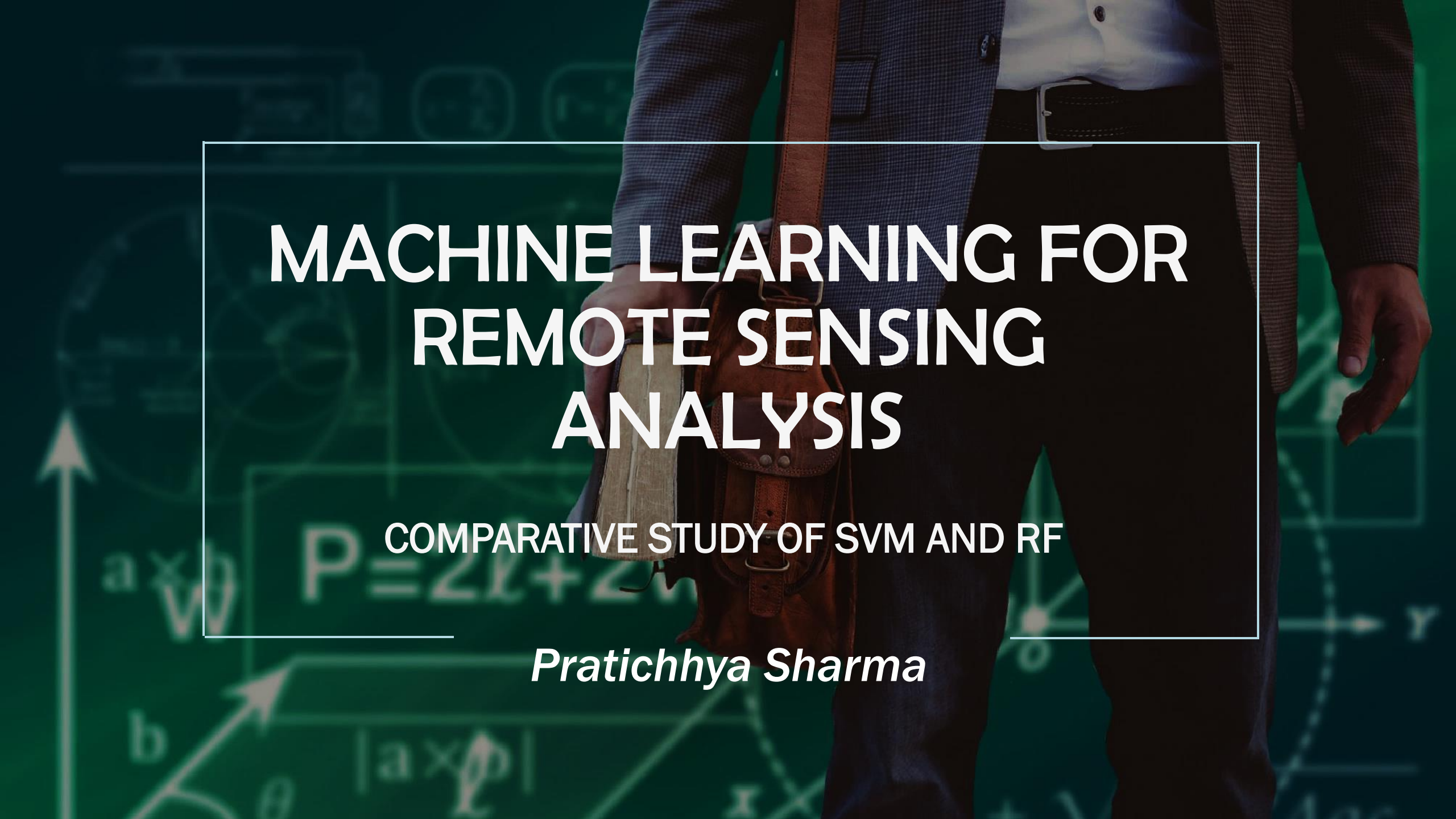
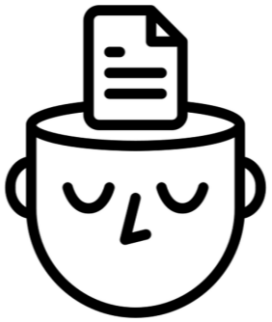


MACHINE LEARNING FOR REMOTE SENSING ANALYSIS

COMPARATIVE STUDY OF SVM AND RF

Pratichhya Sharma

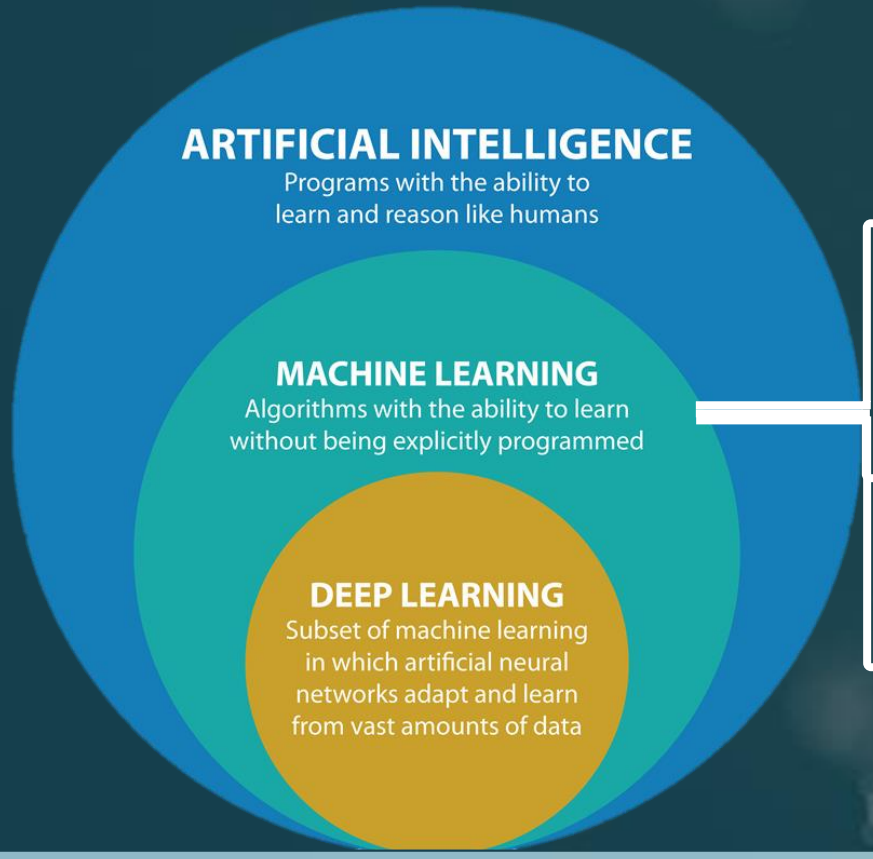
Context



1. Machine Learning
2. Support Vector Machine (SVM)
3. Random Forest
4. Comparative Case Study
5. Conclusion

MACHINE LEARNING

Early definition of Machine Learning by Arthur Samuel: "***Field of Study that gives computer ability to learn without being explicitly programmed.***"



