India. *Advances in Space Research*, 41(11). doi:10.1016/j.asr.2007.05.011
- Ghosh, R. (1991). Reclaiming wastelands of Jharia coalfield, eastern India. *International Journal of Surface Mining, Reclamation and Environment*, 5(4), 185–190.
- Gomes, V., Queiroz, G., & Ferreira, K. (2020). An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sensing*, 12(8). doi:10.3390/rs12081253
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. doi:10.1016/j.rse.2017.06.031
- Griffiths, P., van der Linden, S., Kuemmerle, T., & Hostert, P. (2013). A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5). doi:10.1109/JSTARS.2012.2228167
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73. doi:10.1016/j.eswa.2016.12.035
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160). doi:10.1126/science.1244693
- Hina, P., D., G. R., K., S. A., & Anupma, P. (2015). Impact of Mining Activities on Land Use Land Cover in the Jharia Coalfield, India. In *Coal and peat fires: A global perspective* (Chap. 9, Vol. 3, pp. 263–279). doi:10.1016/B978-0-444-59509-6.00009-0
- Hook, S., & Hulley, G. (2019a). ECOSTRESS Cloud Mask Daily L2 Global 70 m V001 [Data set]. doi:10.5067/ECOSTRESS/ECO2CLD.001
- Hook, S., & Hulley, G. (2019b). ECOSTRESS Land Surface Temperature and Emissivity Daily L2 Global 70 m V001 [Data set]. doi:10.5067/ECOSTRESS/ECO2LSTE.001
- Hook, S., Smyth, M., Logan, T., & Johnson, W. (2019). ECOSTRESS Geolocation Daily L1B Global 70 m V001 [Data set]. doi:10.5067/ECOSTRESS/ECO1BGEO.001
- Hulley, G., Shivers, S., Wetherley, E., & Cudd, R. (2019). New ECOSTRESS and MODIS Land Surface Temperature Data Reveal Fine-Scale Heat Vulnerability in Cities: A Case Study for Los Angeles County, California. *Remote Sensing*, 11(18). doi:10.3390/rs11182136
- Hulley, G. C. [Glynn C], Ghent, D., Göttsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). Land Surface Temperature. In G. C. Hulley & D. Ghent (Eds.), *Taking the temperature of the earth* (Chap. 3, pp. 57–127). doi:https://doi.org/10.1016/B978-0-12-814458-9.00003-4
- Hulley, G. C. [Glynn C.], Göttsche, F. M., Rivera, G., Hook, S. J., Freepartner, R. J., Martin, M. A., ... Johnson, W. R. (2021). Validation and Quality Assessment of the ECOSTRESS Level-2 Land Surface Temperature and Emissivity Product. *IEEE Transactions on Geoscience and Remote Sensing*. doi:10.1109/TGRS.2021.3079879
- Huo, H., Jiang, X., Song, X., Li, Z.-L., Ni, Z., & Gao, C. (2014). Detection of Coal Fire Dynamics and Propagation Direction from Multi-Temporal Nighttime Landsat SWIR and TIR Data: A Case Study on the Rujigou Coalfield, Northwest (NW) China. *Remote Sensing*, 6, 1234–1259. doi:10.3390/rs6021234

