Music and Bands. Optimising Sentinel-2 Feature Space for Improved Crop Biophysical and Biochemical Variables Retrieval Using the Novel Spectral Triad feature selection Algorithm

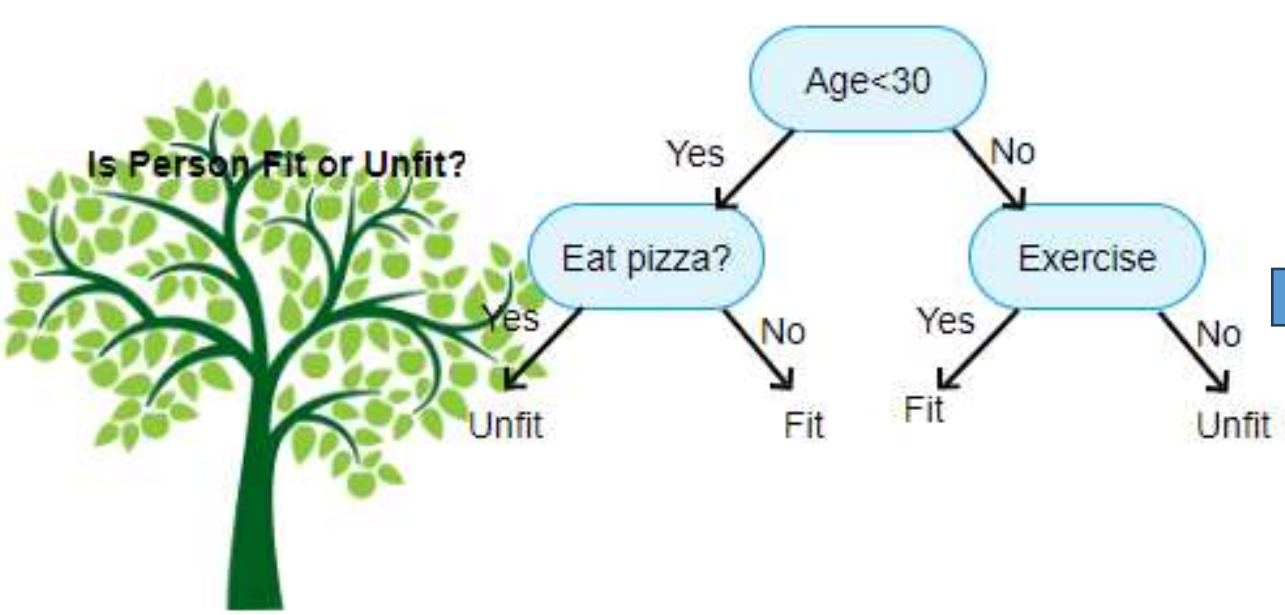
National Space Conference - 30 August – 1 September 2023

