pi1

INTEGRATING OPENSTREETMAP DATA AND SENTINEL-2 IMAGERY FOR CLASSIFYING AND MONITORING INFORMAL SETTLEMENTS

Thesis Defense Presentation

Brenda Ayo (20180802)

MSc. Geospatial Technologies

Supervised by :
Joel Dinis Baptista Ferreira da Silva, PhD

Co-Supervised By:
Prof. Dr. Hanna Meyer,
Ignacio Guerrero

27- 02-2020

pi2

Presentation Outline

Informal Settlements?

"20 Worst Slums", 2017

Slumscapes, 2019

Why do they Grow?

Rapid Population Growth

"This is Place", 2017

Low incomes from agriculture

Better job prospects

The pushing and pulling forces of migration

Poor Governance

"Slumscapes", 2018

Unrecognized Rights of the Urban Poor

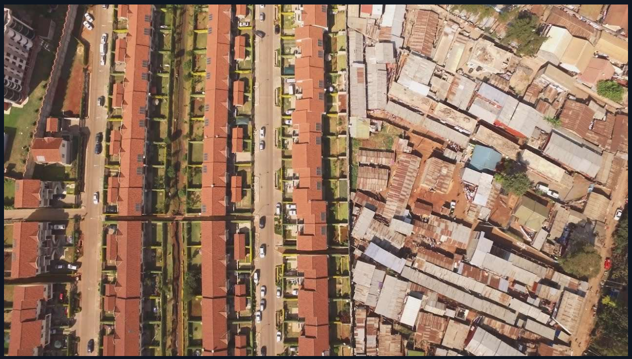
"Providing urban services to the poor attracts Urbanization and Slum growth"

Physical Characteristics of Slums

Buildings
- Small in Size (~50m²)
- Simple regular shapes
- High Density with Minimal Spacing

Slumscapes, 2019

Accessibility
- Short roads with many dead ends
- Narrow , One way



Slumscapes, 2019

Neighborhood
- Near Hazardous Locations (Dumping sites, Railways)
- Near Manufacturing Industries

"Kibera", 2019

Approaches for Mapping Informal Settlements

Census-Based

- Demographics to Map Deprivation and Poverty
- E.g. World Bank Datasets
- UN slum Indicators
- Lack of updated Census Data

Participatory-Based

- Participation of Slum Dwellers to create Spatial and Non-Spatial Data
- Slum Dwellers International (SDI)
- Time Consuming with Limited Coverage

GIS/RS-Based

- Based on Physical Characteristics
- Lack of Vegetation
- VHR satellite Imagery
- High ground coverage

Literature Review

Bradley .J .G et al (2019) - Medellin, Kibera, Makoko etc. , DigitalGlobe, Sentinel-2

- Canonical Correlation Forests (CCFs) prediction on Sentinel-2
- CNN based prediction on DigitalGlobe(30cm) imagery

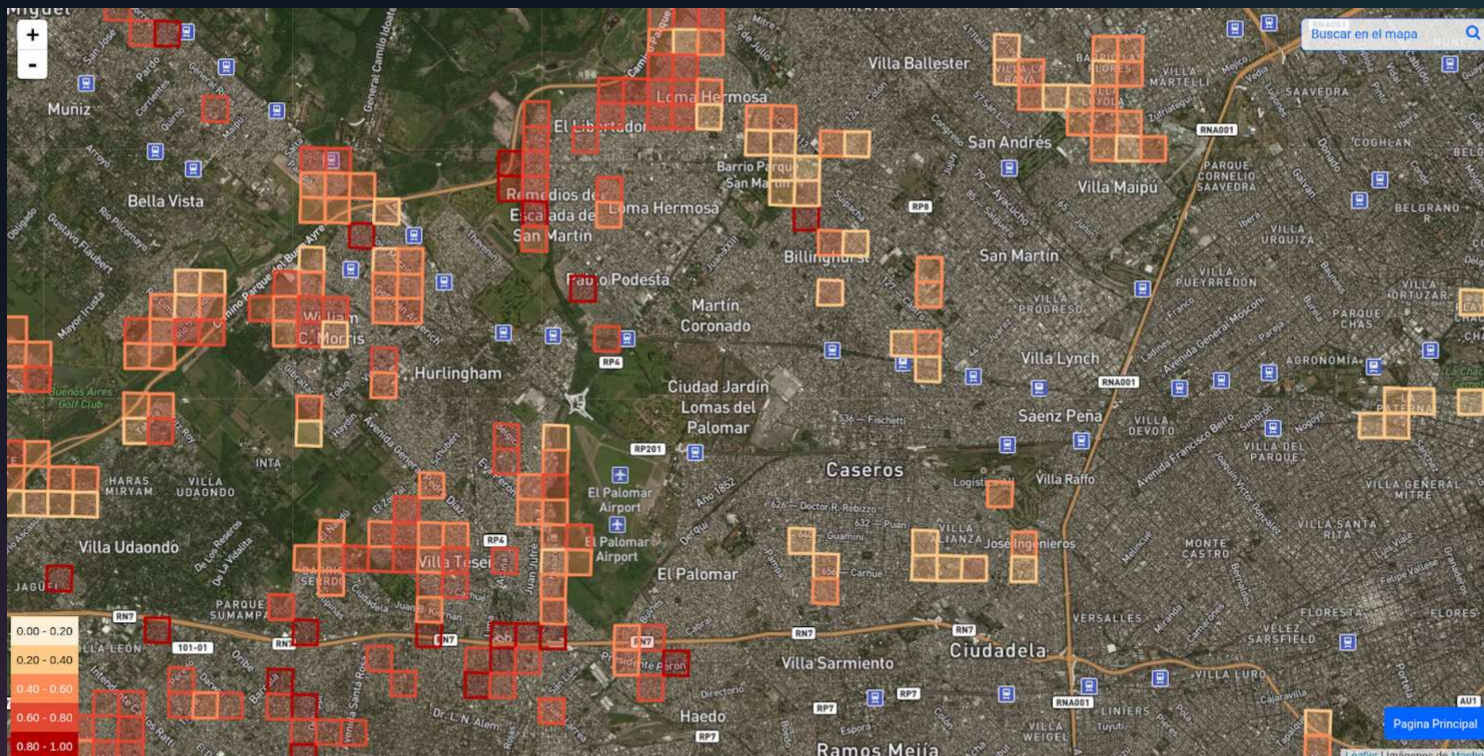
Sentinel-2 | Digital Globe | GT Mask

Bradley .J .G et al , 2019

Literature Review

Federico .B et al (2018) -South America, , Sentinel-2A

- Random Forests Classifier on Image Tiles
- Accuracy Ranging from 75%-87%

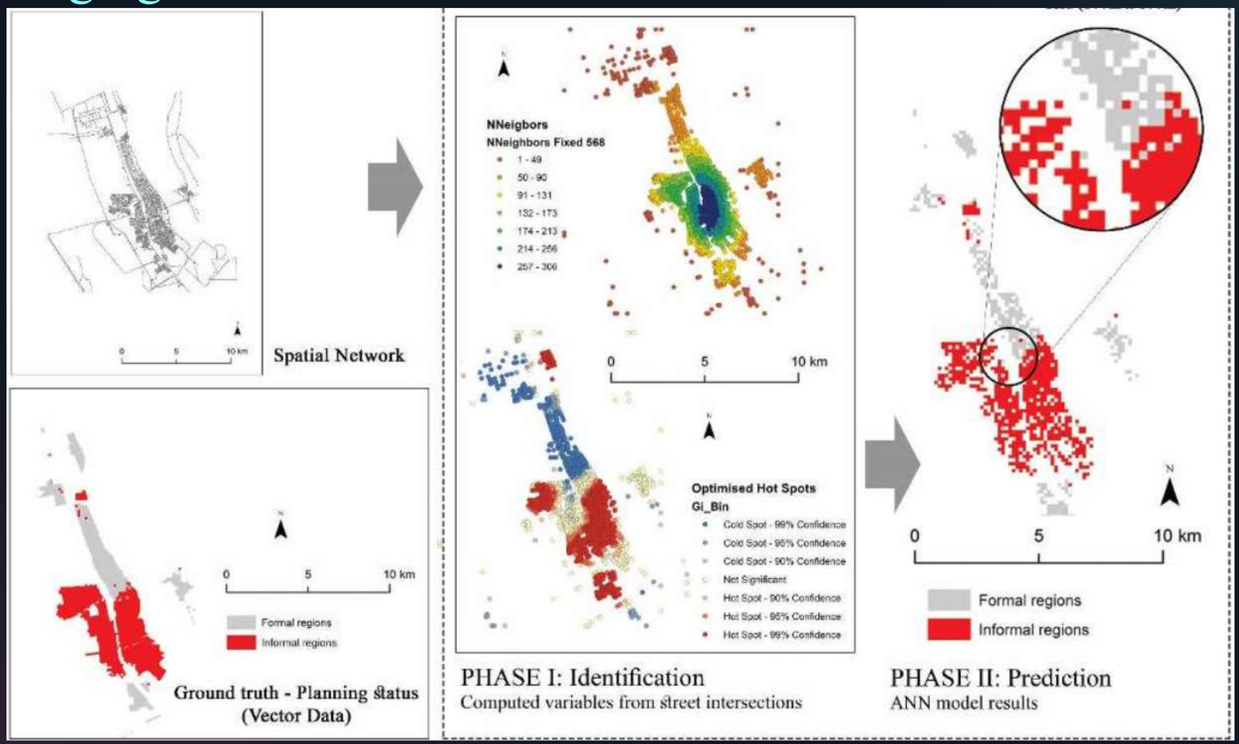


Federico .B et al , 2018

Literature Review

Ibrahim et. al (2018) – Egypt India, OSM Road Intersections

- Spatial Statistics (Hot spot Analysis)
- Muli-Nominal Logistic (MNL) Regression, Artificial Neural Networks
- Accuracy Ranging from 87% to 92%



