

- Kaggle
- Img 2 Biomass Contest
 - background & dataset: img of pasture taken by phone->estimation of the amount of green and withered grass

Model1: Img ->feature ->XGBoost

XGBoost is a decision-tree algorithm and a frequent winner in data science competitions like Kaggle.

Think of it as a "team of top students." The first student solves a problem. The second student focuses on fixing the first one's mistakes. The third student corrects the errors made by the previous two, and so on. Finally, the team combines their insights to produce a highly accurate answer.

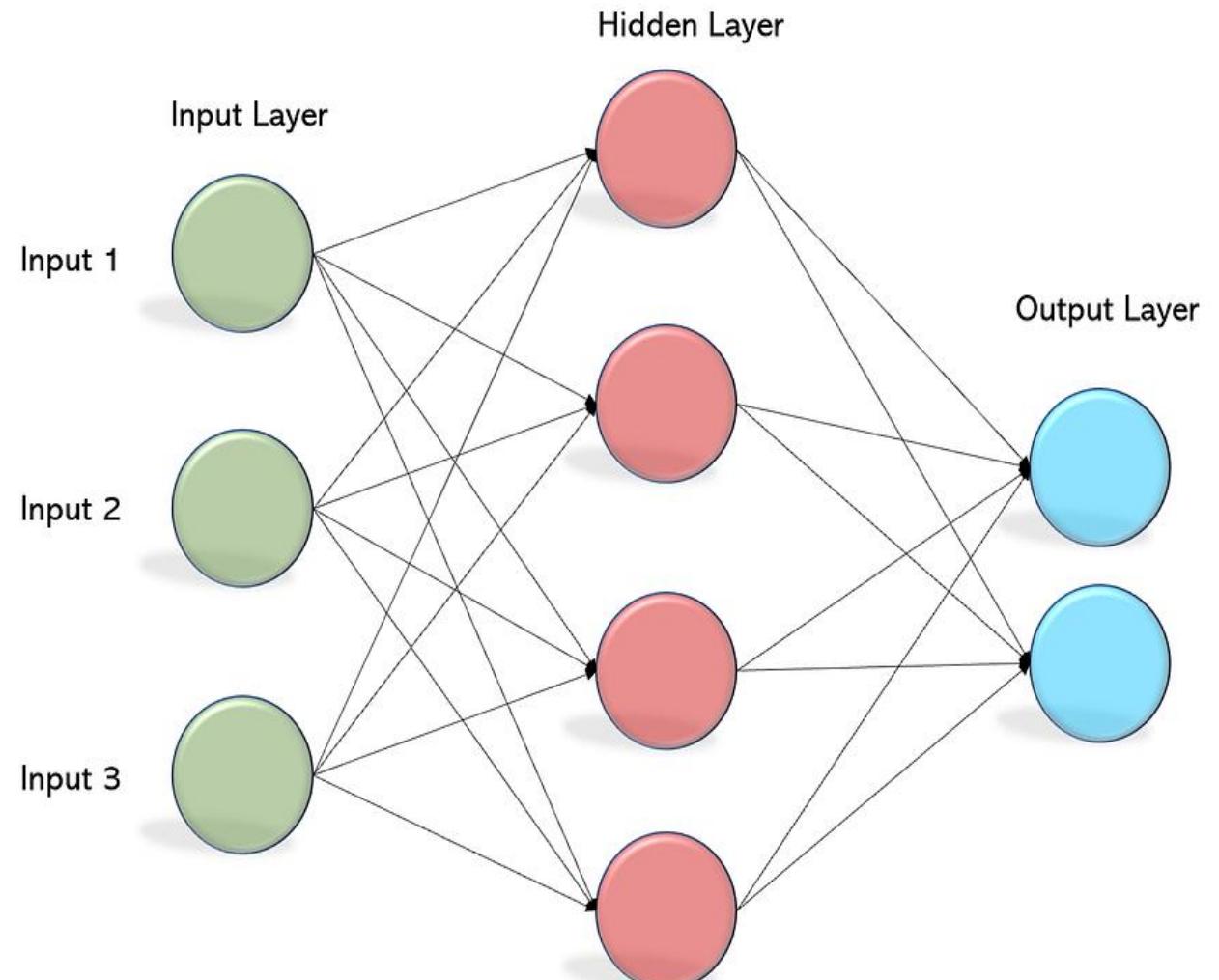
ID1001187975.jpg (3.43 MB)