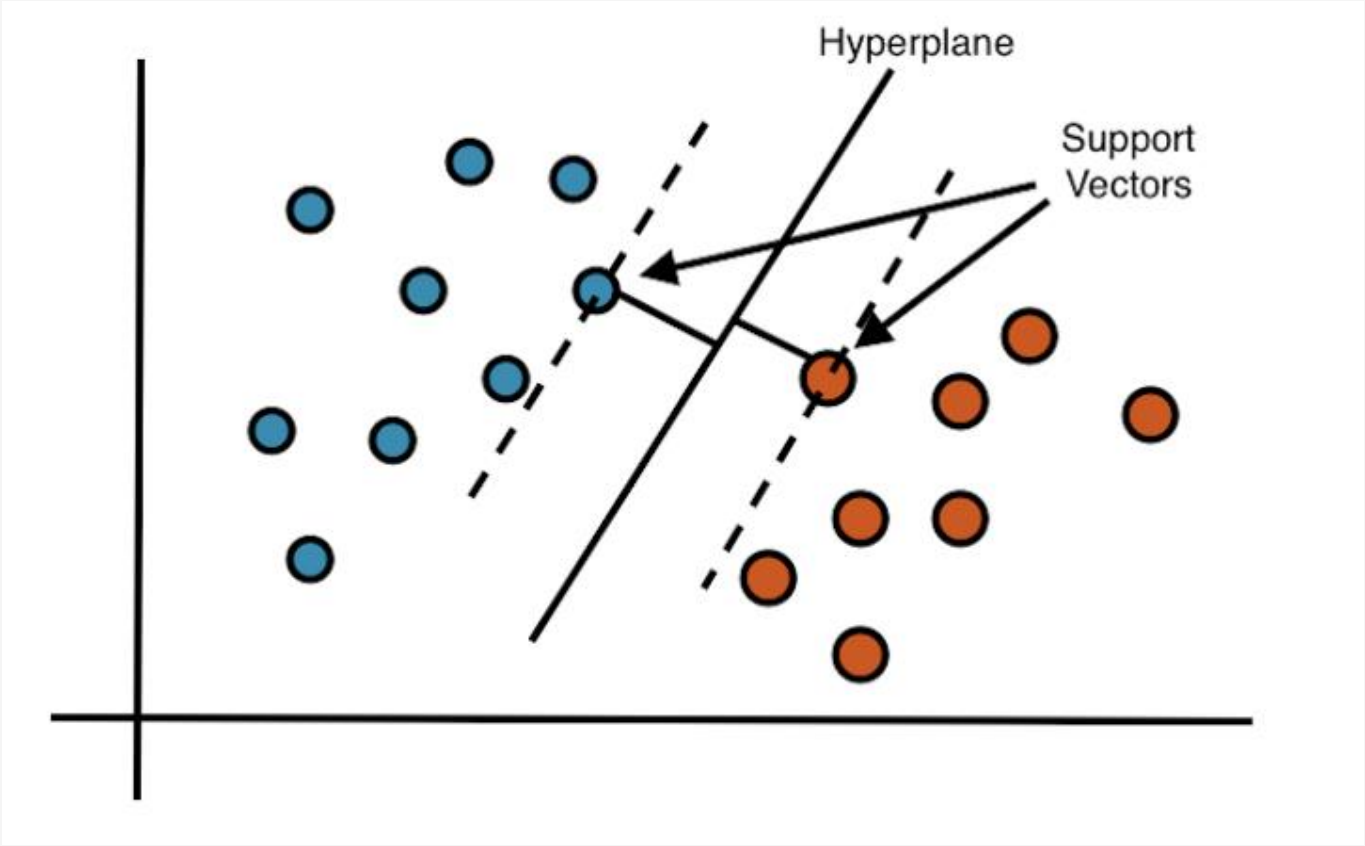
- **Supervised Learning** (Task Driven)
  - Support Vector Machine
  - Random Forest
- **Unsupervised Learning** (Data Driven)
- **Reinforcement Learning** (Learn from errors)

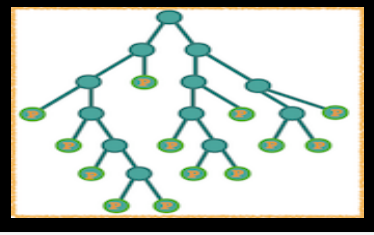
Support Vector Machine

S - Support refers to the extreme values/points in your dataset.

V - Vector refers to the values/points in dataset / feature space.

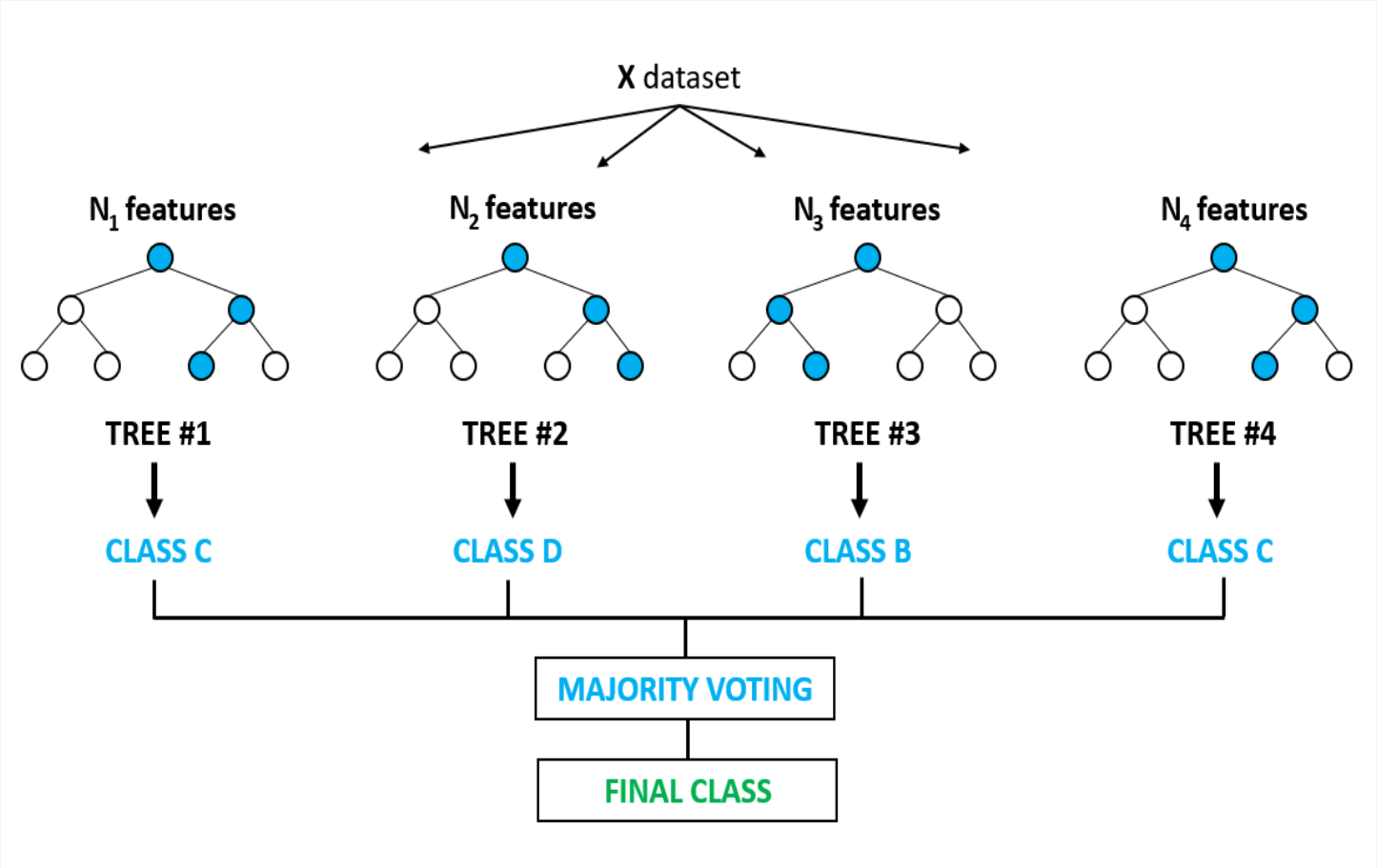
M - Machine refers to the machine learning algorithm that focuses on the support vectors to classify groups of data. This algorithm literally only focuses on the extreme points and ignores the rest of the data.





