

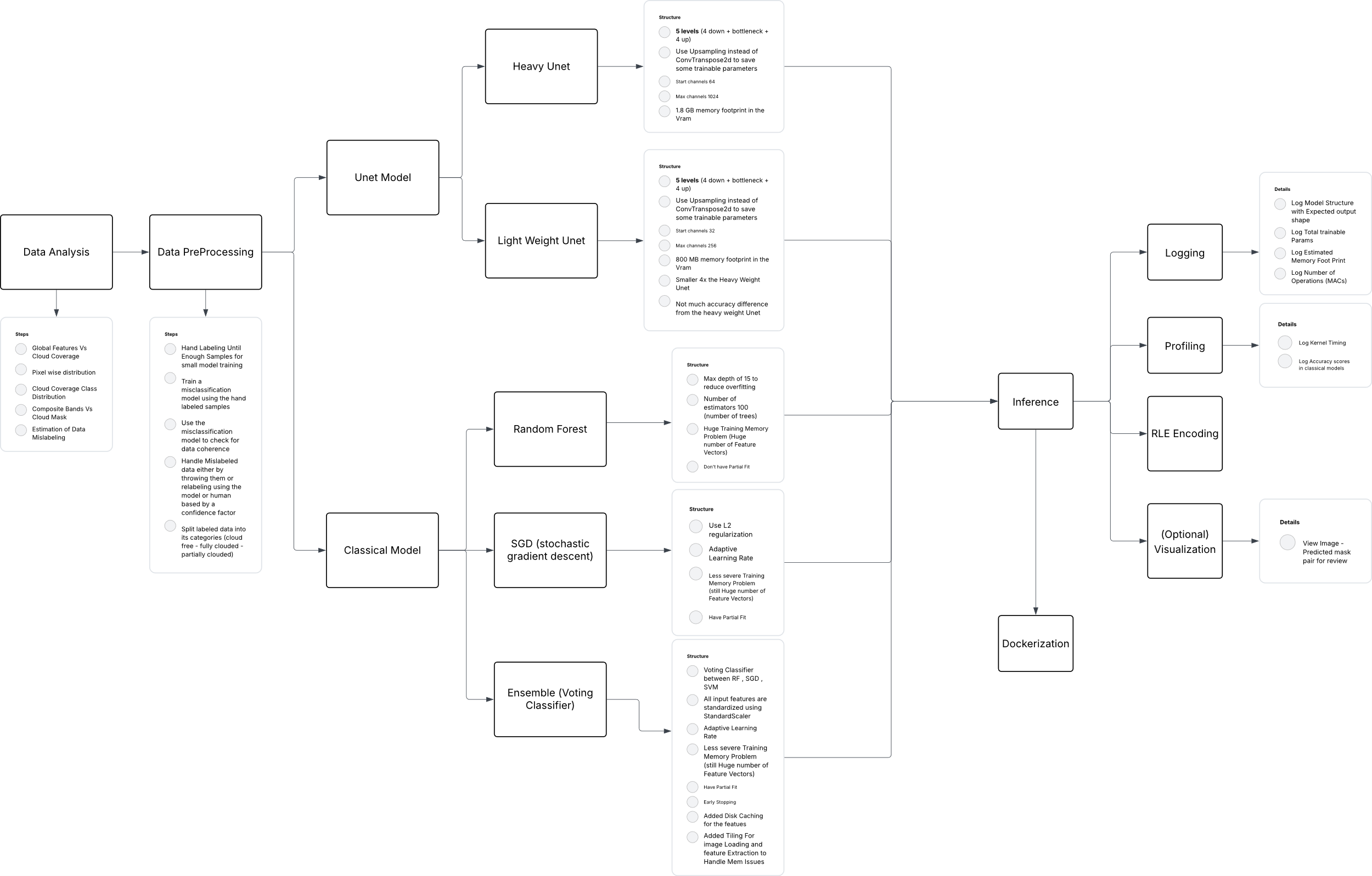
**AI Powered Cloud Masking Project**

|  |  |  |
| --- | --- | --- |
| Name | SEC | BN |
| أحمد أسامة فوزي زهران | 1 | 6 |
| محمود اسامه خطاب | 2 | 14 |
| محمد عادل عز الدين | 2 | 9 |
| كريم علاء ابوالعينين | 2 | 3 |

