# A Consensual Collaborative Learning Method for Remote Sensing Image Classification under Noisy Multi-Labels

Ahmet Kerem Aksoy, Mahdyar Ravanbakhsh, Tristan Kreuziger, Begüm Demir

Remote Sensing Image Analysis (RSiM) Group, TU Berlin

## INTRODUCTION

- ✓ Collecting a large number of reliable training images annotated by multiple land-cover class labels for multi-label classification is time-consuming and costly in remote sensing (RS).
- ✓ To address this problem, publicly available thematic products are often used with zero-labelling-cost [1]. That may result in **noisy training sets**, distorting the learning process.
- ✓ To address this problem, we propose a Consensual Collaborative Multi-Label Learning (**CCML**) method, which can:
  - **identify** the possible **noisy labels** by introducing a novel **ranking function** for identifying reliable labels;
  - estimate the label uncertainty based on the **aggregation** of two **collaborative networks**;
  - be used within **different classification** approaches.

## PROPOSED METHOD

The proposed CCML identifies, ranks and corrects training images with noisy multi-labels through four main modules: 1) discrepancy module; 2) group lasso module; 3) flipping module; and 4) swap module.

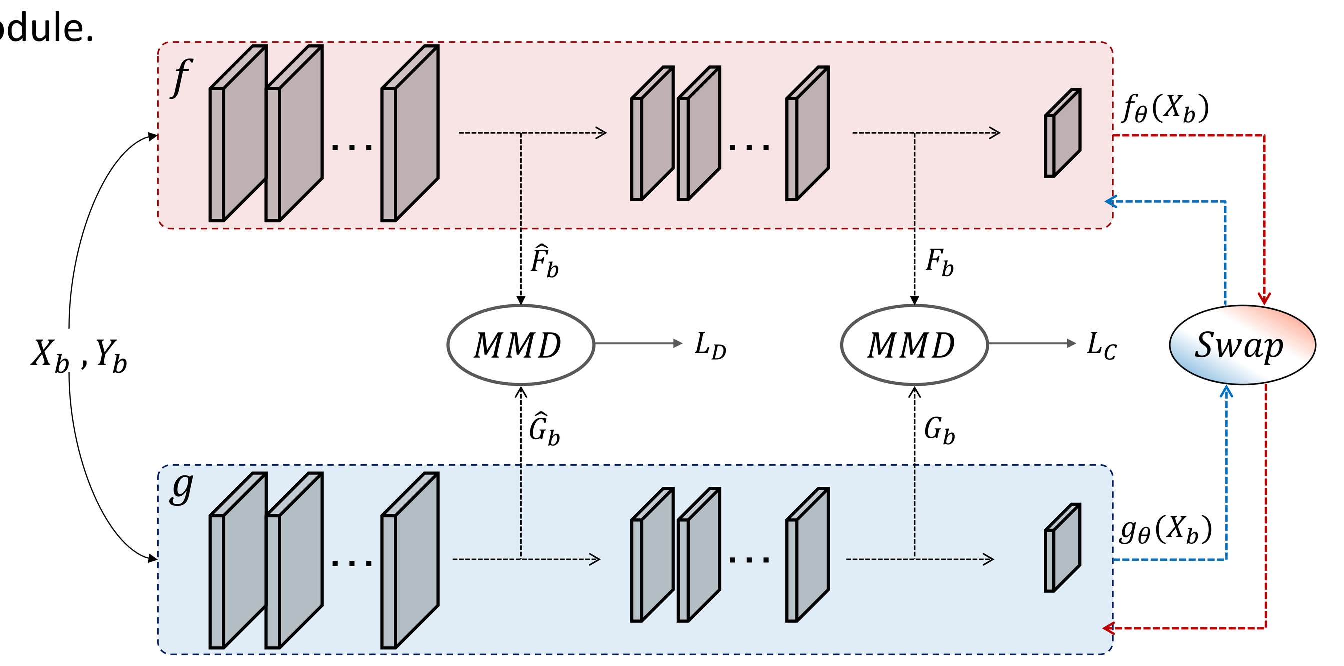


**Fig 1.** Block diagram of the training phase of the proposed CCML. $f$ and $g$ represent the two collaborative networks with parameters $\theta$ and $\hat{\theta}$, respectively.

**Discrepancy module** aims at forcing the two networks to learn diverse features, while achieving consistent predictions. It includes two loss functions:

- ✓ Disparity loss ($L_D$) ensures that the networks learn distinct features;
- ✓ Consistency loss ($L_C$) ensures that the two networks produce similar predictions.