Model2: Img ->feature ->MLP

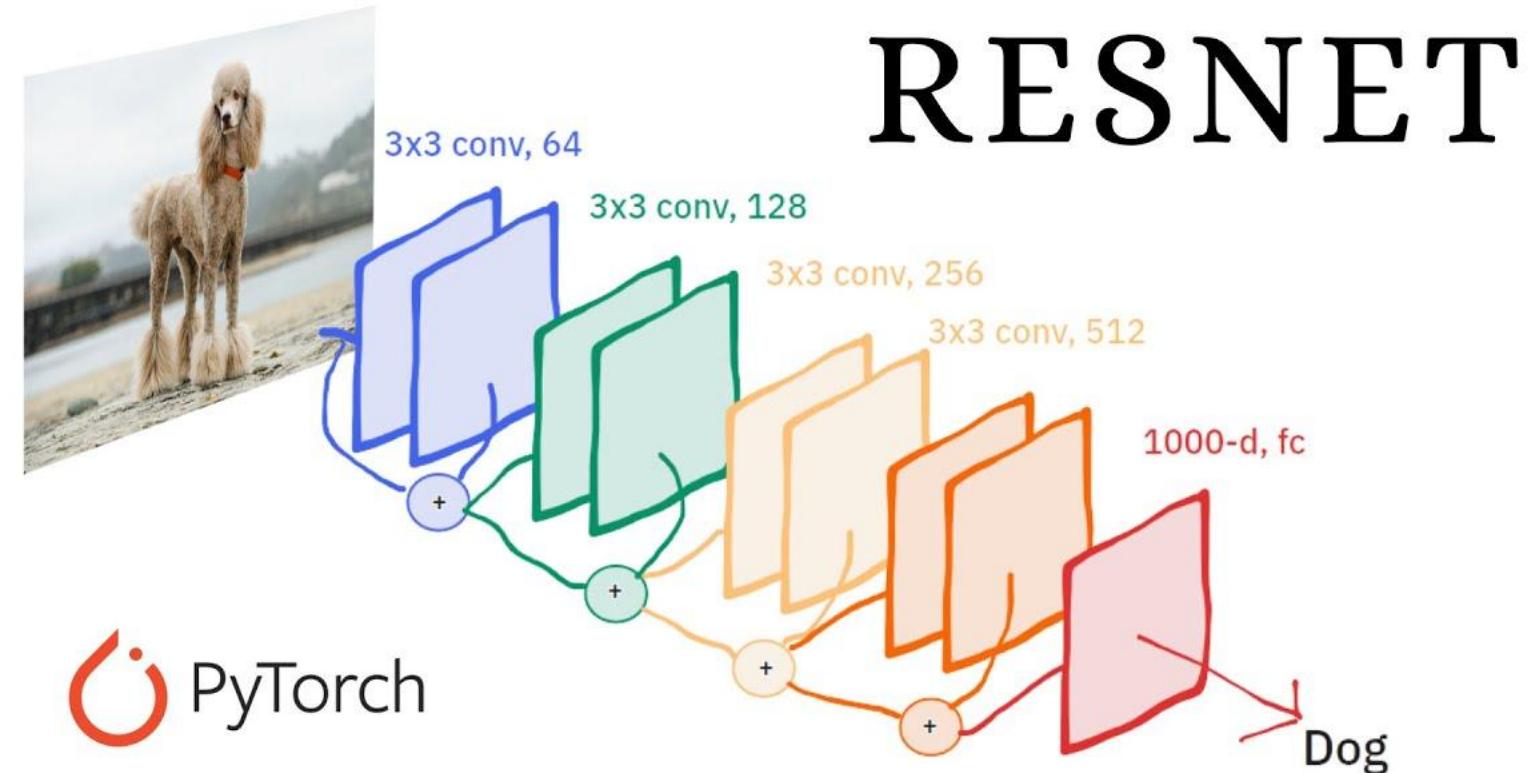
A multilayer perceptron (MLP) neural network model was employed in this study. It is one of the simplest forms of an ANN.