UNIVERSITY OF JOHANNESBURG

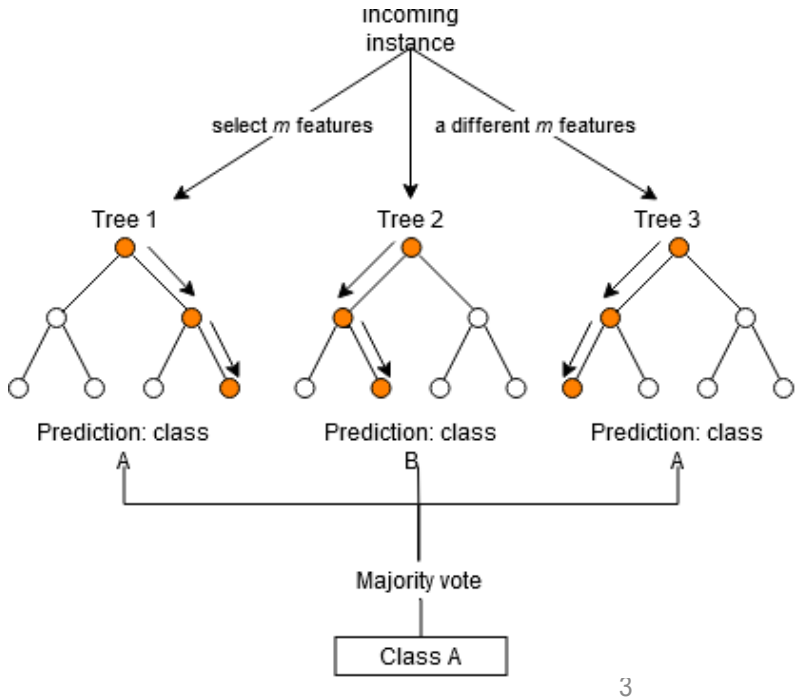
Contents

Background

Decision Trees



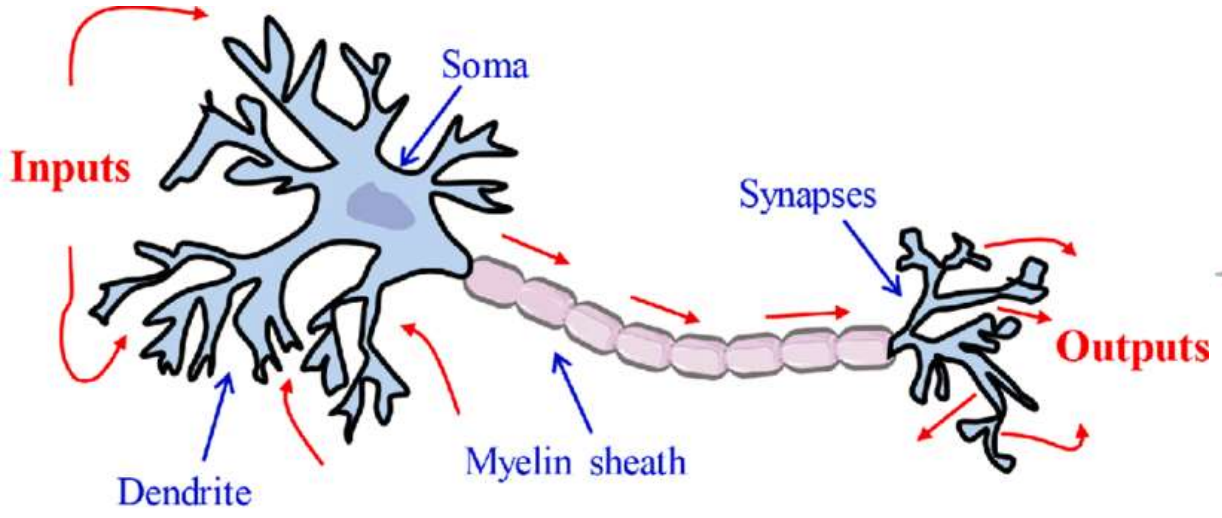
Random Forest



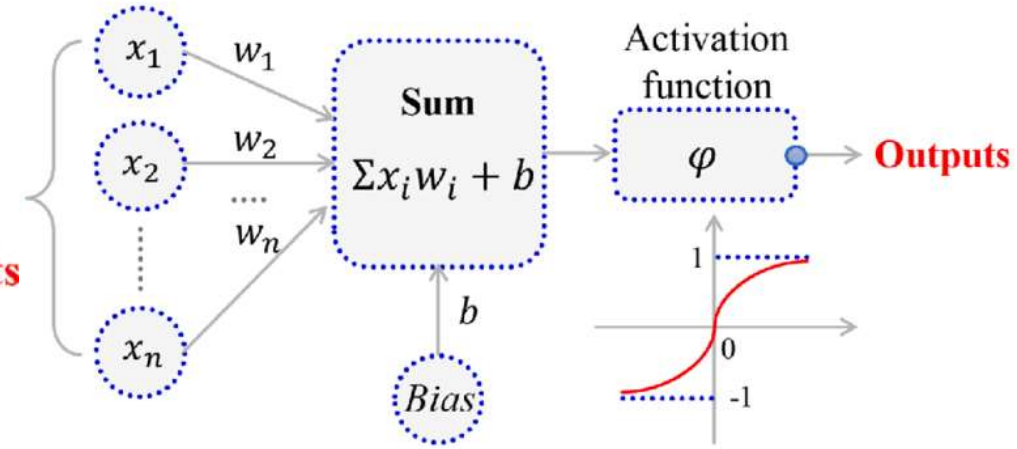
https://iprathore71.medium.com/complete-guide-to-decision-tree-cee0238128d