Ibrahim et. al, 2018

Research Gap

- Small Study Areas with Knowledge of Slum existence
- Inadequate data at scales where slums can be delineated at City Level
- OSM Building Outlines combined with Machine learning to generate Informal Settlement Regions
- Sentinel-2 Imagery to monitor growth of Slums over time

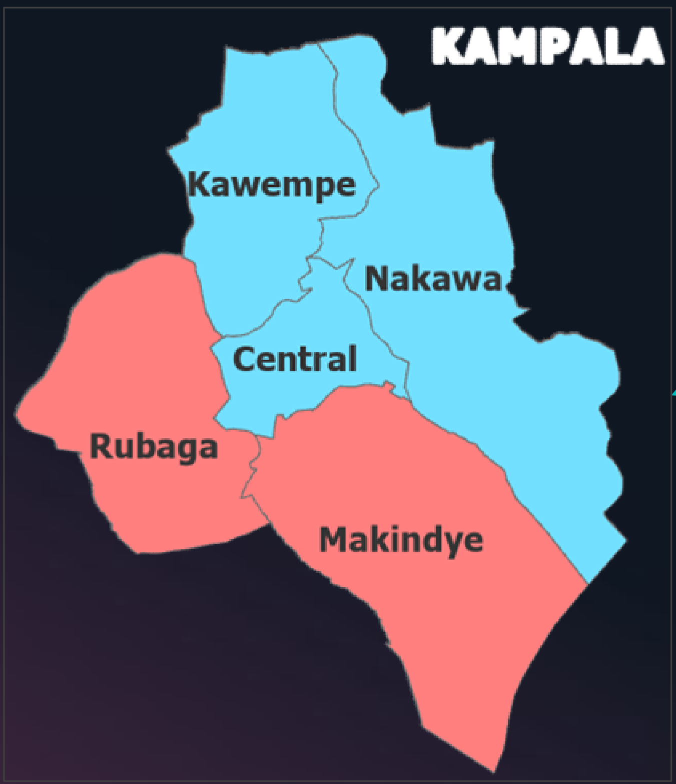
Research Aim

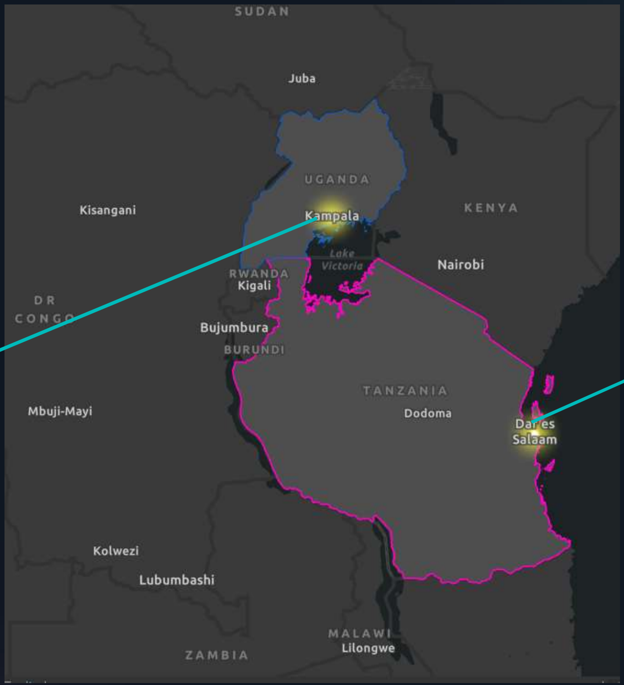
Integrate OSM data and sentinel-2 imagery for classifying and monitoring the growth of informal settlements

Research Questions:

1. How can buildings outline characteristics be used to differentiate Informal settlements from Formal Settlements?
2. Is it possible to predict informal areas in a city by understanding housing informality in other cities of similar context using buildings outline characteristics and machine learning?
3. What is the most appropriate Machine Learning technique based on accuracy to predict informal areas in a city based on buildings outline characteristics?
4. How can we exploit the potential of freely available Sentinel-2 satellite imagery with advanced machine learning to estimate the growth of Informal settlements in an Area?

Study Area


KAMPALA
Kawempe
Nakawa
Central
Rubaga
Makindye


SUDAN
Juba
UGANDA
Kampala
KENYA
Kisangani
Lake Victoria
Nairobi
RWANDA
Kigali
DR CONGO
Bujumbura
BURUNDI
TANZANIA
Dodoma
Dar es Salaam
Mbuji-Mayi
Kolwezi
Lubumbashi
MALAWI
Lilongwe
ZAMBIA

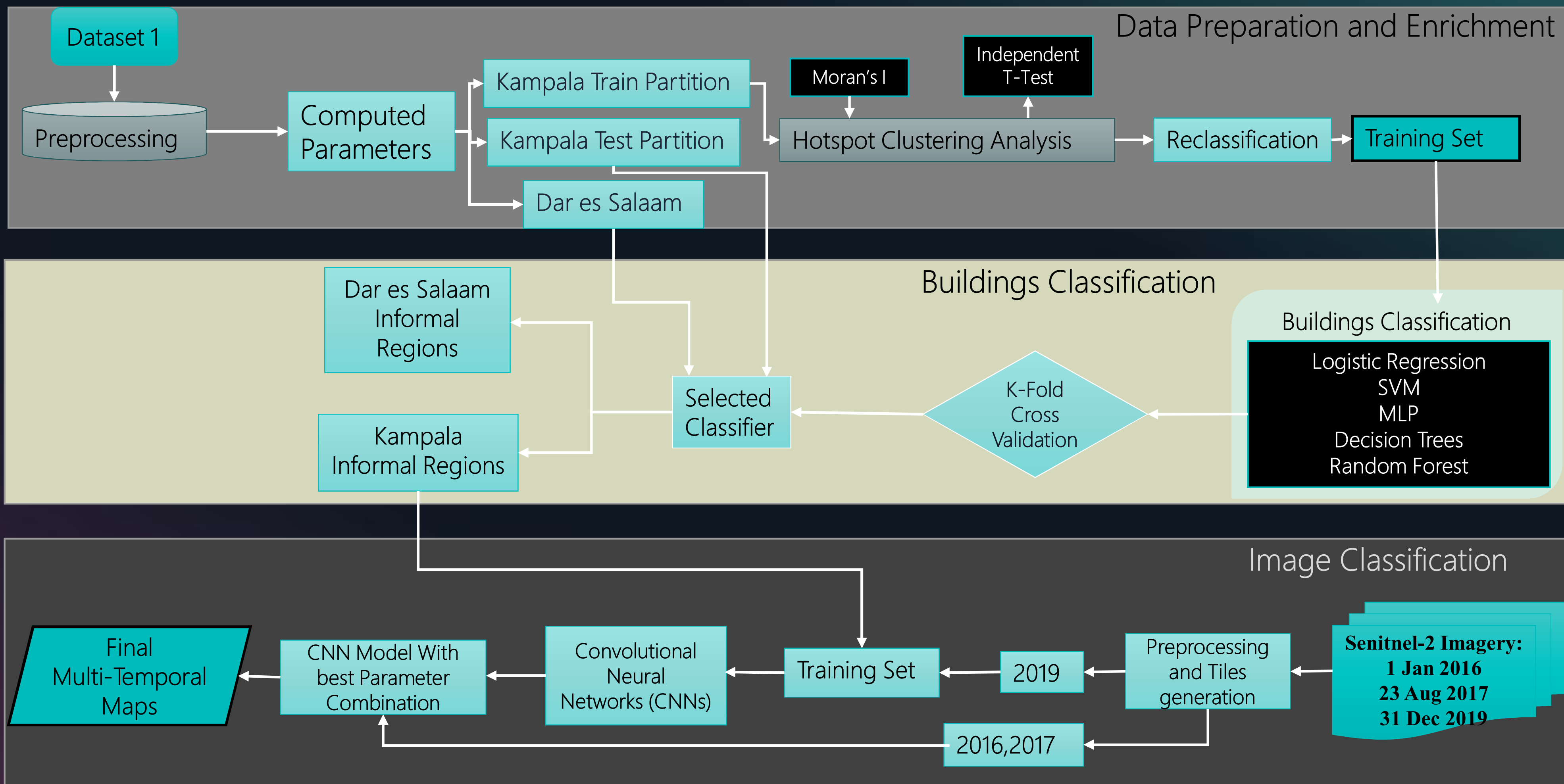

DAR ES SALAAM
Kinondoni
Ilala
Temeke

Datasets

City	# Buildings
Kampala (training partition)	214,885
Kampala (test partition)	214,191
Dar es Salaam	971,008

