Multi-sensor Model for Earth Observation Robust to Missing Data via Sensor Dropout and Mutual Distillation

Francisco Mena, Dino Ienco, Cássio Dantas, Roberto Interdonato, Andreas Dengel

▶ **To cite this version:**

Francisco Mena, Dino Ienco, Cássio Dantas, Roberto Interdonato, Andreas Dengel. Multi-sensor Model for Earth Observation Robust to Missing Data via Sensor Dropout and Mutual Distillation. IEEE Access, 2025, 13, pp.83930-83943. ⟨10.1109/ACCESS.2025.3568706⟩. ⟨hal-05063111⟩

HAL Id: hal-05063111

https://hal.science/hal-05063111v1

Submitted on 14 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier None

Multi-sensor Model for Earth Observation Robust to Missing Data via Sensor Dropout and Mutual Distillation

FRANCISCO MENA[1,2] (Graduate Student Member), DINO IENCO[3,5] (Member), CÁSSIO F. DANTAS[3,5], ROBERTO INTERDONATO[4,5] and ANDREAS DENGEL[1,2]

[1]University of Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany
[2]German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[3]INRAE, UMR TETIS, University of Montpellier, Montpellier, France
[4]CIRAD, UMR TETIS, University of Montpellier, Montpellier, France
[5]Inria, University of Montpellier, Montpellier, France

Corresponding author: Francisco Mena (e-mail: f.menat@rptu.de).

F. Mena acknowledges support through a scholarship of the University of Kaiserslautern-Landau.

ABSTRACT Multi-sensor data has become a foundation of Earth Observation (EO) research, offering models with enhanced accuracy via optimal fusion strategies. However, the unavailability of sensor data at the regional or country scale during inference can significantly undermine model performance. The literature explores diverse approaches to increasing model robustness to missing sensor scenarios, i.e., to reducing the decline in accuracy caused by missing data at inference time. Nevertheless, most of them have suboptimal behavior when a single-sensor is available for prediction. To address this challenge, we propose a novel method for multi-sensor modeling, Decision-level Sensor Dropout with mutual distillation (DSensD+). This employs a decision-level fusion, ignoring predictions from missing sensors and incorporating the Sensor Dropout (SensD) technique. Unlike works that use the SensD at the input or feature level, we use it at the decision level. Moreover, we include a mutual distillation strategy to improve the robustness. From a practical viewpoint, the additional components in the DSensD+ method are incorporated only for the training phase. During inference, it operates as a standard decision-level fusion model that ignores missing sensors. We validate our method on three EO datasets, spanning binary, multi-class, and multi-label classification tasks for crop- and tree-mapping related applications. Notably, DSensD+ outperforms several state-of-the-art methods, achieving consistent improvements across moderate (single-sensor missing) and extreme (single-sensor available) conditions, as well as with full-sensor data. These results demonstrate the robustness of DSensD+ and highlight the effectiveness of our method for the missing sensor problem, advancing the field of multi-sensor modeling in EO.

INDEX TERMS Earth Observation, Multi-sensor model, Missing sensor data, Deep learning, Robustness.

I. INTRODUCTION

Accessing multi-sensor data is becoming a common setting in the Earth Observation (EO) domain. The advances in instruments and the accessibility to data from diverse remote sensors have supported the research of deep learning models for multi-sensor data analysis [1]. Multi-sensor data can be used to learn common features and transfer knowledge between sensor-dedicated models [2] or to improve the predictive performance of deep learning models [3]. We focus on the latter in this work, which consists of designing an effective way to fuse multi-sensor data. Thus, evidence from literature shows that fusing multi-sensor data improves the model performance compared to single-sensor models [3]. However, accessing multiple sensor data during both the training and inference stages could be infeasible in scenarios characterized by operational constraints and requirements.

Missing data is an inherent phenomenon in the EO domain [4]. This is because data collection occurs under operational constraints in real-world environments, where different situations and human decisions may affect its consistent and

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

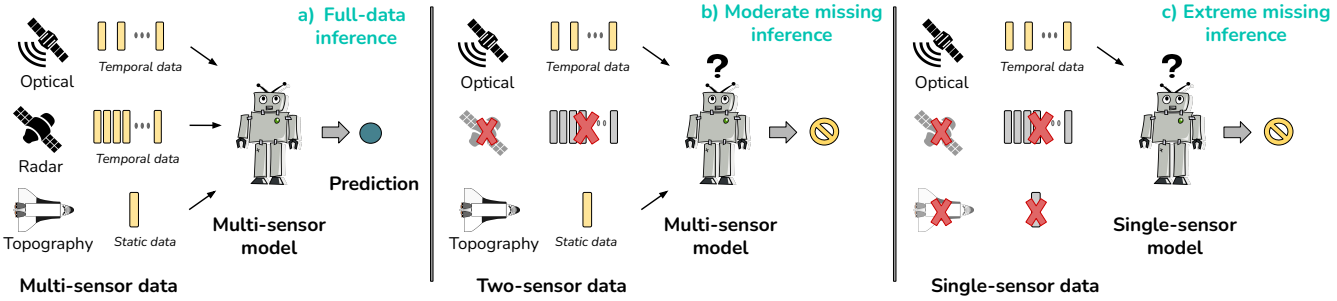


FIGURE 1. Illustration of inference cases with missing sensor data. The multi-sensor model is usually trained in the first scenario of full-sensor data with optical, radar, and topography (a). In missing sensor scenarios, b) moderate sensor missingness corresponds to predictions with a single-sensor missing (radar in this case), and c) extreme sensor missingness corresponds to single-sensor predictions where all others are missing except one (optical in this case).