- Huo, H., Ni, Z., Gao, C., Zhao, E., Zhang, Y., Lian, Y., ... Cui, T. (2015). A Study of Coal Fire Propagation with Remotely Sensed Thermal Infrared Data. *Remote Sensing*, 7(3). doi:10.3390/rs70303088
- Isidro, C., McIntyre, N., Lechner, A., & Callow, I. (2017). Applicability of Earth Observation for Identifying Small-Scale Mining Footprints in a Wet Tropical Region. *Remote Sensing*, 9(9). doi:10.3390/rs9090945
- J., J. E., & Cochran, W. G. (1978). Sampling Techniques. *Technometrics*, 20(1). doi:10.2307/1268167
- Jain, P. K. (2021). Impact of lockdown on the mining industry in India. *Mineral Economics*, (34), 331–335. doi:10.1007/s13563-021-00263-6
- Karan, S. K., Samadder, S. R., & Maiti, S. K. (2016). Assessment of the capability of remote sensing and GIS techniques for monitoring reclamation success in coal mine degraded lands. *Journal of Environmental Management*, 182. doi:10.1016/j.jenvman.2016.07.070
- Kollert, A., Bremer, M., Löw, M., & Rutzinger, M. (2021). Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region. *International Journal of Applied Earth Observation and Geoinformation*, 94. doi:10.1016/j.jag.2020.102208
- Krehbiel, C. (2021). ECOSTRESS Swath to Grid Conversion Script. Retrieved from https://git.earthdata.nasa.gov/projects/LPDUR/repos/ecostress_swath2grid/browse
- Kuenzer, C. (2015). Remote and In situ Mapping of Coal Fires: Case Studies from China and India. In *Coal and peat fires: A global perspective* (Vol. 3, pp. 57–93). doi:10.1016/B978-0-444-59509-6.00003-X
- Kumar, L., & Mutanga, O. (2018). Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. doi:10.3390/rs10101509
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11). doi:10.18637/jss.v036.i11
- Lee, J., Cardille, J., & Coe, M. (2018). BULC-U: Sharpening Resolution and Improving Accuracy of Land-Use/Land-Cover Classifications in Google Earth Engine. *Remote Sensing*, 10(9). doi:10.3390/rs10091455
- Li, F., Yang, W., Liu, X., Sun, G., & Liu, J. (2018). Using high-resolution UAV-borne thermal infrared imagery to detect coal fires in Majiliang mine, Datong coalfield, Northern China. *Remote Sensing Letters*, 9(1). doi:10.1080/2150704X.2017.1392632
- Li, X. [Xianju], Chen, W., Cheng, X., & Wang, L. (2016). A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. *Remote Sensing*, 8(6). doi:10.3390/rs8060514
- Li, X. [Xing], Xiao, J., Fisher, J. B., & Baldocchi, D. D. (2021). ECOSTRESS estimates gross primary production with fine spatial resolution for different times of day from the International Space Station. *Remote Sensing of Environment*, 258, 112360. doi:10.1016/J.RSE.2021.112360
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote Sensing and Image Interpretation* (7th ed.). Hoboken, NJ: John Wiley & Sons.
- Liu, J., Wang, Y., Yan, S., Zhao, F., Li, Y., Dang, L., ... Peng, B. (2021). Underground Coal Fire Detection and Monitoring Based on Landsat-8 and Sentinel-1 Data Sets in Miqian Fire Area, XinJiang. *Remote Sensing*, 13(6). doi:10.3390/rs13061141
- López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250, 113–141. doi:10.1016/J.INS.2013.07.007

- Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, 25(12). doi:10.1080/0143116031000139863
- McKenna, P. B., Lechner, A. M., Phinn, S., & Erskine, P. D. (2020). Remote Sensing of Mine Site Rehabilitation for Ecological Outcomes: A Global Systematic Review. *Remote Sensing*, 12(21). doi:10.3390/rs12213535
- Mishra, R. K. [R K], Bahuguna, P. P., & Singh, V. K. (2011). Detection of coal mine fire in Jharia Coal Field using Landsat-7 ETM+ data. *International Journal of Coal Geology*, 86(1). doi:10.1016/j.coal.2010.12.010
- Mishra, R. K. [Rakesh Kumar], Pandey, J. K., Pandey, J., Kumar, S., & Roy, P. N. S. (2020). Detection and Analysis of Coal Fire in Jharia Coalfield (JCF) Using Landsat Remote Sensing Data. *Journal of the Indian Society of Remote Sensing*, 48(2). doi:10.1007/s12524-019-01067-6
- Mujawdiya, R., Chatterjee, R. S., & Kumar, D. (2020). MODIS land surface temperature time series decomposition for detecting and characterizing temporal intensity variations of coal fire induced thermal anomalies in Jharia coalfield, India. *Geocarto International*. doi:10.1080/10106049.2020.1818853
- Mukherjee, J., Mukherjee, J., Chakravarty, D., & Aikat, S. (2019). A Novel Index to Detect Open-cast Coal Mine Areas From Landsat 8 OLI/TIRS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(3). doi:10.1109/JSTARS.2019.2896842
- Müller, H., Rufin, P., Griffiths, P., Siqueira, A. J. B., & Hostert, P. (2015). Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment*, 156. doi:10.1016/j.rse.2014.10.014
- Naboureh, A., Ebrahimi, H., Azadbakht, M., Bian, J., & Amani, M. (2020). RUESVMs: An Ensemble Method to Handle the Class Imbalance Problem in Land Cover Mapping Using Google Earth Engine. *Remote Sensing*, 12(21). doi:10.3390/rs12213484
- NASA/JPL-Caltech. (2020). NASA's ECOSTRESS Monitors California's Apple Fire From Space. Retrieved from <https://www.nasa.gov/feature/jpl/nasas-ecostress-monitors-californias-apple-fire-from-space>
- Nascimento, F. S., Gastauer, M., Souza-Filho, P. W. M., Nascimento, W. R., Santos, D. C., & Costa, M. F. (2020). Land Cover Changes in Open-Cast Mining Complexes Based on High-Resolution Remote Sensing Data. *Remote Sensing*, 12(4). doi:10.3390/rs12040611
- National Remote Sensing Centre. (2012). *National Land Use Land Cover Mapping using Multi-temporal Satellite Data Technical Manual (2nd Cycle)*. National Remote Sensing Centre. Hyderabad.
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148. doi:10.1016/j.rse.2014.02.015
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129. doi:10.1016/j.rse.2012.10.031
- Pandey, J., Kumar, D., Panigrahi, D. C., & Singh, V. K. (2017). Temporal transition analysis of coal mine fire of Jharia coalfield, India, using Landsat satellite imageries. *Environmental Earth Sciences*, 76(12), 439. doi:10.1007/s12665-017-6765-8
- Pei, W., Yao, S., Knight, J. F., Dong, S., Pelletier, K., Lian, •., ... Klassen, J. (2017). Mapping and detection of land use change in a coal mining area using object-based image analysis. *Environmental Earth Sciences*, 76. doi:10.1007/s12665-017-6444-9
- Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R. V., Clinton, N. E., ... Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia