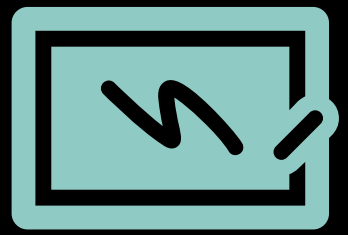
Random Forest

- **Random Forest Algorithm** is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement.
- The trees protect each other from their individual errors resulting in highly accurate result.



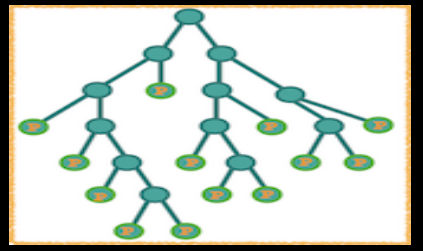
Strength and Weakness

Support Vector Machine		Random Forest	
Strength	Weakness	Strength	Weakness
• Can model complex dimensions; • Less memory space	• Requires lots of processing power; • SVM is not suitable for large datasets. • SVM is not suitable for imbalanced datasets.	• Gives good accuracy results; • It has automatic feature selection; • Can handle missing data and imbalanced classes.	• It's hard to interpret and won't perform well if a bad sets of features are given. • It is slow in generating predictions because it has multiple decision trees.

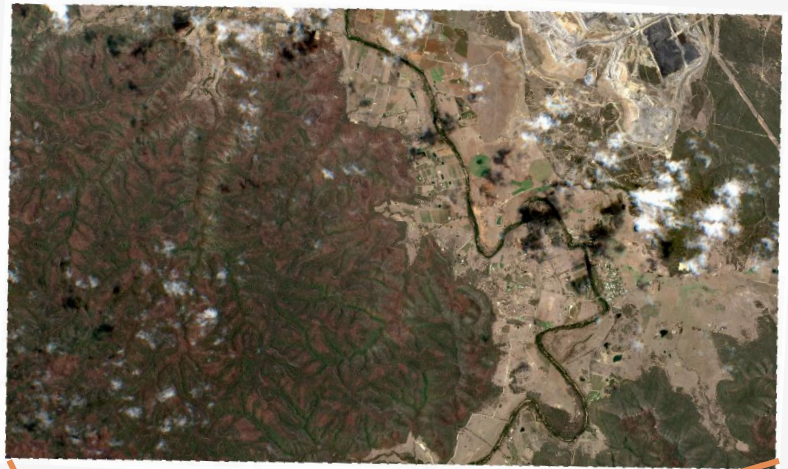


A Comparative Case Study on use of SVM and RF for Burnt Area Mapping in Southern Australia using Sentinel 2 Imagery