**Group Lasso Module** aims at identifying:

- ✓ potentially noisy labels in the training set by using the predictions of the two networks.
- ✓ the type of label noise by computing a sample-wise ranking loss as:

$$\mathrm{Lasso}_f(\mathrm{x}_i) = \underbrace{\alpha \sum_{\hat{c}=a+1}^{m} \sqrt{\sum_{c=1}^{a} \epsilon}}_{\text{aggregated loss based on missing labels}} + \underbrace{\beta \sum_{c=b+1}^{m} \sqrt{\sum_{\hat{c}=1}^{b} \epsilon}}_{\text{aggregated loss based on wrong labels}}, \qquad \epsilon = E^2_{c,\hat{c}}(\mathbf{x}_i)$$

assigned labels

unassigned labels

ranking error

ranking loss for network $f$

ranking error function

Information about potential noisy labels is provided through a **ranking error function** as:

$$E_{c,\hat{c}}(\mathbf{x}_i) = \max(0, 2[f_{\hat{c}}(\mathbf{x}_i) - f_c(\mathbf{x}_i)] + 1)$$

| Missing label | Wrong label | $f_{\hat{c}}$ | $f_c$ | $E_{c,\hat{c}}$ |
|---|---|---|---|---|
| ✗ | ✗ | 0 | 1 | 0 |
| ✓ | ✗ | 0 | 0 | +1 |
| ✗ | ✓ | 1 | 1 | +1 |
| ✓ | ✓ | 1 | 0 | +3 |

**Flipping module** aims at flipping the identified noisy labels and includes:

- ✓ Noisy class selector (NCS) receives the ranking loss from two networks and identifies the samples with higher uncertainty;
- ✓ Noisy class flipper (NCF) selects the labels with the largest ranking losses to apply the flipping.

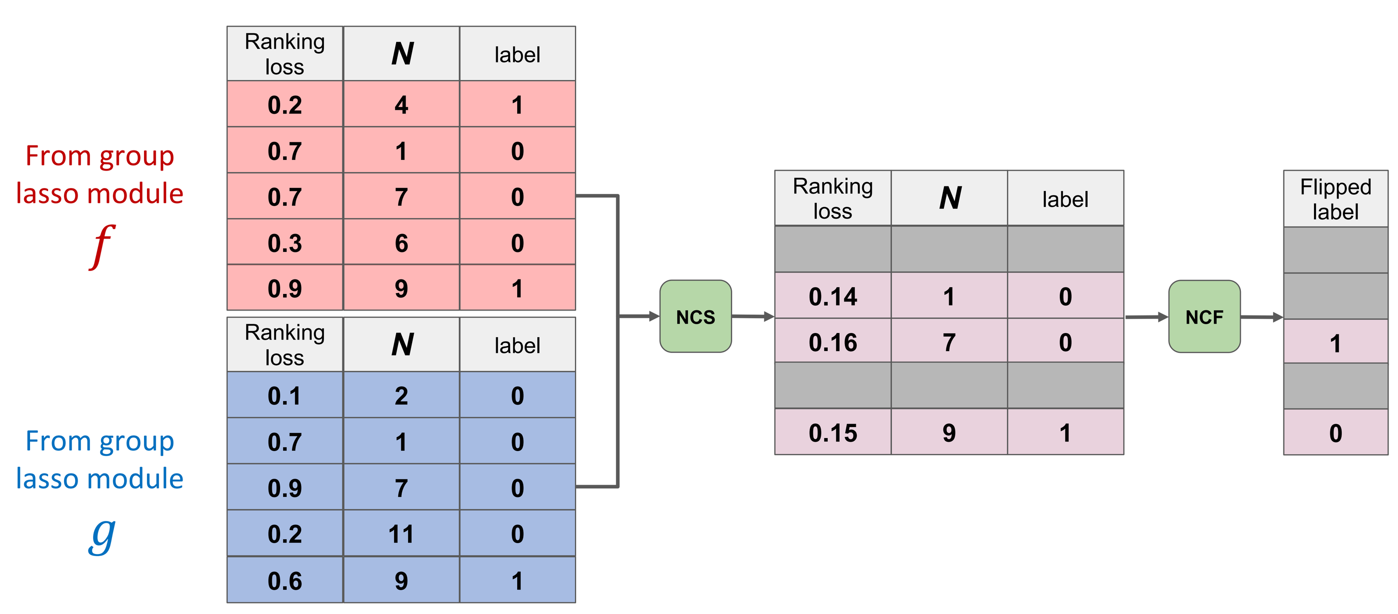


**Fig 2.** A qualitative example of flipping noisy labels in the flipping module. ***N*** implies the potential noisy class indexes, output of ranking error function.