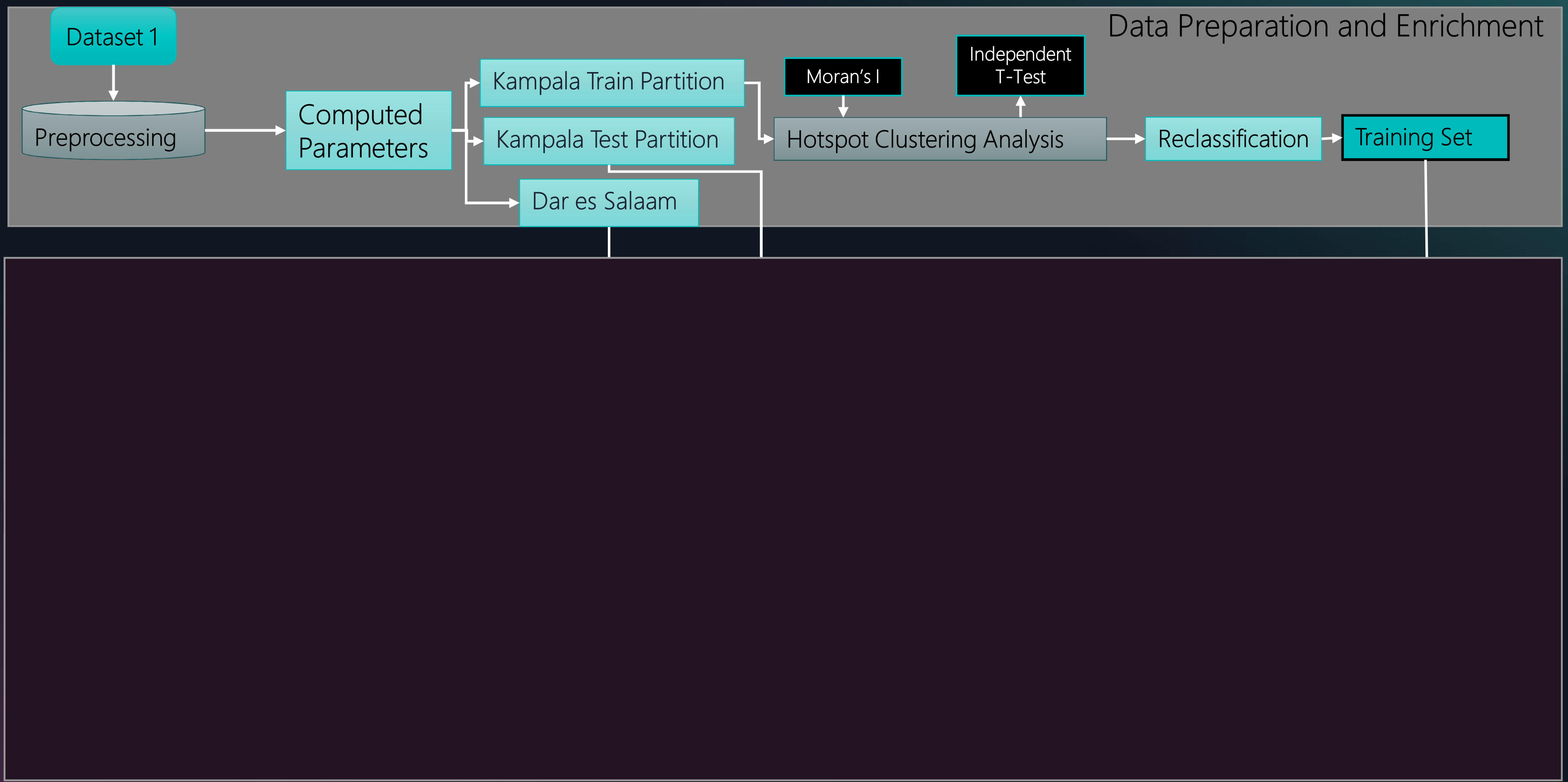
METHODOLOGY

Flowchart



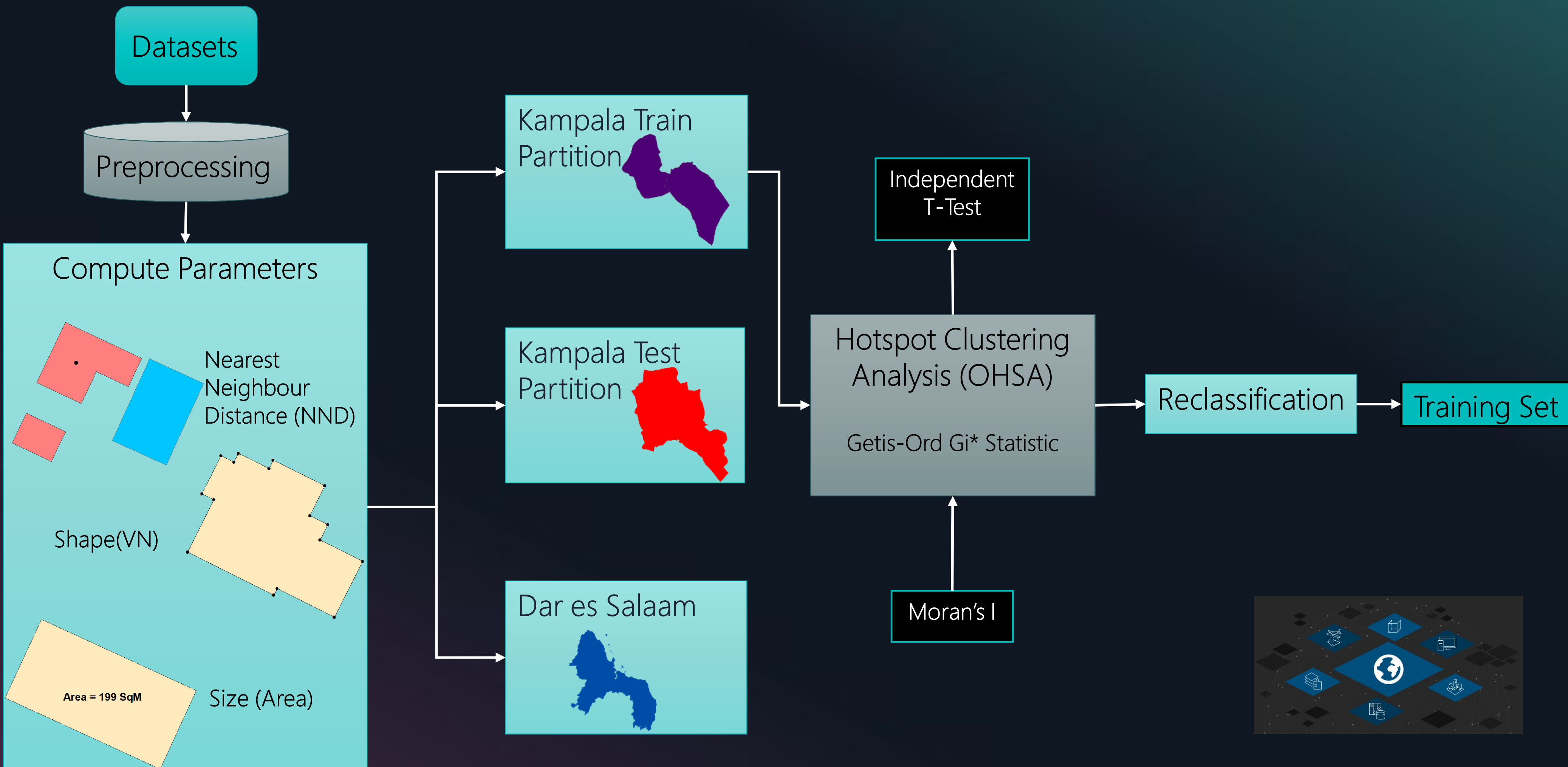
Data Preparation and Enrichment
Dataset 1
Preprocessing
Computed Parameters
Kampala Train Partition
Kampala Test Partition
Dar es Salaam
Moran's I
Independent T-Test
Hotspot Clustering Analysis
Reclassification
Training Set
Buildings Classification
Dar es Salaam Informal Regions
Kampala Informal Regions
Selected Classifier
K-Fold Cross Validation
Buildings Classification
Logistic Regression
SVM
MLP
Decision Trees
Random Forest
Image Classification
Final Multi-Temporal Maps
CNN Model With best Parameter Combination
Convolutional Neural Networks (CNNs)
Training Set
2019
Preprocessing and Tiles generation
Senitnel-2 Imagery:
1 Jan 2016
23 Aug 2017
31 Dec 2019
2016,2017