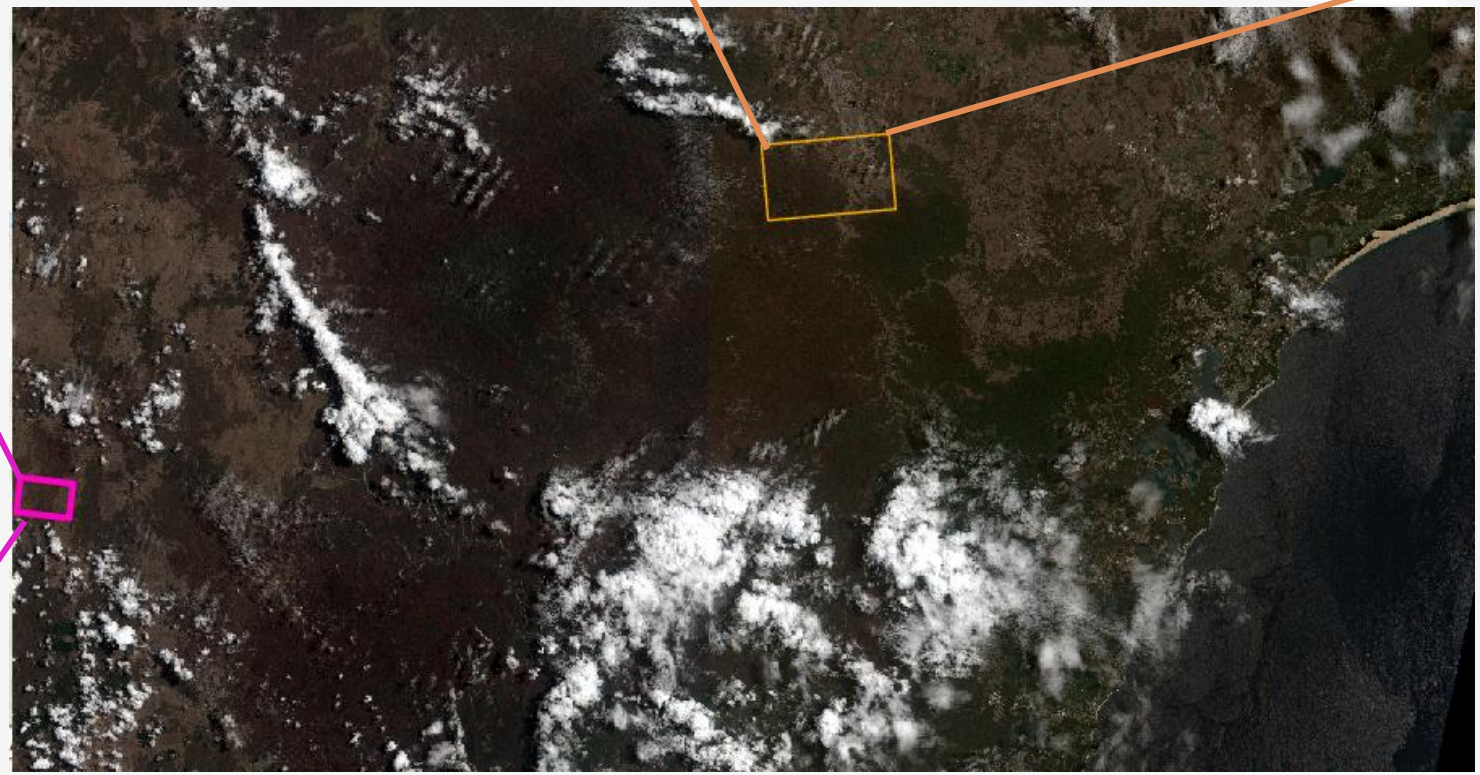



Dataset

Imagery: **Sentinel 2A**
Used as: **Training, Validation and Testing sets**
Date of Acquisition: 2020-01-20
Source: https://scihub.copernicus.eu/dhus/#/home

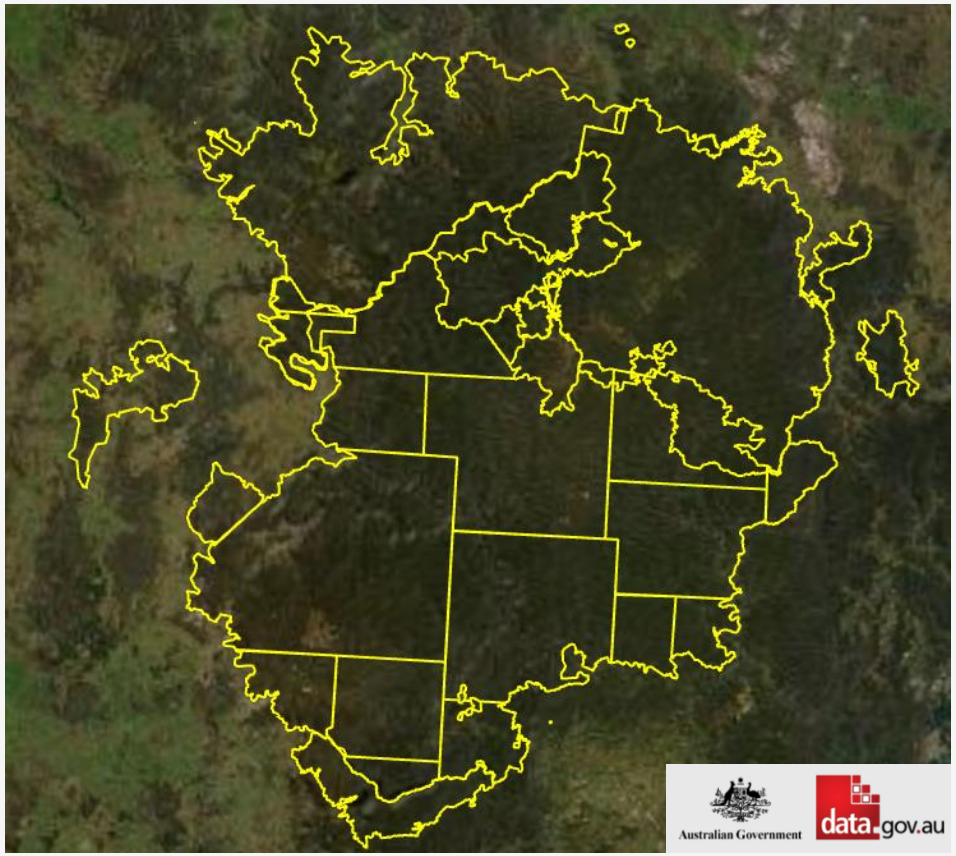


Imagery: **Sentinel 2A**
Used as : **Inferencing** over different area
Date of Acquisition: 2020-01-20
Source: https://scihub.copernicus.eu/dhus/#/home