global availability. Thus, different configurations of missing data can be found in the EO domain, such as spectral, spatial, temporal, or sensor-wise [4]. For instance, optical sensors on board satellites can be affected by clouds, partially occluding some regions [5]. Moreover, data from a sensor can be affected by failures, errors, or limited spatial coverage, being partially or entirely missing during inference. For instance, the Landsat 7 satellite problem faced after 2003 (ETM+ SLC-off) [6], the NAIP satellite that operates only in the US, and the Sentinel-1b satellite that stopped its operation in 2021 [7].

In the deep learning field, the lack of sensors (or modalities) has been shown to greatly affect model performance [2]. Even recent models, such as transformers, are not naturally robust to this setting [8]. Similar results are observed in the EO domain. As expected, when the percentage of missing information in the input data increases (at spectral, spatial, temporal, or sensor dimensions), the model predictions get worse [9], [10], [5], [11]. This decline in predictive performance is associated with the robustness of the models to deal with missing data at the inference stage. For instance, Ofori-Ampofo et al. [10] show that additional sensors can increase the model robustness to cloudy conditions in optical Satellite Image Time Series (SITS) for crop classification. Similarly, Garnot et al. [5] observe that different fusion strategies make multi-sensor models more robust to missing optical sensor data. Moreover, Mena et al. [11] notice that simple data processing techniques that ignore the data from missing sensors improve model robustness compared to data imputation in vegetation-related applications.

Various multi-sensor models robust to missing sensors have been explored in the EO domain. Here, we identify two main missing sensor scenarios for inference, moderate sensor missingness (single-sensor missing) and extreme sensor missingness (single-sensor available), as illustrated in Fig. 1. These multi-sensor models can involve dedicated training, i.e., predefining a specific sensor to be missing (or be available) during inference [2]. For instance, the DisOptNet [12] model is pre-trained on optical images and then transferred to radar images via a hallucination branch. Similarly, Kampffmeyer et al. [13] use a hallucination branch from the optical image to imitate the missing depth image. On the other hand, the non-dedicated training considers models designed for arbitrary missing sensor scenarios. For instance, recent works consider simply sharing weight layers [14], or using the Sensor Dropout (SensD) technique. The SensD simulates missing sensor data randomly during training, which has been shown to improve robustness in multi-sensor images [15], multi-sensor time-series [16], or improve pre-training in multi-sensor self-supervision [17], [18], [19]. Moreover, Xie et al. [20] use a semi-supervised mutual distillation strategy where the predictions of modality-dedicated models are used to learn the common patterns in aerial images and point cloud data. Furthermore, in the direction of Geospatial Foundational Model (GeoFM), research efforts have explored sensor-agnostic approaches, such as OmniSat [21], and AnySat [22].

Multi-sensor models from the literature have some limitations in their robustness. As observed in [23], [14], [24], these models tend to improve robustness in moderate sensor missingness while disregarding the extreme cases. To address this issue, we introduce a multi-sensor model that is not only robust to moderate missing sensor scenarios but also to extreme cases. Our method employs a decision-level fusion strategy ignoring predictions from missing sensors and including the SensD technique, but differently from methods in the related literature, it is applied at the decision-level. As we observe in our experiments, this Decision-level Sensor Dropout (DSensD) method is incapable of overcoming all the aforementioned limitations. To this end, we consider a mutual distillation mechanism to further improve the modeling and to guide the single-sensor predictions toward a multi-sensor consensus. We named this method Decision-level Sensor Dropout with mutual distillation (DSensD+). The mutual distillation process [25] has been used in the EO domain by Gbodjo et al. [26] to improve the predictions of a multi-sensor model in land-cover classification. However, we use it directly at the decision-level, including a temperature hyper-parameter to guide the distillation process [27].

The motivation for using the decision-level SensD comes from evidence in the literature that approaches that fuse information at the last stages of the process, i.e. close to the output-level, are more robust to missing and noisy data [28],

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

[24]. For instance, Garnot et al. [5] show the effectiveness of decision-level dropout for missing images in SITS compared to feature and input levels. Moreover, Hong et al. [29] and Mena et al. [11] show the same robustness behavior, but when sensors are missing in multi-sensor data scenarios.

We validate our method in three EO datasets, considering binary, multi-class, and multi-label classification tasks involving crop and tree-mapping related applications. These datasets consider different sensor data types, such as time series data (SITS) and single-date images. During inference, we evaluate the models with fully or partially missing sensor data in the two aforementioned scenarios, moderate and extreme sensor missingness, shown in Fig. 1. Our method outperforms several recent approaches from the literature, as well as single-sensor baselines, across the considered scenarios. Our consistent improvements in various missing sensor scenarios clearly highlight the benefits of our approach in the considered classification tasks.