Flowchart

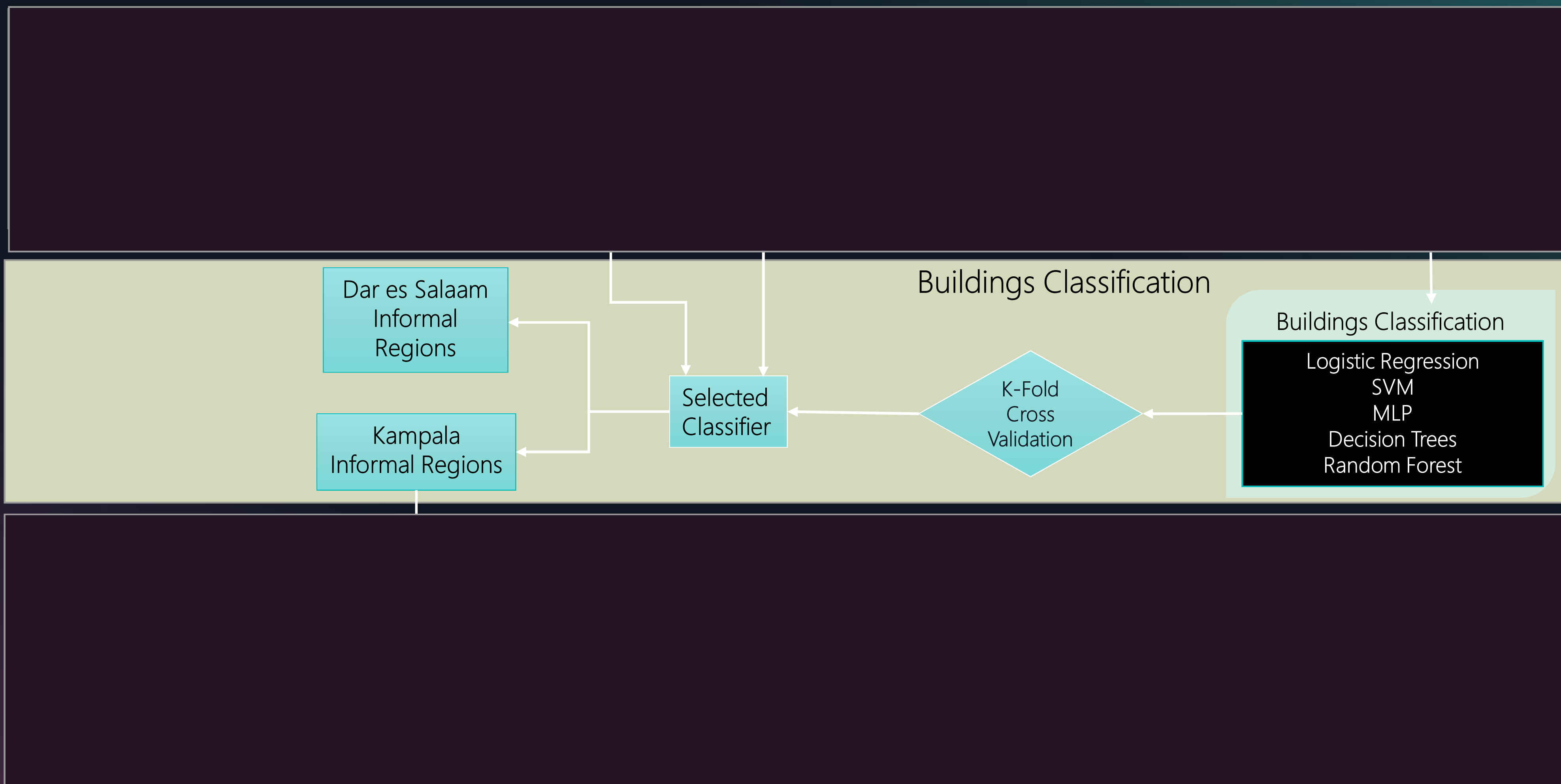


Data Preparation and Enrichment
Dataset 1
Preprocessing
Computed Parameters
Kampala Train Partition
Kampala Test Partition
Dar es Salaam
Moran's I
Independent T-Test
Hotspot Clustering Analysis
Reclassification
Training Set

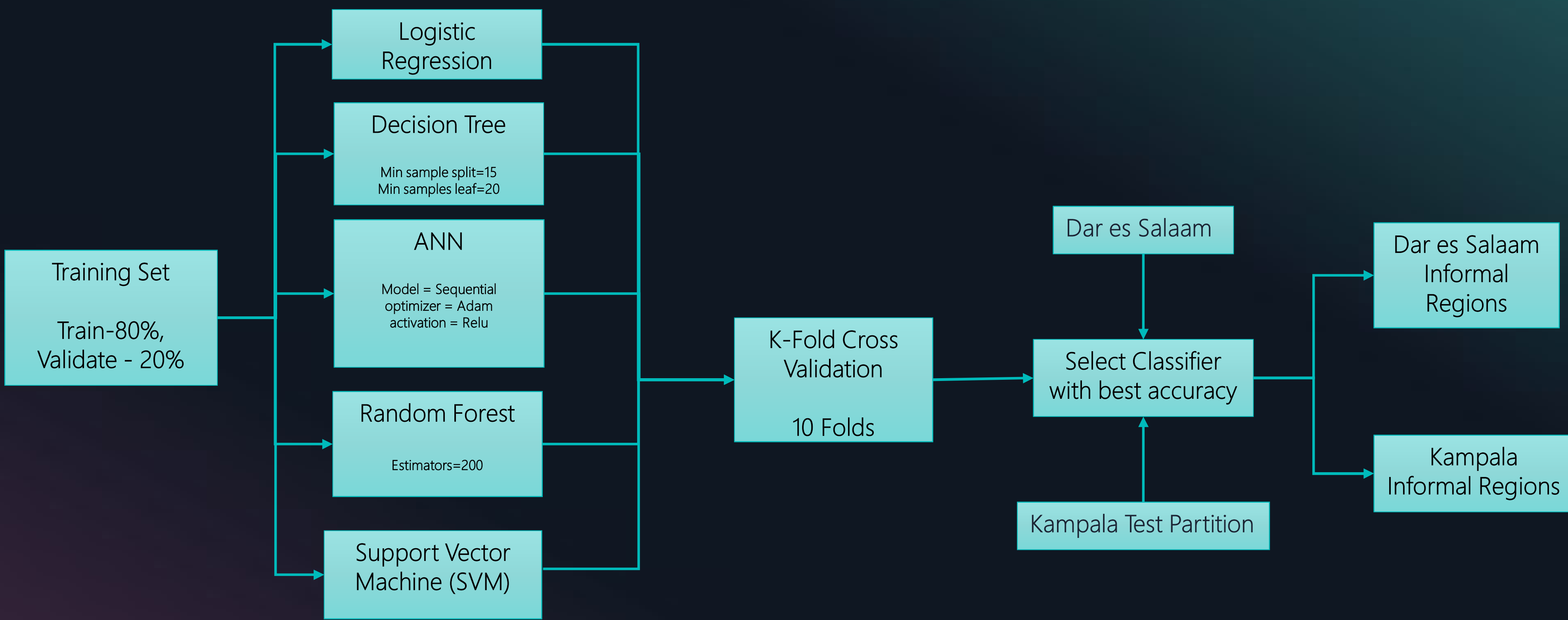
Phase I: Data Preparation and Enrichment



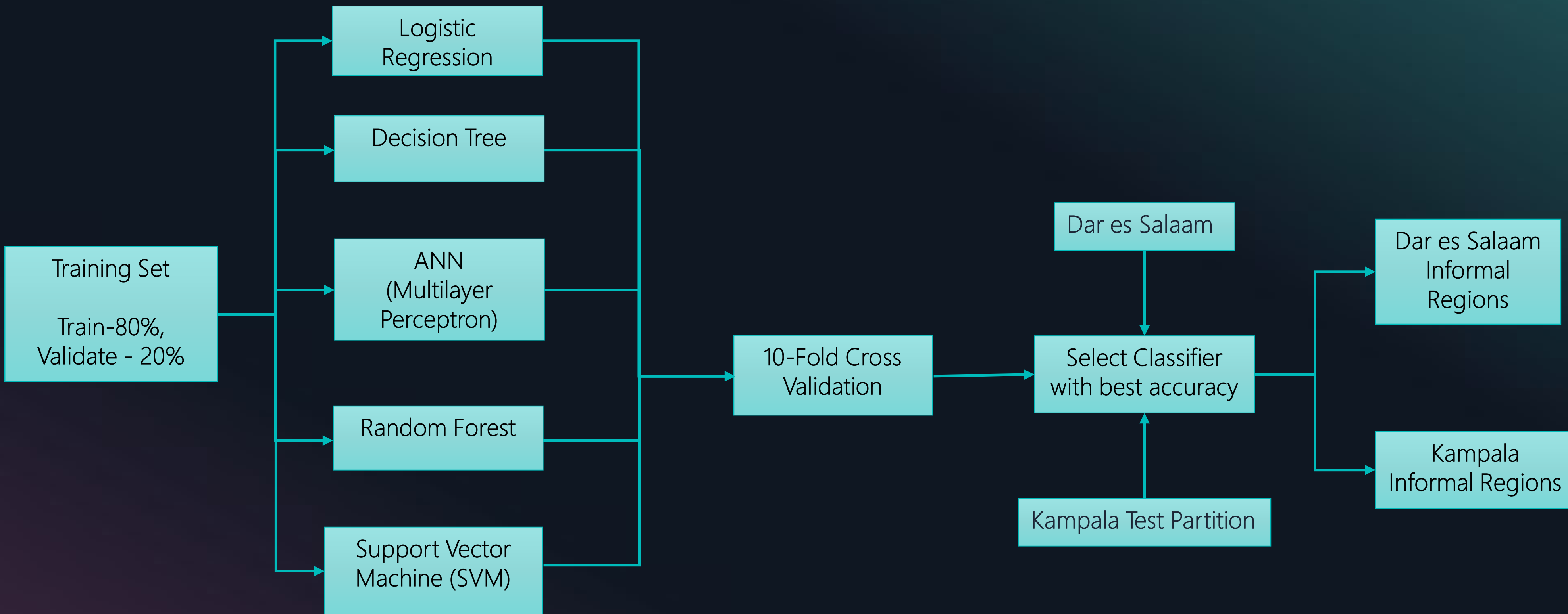
Flowchart


Buildings Classification
Dar es Salaam Informal Regions
Kampala Informal Regions
Selected Classifier
K-Fold Cross Validation
Buildings Classification
Logistic Regression
SVM
MLP
Decision Trees
Random Forest