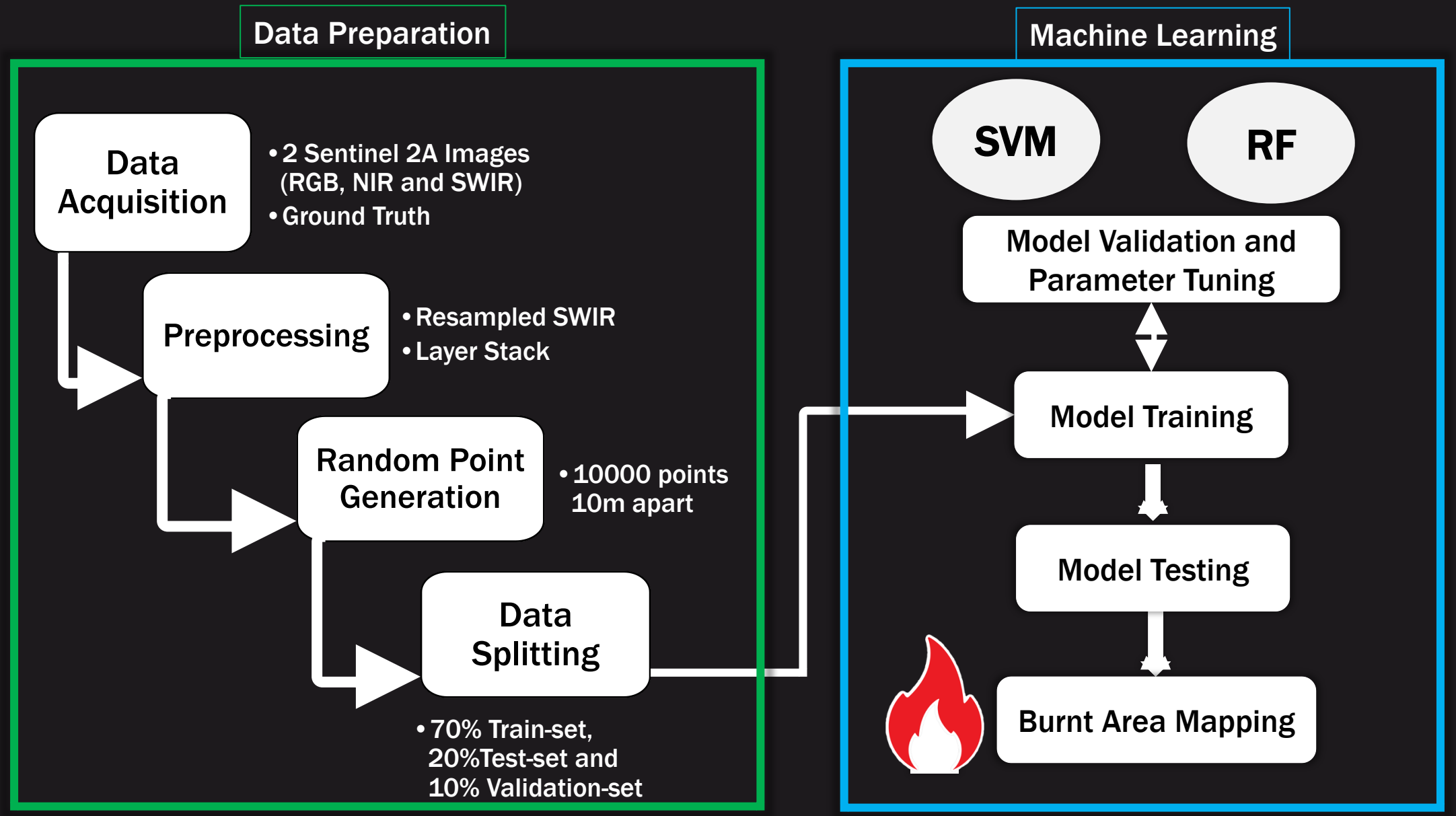
Dataset

Shapefile: **Ground Truth**
Date of Acquisition: 2020-01-15
Source: https://data.gov.au/data/dataset/201920fy-bushfire-boundaries

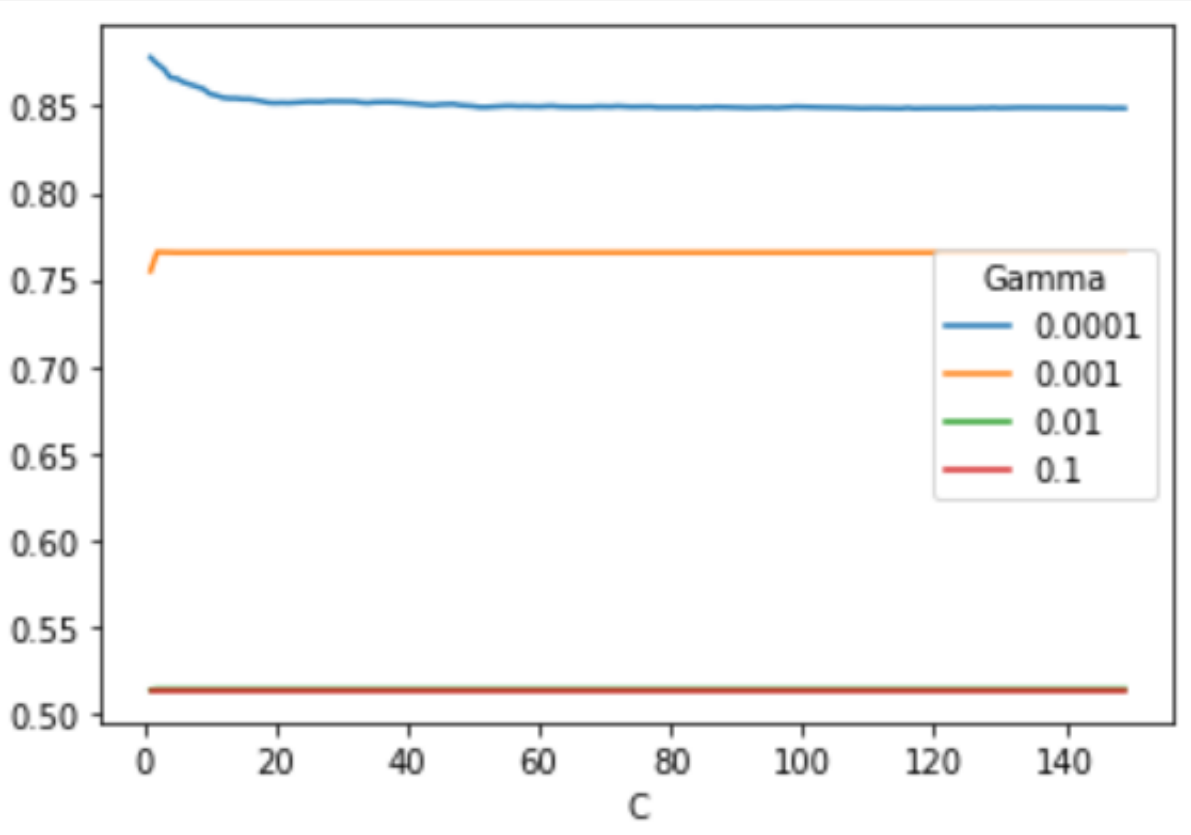


Field	Value
Title	2019-20 Financial Year Bushfire Boundaries
Type	Dataset
Language	English
Licence	Creative Commons Attribution 4.0 International
Data Status	active
Update Frequency	never
Landing Page	https://data.gov.au/data/dataset/f91c2ffd-a0a0-4042-a8ac-71757a6ef727
Date Published	2020-07-01
Date Updated	2020-07-15
Contact Point	National Bushfire Recovery Agency +61 2 6228 6300 reporting@bushfirerecovery.gov.au
Temporal Coverage	2019-07-01 - 2020-06-30
Geospatial Coverage	Australia
Jurisdiction	Commonwealth of Australia
Data Portal	data.gov.au
Publisher/Agency	National Bushfire Recovery Agency
Geospatial Topics	Boundaries