- two hidden layers (64 nodes each) and one output
- activation function - rectified linear unit and the
- optimisation algorithm - Adam
- learning rate of 0.001
- 5 epochs.



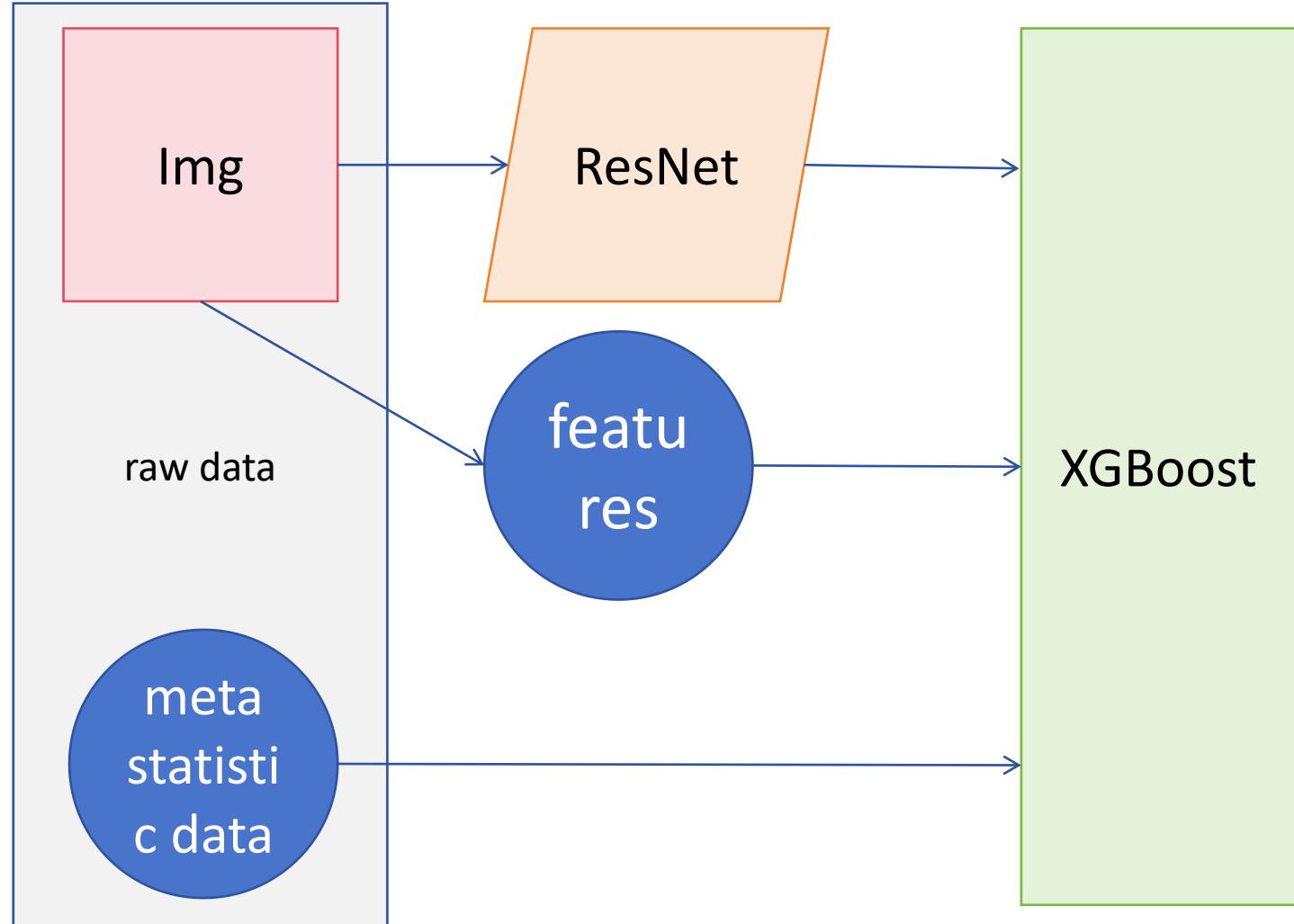
Model3: Img ->ResNet

Images are processed by ResNet, where the output of the final convolutional layer serves as an embedding to model pasture biomass.



Model4: Fusion model

Handcrafted features, metadata, and ResNet embeddings (extracted from the final conv layer) are concatenated and input into XGBoost.