Phase 2: Buildings Classification



Phase 2: Buildings Classification



Pandas

Flowchart

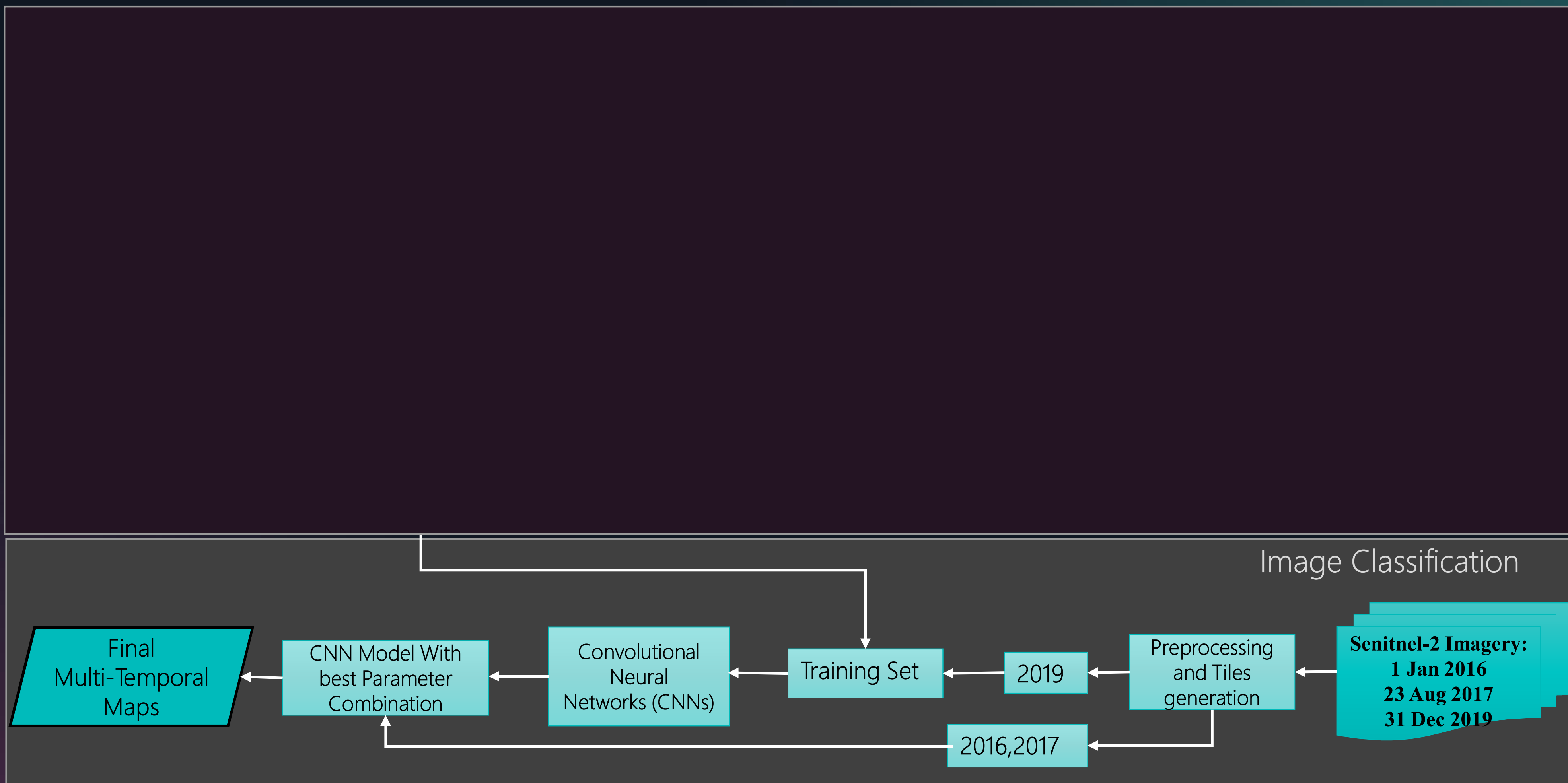
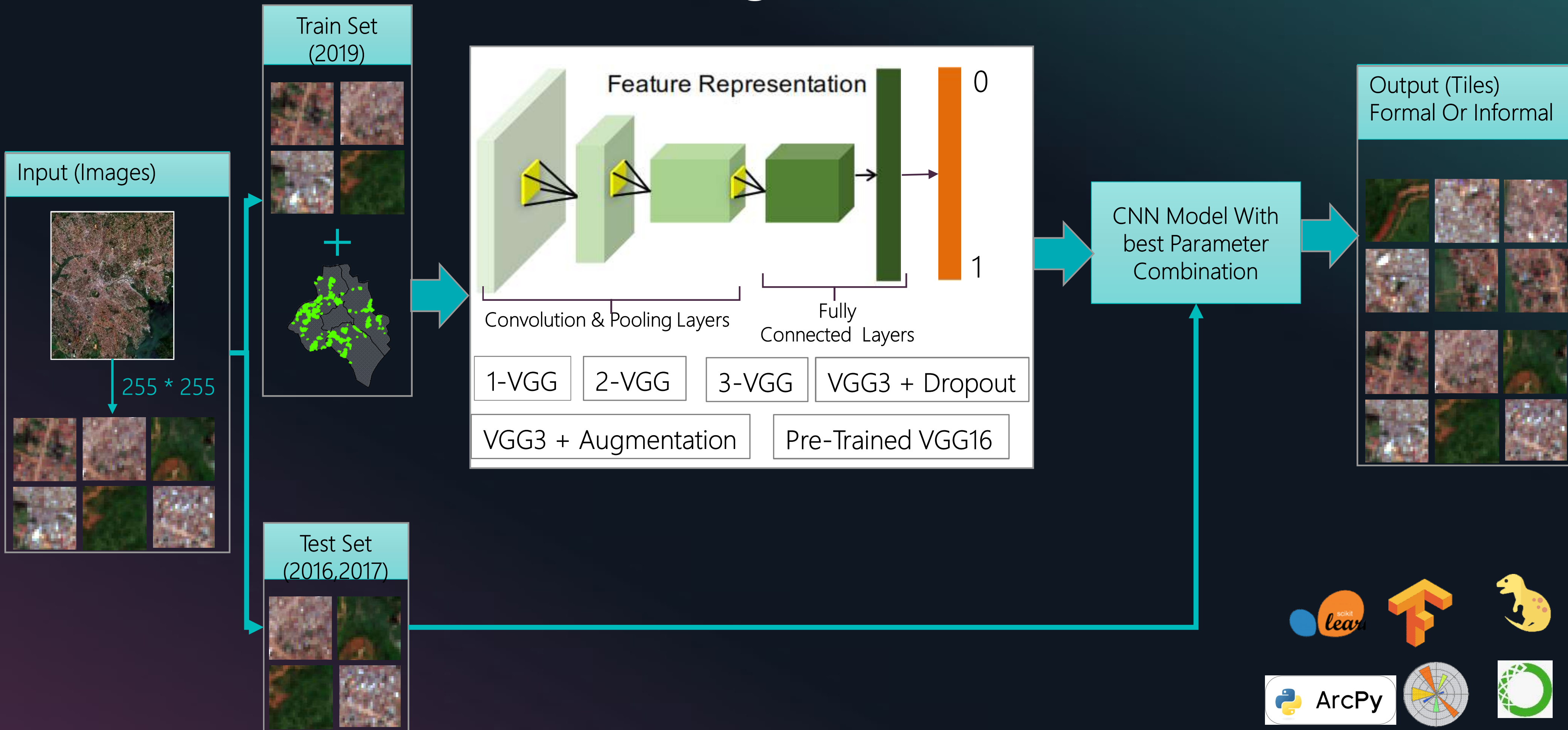


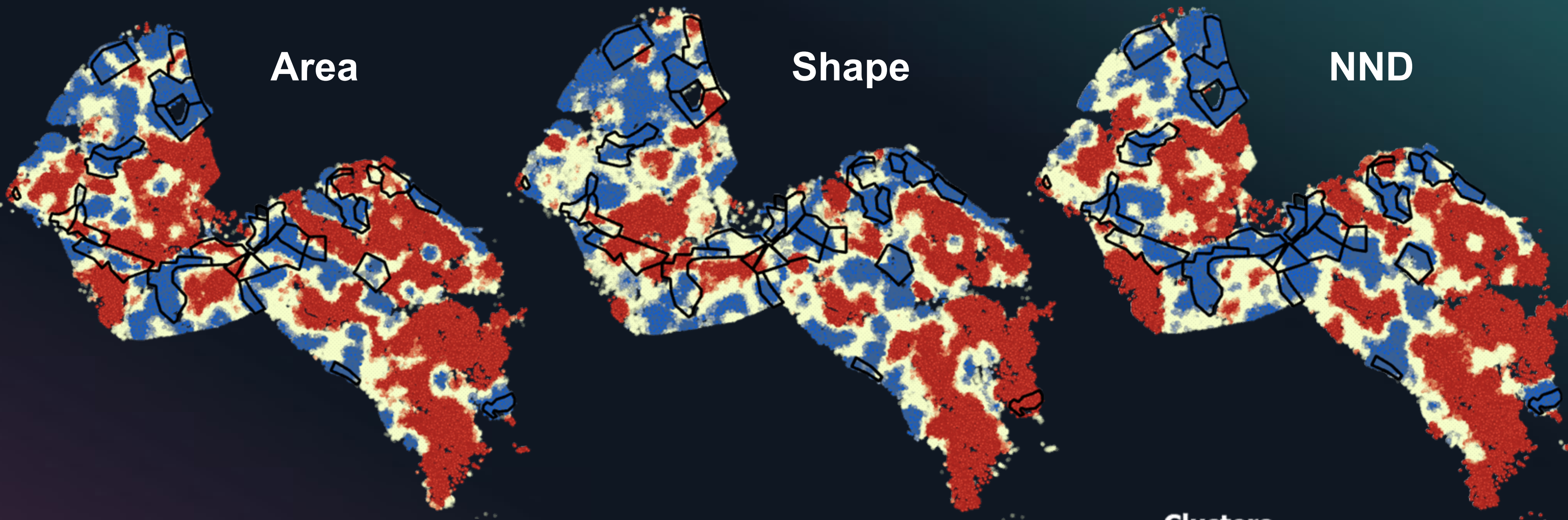
Image Classification
Senitnel-2 Imagery:
1 Jan 2016
23 Aug 2017
31 Dec 2019
Preprocessing and Tiles generation
2019
2016,2017
Training Set
Convolutional Neural Networks (CNNs)
CNN Model With best Parameter Combination
Final Multi-Temporal Maps