Under Supervision of:

Eng. Noran Hany

# Project Pipeline

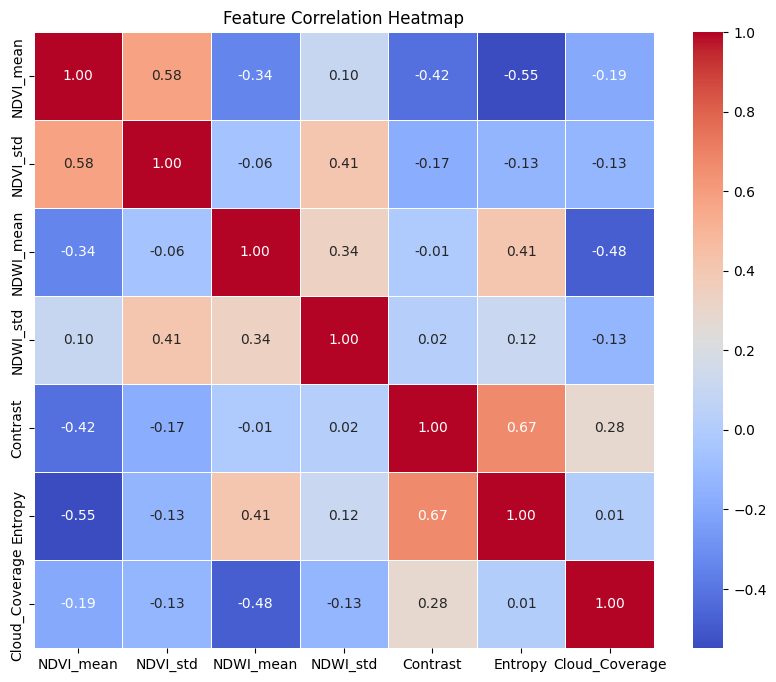


Preprocessing Module

* Hand Labelling Until Enough Samples for small model training
* Train a misclassification model (Unet model) using the hand labelled samples
* Use the misclassification model to check for data coherence
* Handle Mislabelled data either by throwing them or relabelling using the model or human based by a confidence factor (Nearly ¼ of the data is mislabelled)
* Split labelled data into its categories (cloud free - fully clouded - partially clouded)
* For Image and Mask Loading (Normalization is applied to standardize the pixel value distributions)
* Decreasing number of bands isn’t needed as the 4 is low already and all the 4 bands will be useful in some composite band creation

Exploratory Data Analysis

* The main goals of the EDA were:
  + To understand the distribution of cloud coverage across the dataset.
  + To inspect the quality and structure of the input images and masks.
  + To identify any anomalies, corrupted samples, or labelling issues.
* Global Features Vs Cloud Coverage



* Cloud Coverage is positively correlated with Contrast and negatively correlated with water index
* Pixel wise distribution
* Cloud Coverage Class Distribution

A graph and diagram of a graph

AI-generated content may be incorrect.

* Composite Bands Vs Cloud Mask
* Estimation of Data Mislabelling
* By using a combination of composite bands, it was found that it was near enough to the cloud mask to be used to estimate the data mislabelling effect

Model Selection & Training Module

## Classical Models

Random Forest (RF)

* Based on literature review, Random Forest (RF) was identified as a strong candidate for pixel-wise classification
* Configurations:
  + Maximum tree depth: 15 (to prevent overfitting).
  + Number of estimators (trees): 100.
* Challenges:
  + High memory consumption, as Random Forest cannot perform partial fitting.
  + Impractical to load all pixel-level feature vectors into memory at once, making training infeasible.