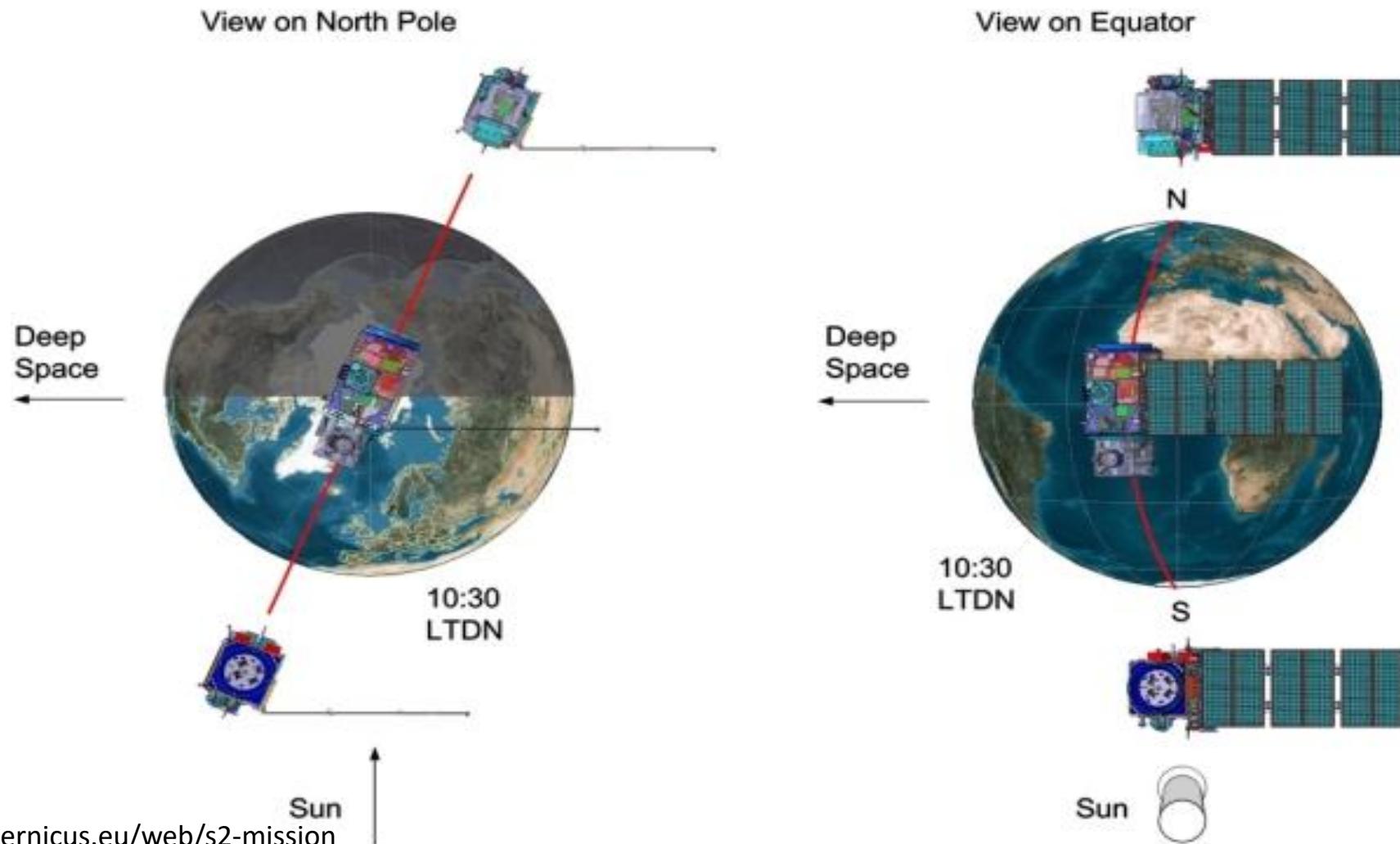


Model5: [Img ->feature + Remote Sensor->feature]->MLP

The screenshot shows the Digital Earth Australia (DEA) map interface. The map displays a satellite view of a rural landscape in New South Wales (NSW), Australia. Key features visible include roads labeled "Raymond Terrace Rd", "Scotch Creek Rd", "Woodberry Rd", "Alnwick Rd", and "Valleys Creek Rd". A body of water labeled "Swamp" is also present. The map includes a legend for "DEA Surface Reflectance (Sentinel-2A MSI)" dated "11/07/2015, 17:00:00". On the left side, there is a sidebar with various controls: a search bar for "NSW", buttons for "Explore data" and "Upload data", a "DATASETS (1)" section with "Disable All", "Collapse All", and "Remove All" options, a "Time" slider set to "11/07/2015, 17:00:00", a "Filter by location" button, and a "Styles" dropdown currently set to "False colour - Green, SWIR, ...". The top navigation bar includes links for "About", "Related Maps", "Map Settings", "Help", "Story", "Share / Print", and a search icon. A blue callout box in the upper right corner contains the text "Find the Tour, how-to videos & other help content here" and a "Got it, thanks!" button. The bottom of the screen shows a timeline from "29/12/2017" to "28/12/2021" and standard map controls (zoom, orientation, etc.).