Background…

Biological Neuron



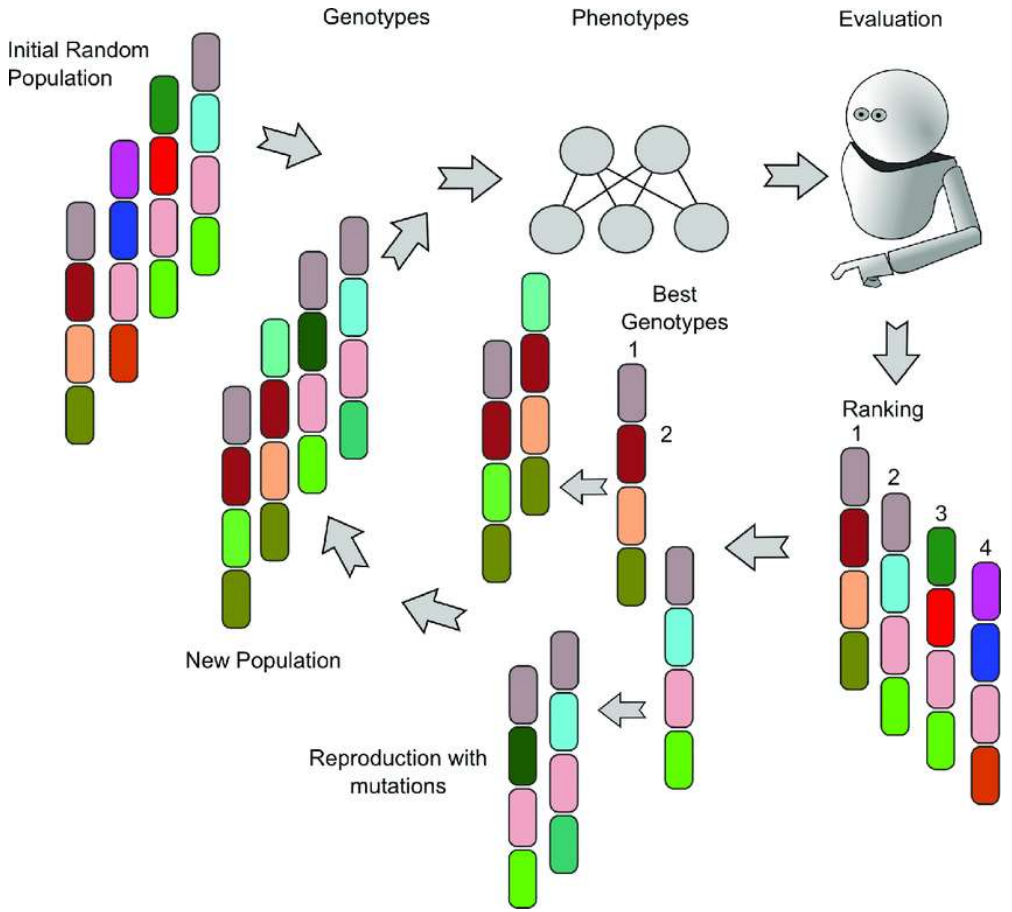
Artificial Neuron

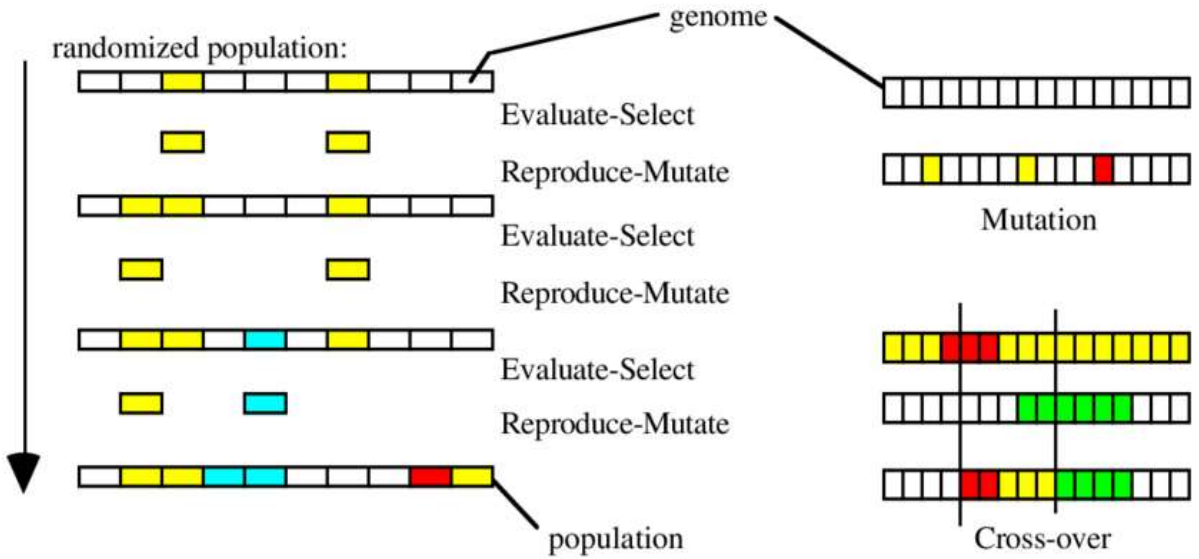


https://www.xenonstack.com/blog/artificial-neural-network-applications

Background…

Genetic Algorithms





Objectives

- To optimize the retrieval of biophysical and biochemical variables of agricultural crops using feature selection techniques by identify and select relevant remotely sensed covariates from Sentinel-2 data.