From a practical perspective, our method offers a key advantage: the components introduced to enhance robustness, SensD and mutual distillation, are only used during the training phase. During inference, it performs a direct decision-level fusion, simply ignoring predictions from any missing sensors. Our source code will be released upon acceptance.

Our work is organized as follows. In Section II, the related literature is discussed. The proposed method is introduced in Section III, the experimental assessment and the comparison with state-of-the-art methods are conducted in Section IV. Additionally, in Section V we discuss the key factors characterizing the behavior of our method and, at the same time, we perform an ablation analysis to disentangle the different components on which DSensD+ is built. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

A. LEARNING WITH MULTI-SENSOR DATA

Multi-sensor or multi-modal models have proven to be effective in the deep learning field [8], [2]. Gathering different sources of information to improve predictive performance is an inherent data modeling technique. Neural network models implement several fusion strategies [3], such as input-level, feature-level, and decision-level fusion, where the name suggests the level at which the fusion process takes place. Nowadays, full transformer-based models are used to improve multi-sensor modeling via multi-head attention mechanism [30]. However, it has been shown that multi-modal transformers are not naturally robust to missing modalities [8].

In the EO domain, there has been a recent increase in methods using multi-sensor data [1], [3]. For instance, Kussul et al. [31] use multi-spectral and radar images in land-use classification with a CNN model employing input-level fusion. However, several works have shown that using sensor-dedicated layers improves the performance [29], [5]. For example, Audebert et al. [32] introduce sensor-dedicated encoders for multi-spectral images and topographic maps, fusing across several layers in the model. Additionally, research in crop-type classification has shown that using specialized encoder architectures for SITS benefits the fusion process [10], [33]. Nowadays, there have been efforts in designing GeoFM using multi-sensor data. For instance, OmniSat [21] and AnySat [22] are self-supervised models employing feature-level fusion with sensor-dedicated encoders. These models require just fine-tuning the pretrained weights corresponding to the available sensors for a particular downstream task.

B. SIMULATE MISSING SENSOR DATA

Simulating missing data has become a widely adopted practice in the deep learning field. Utilized in various contexts, ranging from standard regularization to objective functions, e.g. reconstruction [34]. Here, the well-known dropout operator [35] used at the input can be viewed as a data augmentation technique. For instance, tokens can be masked out in text data to learn how to reconstruct them [36]. Similarly, time-steps can be masked out so that models learn to impute a signal [37]. Furthermore, modalities can be randomly dropped to increase model robustness, as shown in [30].

In the EO domain, models like SatMAE [17], Presto [18], and OmniSat [21] mask out part of the input data to set up a reconstruction task for self-supervised learning. Furthermore, the dropout can be applied to different dimensions of the EO data. For instance, Fasnacht et al. [9] introduce a spectral dropout to increase the model's robustness to missing spectral bands for hyper-spectral image segmentation. In the same task, Haut et al. [38] use spatial dropout (random occlusion) as a data augmentation technique, while Garnot et al. [5] present temporal dropout in SITS for crop-type segmentation. Furthermore, the SensD has been used to learn inter-sensor representations [39], to assess sensor contributions [40], and to increase model robustness to missing sensor data [14]. Most of the dropout usages as data augmentation are deployed at the input-level. Nonetheless, SensD can be applied at the feature level to avoid overfitting to dominant sensors [19] or to increase robustness to missing sensor data [15]. Up to our knowledge, there is no exploration of the SensD technique at the decision-level in the EO domain.

C. KNOWLEDGE AND MUTUAL DISTILLATION

Knowledge distillation is a process used to transfer knowledge from a complex model (teacher) to a simpler one (student) [27]. The transfer can be done offline, i.e. pre-train the teacher model and then learn the student one, or online, with a teacher updated from the moving average weights of the student [41]. This process has been adopted for multi-sensor models, e.g. with a multi-sensor teacher and single-sensor student [42], or from one sensor to another [43], known as cross-modal knowledge distillation. Moreover, the transfer can be done in a teacher-free setting between intermediate layers of the same model, known as self-distillation [44]. When this teacher-free framework uses inter-sensor/modal relations in multi-sensor/modal data, it is known as mutual distillation [25]. Self- and mutual distillation improve the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

performance without increasing the model parameters, as shown in Black et al. [45].

In the EO domain, knowledge distillation with multi-sensor data has been used for different purposes. For instance, Wang et al. [39] use an online knowledge distillation for self-supervised learning with optical and radar images in an input-level fusion model. Similarly, Astruc et al. [22] use it in a GeoFM. Bakalos et al. [46] use an offline knowledge distillation with a radar-elevation teacher model implementing input-level fusion and a radar-only student model. In addition, Kang et al. [12] use an offline approach in the cross-sensor distillation from an optical-sensor model to a radar-sensor model. The self-distillation between the multi-sensor prediction and per-sensor predictions has been used by Gbodjo et al. [26] to improve land-cover mapping. Furthermore, Xie et al. [20] have used mutual distillation to learn per-sensor models (optical images and point clouds) in a semi-supervised co-learning framework.