Model5: [Img ->feature + Remote Sensor->feature]->MLP

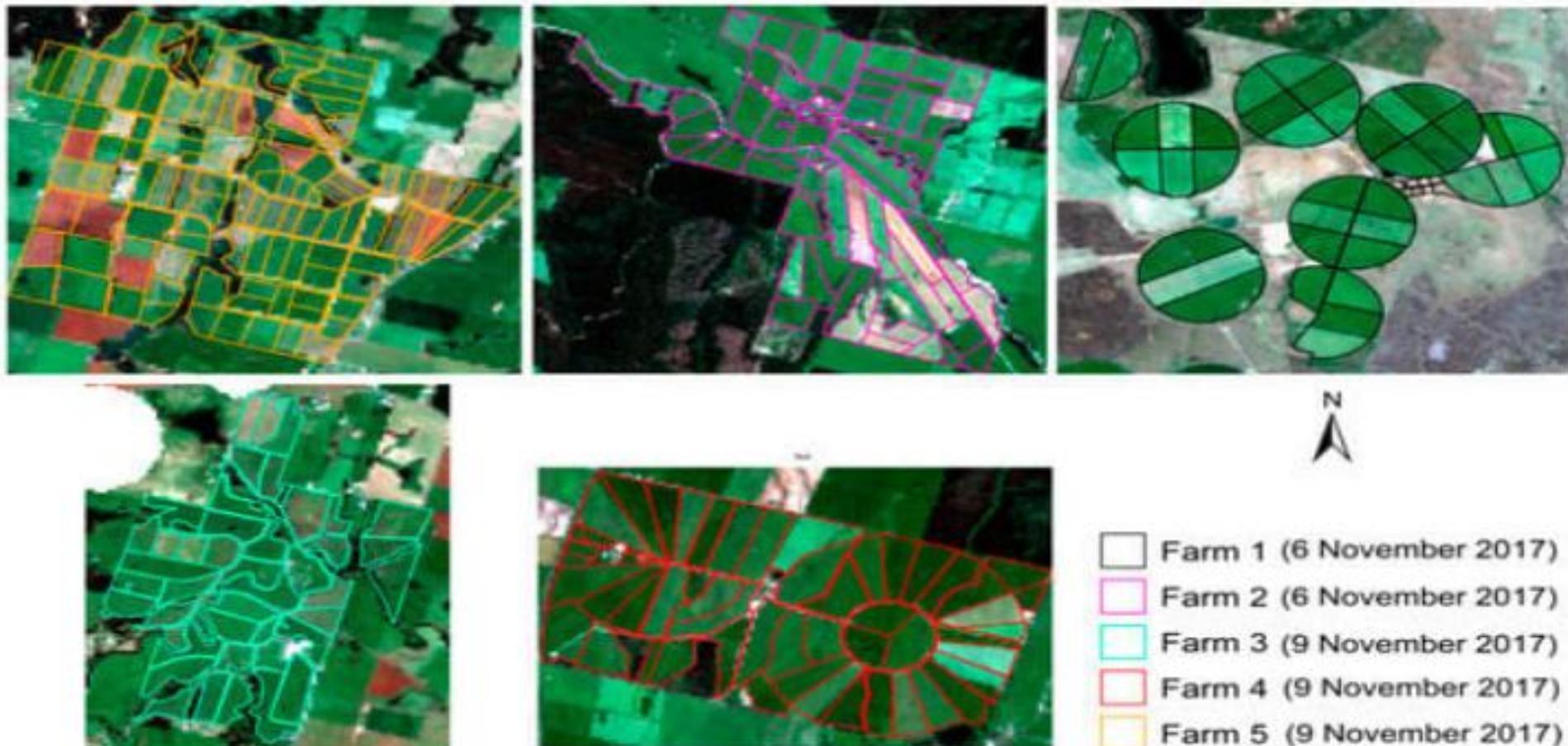
◆ Remote sensor: How it works



Model5: [Img ->feature + Remote Sensor->feature]->MLP

◆ Data: How to use

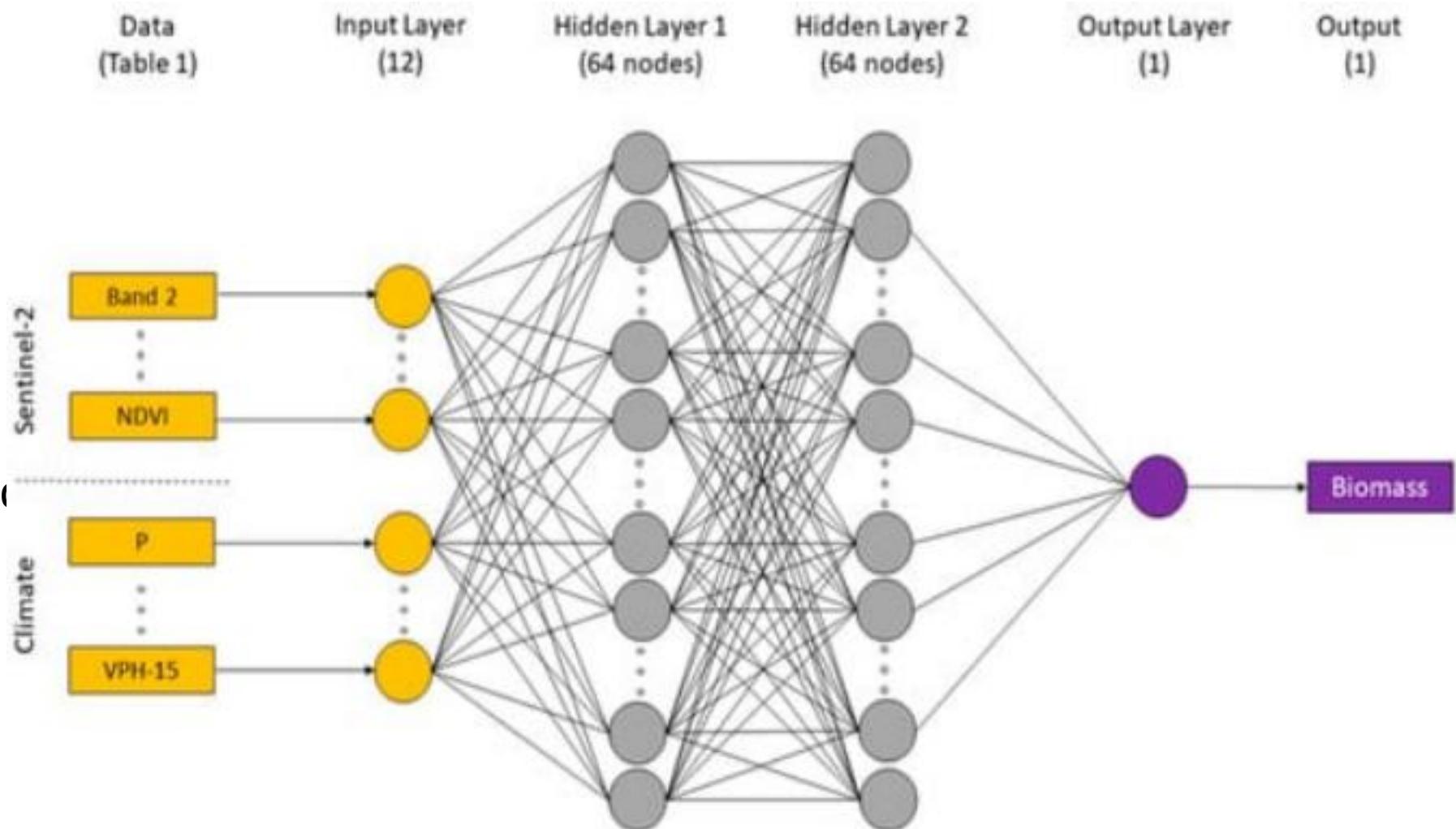
coordinate (shape)
time
radio range
resolution
satellite
cloud cover



Model5: [Img ->feature + Remote Sensor->feature]->MLP

Data (Table 1)	Input Layer (12)	Hidden Layer 1 (64 nodes)	Hidden Layer 2 (64 nodes)	Output Layer (1)	Output (1)
-------------------	---------------------	------------------------------	------------------------------	---------------------	---------------