Stochastic gradient descent (SGD)

* SGD classifier was considered due to its ability to incrementally learn using partial\_fit
* Configurations:
  + Regularization: L2 penalty.
  + Adaptive learning rate.
* Less severe memory problems

Ensemble Model (Voting Classifier)

* An ensemble model combining Random Forest, SGD, and SVM classifiers using soft voting was developed.
* Key improvements:
  + Standardization of features using StandardScaler.
  + Adaptive learning rate.
  + Partial fitting capability (mainly through SGD).
  + Early stopping to prevent overfitting and reduce training time.
  + Disk caching for intermediate feature vectors to overcome RAM limitations.
  + Image tiling during feature extraction and loading to manage memory efficiently.

Deep Learning Models

Heavy Weight Unet

* Architecture:
  + 5 levels (4 down-sampling layers + bottleneck + 4 up-sampling layers).
  + Up-sampling layers were used instead of ConvTranspose2d to reduce trainable parameters.
  + Maximum number of feature channels: 1024.
* Challenges:
  + Very large memory footprint (~1.8 GB VRAM required).

Light Weight Unet

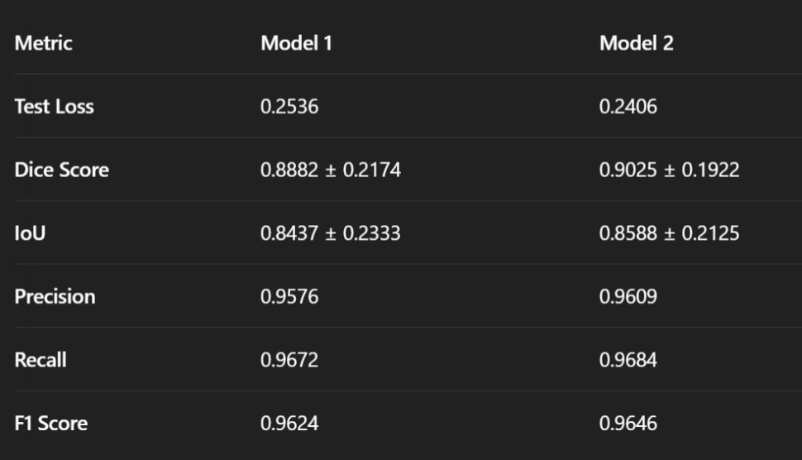
* Architecture:
  + Same as the heavyweight U-Net (5 levels).
  + Maximum number of feature channels reduced to 256.
* Advantages:
  + Significantly reduced VRAM usage (~800 MB).
  + Minimal drop in accuracy compared to the heavyweight version.

Deep Lab V3

* DeepLabV3 was considered based on its strong performance in semantic segmentation tasks.
* However, implementation was not completed due to time constraints.

Performance Analysis Module

* In the training of all models a validation set was used to prevent overfitting and evaluate the model learning process
* In the Unet training it was used for best threshold calculation too
* Ensemble Model has an accuracy of 86.8 % on test set with an F1 score of 0.86
* Heavy Weight Unet has test dice score of 0.905
* Light Weight Unet has test dice score of 0.89



model 1 is the light weight Unet and model 2 is the heavy weight Unet

Note: this kind of low score is because we relabelled the data and make a tight cloud prediction criterion

* A Looser (trained on loos data) light weight Unet model score 0.93 dice score

Additional Developed Modules

Custom Profiling and Logging

* Log Model Structure with Expected output shape
* Log Total trainable Parameters
* Log Estimated Memory Footprint
* Log Number of Operations (MACs)
  + Calculates and logs the number of Multiply-Accumulate Operations (MACs), which serve as a proxy for the computational complexity of the model
* Log Kernel Timing
* Log Accuracy scores in classical models

Prediction Visualization

* View Image - Predicted mask pair for review
  + Displays a side-by-side comparison of the original image and the model’s predicted mask. This visual comparison allows for easy detection of discrepancies between the ground truth and the model's prediction

Enhancements and Future Work

* It was found that the model weakness is free clouded mask as it does not get much penalty because dice score only care about foreground similarity so we could make a stricter training criterion
* Testing deep lab v3

Workload Division

|  |  |  |
| --- | --- | --- |
| Name | Code | Tasks |
| احمد اسامه فوزي زهران | 9210026 |  |
| محمود اسامه خطاب |  |  |
| محمد عادل عز الدين |  |  |
| كريم علاء ابوالعينين |  |  |