In our work, we explore the intersection of the previous points, with the aim of improving robustness to missing sensor data at the inference stage. Contrary to works that use the SensD technique in multi-sensor models at the input or feature level, here, we explore the SensD technique at the decision-level. More precisely, we introduce a mutual distillation process which, to the best of our knowledge, has not yet been investigated for increasing the robustness of multi-sensor models in the EO domain.

III. METHODOLOGY

To address the problem of missing sensor data, we introduce two methods (DSensD and DSensD+) based on the SensD technique with a particular focus on the extreme missing sensor scenario, as shown in Fig. 1c. First, we introduce an overview of the SensD technique (in Sec. B), then we describe how we adapt this technique to work at the decision-level by ignoring missing sensors (DSensD, in Sec. C). Furthermore, we discuss how we improve this method by including two additional loss functions, one based on the full-sensor data and another based on mutual distillation (DSensD+, in Sec. D).

A. NOTATION

The multi-sensor modeling scenario consists of having N training samples of paired sensors and corresponding ground truth labels as $\mathbb{D} = \{\mathbb{X}^{(i)}, y^{(i)}\}_{i=1}^{N}$. For each sample i, the label is $y^{(i)} \in \{1, 2, \dots, K\}$ and multi-sensor data is $\mathbb{X}^{(i)} = \{\boldsymbol{X}_s^{(i)}\}_{s\in\mathbb{S}}$, with $\mathbb{S}$ a set describing all sensors available for training. During inference, any subset of the sensors used for training can be available, i.e. $\tilde{\mathbb{S}} \subseteq \mathbb{S}$. Thus, the objective is to find a model $\mathcal{G}$ that approximates the label, regardless of the available sensors $\tilde{\mathbb{S}}$. Our setting considers sensor data $\boldsymbol{X}_s$ with temporal or static features, and models generating ***logits*** as prediction, $\hat{\boldsymbol{y}}^{(i)} \in \mathbb{R}^K$. Then, the learning process optimizes a loss function of the form $\mathcal{L}(y^{(i)}, \hat{\boldsymbol{y}}^{(i)})$. Here, we consider the cross entropy loss as $\mathcal{L}(p, \boldsymbol{q}) = -\sum_{k=1}^{K} \mathbb{1}(p = k) \log(\sigma(\boldsymbol{q})_k)$, where $\sigma(\cdot)$ is the softmax function, and $\mathbb{1}(\cdot)$ denotes the indicator function, which is equal to 1 if the argument holds and 0 otherwise.

We consider a full-data training scenario with arbitrary missing sensors at inference time. This framework is more general than dedicated training (i.e., pre-defined missing or available sensor) in the literature [13], [12], [46].

B. OVERVIEW ON SENSOR DROPOUT

To fuse multi-sensor data, the well-known feature-level fusion can be used [3]. This strategy uses a dedicated encoder per sensor $\mathcal{G}_s^{\text{enc}}$ that extracts a high-level representation as

$$\boldsymbol{z}_s^{(i)} = \mathcal{G}_s^{\text{enc}}(\boldsymbol{X}_s^{(i)}) \;\; \forall s \in \mathbb{S}, \tag{1}$$

with $\boldsymbol{z}_s^{(i)} \in \mathbb{R}^d$ the extracted features of a sensor s. Before merging the extracted features, the Sensor Dropout (SensD) technique is used to increase the robustness of the model to missing sensors, as in [14], [19], [15]. This technique simulates missing sensor data during training via random sensor dropping. For example, replacing features associated to missing sensors with zeros, known as zero-out [39], [14]: $\tilde{\boldsymbol{z}}_s^{(i)} = (1 - d_s^{(i)}) \cdot \boldsymbol{z}_s^{(i)}$, with $d_s^{(i)} \sim \text{Bern}(\alpha)$ the binary decision (based on the dropout ratio α) if the sensor s is masked out in the sample i. Another way is to replace features of dropped sensors with a learnable token $\boldsymbol{t}$, as in [21]: $\tilde{\boldsymbol{z}}_s^{(i)} = (1 - d_s^{(i)}) \cdot \boldsymbol{z}_s^{(i)} + d_s^{(i)} \cdot \boldsymbol{t}$. Moreover, the features can be ignored when using a dynamic aggregation function $\text{merge}(\cdot)$, such as with average [16] or masked attention [19], which can be expressed by

$$\boldsymbol{z}_{\text{miss}}^{(i)} = \text{merge}\left(\left\{d_s^{(i)}, \boldsymbol{z}_s^{(i)}\right\}_{s\in\mathbb{S}}\right), \tag{2}$$

where $\boldsymbol{z}_{\text{miss}}^{(i)} \in \mathbb{R}^{d_{\text{fus}}}$ is the fused representation with missing sensors, and $\text{merge}(\cdot)$ has two arguments, the per-sensor binary decision to mask out $d_s^{(i)}$ and the per-sensor features $\boldsymbol{z}_s^{(i)}$. Finally, the fused representation is used to obtain the prediction of the model: $\hat{\boldsymbol{y}}^{(i)} = \mathcal{G}^{\text{head}}(\boldsymbol{z}_{\text{miss}}^{(i)})$. This approach enforces the prediction head to be resilient to the lack of sensor information, as some sensor data may not be available (after the SensD). In the following, we introduce an alternative that uses an ensemble of sensor-dedicated models.