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Spectral band selection for vegetation properties retrieval using Gaussian processes regression

Jochem Verrelst[a,*], Juan Pablo Rivera[a,b], Anatoly Gitelson[c], Jesus Delegido[a], José Moreno[a], Gustau Camps-Valls[a]

[a] Image Processing Laboratory (IPL), Parc Científic, Universitat de València, 46980 Paterna, València, Spain
[b] Departamento de Oceanografía Física, CICESE, 22860 Ensenada, Mexico
[c] Israel Institute of Technology, Technion, Haifa, Israel

EVALUATING MACHINE LEARNING REGRESSION ALGORITHMS FOR OPERATIONAL RETRIEVAL OF BIOPHYSICAL PARAMETERS: OPPORTUNITIES FOR SENTINEL

Jochem Verrelst[1], J.P. Rivera[1], L. Alonso[1], L. Guanter[2], and J. Moreno[1]

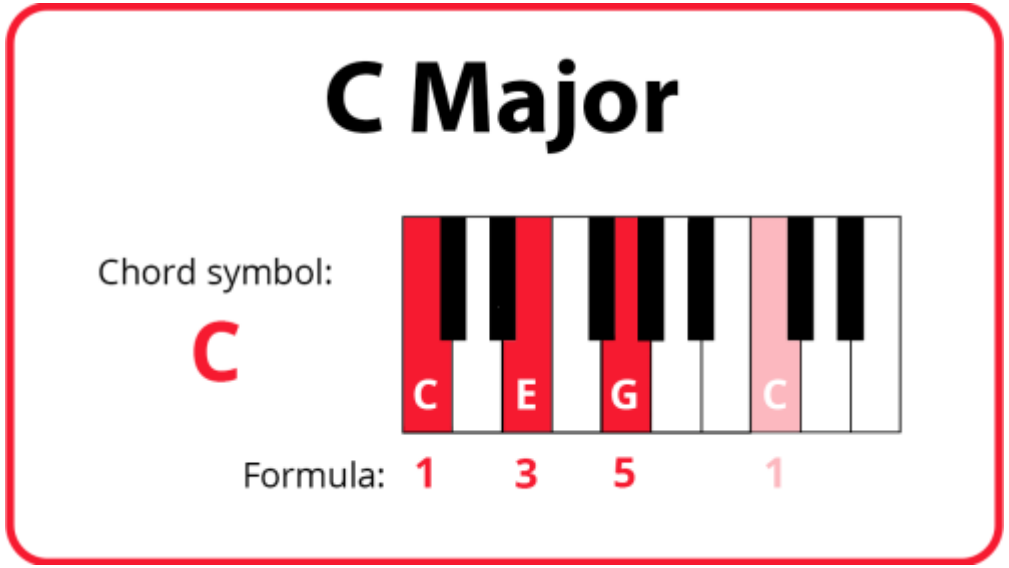
[1] Image Processing Laboratory (IPL), University of Valencia, Spain, jochem.verrelst@uv.es
[2] Institute for Space Sciences, Free University of Berlin, Germany

Objectives…

- To achieve this, we propose a novel Spectral Triad feature selection (STfs) algorithm inspired by music theory.
- Compare STfs to the entire MSI feature space (i.e., no variable selection) and a popular wrapper-based algorithm, i.e., Recursive Feature elimination, coupled with Random Forest (RF-RFE) in terms of their accuracy and training times in retrieving LAI, LCab, and CCC.
- Assess the consistency in the selected feature subsets between STfs and RF-RFE in a semi-arid agricultural landscape, i.e., Bothaville, South Africa.

Introduction to Spectral Triad feature selection

- STfs algorithm is based on music theory
- Major chord formula: 1-3-5, on a diatonic scale
- A chord consists of three or more pitches (or notes) played simultaneously



Intro. to Spectral Triad feature selection…

- The intervals of the major chords (on a diatonic scale) inspire the formulation of the STfs algorithm because their formula remains the same regardless of the root note and
- Ensures a pattern where there is a maximal separation between the half steps (i.e., adjacent variables) in multiple octaves.
- In the context of the STfs algorithm, this interval ensures that each spectral triad contains variables that are adequately separated (i.e., spectrally) from each other, thus avoiding excessive collinearity in evaluating the triad.

Intro. to Spectral Triad feature selection…

- STfs uses the major triad chord formula, where 1-3-5 represents the starting variable (root note, Rn), and the third and the firth variables.
- The Rn refers to the starting variable (an index of the explanatory variables in the dataset), which is 1 by default.
- The octaves are the search iterations or sequences (progressions) that stop when there are no more unique spectral triads to evaluate or all the explanatory variables in the dataset are evaluated (i.e., equivalent to when all notes on the diatonic scale are played).