**Swap Module** aims at exchanging the ranking information between the networks. To this end, it:

- ✓ is inserted between the two collaborative networks;
- ✓ takes the Binary Cross Entropy (BCE) and ranking losses into consideration to eliminate the detected noisy samples from back-propagation.

$$B_i^f = L_f(\mathbf{x}_i) + \gamma Lasso_f(\mathbf{x}_i)$$

swapping ranking loss for network $f$

BCE loss

Sample from mini-batch

$$Loss_f^b = \frac{1}{R}\sum_{i=1}^{R} L_f(\mathbf{x}_i) + \lambda_1 L_c + \lambda_2 L_D$$

BCE loss for R samples associated with low loss values identified by network $g$

disparity loss

consistency loss

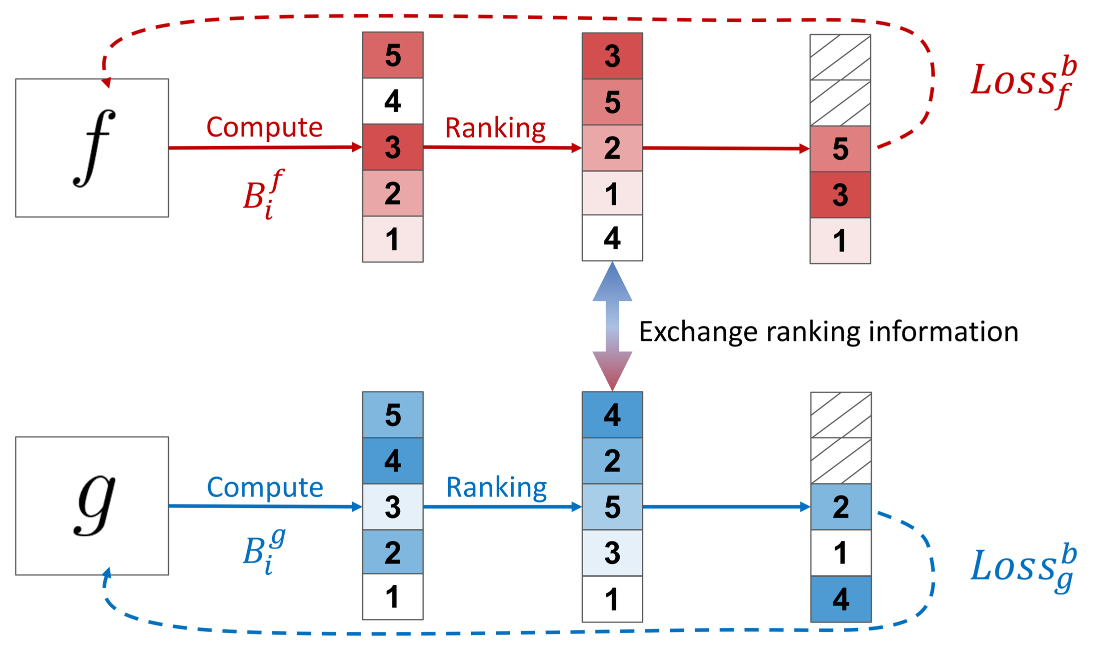


**Fig 3.** A qualitative example to describe the swap module. The two networks exchange the ranking information.

## DATASET AND EXPERIMENTAL RESULTS

- ✓ Experiments have been carried out on the Ireland subset of the **BigEarthNet** [2] benchmark archive, consisting of 15,894 Sentinel-2 images.
- ✓ Two architectures **ResNet** [3] and **DenseNet** [4] were used as **baselines** for comparison.
- ✓ Noise injection is applied by random selection of *n*% of samples from each mini-batch, and flipping randomly *n*% of the labels from the selected samples.