C. DECISION-LEVEL SENSOR DROPOUT

Inspired by the ensemble framework, we propose to leverage the SensD technique at the decision-level in a multi-sensor model implementing the decision-level fusion. Hence, we expect greater flexibility in ignoring the information from unavailable sensors. Considering a dedicated per sensor model $\mathcal{G}_s = \mathcal{G}_s^{\text{head}} \circ \mathcal{G}_s^{\text{enc}}$, composed of an encoder and prediction head, that provides predictions in the form of logits, the decision-level fusion model is expressed by

$$\hat{\boldsymbol{y}}_s^{(i)} = \text{normalize}(\mathcal{G}_s(\boldsymbol{X}_s^{(i)})) \;\; \forall s \in \mathbb{S}, \tag{3}$$

$$\hat{\boldsymbol{y}}_{\text{full}}^{(i)} = \frac{1}{|\mathbb{S}|} \sum_{s\in\mathbb{S}} \hat{\boldsymbol{y}}_s^{(i)}, \tag{4}$$

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

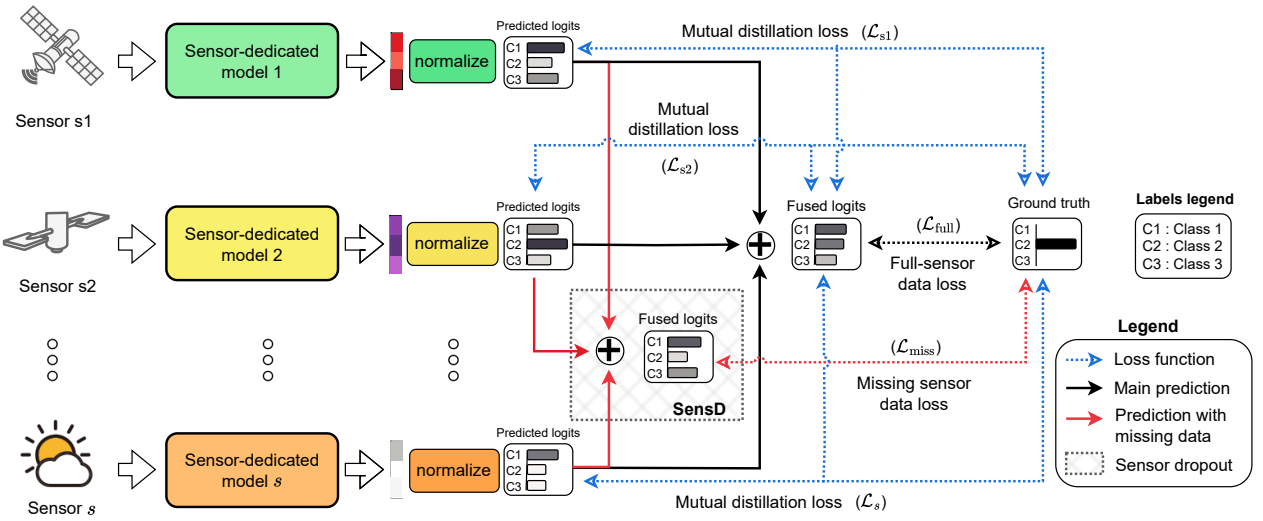


FIGURE 2. Illustration of the Decision-level Sensor Dropout with mutual distillation (DSensD+) method during training, where the fusion, SensD, and mutual distillation occur at the decision-level. During inference, the forward pass of the multi-sensor model corresponds to only the dark arrows.

with $\mathrm{normalize}(\cdot)$ the z-score normalization function, $\hat{\boldsymbol{y}}_s^{(i)} \in \mathbb{R}^K$ the normalized predicted logits for a sensor s, and $\hat{\boldsymbol{y}}_{\text{full}}^{(i)} \in \mathbb{R}^K$ the fused prediction (via average) of the multi-sensor data. We found that normalizing the logits prior to fusion enhances the performance of our decision-level fusion model, which is consistent with recent findings from the literature [47]. This is because it avoids optimizing the magnitude of the logits, focusing on just the directions. In this context, the SensD technique, which replaces missing input data with zeros or learnable tokens, proves to be less effective as it introduces a strong bias in the model decision, which cannot be recovered in subsequent parts of the method. Therefore, we apply the SensD by ignoring the predictions of missing sensors from the merge function, given by

$$\hat{\boldsymbol{y}}_{\text{miss}}^{(i)} = \frac{1}{\sum_{s\in\mathbb{S}}(1-d_s^{(i)})} \sum_{s\in\mathbb{S}} (1-d_s^{(i)}) \cdot \hat{\boldsymbol{y}}_s^{(i)} \tag{5}$$