- using Landsat and Google Earth Engine. *PLoS ONE*, 13(7). doi:10.1371/journal.pone.0197758
- Petropoulos, G. P., Partsinevelos, P., & Mitraka, Z. (2013). Change detection of surface mining activity and reclamation based on a machine learning approach of multi-temporal Landsat TM imagery. *Geocarto International*, 28(4). doi:10.1080/10106049.2012.706648
- Poulos, H. M., Barton, A. M., Koch, G. W., Kolb, T. E., & Thode, A. E. (2021). Wildfire severity and vegetation recovery drive post-fire evapotranspiration in a southwestern pine-oak forest, Arizona, USA. *Remote Sensing in Ecology and Conservation*. doi:10.1002/rse2.210
- Prakash, A., Gens, R., Prasad, S., Raju, A., & Gupta, R. (2013). Coal Fires in the Jharia Coalfield, India. In G. Stracher, A. Prakash, & Sokol E.V (Eds.), *Coal and peat fires: A global perspective* (Chap. 10, pp. 564–). doi:10.1016/B978-0-444-59412-9.00010-7
- Prakash, A., & Gupta, R. P. (1998). Land-use mapping and change detection in a coal mining area - a case study in the Jharia coalfield, India. *International Journal of Remote Sensing*, 19(3). doi:10.1080/014311698216053
- Prakash, A., & Gupta, R. P. (1999). Surface fires in Jharia coalfield, India-their distribution and estimation of area and temperature from TM data. *International Journal of Remote Sensing*, 20(10). doi:10.1080/014311699212281
- Pu, D. C., Sun, J. Y., Ding, Q., Zheng, Q., Li, T. T., & Niu, X. F. (2020). Mapping urban areas using dense time series of Landsat images and Google Earth Engine. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W10. doi:10.5194/isprs-archives-XLII-3-W10-403-2020
- QGIS Development Team. (2021). QGIS Geographic Information System. Retrieved from <https://www.qgis.org>
- R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org>
- Raju, A., Gupta, R. P., & Prakash, A. (2013). Delineation of coalfield surface fires by thresholding Landsat TM-7 day-time image data. *Geocarto International*, 28(4). doi:10.1080/10106049.2012.710651
- Ranjan, A. K., Sahoo, D., & Gorai, A. K. (2021). Quantitative assessment of landscape transformation due to coal mining activity using earth observation satellite data in Jharsuguda coal mining region, Odisha, India. *Environment, Development and Sustainability*, 23, 4484–4499. doi:10.1007/s10668-020-00784-0
- Saini, V., Gupta, R. P., & Arora, M. K. (2016). Environmental impact studies in coalfields in India: A case study from Jharia coal-field. *Renewable and Sustainable Energy Reviews*, 53. doi:10.1016/j.rser.2015.09.072
- Shariffar, A., Sarmadian, F., & Minasny, B. (2019). Mapping imbalanced soil classes using Markov chain random fields models treated with data resampling technique. *Computers and Electronics in Agriculture*, 159. doi:10.1016/j.compag.2019.03.006
- Silvestri, M., Romaniello, V., Hook, S., Musacchio, M., Teggi, S., & Buongiorno, M. F. (2020). First Comparisons of Surface Temperature Estimations between ECOSTRESS, ASTER and Landsat 8 over Italian Volcanic and Geothermal Areas. *Remote Sensing*, 12(1). doi:10.3390/rs12010184
- Singh, A. (1989). Digital change detection techniques using remotely-sensed data Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6), 989–1003. doi:10.1080/01431168908903939
- Coal Fires in the Jharia Coalfield, India. (2013). In G. B. Stracher, A. Prakash, & E. V. Sokol (Eds.), *Coal and peat fires: A global perspective* (Chap. 10, Vol. 2, pp. 153–177). doi:<https://doi.org/10.1016/B978-0-444-59412-9.00010-7>