Intro. to Spectral Triad feature selection…

- For each triad, an evaluation criterion is used to rank the explanatory variables where the best single variable in a triad is selected.
- The evaluation criteria in the filter mode of STfs (Algorithm 1) can be an Entropy-based measure (e.g., Information gain) or ReliefF while in wrapper mode, it uses an explainable machine leaning algorithm such as Multiple Linear Regression (MLR) or RF Variable Importance measure (Percent Increase in Mean Squared Error, %IncMSE).
- The STfs algorithm then returns an optimal subset of features *P*, i.e., optimal features selected from each triad, which is fewer than the original input features.

Intro. to Spectral Triad feature selection…

Algorithm 1. Spectral Triad feature selection

(1) **for** Number of covariates *n* **do**

(2) Choose the Root note R*n* (from 1 to *n* of variables),

(3) Based on the R*n*, select the third R_3 and fifth R_5 notes (i.e., variables) to form a triad T_1,

(4) Evaluate the relationship of T_1 to the response variable y using ReliefF as an evaluation criterion OR Train the MLR algorithm or Random Forest Model using T_1 to T_n and the relevant response variable y,

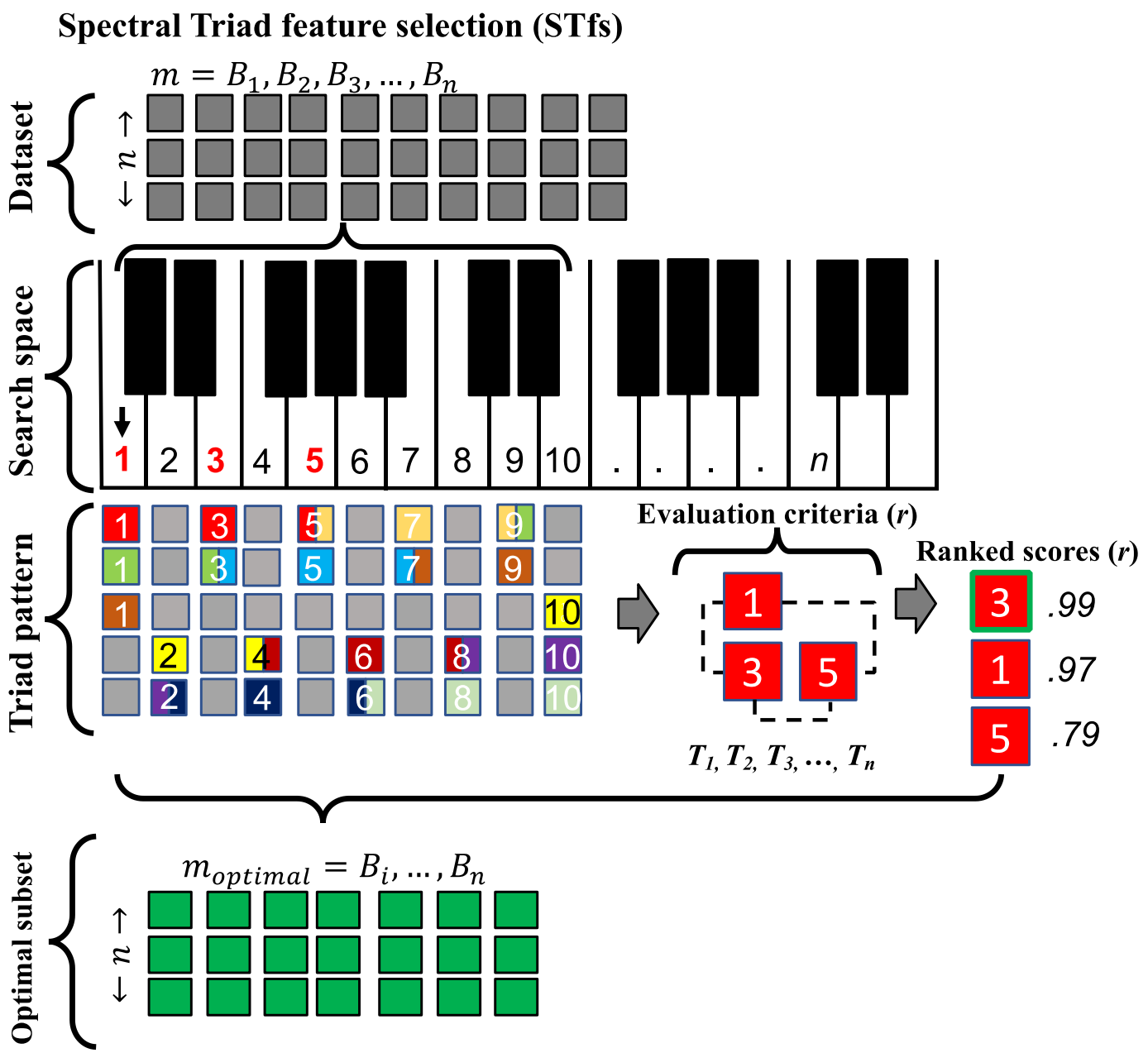
(5) Rank and select the most significant (i.e., influential) variable R_i, i.e., one with the lowest p-value of F-Statistic or highest importance score,