with $d_s^{(i)} \sim \mathrm{Bern}(\alpha)$ the randomly drawn binary decision if the sensor s is masked out, and $\hat{\boldsymbol{y}}_{\text{miss}}^{(i)} \in \mathbb{R}^K$ the fused prediction with random sensors dropped. To have an unbiased sampling of the dropped sensors (i.e. all configurations are equally likely), we use $\alpha = 0.5$, while an in-depth analysis of this α hyper-parameter is shown in the experiments. Thus, if $\mathbb{S} = \{\text{optical}, \text{radar}\}$, then $\boldsymbol{d}^{(i)} = [d_{\text{optical}}^{(i)}, d_{\text{radar}}^{(i)}]$ can be either $[0,1]$, $[1,0]$, or $[1,1]$. Hence, this method implementing decision-level fusion, named **Decision-level Sensor Dropout (DSensD)**, optimizes a loss function using the predictions with randomly missing sensors per sample (or per batch as in [14]), expressed by

$$\mathcal{L}_{\text{miss}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(y^{(i)},\ \hat{\boldsymbol{y}}_{\text{miss}}^{(i)}\right) . \tag{6}$$

However, the learning of multi-sensor models over $\mathcal{L}_{\text{miss}}$ has a drawback in worsening the full-sensor predictions, $\hat{\boldsymbol{y}}_{\text{full}}^{(i)}$, as shown in the literature [23], [14]. This is because the full-sensor data is just one among all missing sensor cases. The number of possible combinations is given by the power set of all sensors, $\mathcal{P}(\mathbb{S})$, minus the no-sensor case, i.e., $2^{|\mathbb{S}|} - 1$.

D. IMPROVING DECISION-LEVEL SENSOR DROPOUT

To improve the drawbacks of optimizing multi-sensor models on just $\mathcal{L}_{\text{miss}}$, Eq. (6), we consider the previous formulation but include both losses. The $\mathcal{L}_{\text{miss}}$ that is based on predictions with randomly missing data, and another based on the predictions with full-sensor data $\hat{\boldsymbol{y}}_{\text{full}}^{(i)}$, Eq. (4), given by

$$\mathcal{L}_{\text{full}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(y^{(i)},\ \hat{\boldsymbol{y}}_{\text{full}}^{(i)}\right) . \tag{7}$$

The goal is that the multi-sensor model has access to data patterns of the affected scenario (with missing data), as well as the full-data scenario (no missing data). Furthermore, the ensemble of sensor-dedicated models can be guided into the common knowledge, as in [25]. To this end, we incorporate a mutual distillation loss term based on the prediction of each sensor-dedicated model $\hat{\boldsymbol{y}}_s^{(i)}$, Eq. (3), the fused prediction $\hat{\boldsymbol{y}}_{\text{full}}^{(i)}$, Eq. (4), and the real label $y^{(i)}$. Thus, the mutual distillation loss on each sensor s corresponds to

$$\mathcal{L}_s = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(y^{(i)},\ \hat{\boldsymbol{y}}_s^{(i)}\right) + \lambda \cdot \mathcal{L}_{\text{kd}}\left(\hat{\boldsymbol{y}}_{\text{full}}^{(i)},\ \hat{\boldsymbol{y}}_s^{(i)}; \tau\right), \tag{8}$$

with $\mathcal{L}_{\text{kd}}$ the knowledge distillation function parametrized by a temperature $\tau > 0$, and λ a weighting factor on this. Here, we use the Kullback–Leibler divergence as $\mathcal{L}_{\text{kd}}(\hat{\boldsymbol{y}}_{\text{full}}^{(i)},\ \hat{\boldsymbol{y}}_s^{(i)}; \tau) = \mathcal{D}_{\text{KL}}(\sigma(\hat{\boldsymbol{y}}_{\text{full}}^{(i)}/\tau), \sigma(\hat{\boldsymbol{y}}_s^{(i)}/\tau))$,where

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

$\mathcal{D}_{\text{KL}}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{k=1}^{K} \boldsymbol{p}_k \log(\boldsymbol{p}_k / \boldsymbol{q}_k)$. We divide the logits by the temperature hyper-parameter τ before the softmax function $\sigma(\cdot)$, as in [27]. In this way, the $\mathcal{L}_{\text{kd}}$ function aims to transfer the knowledge between the temperature-softened per-sensor prediction ($\hat{\boldsymbol{y}}_s^{(i)}$) and the consensus prediction ($\hat{\boldsymbol{y}}_{\text{full}}^{(i)}$). Preliminary experimentation led us to find that τ^2 is an appropriate value for the weight λ in Eq. (8), as also suggested in [27], [45], while an in-depth analysis on the sensitivity of our method to the τ hyper-parameter is shown in the experiments. The final loss function optimized by our multi-sensor model is the following

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{miss}} + \mathcal{L}_{\text{full}} + \frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}} \mathcal{L}_s \ . \tag{9}$$

We consider a uniform sum of the loss terms as it offers the best balance and trade-off between all components that the model has to optimize. A schema of this method, named **Decision-level Sensor Dropout with mutual distillation (DSensD+)**, is depicted in Fig. 2.