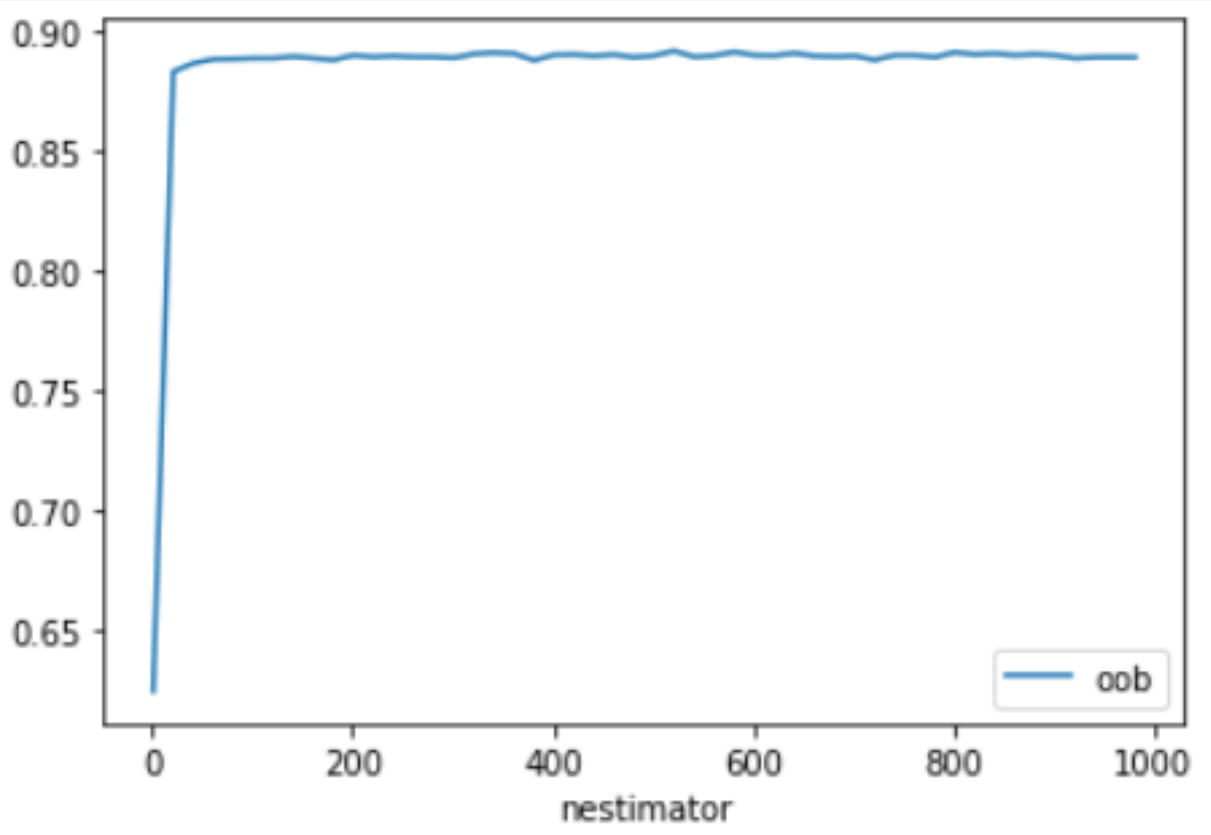
METHODOLOGY



Parameter Tuning



SVM with varying C (1,150,1) & Gamma

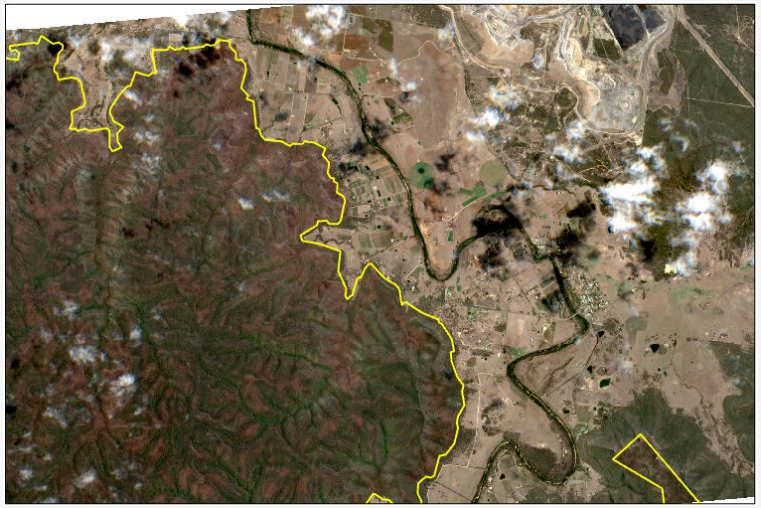


Random Forest with varying n-estimator (1,1001,20)

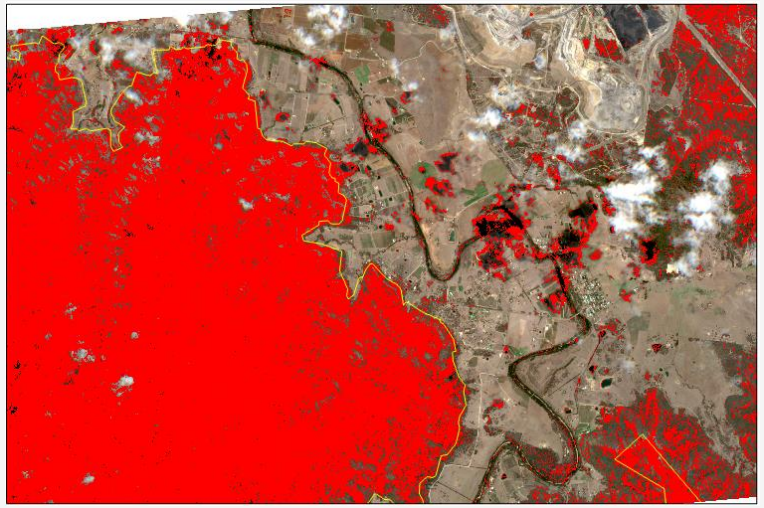
Model Evaluation

Method	Parameters	Overall Accuracy	F1 Score	Precision	Recall
SVM (Linear)	Default	84.59%	0.8271	0.8271	0.8581
SVM (RBF)	Default	88.09 %	0.9236	0.9236	0.8196
SVM (RBF)	C=1 Gamma= 0.0001	88.34%	0.8637	0.8637	0.8988
Random Forest	Default	89.34%	0.9011	0.9011	0.8738
Random Forest	n_estimator = 521	89.74 %	0.9087	0.9087	0.8717