Noise injection with the rate of 50%

0 synthetically injected missing labels

1 synthetically injected wrong labels

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_1$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| $y_2$ | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| $y_3$ | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| $y_4$ | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| $y_5$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| $y_6$ | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |

**Table 1.** Results obtained by the proposed CCML and the baseline architecture ResNet.

| Injected Noise Rate | Precision (%) | | Recall (%) | | F1 (%) | |
|---|---|---|---|---|---|---|
| | Baseline (ResNet) | Proposed CCML | Baseline (ResNet) | Proposed CCML | Baseline (ResNet) | Proposed CCML |
| 20% | 87.8 | **90.2** | **68.7** | **68.7** | 77.1 | **78** |
| 30% | 84 | **88.2** | 67.2 | **68.9** | 74.7 | **77.4** |
| 40% | 76.4 | **88.4** | 65.1 | **69.3** | 70.3 | **77.7** |
| 50% | 62.5 | **87.5** | 57.6 | **62.1** | 60 | **72.6** |

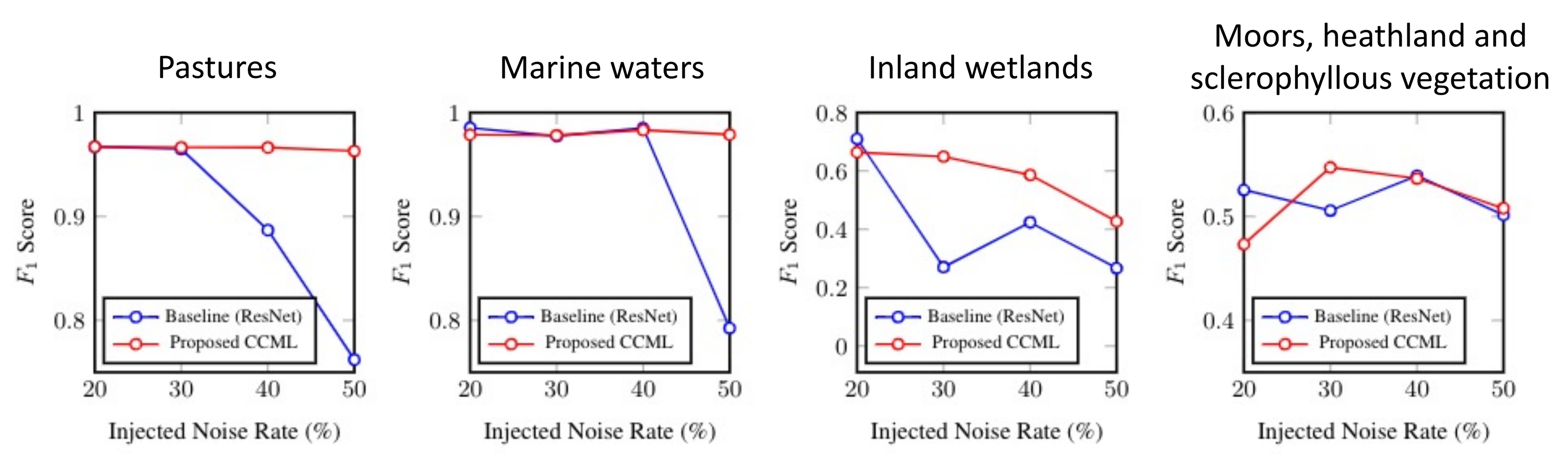


**Fig 4.** Different noise rates versus class based F1 scores obtained by the ResNet baseline and the proposed CCML for four selected classes.

## CONCLUSION

The proposed CCML is promising since it:

- ✓ is able to **automatically** identify two different types of **multi-label noise** (i.e., missing and wrong class label annotations) without making any prior assumption;
- ✓ achieves **high accuracy** under a high (synthetically added) multi-label noise rates;
- ✓ is **architecture-independent**, and thus can be used within different network architectures.

As a future development, we plan to extend the proposed CCML with an **adaptive ranking loss function** to adjust the amount of sample removal.

## ACKNOWLEDGEMENT

This work is supported by the German Ministry for Education and Research as **BIFOLD** - Berlin Institute for the Foundations of Learning and Data (01IS18025A).

## REFERENCES

[1] C. Paris and L. Bruzzone, "A Novel Approach to the Unsupervised Extraction of Reliable Training Samples from Thematic Products," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 1930-1948, 2021, doi: 10.1109/TGRS.2020.3001004.

[2] G. Sumbul, A. d. Wall, T. Kreuziger, F. Marcelino, H. Costa, P. Benevides, M. Caetano, B. Demir, V. Markl, "BigEarthNet-MM: A Large Scale Multi-Modal Multi-Label Benchmark Archive for Remote Sensing Image Classification and Retrieval", IEEE Geoscience and Remote Sensing Magazine, 2021, doi: 10.1109/MGRS.2021.3089174.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[4] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261–2269, 2017.