IV. EXPERIMENTS

A. DATASETS

We use three classification benchmarks to validate our proposal, i.e., two variants of the CropHarvest dataset [48] and the TreeSatAI-TS dataset [21]. These include sensor data with static (single-date) and temporal information, as well as image and time-series data. The summary of the sensors used in each dataset is reported in Table 1.

1) CropHarvest

We use the CropHarvest dataset for crop recognition, which provides four sensor data [48]. The **CropHarvest binary (CropH-b)** involves a cropland (binary) classification task. The dataset has 69,800 samples around the globe, collected between 2016 and 2021. Each sample has three temporal sensor data at $10[m]$ of spatial resolution: multi-spectral optical SITS, radar SITS, and weather data. Besides, the samples have one static sensor data, the topographic information interpolated to the same spatial resolution as the other sensor data.

In addition, we use a multi-class version of CropHarvest, named **CropHarvest multi (CropH-m)**. This is a subset with 29,642 samples and 10 crop types to distinguish. Besides, the same sensor features from the CropH-b dataset are used.

Since there is no test partition in this dataset, we use a standard 10-fold cross-validation for assessing the model's performance.

2) TreeSatAI-TS

We use the TreeSatAI dataset version extended with temporal sensor data [21], named **TreeSatAI-Time-Series (TreeSatAI-TS)**. This involves a multi-label classification task between 15 tree species. There are 38,520 samples for training, 6,810 for validation, and 5,044 for testing, collected between 2017 and 2020 in Germany. Each sample has two temporal sensor data at $10[m]$ spatial resolution: optical and radar SITS. Additionally, a single-date image is provided at a high spatial resolution of $0.2[m]$. This aerial-based optical image captures RGB and infrared bands.

TABLE 1. Overview of sensor data available in each dataset.

Dataset	Sensor	Bands	Spatial	Temporal
CropH-b & CropH-m	optical SITS	11	$10[m]$	1/month (1 year)
	radar SITS	2	$10[m]$	1/month (1 year)
	weather	2	$10[m]$	1/month (1 year)
	topographic	2	$10[m]$	None
TreeSatAI-TS	optical SITS	12	$10[m]$	∼10/month (1 year)
	radar SITS	2	$10[m]$	∼20/month (1 year)
	aerial	4	$0.2[m]$	None

B. EXPERIMENTAL SETUP

1) Competing Methods

We include the results of a model trained individually on each sensor data as single-sensor baselines. Furthermore, we consider seven recent multi-sensor models from the EO domain that can manage binary, multi-class, and multi-label classification tasks. **FSensD** [19], a model equipped with cross-attention (transformer-based) that uses SensD during training. We use the supervised version described in [19] that masks out the dropped sensors of the attention. **OOD-f** [28], a concatenation-based model that simulates all cases of missing sensors during training (with zero-out). We extended the two-sensor formulation in [28] to a multi-sensor setting. **FCoM-av** [16], a method simulating all cases of missing sensors during training by ignoring the missing features with average as aggregation. **TIMML** [15], a method with cross-attention (transformer-based) that uses SensD (with zero-out) and auxiliary prediction losses per sensor during training. While the previous methods employ a feature-level fusion, **ESensI** [14] is a sensor-invariant ensemble that uses shared layers in the per-sensor prediction head and sensor encoding. Moreover, **Anysat** [22], is a transformer-based GeoFM that works with several sensors, including aerial data, optical, and radar SITS. However, weather and topography from CropHarvest datasets are not included in the Anysat model. This model is pre-trained with a teacher-student distillation process and self-supervised contrastive learning [22]. **Galileo** [49], is a recent transformer-based GeoFM based on per sensor encoder and self-supervised contrastive learning.

In addition, we include **ISensD** [14] for the CropHarvest datasets. This method uses an input-level fusion strategy with SensD during training (zero-out). We refrain from deploying the **ISensD** method for the TreeSatAI-TS benchmark since its application for this dataset is not straightforward due to the differences in sensor characteristics, both spatial and temporal resolutions, as shown in Table 1. We were unable to include the Galileo GeoFM for the TreeSatAI-TS data based on computational resources. Nevertheless, we include the results of the **OmniSat** GeoFM [21] in the TreeSatAI-TS dataset. Similar to other GeoFM, **OmniSat** is based on self-supervised contrastive learning.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

TABLE 2. Weighted F1 scores when different sensor combinations are available for inference in the CropH-b dataset. The **best** and second-best values are highlighted. *Averaged among the available cases.