- Stracher, G. B., & Taylor, T. P. (2004). Coal fires burning out of control around the world: thermodynamic recipe for environmental catastrophe. *International Journal of Coal Geology*, 59(1-2). doi:10.1016/j.coal.2003.03.002
- Sun, A. Y., & Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters*, 14(7). doi:10.1088/1748-9326/ab1b7d
- Sunar, F. (1998). An analysis of changes in a multi-date data set: A case study in the Ikitelli area, Istanbul, Turkey. *International Journal of Remote Sensing*, 19(2). doi:10.1080/014311698216215
- Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations-A Review. *Remote Sensing*, 12, 1135. doi:10.3390/rs12071135
- Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164. doi:10.1016/j.isprsjprs.2020.04.001
- Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., ... Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144. doi:10.1016/j.isprsjprs.2018.07.017
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. doi:10.1016/0034-4257(79)90013-0
- Waldner, F., Chen, Y., Lawes, R., & Hochman, Z. (2019). Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sensing of Environment*, 233. doi:10.1016/j.rse.2019.111375
- Wang, M., Zhang, Z., Hu, T., Wang, G., He, G., Zhang, Z., ... Liu, X. (2020). An Efficient Framework for Producing Landsat-Based Land Surface Temperature Data Using Google Earth Engine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13. doi:10.1109/JSTARS.2020.3014586
- Xue, J., Anderson, M. C., Gao, F., Hain, C., Sun, L., Yang, Y., ... Schull, M. (2020). Sharpening ECOSTRESS and VIIRS land surface temperature using harmonized Landsat-Sentinel surface reflectances. *Remote Sensing of Environment*, 251. doi:10.1016/j.rse.2020.112055
- Zhang, J. (2004). *Spatial and statistical analysis of thermal satellite imagery for extraction of coal fire related anomalies* (Doctoral dissertation, Technical University Vienna, Austria).

Appendix A

Table A.1: Time stamp and corresponding ECOSTRESS granule ID of ECO2LSTE, ECO1BGEO and ECO2CLD data products for the study area. Filenames of the products and processed images can be identified with the granule ID

Date Time UTC	Granule ID
2019-01-26 10:25:10	03166_007_20190126T102510
2019-01-30 8:35:51	03227_006_20190130T083551
2019-02-02 7:38:40	03273_006_20190202T073840
2019-03-03 5:29:38	03722_007_20190303T052938
2019-03-10 2:45:52	03829_008_20190310T024552
2019-05-27 10:18:51	05046_003_20190527T101851
2019-06-07 6:05:00	05214_006_20190607T060500
2019-06-11 4:23:25	05275_006_20190611T042325
2019-06-15 2:42:25	05336_005_20190615T024225
2019-06-18 1:49:53	05382_005_20190618T014953
2019-07-02 19:53:54	05611_005_20190702T195354
2019-07-06 18:13:32	05672_005_20190706T181332
2019-07-10 16:33:25	05733_005_20190710T163325
2019-07-17 14:01:14	05840_005_20190717T140114
2019-07-25 10:40:36	05962_001_20190725T104036
2019-08-09 4:46:32	06191_004_20190809T044632
2019-08-12 3:55:22	06237_005_20190812T035522
2019-08-30 20:30:44	06527_005_20190830T203044
2019-09-03 18:52:48	06588_006_20190903T185248
2019-09-07 17:14:11	06649_006_20190907T171411
2019-09-09 2:14:24	06670_017_20190909T021424
2019-09-11 15:36:02	06710_005_20190911T153602
2019-09-15 13:57:15	06771_006_20190915T135715
2019-09-18 13:08:33	06817_006_20190918T130833
2019-09-22 11:31:36	06878_005_20190922T113136
2019-09-30 8:17:39	07000_003_20190930T081739
2019-10-08 5:02:44	07122_002_20191008T050244
2019-10-26 21:43:25	07412_006_20191026T214325
2019-10-29 20:54:10	07458_005_20191029T205410
2019-10-30 20:05:53	07473_005_20191030T200553
2019-11-01 5:05:53	07494_011_20191101T050553
2019-11-03 18:28:09	07534_006_20191103T182809
2019-11-06 17:38:46	07580_005_20191106T173846
2019-11-10 16:01:51	07641_005_20191110T160151
2019-11-12 1:02:54	07662_012_20191112T010254