Phase 3: Image Classification



RESULTS AND ANALYSIS

Hotspot Clustering Analysis

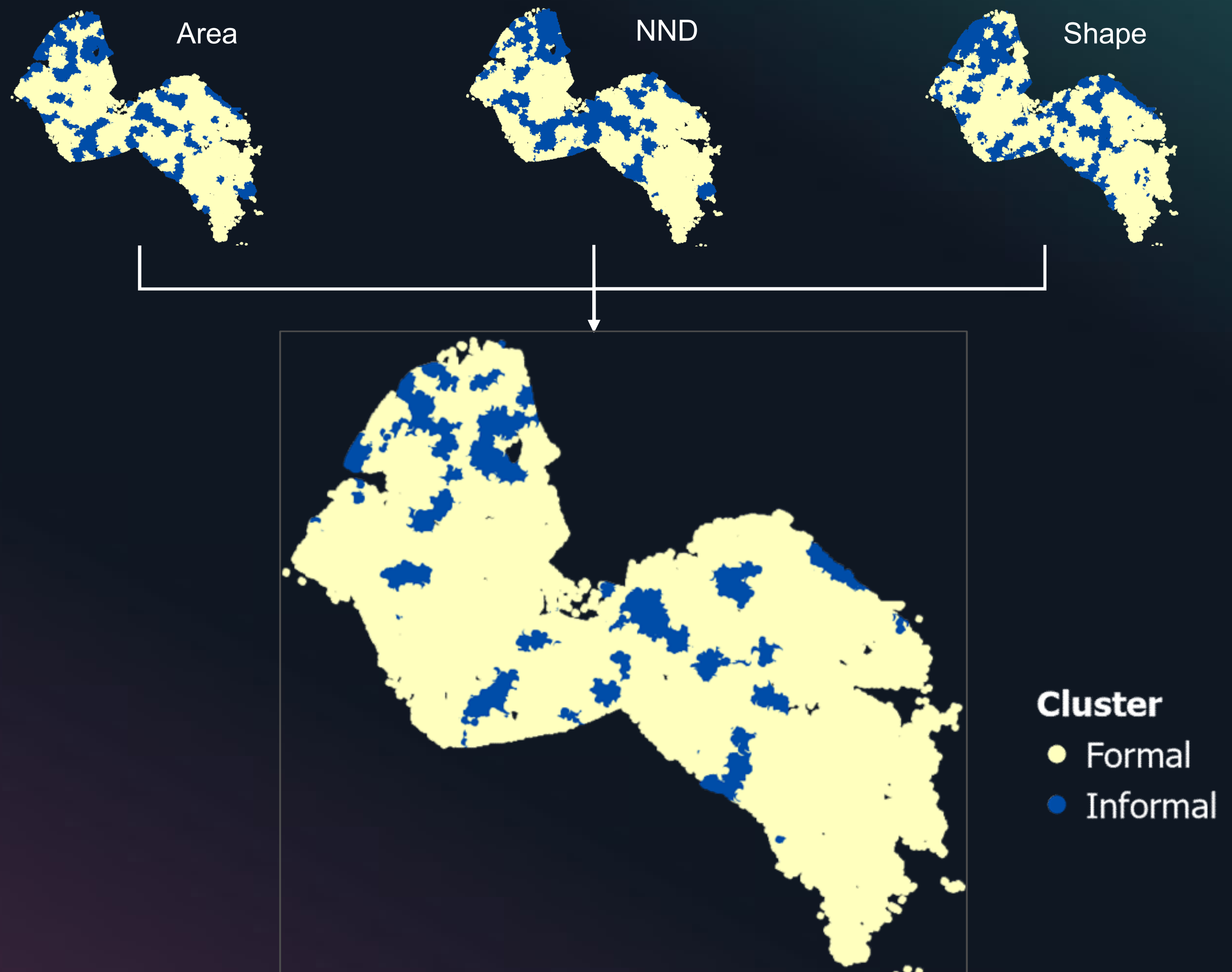


Variable	Moran's Index	Expected Index	Variance	z-score	p-value
NND	0.2576	-0.00002	0.0001	21.8482	0.0000
Area	0.2788	-0.00002	0.0001	23.6968	0.0000
Shape	1.3368	-0.00002	0.0001	113.2856	0.0000

Independent Sample T-Test Analysis

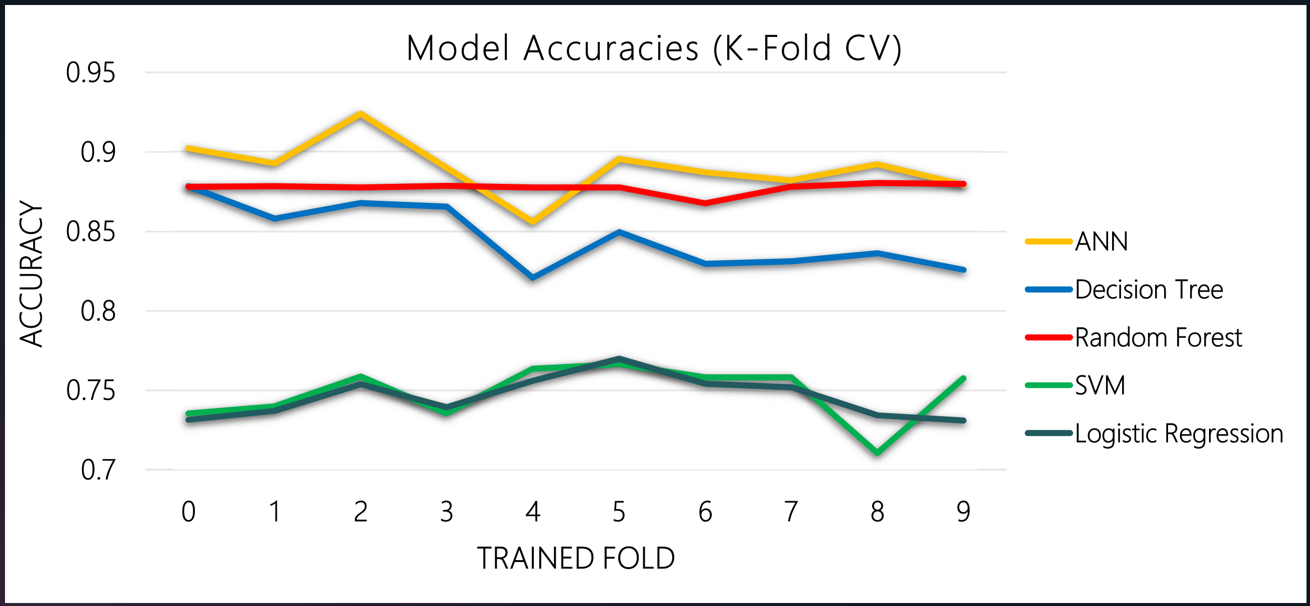
		SIZE(AREA)				
Group		N	Mean	Std. Deviation	Std. Error Mean	Sig. (2-tailed)
GiZScore	ColdSpot(Informal)	104884	-4.4855	2.0226	0.0062	0.000
	HotSpot (Formal)	36868	3.6858	1.8595	0.0096	
Nneighbors	ColdSpot(Informal)	104884	958.00	302.407	0.934	0.000
	HotSpot (Formal)	36868	363.13	129.434	0.674	
		NEAREST NEIGHBOUR DISTANCE (NND)				
Group		N	Mean	Std. Deviation	Std. Error Mean	Sig. (2-tailed)
GiZScore	ColdSpot(Informal)	123467	-5.6911	2.8813	0.0082	0.000
	HotSpot (Formal)	41942	5.1848	3.3074	0.0161	
Nneighbors	ColdSpot(Informal)	123467	936.37	281.848	0.802	0.000
	HotSpot (Formal)	41942	344.97	119.569	0.584	
		SHAPE (VERTEX NUMBER)				
Group		N	Mean	Std. Deviation	Std. Error Mean	Sig. (2-tailed)
GiZScore	ColdSpot(Informal)	129377	-4.8803	1.9625	0.0054	0.000
	HotSpot (Formal)	34552	5.8729	4.6397	0.0249	
Nneighbors	ColdSpot(Informal)	129377	850.26	319.712	0.889	0.000
	HotSpot (Formal)	34552	512.44	286.337	1.540	