(6) Repeat steps 3 to 5, until all triads T_n are evaluated,

(7) Return all significant or important variables from T_1 to T_n with their triad of origin T_0 and p-value or importance scores,

(8) End **for**

Intro. to Spectral Triad feature selection…



Study area

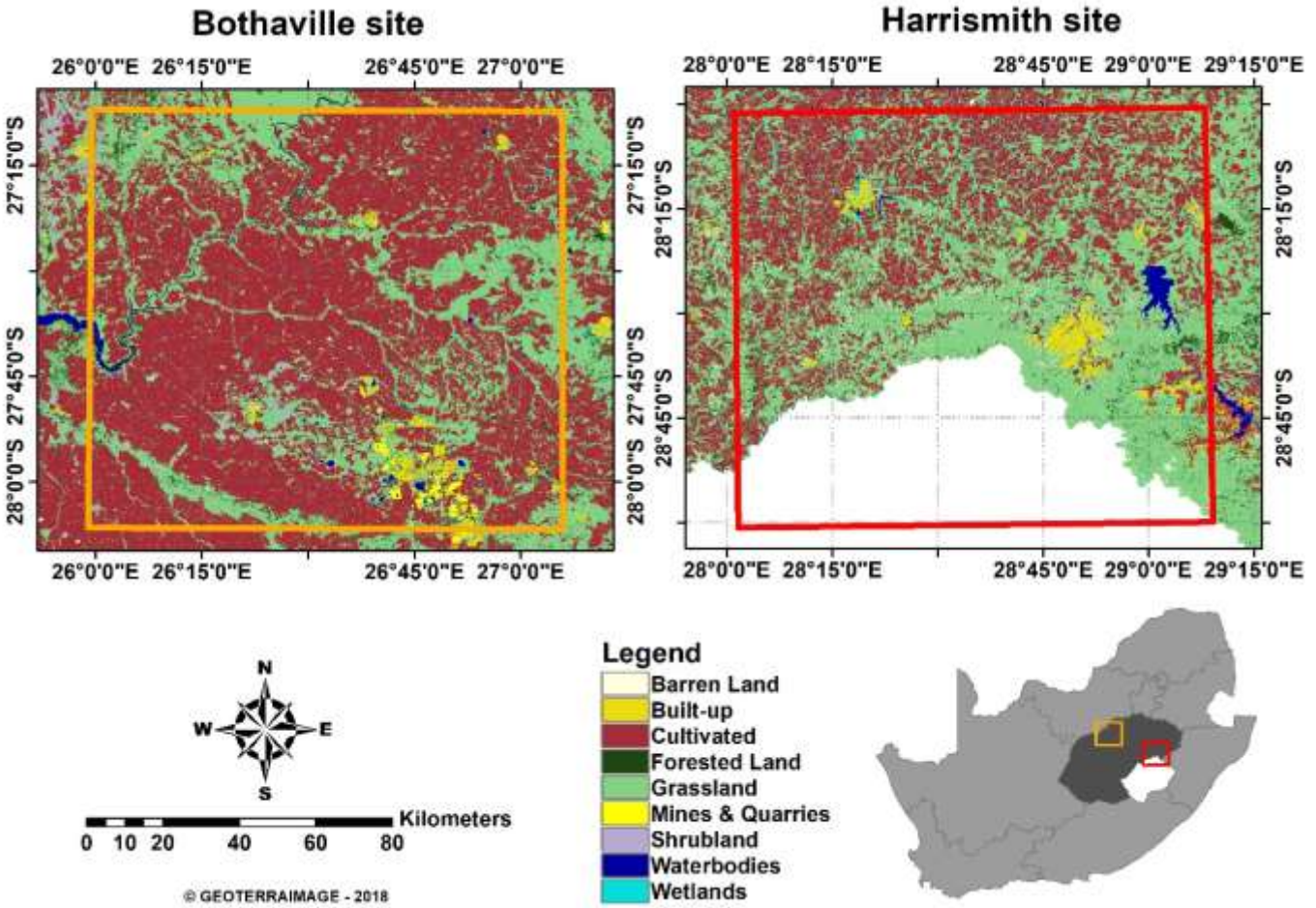
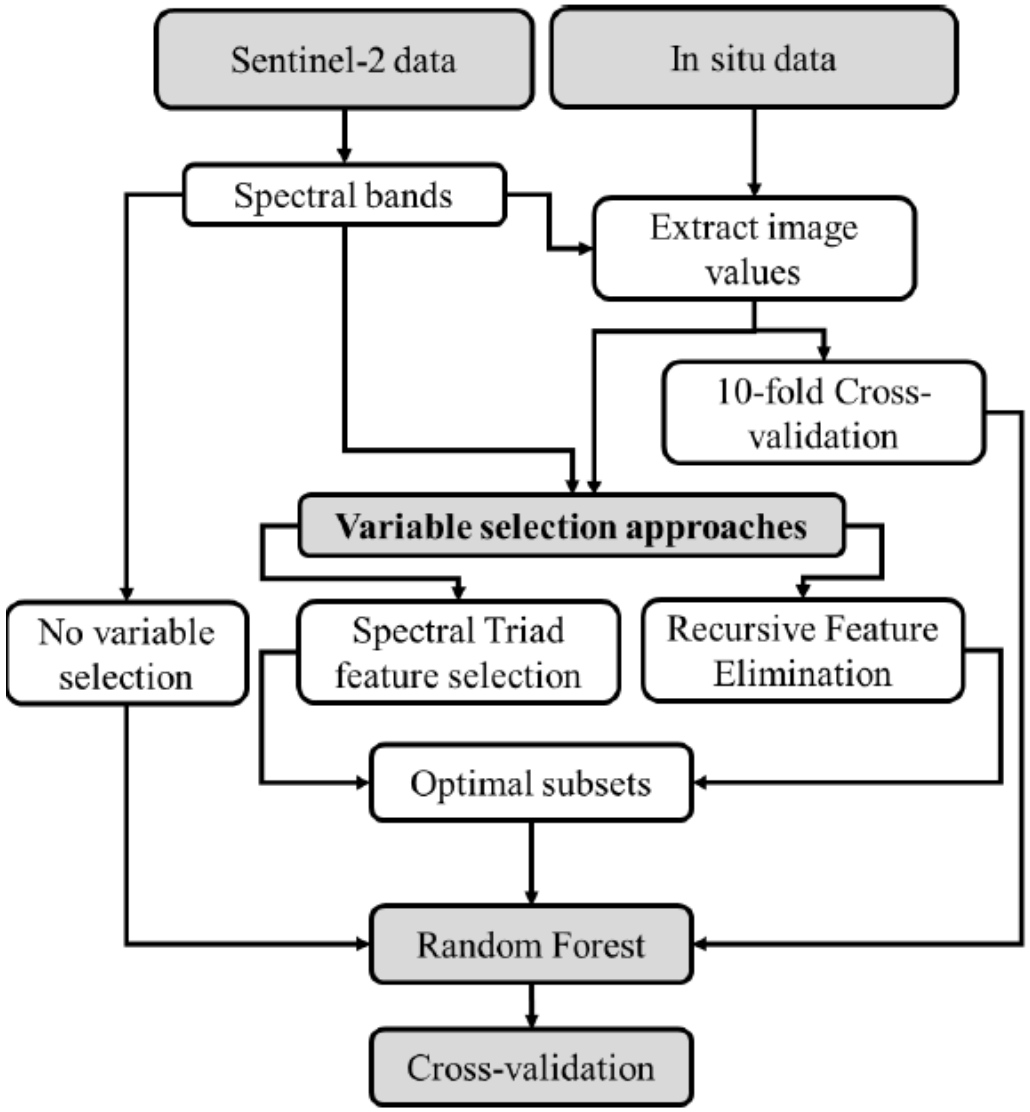


