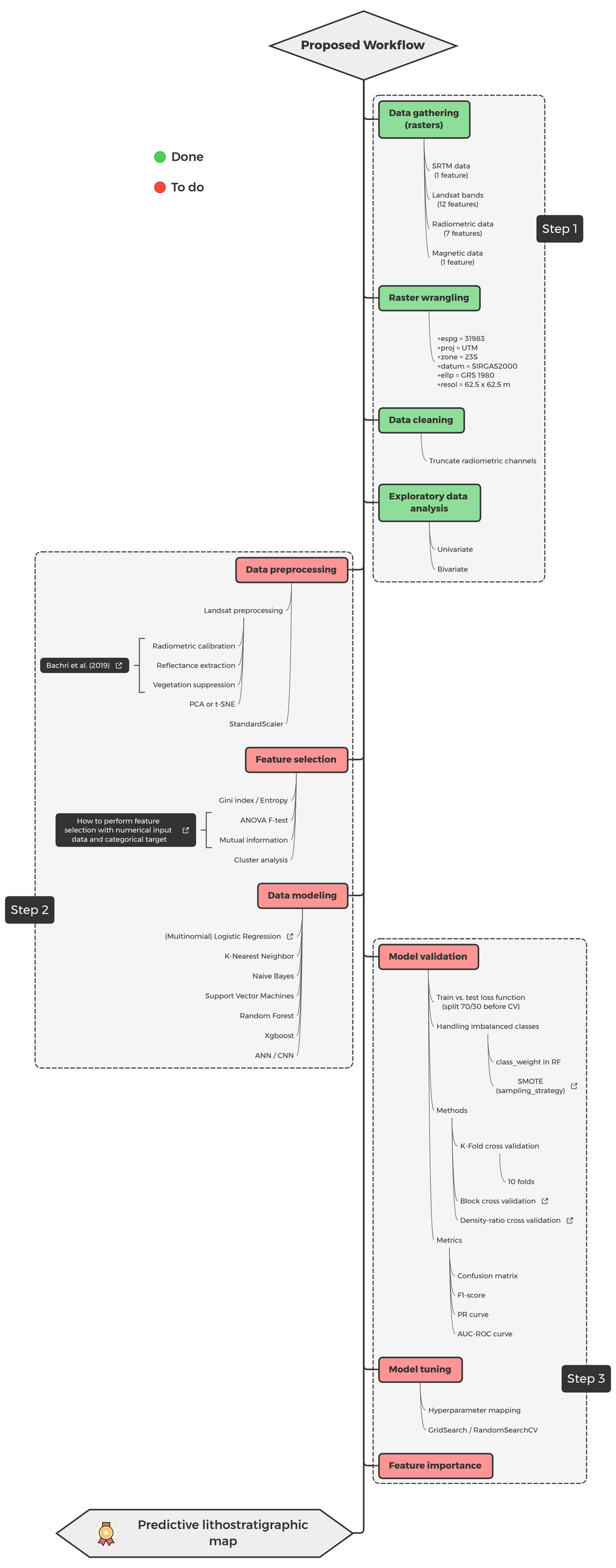
Proposed Workflow
Done
To do
Data gathering (rasters)
SRTM data (1 feature)
Landsat bands (12 features)
Radiometric data (7 features)
Magnetic data (1 feature)
Step 1
Raster wrangling
+espg = 31983
+proj = UTM
+zone = 23S
+datum = SIRGAS2000
+ellp = GRS 1980
+resol = 62.5 x 62.5 m
Data cleaning
Truncate radiometric channels
Exploratory data analysis
Univariate
Bivariate
Data preprocessing
Landsat preprocessing
Bachri et al. (2019)
Radiometric calibration
Reflectance extraction
Vegetation suppression
PCA or t-SNE
StandardScaler
Feature selection
How to perform feature selection with numerical input data and categorical target
Gini index / Entropy
ANOVA F-test
Mutual information
Cluster analysis
Data modeling
Step 2
(Multinomial) Logistic Regression
K-Nearest Neighbor
Naive Bayes
Support Vector Machines
Random Forest
Xgboost
ANN / CNN
Model validation
Train vs. test loss function (split 70/30 before CV)
Handling imbalanced classes
class_weight in RF
SMOTE (sampling_strategy)
Methods
K-Fold cross validation
10 folds
Block cross validation
Density-ratio cross validation
Metrics
Confusion matrix
F1-score
PR curve
AUC-ROC curve
Model tuning
Step 3
Hyperparameter mapping
GridSearch / RandomSearchCV
Feature importance
Predictive lithostratigraphic map