Reclassification



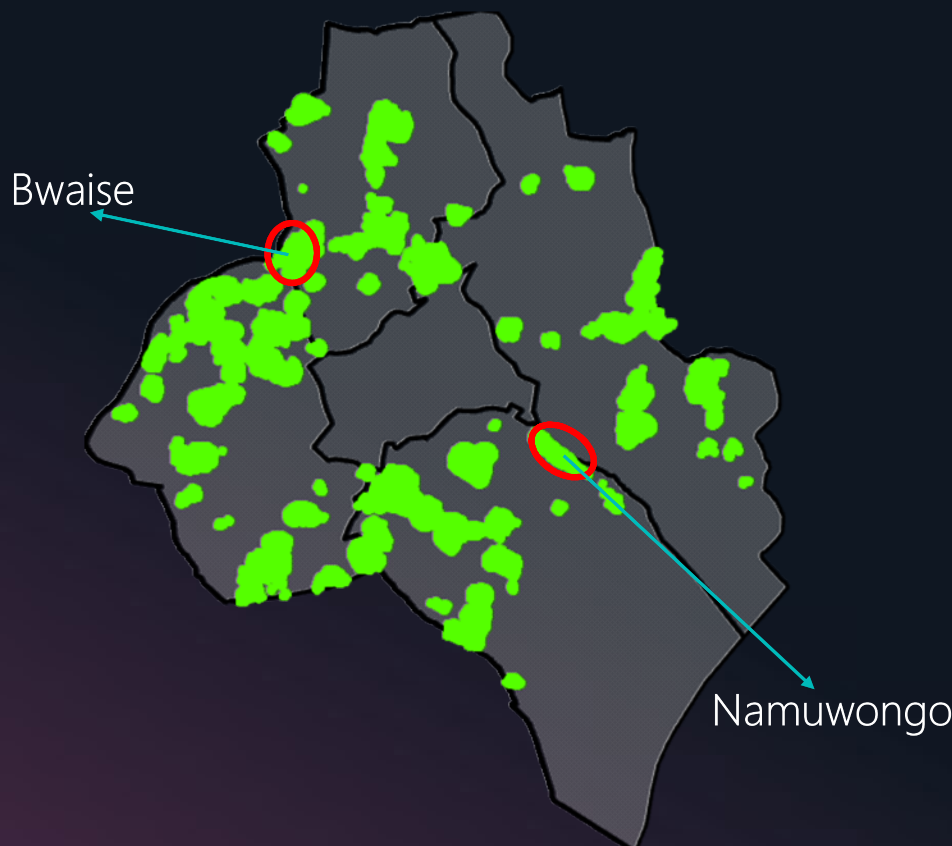
MLTs on Building Characteristics

Classifier	Accuracy	Precision	Recall	F1-Score
Logistical Regression	0.74	0.54	0.74	0.62
Decision Tree	0.84	0.84	0.84	0.79
Random Forest	0.89	0.88	0.89	0.88
ANN(Multilayer Perceptron)	0.81	0.79	0.80	0.77
Support Vector Machine	0.75	0.56	0.75	0.64



Informal Settlements

Kampala



Dar es Salaam

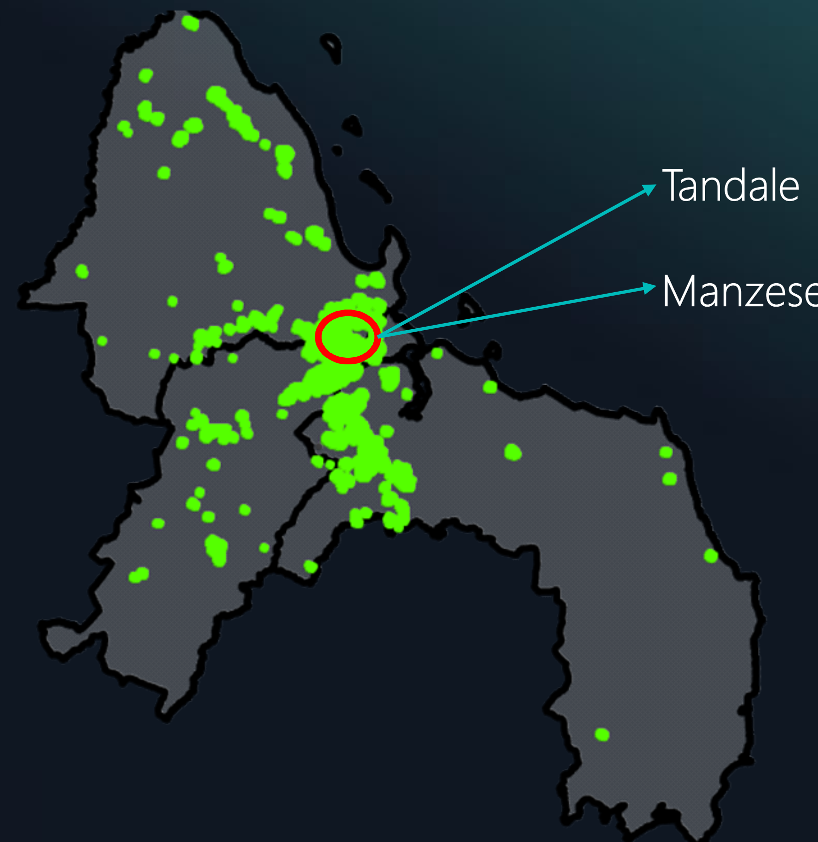
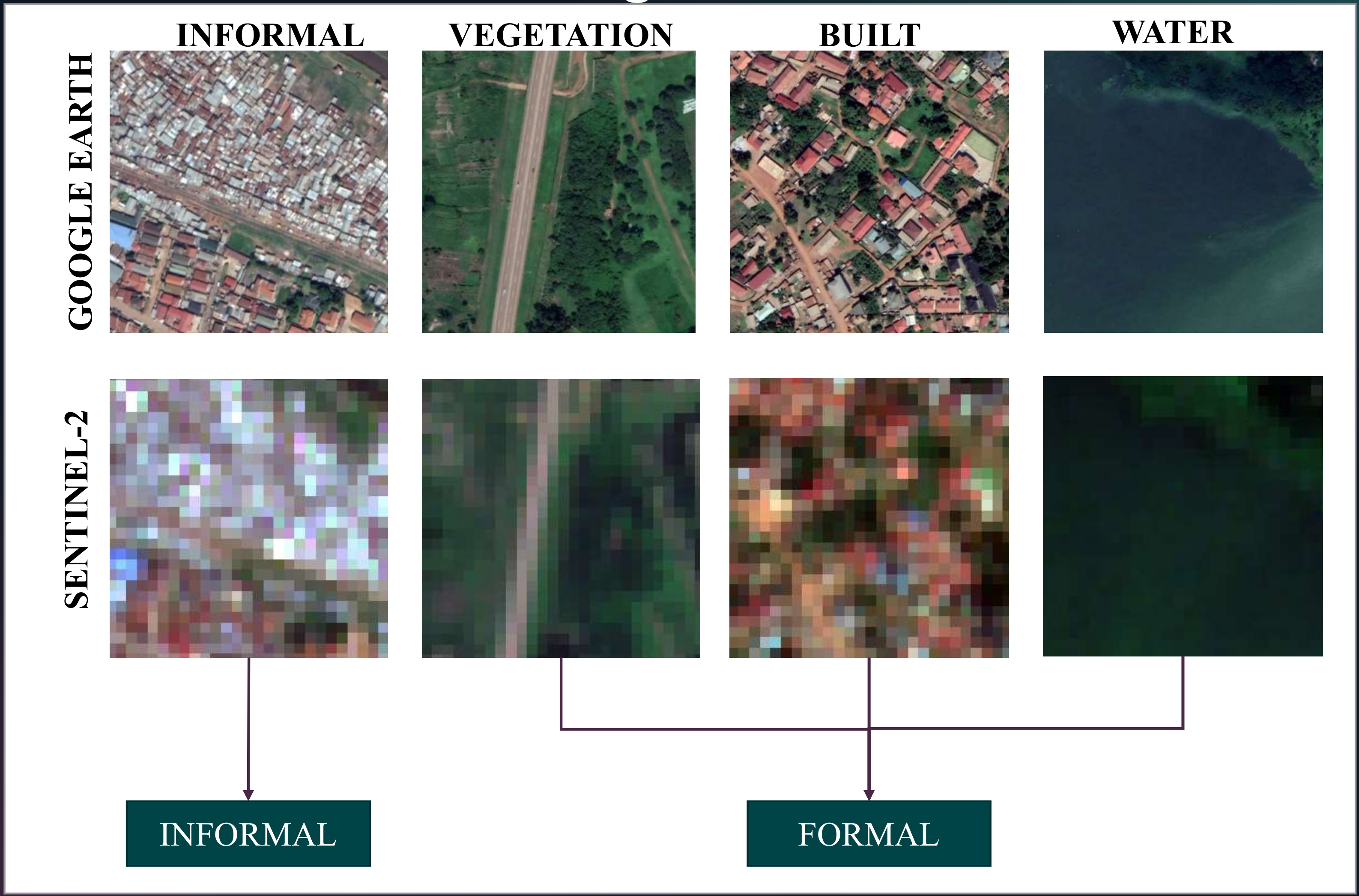