Table 4–1. Characteristics of the study area.

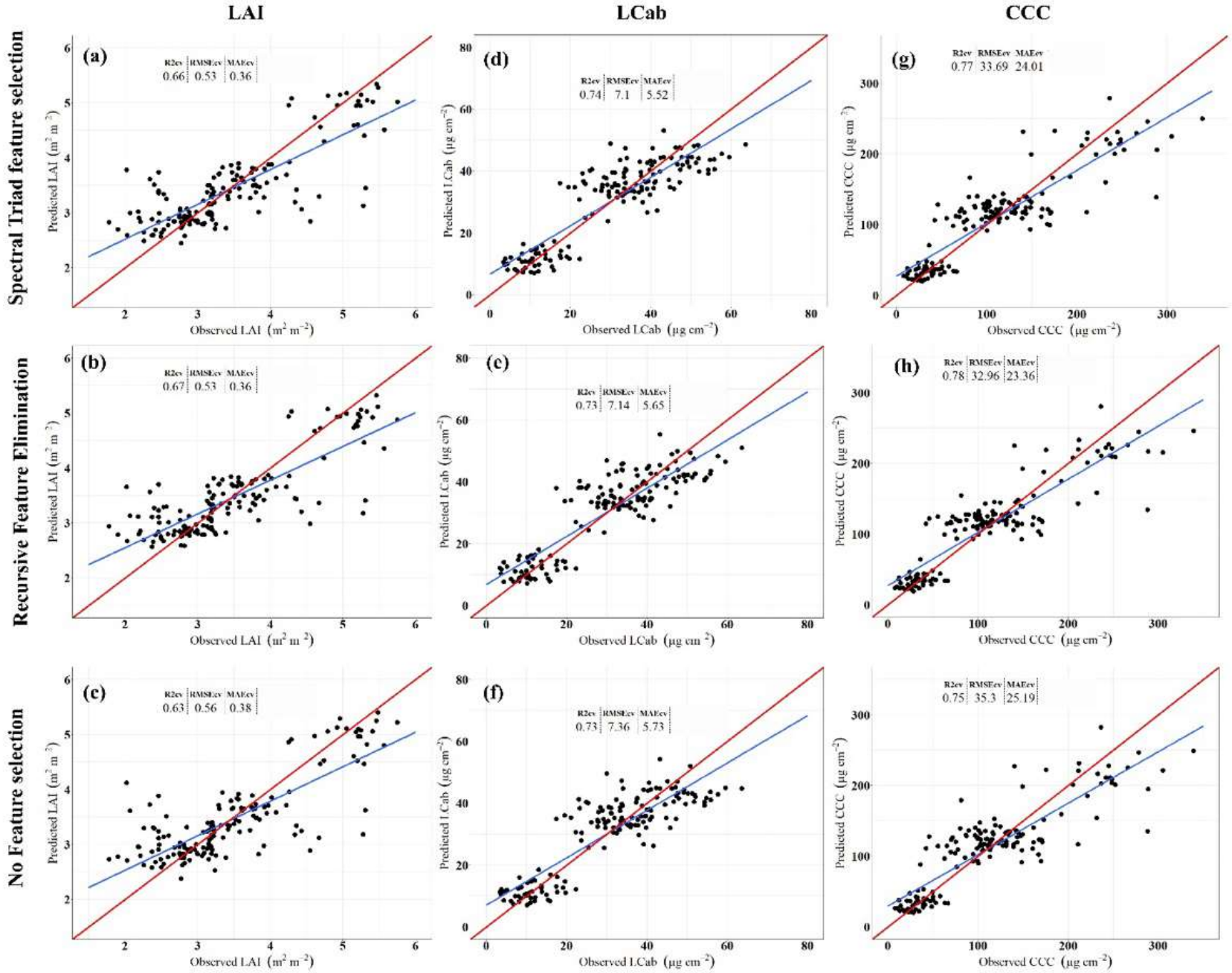
Site characteristics	Bothaville	Harrismith
Mean annual Temperature	18°C	19.2°C
Mean annual Rainfall	584 mm	115 mm
Soil Types	loamy to sandy-loamy	Sandy to sandy-loamy
Slope	Flat slope	Undulating slopes
Main crop types	Maize, Beans, and Sunflower	Maize and Beans
Crop calendar	December to May/June	December to May/June

Tests and results

- STfs was tested under 2 modes, i.e., Filter and Wrapper
- Filter mode Evaluation criteria : ReliefF algorithm,
- Wrapper mode Evaluation criteria: MLR and RF
- Random Forest-Recursive Feature Elimination (RFE) and All bands were used as a benchmark



Tests & Results…



Summary & Conclusion, Future works

- This study proposed the novel Spectral Triad feature selection (STfs) for optimising the spectral feature spaces of quasi-hyperspectral sensors, which are characterised by many typical broad-band and strategically located narrow-band wavebands.
- The performance of the proposed feature selection technique was benchmarked against the well-established Recursive Feature Elimination (RFE) coupled with a machine learning algorithm, i.e., Random Forest (RF), and the entire MSI spectral feature space, in retrieving essential biophysical and biochemical variables for agricultural crops.
- Overall, the results demonstrated that the MSI feature space could be optimised through feature selection, which results in slightly better or equivalent retrieval accuracies to the entire MSI feature space. Moreover, the spectral bands selected by STfs were consistent with those selected by a well-established algorithm, i.e., RF-RFE.

- **R-package development underway, integrated other chord formula >>> Public release**

Thank You

Mahlatse Kganyago, Clement Adjorlolo, Paidamwoyo Mhangara

mahlatsek@uj.ac.za

UNIVERSITY OF JOHANNESBURG