Continued on next page

Table A.1 – continued from previous page

Date Time UTC	Granule ID
2019-11-18 12:48:28	07763_006_20191118T124828
2019-11-22 11:11:32	07824_006_20191122T111132
2019-11-26 9:34:20	07885_003_20191126T093420
2019-11-27 18:35:27	07906_009_20191127T183527
2019-12-01 16:58:31	07967_008_20191201T165831
2019-12-03 7:08:51	07992_005_20191203T070851
2019-12-22 23:03:04	08297_006_20191222T230304
2020-01-06 17:28:28	08526_005_20200106T172828
2020-01-09 1:41:59	08562_009_20200109T014159
2020-01-14 14:18:30	08648_005_20200114T141830
2020-01-16 22:32:11	08684_009_20200116T223211
2020-01-18 12:43:27	08709_006_20200118T124327
2020-01-22 11:08:16	08770_006_20200122T110816
2020-01-26 9:34:26	08831_006_20200126T093426
2020-02-11 3:20:14	09075_004_20200211T032014
2020-02-15 1:46:48	09136_006_20200215T014648
2020-02-22 22:39:35	09258_006_20200222T223935
2020-03-01 19:32:10	09380_006_20200301T193210
2020-03-05 17:57:49	09441_005_20200305T175749
2020-03-05 17:58:41	09441_006_20200305T175841
2020-03-17 13:16:44	09624_006_20200317T131644
2020-03-21 11:43:23	09685_006_20200321T114323
2020-03-23 19:58:08	09721_006_20200323T195808
2020-03-27 18:25:43	09782_006_20200327T182543
2020-04-29 20:16:27	10295_005_20200429T201627
2020-05-03 18:42:08	10356_005_20200503T184208
2020-05-11 15:33:36	10478_006_20200511T153336
2020-05-19 12:25:00	10600_006_20200519T122500
2020-05-21 20:38:22	10636_009_20200521T203822
2020-06-19 23:48:30	11088_002_20200619T234830
2020-06-30 19:52:09	11256_005_20200630T195209
2020-07-20 11:59:56	11561_005_20200720T115956
2020-08-05 5:42:57	11805_005_20200805T054257
2020-08-09 4:09:31	11866_006_20200809T040931
2020-08-13 2:35:20	11927_004_20200813T023520
2020-08-24 21:54:59	12110_006_20200824T215459
2020-08-28 20:21:28	12171_001_20200828T202128
2020-09-17 12:34:10	12476_006_20200917T123410
2020-09-21 11:01:14	12537_005_20200921T110114
2020-09-27 17:44:11	12634_011_20200927T174411
2020-09-29 7:57:05	12659_006_20200929T075705
2020-10-03 6:24:42	12720_006_20201003T062442
2020-10-26 21:01:21	13086_002_20201026T210121
2020-11-03 17:53:01	13208_004_20201103T175301
2020-11-07 16:18:12	13269_003_20201107T161812
2020-11-17 21:24:37	13427_004_20201117T212437

Continued on next page

Table A.1 – continued from previous page

Date Time UTC	Granule ID
2020-11-23 10:04:11	13513_005_20201123T100411
2020-11-27 8:30:58	13574_006_20201127T083058
2020-12-13 2:16:55	13818_006_20201213T021655

Table A.2 Land cover mapping scheme used in this study

Level - 1	description
Barren land	Barren rocky, sandy areas, scrub land
Built up	Urban, rural and industrial
Coal overburden	Barren overburden dumps
Coal mine	Coal quarry, coal dump, advance quarry site
Agriculture	Cropland, fallow land
Forest	Deciduous forests, scrub forests
Water bodies	River, ponds, streams, canals, reservoir, water filled quarry

Table A.3 Citation and references to scripts and datasets used in this study

Item no	Description	DOI
1	Reference samples, training and validation data	10.6084/m9.figshare.15181098
2	Classified and coal mine change maps	10.6084/m9.figshare.15181122
3	Study area shapefiles	10.6084/m9.figshare.15181143
4	Coal fire locations for validation	10.6084/m9.figshare.15181164
5	ECOSTRESS LST images	10.6084/m9.figshare.15181173
6	Coal fire and change maps	10.6084/m9.figshare.15181200
7	GEE scripts of app and analysis	10.6084/m9.figshare.16571268