Image Tiles

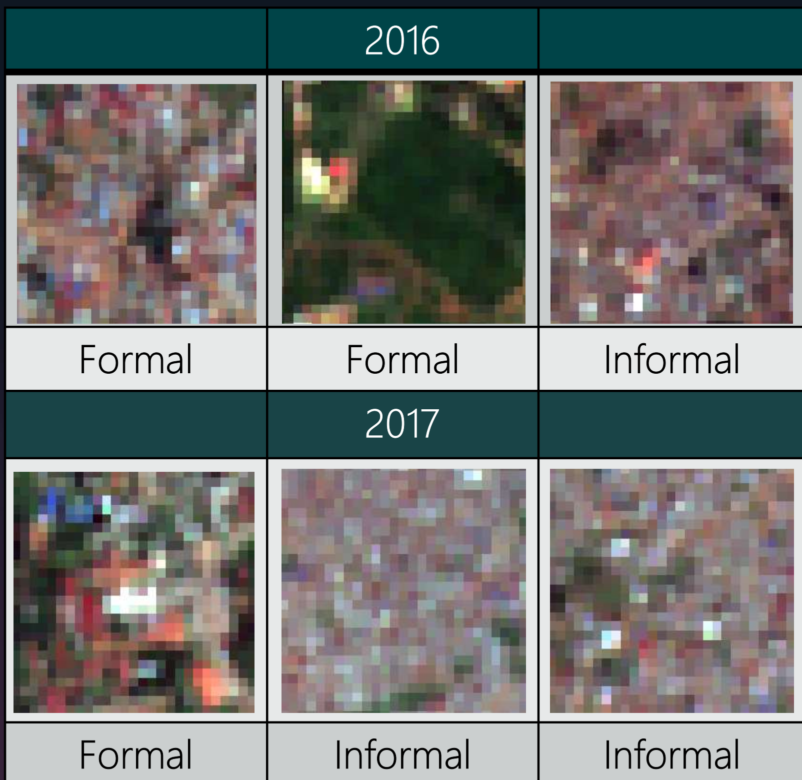


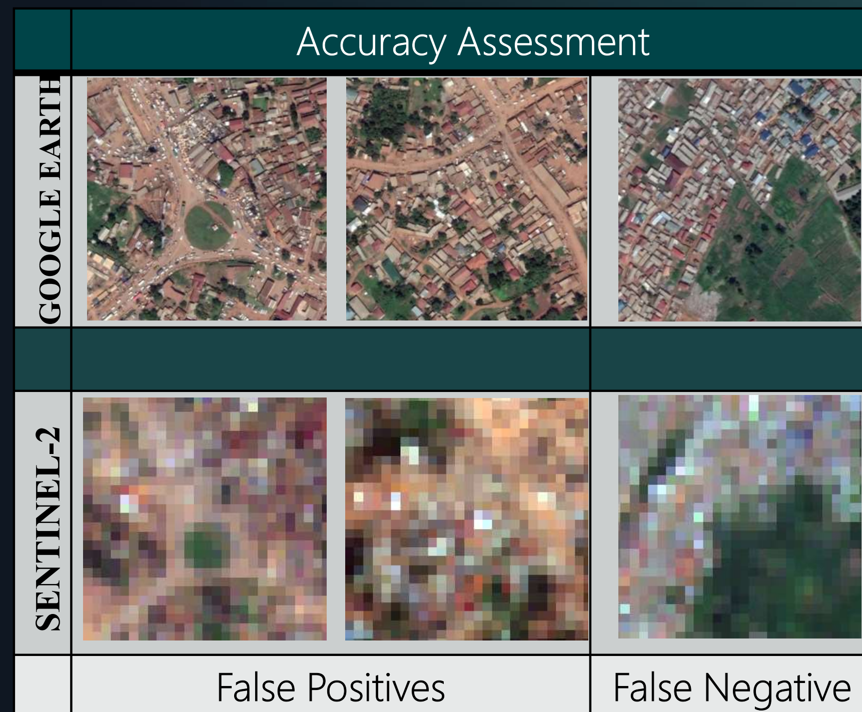
INFORMAL
VEGETATION
BUILT
WATER
GOOGLE EARTH
SENTINEL-2
INFORMAL
FORMAL

CNN Models

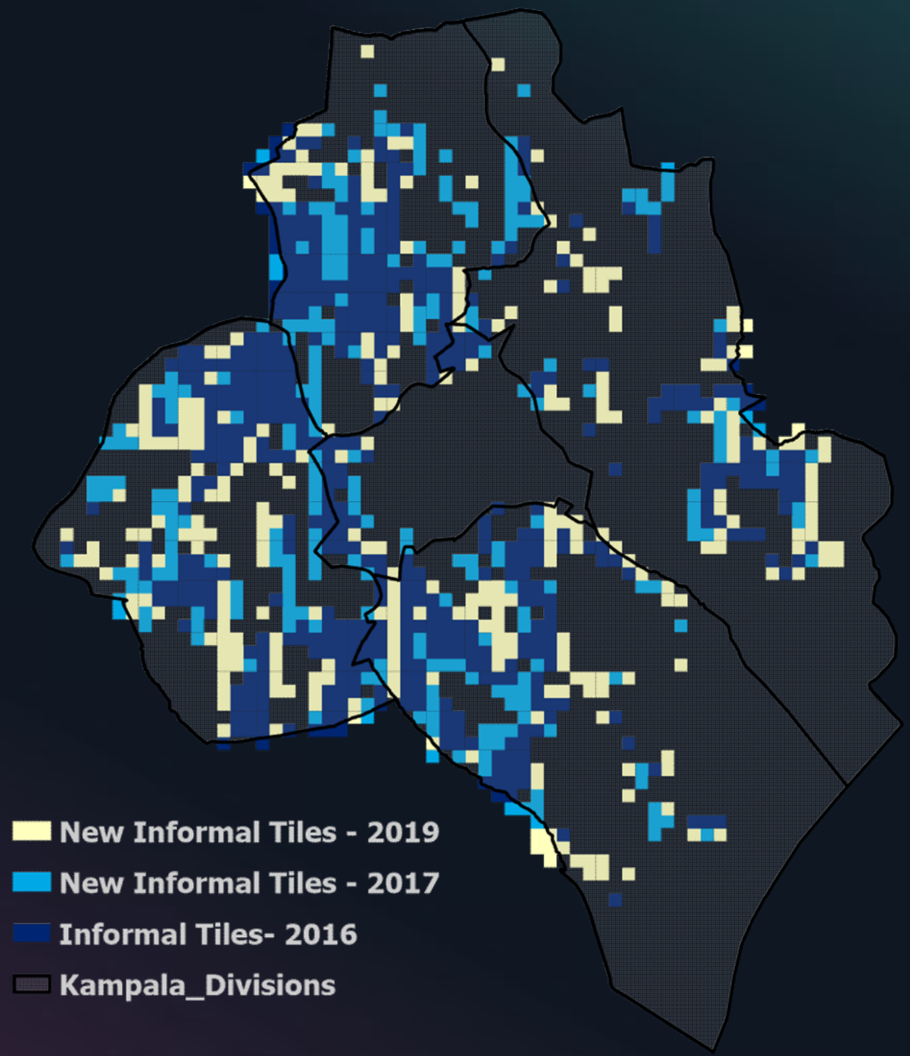
CNN Model	Accuracy (%)	Precision	Recall	f1-score
1-VGG	84.547	0.704	0.70	0.676
2-VGG	84.327	0.6395	0.74	0.6414
3-VGG	86.534	0.7271	0.77	0.6085
Baseline VGG3 + Dropout	80.353	0.774	0.73	0.6965
Baseline VGG3 + Data Augmentation	90.3	0.8214	0.79	0.7545
Pre-Trained VGG16	68.433	0.5915	0.64	0.5117

Image Tiles Classification





CHANGE IN INFORMALITY IN KAMPALA FROM 2016 TO 2019



DISCUSSION & CONCLUSION

Discussion

- 25.25%- Kampala , 19.44%- Dar-es Salaam Buildings are in Informal Settlements
- Award of Legal tenure to slum Dwellers in Tanzania could be Reducing Informality (MKURABITA Programme)
- CNN on sentinel-2A Image tiles gave 90.3%Accuracy vs 87% using Random Forest in (Federico .B et al, 2018)
- Building Outlines vs Road Intersections (Mohamed et al, 2018)

Limitations

- Limited Reference data to Assess Generated Informal Regions
- Limited OSM Buildings data to test Model on Other Cities Outside East Africa

Conclusions

- Hotspot clustering analysis on Building Characteristics can differentiate Informal from formal areas
- The transferability of a Trained classifier to predict informal settlements in new datasets in the same city and in a different city is possible
- Random Forest classifier – For Building Feature Characteristics
- Sentinel-2 image tiles for informal settlement monitoring when coupled with Advanced Machine Learning Algorithms like CNNs.

Future Works

- Combine OSM features (Roads and accessibility + Building Outlines)
- Digital Surface Model (DSM)


THANK YOU!

pi4

References

- “Aerial view of Makoko Slum in Lagos Nigeria.png - Wikimedia Commons.”

https://commons.wikimedia.org/wiki/File:Aerial_view_of_Makoko_Slum_in_Lagos_Nigeria.png

- “The 20 Worst Slums in Africa.”

http://www.africaranking.com/20-worst-slums-in-africa/6

- http://www.thisisplace.org/i/?id=bde34078-ccb7-42c7-b949-b9139d65b47c
- http://www.thisisplace.org/i/?id=221ac05d-2ab5-474e-95fc-814a0c1fd48b
- https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanizationprospects.html. [Accessed: 28-May-2019].
- “Introduction to urban poverty | International Institute for Environment and Development.” [Online]. Available: https://www.iied.org/introduction-urban-poverty. [Accessed: 28-May-2019].
- Ibrahim MR, Titheridge H, Cheng T, Haworth J. predictSLUMS: A new model for identifying and predicting informal settlements and slums in cities from street intersections using machine learning. Comput Environ Urban Syst. 2019;76:31–56.
- Detecting informal settlements in South America: How I built it [Internet]. [cited 2020 Feb 20]. Available from: https://blog.mapbox.com/detecting-informal-settlements-in-south-america-how-i-built-it-cb139a870816
- Gram-Hansen BJ, Azam F, Helber P, Coca-Castro A, Bilinski P, Varatharajan I, et al. Mapping informal settlements in developing countries using machine learning and low resolution multi-spectral data. AIES 2019 - Proc 2019 AAAI/ACM Conf AI, Ethics, Soc. 2019;361–8.