For each sample, the median value of all Sentinel- 2 bands is combined with the engineered features and metadata, then passed into an MLP.



Result

Model Input + Algorithm	Target: GDM_g (R2) (Train / Val)	Target: Dry_Total_g (R2) (Train / Val)	Training time (sec)
Images + ResNet	-0.4491 / -1.2452	-0.6701 / -1.2970	267.15 (Slowest)
Hand-crafted + Meta + XGBoost	0.9997 / 0.7140 	0.9998 / 0.6618 	251.00
Hand-crafted + Meta + MLP	0.5519 / 0.3956	0.4625 / 0.1569	112.91 (Fastest)
Sentinel-2 + Features + MLP	0.3524 / 0.5116	0.3898 / 0.2550	126.72
Fusion (ResNet + XGBoost)	0.9511 / 0.6518	0.9999 / 0.5423	182.58

Result

- Best Model (Model 2): By achieving the highest validation R^2 on both GDM_g and Dry_Total_g, Model 2 demonstrates that Structured Features paired with XGBoost is the optimal approach for this data.
- Value of Satellite Data: Comparing the MLP models, adding Sentinel-2 data (Model 4) raised the GDM_g score from 0.3956 to 0.5116 over the baseline (Model 3), validating the utility of satellite inputs.
- Image Data Issues: Model 1 (ResNet) failed to converge, resulting in negative scores. Consequently, the fusion model (Model 5) underperformed Model 2, as the image data effectively acted as noise.

Further Work: other remote sensor data,
network(DCN)