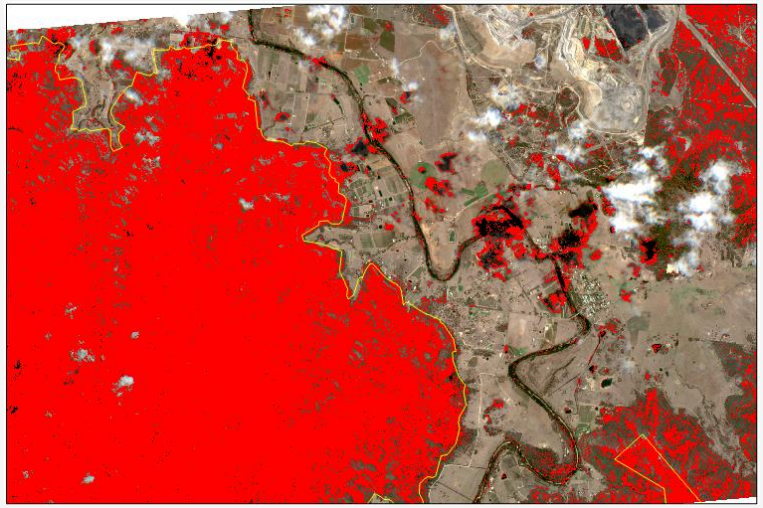
Testing SVM and RF result



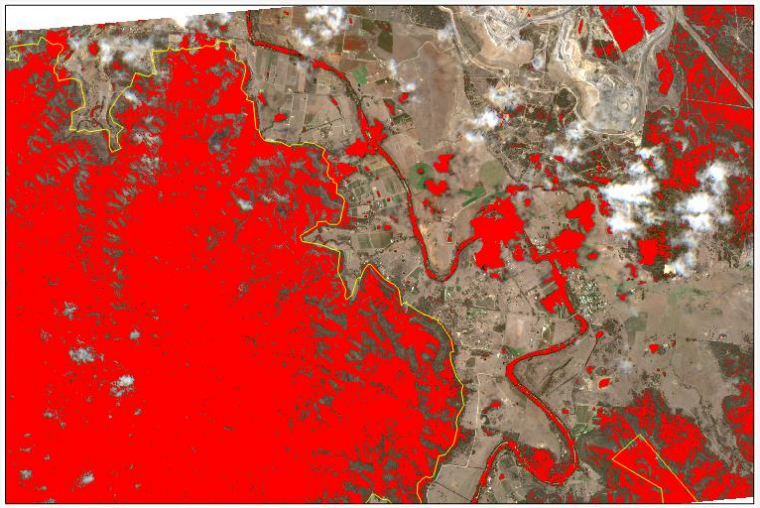
Reference Imagery



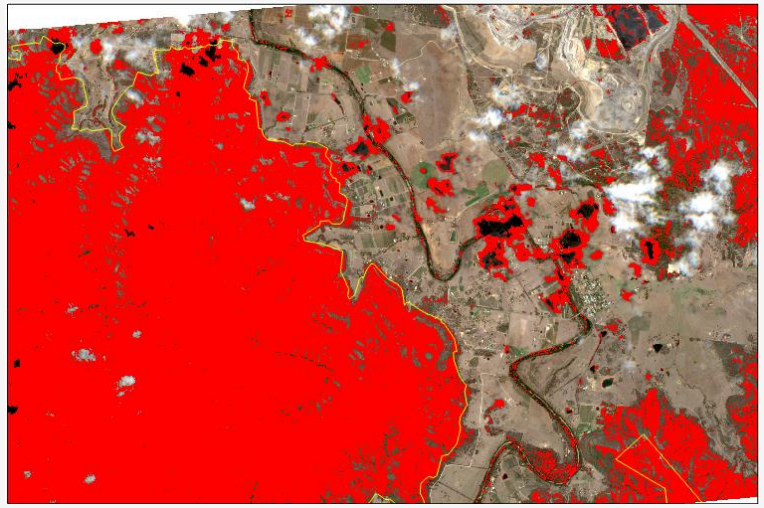
Random Forest



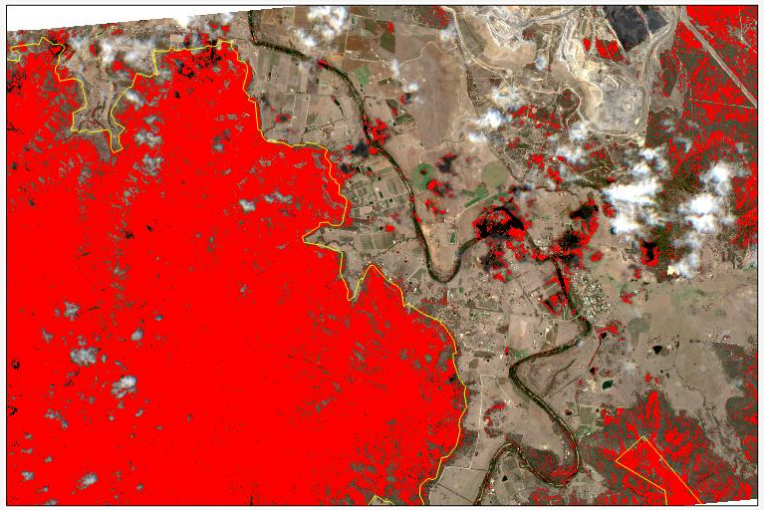
Random Forest (n_estimator=521)



SVM (Linear Kernel)

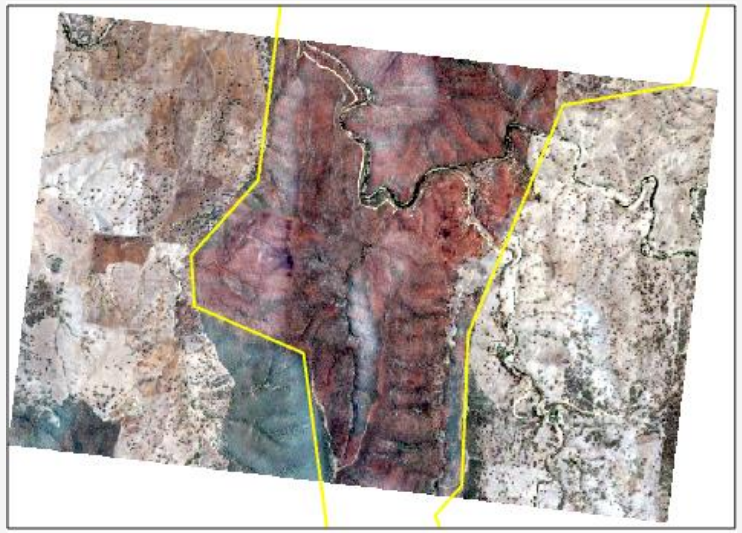


SVM (RBF Kernel)

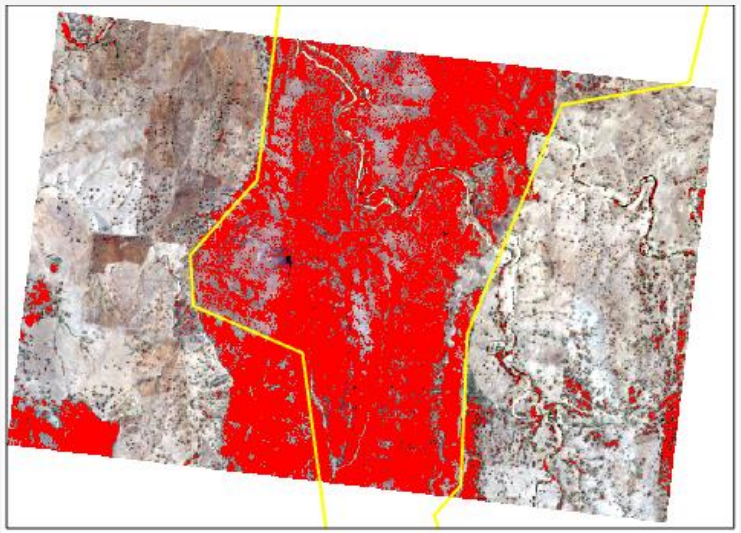


SVM (RBF(c=1, gamma=0.0001))

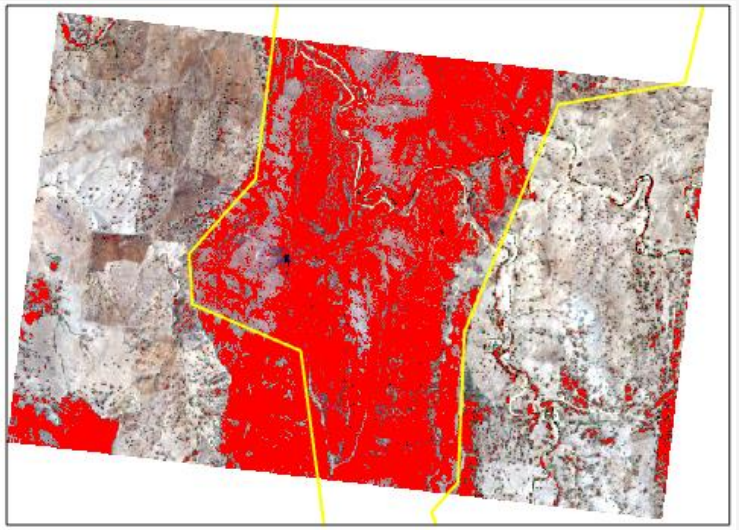
Testing SVM and RF result over different area



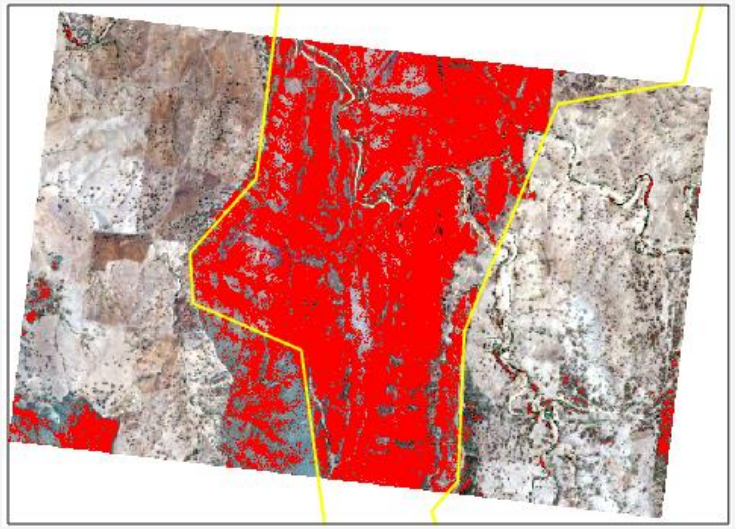
Reference Imagery



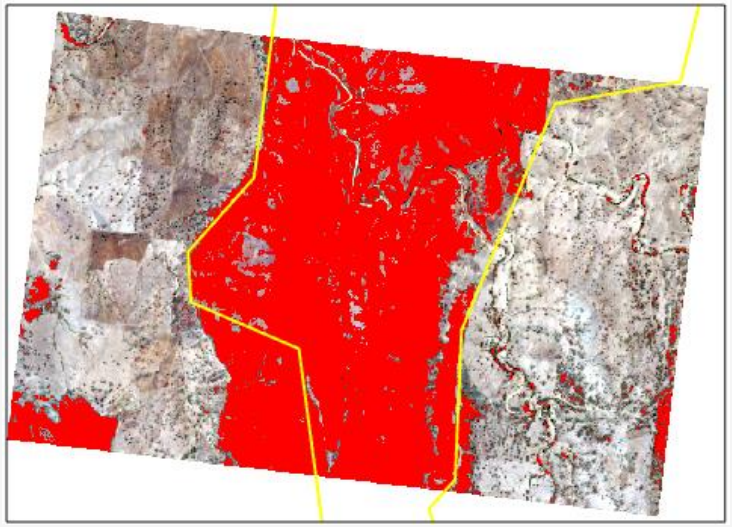
Random Forest



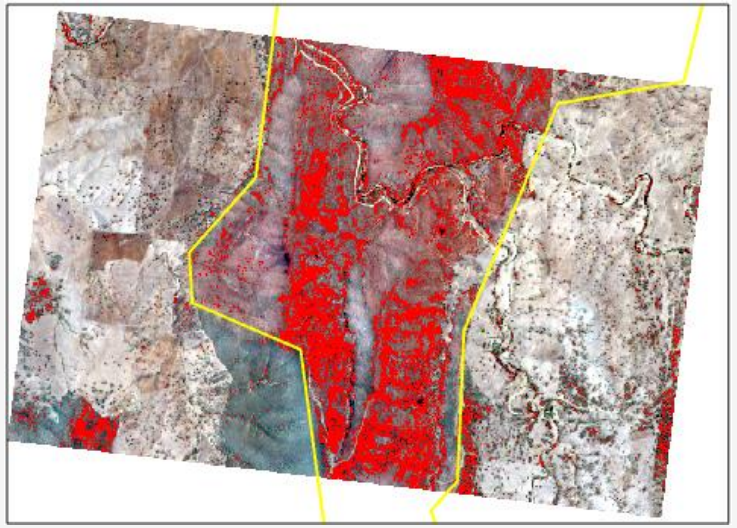
Random Forest (n_estimator=521)



SVM (Linear Kernel)



SVM (RBF Kernel)



SVM (RBF(c=1, gamma=0.0001))

Conclusion

- Each algorithm is different and is tailored based on the data available, the context of the domain problem, and other external/internal constraints.
- If you want to Classify dataset with extreme values and outliers for classification problems, SVM algorithm is the way to go.
- If you have fairly balanced dataset and want to perform classification then Random Forest could be a better solution.
- My model could be better improved by getting more training examples.

References

Belgiu, Mariana, and Lucian Drăgu. 2016. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 114:24–31.

Blaschke, T. 2010. “Object Based Image Analysis for Remote Sensing.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1):2–16.

Mitra, Pabitra, B. Uma Shankar, and Sankar K. Pal. 2004. “Segmentation of Multispectral Remote Sensing Images Using Active Support Vector Machines.” *Pattern Recognition Letters* 25(9):1067–74. doi: 10.1016/j.patrec.2004.03.004.

Salehi, Bahram, Yun Zhang, Ming Zhong, and Vivek Dey. 2012. “Object-Based Classification of Urban Areas Using VHR Imagery and Height Points Ancillary Data.” *Remote Sensing* 4(8):2256–76. doi: 10.3390/rs4082256.

Zhang, Hui, Jason E. Fritts, and Sally A. Goldman. 2008. “Image Segmentation Evaluation: A Survey of Unsupervised Methods.” *Computer Vision and Image Understanding* 110(2):260–80. doi: 10.1016/j.cviu.2007.08.003