Sensor combinations														
optical SITS	radar SITS	weather	topo-graphy	Single	ISensD	FSensD	OOD-f	FCoM-av	TIMML	ESensI	Anysat	Galileo	DSensD	DSensD+
✓	✓	✓	✓		81.8	81.7	**84.7**	**84.7**	82.9	81.5	81.2	82.3	82.6	83.6
-	✓	✓	✓		68.9	78.4	**82.4**	81.7	79.9	78.4			79.7	80.3
✓	-	✓	✓		81.2	81.5	84.1	**84.5**	82.7	82.5			82.6	83.6
✓	✓	-	✓		81.1	79.6	81.3	**82.9**	81.0	79.8			80.6	82.1
✓	✓	✓	-			81.9	**84.3**	**84.3**	82.9	81.1			82.5	83.5
✓	-	-	-	81.8	80.5	79.6	79.5	81.6	80.9	80.5	**82.0**	81.3	80.2	**82.0**
-	✓	-	-	71.4	36.9	68.9	69.3	64.4	70.0	70.9	70.4	71.3	70.9	**71.8**
-	-	✓	-	77.7	56.2	75.3	77.1	75.2	76.1	77.4		78.1	76.5	**78.2**
-	-	-	✓	67.5		46.1	67.1	65.8	64.2	66.5		**69.2**	67.8	68.4
Average				74.6	68.5*	74.8	75.5	78.3	77.8	77.6	77.9*	76.4*	78.2	**79.3**

TABLE 3. Weighted F1 scores when different sensor combinations are available for inference in the CropH-m dataset. The **best** and second-best values are highlighted. *Averaged among the available cases.

Sensor combinations														
optical SITS	radar SITS	weather	topo-graphy	Single	ISensD	FSensD	OOD-f	FCoM-av	TIMML	ESensI	Anysat	Galileo	DSensD	DSensD+
✓	✓	✓	✓		71.8	69.2	73.3	**76.7**	72.4	70.8	71.2	69.7	74.1	76.5
-	✓	✓	✓		35.1	60.4	65.1	**66.6**	63.1	59.5			63.8	65.9
✓	-	✓	✓		70.3	67.8	69.7	74.6	71.0	69.7			72.6	**75.4**
✓	✓	-	✓		71.1	68.8	72.0	76.2	72.7	72.5			74.0	**76.4**
✓	✓	✓	-			69.2	73.2	76.5	72.5	70.6			74.4	**76.8**
✓	-	-	-	71.0	69.6	67.4	69.2	73.8	71.5	72.0	70.9	70.5	72.4	**75.3**
-	✓	-	-	56.0	10.9	46.1	51.4	53.6	53.6	54.6	51.5	50.4	53.7	**57.3**
-	-	✓	-	46.7	10.1	41.5	42.6	42.8	44.3	45.0		48.3	46.1	**48.4**
-	-	-	✓	28.0		15.7	14.3	19.1	2.6	14.8		30.1	25.3	**31.5**
Average				50.4	52.7*	56.2	59.0	62.2	58.2	58.8	64.6*	53.8*	61.8	**64.9**

Following standard practices in the literature [22], [21], the pre-trained models are fine-tuned specifically for each single-sensor, i.e., they employ dedicated training.

2) Implementation

We adopt a z-score normalization to the input data to scale different magnitudes. Besides, we use standard encoders for each sensor data. For all temporal sensor data, we use TempCNN, a 1D convolutional network that involves convolutions over the time dimension [50]. For the pixel-wise topographic information (in the CropHarvest data), we use a standard MLP, while for the aerial image (in the TreeSatAI-TS data), we use ResNet-50, a 2D convolutional neural network with skip connections [51]. For all encoders (apart from ResNet), we use two hidden layers and a final embedding layer of 128 units with 20% of dropout ratio. On top of the encoders, we use an MLP of 128 units, with dropout, batch normalization, and a final linear layer for the prediction. All other architecture hyper-parameters of the competitors are set as in the original proposals. For optimization, we use the Adam optimizer with a learning rate of 0.001, while the original optimizer is used in the competitors. For a fair comparison, we train all the competing methods over 100 epochs with a batch size of 128 and an early stopping criterion with a patience of 5. In our model, the stopping criterion is applied over the full-sensor data loss, i.e. Eq. (7).

3) Evaluation

We train all competing methods in a full-sensor data setup, simulating missing sensors during inference to assess the model robustness, i.e. the drop in predictive performance with missing data. We experiment with two cases of *sensor missingness*, see Fig. 1 for an illustration: i) moderate sensor missingness, when only one sensor is missing, and ii) extreme sensor missingness, when all sensors are missing except one. Moreover, we include the results with no missing data for reference.

For evaluation, we use the weighted F1 score across all experiments (similar results were observed with other performance metrics). For each method, we report results averaged over five runs.

C. RESULTS WITH WHOLE MISSING SENSORS

We show the main results with missing sensors in Tables 2, 3, and 4, for CropH-b, CropH-m, and TreeSatAI-TS dataset, respectively. This evaluation assesses the predictive performance when the data from specific sensors is missing in all samples at the inference stage.

In the CropH-b data, we notice that the models OOD-f and FCoM-av obtain the best results in the moderate sensor missingness. However, our DSensD+ model obtains the second-best result in some of these moderate missing sensor cases, such as when the weather or topography sensor is missing. Furthermore, in extreme sensor missingness, our DSensD+ model outperforms all competitors when different

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

TABLE 4. Weighted F1 scores when different sensor combinations are available for inference in the TreeSatAI-TS dataset. The **best** and second-best values are highlighted. † values are from the original paper. *Averaged among the available cases.

Sensor combinations												
optical SITS	radar SITS	aerial	Single	FSensD	OOD-f	FCoM-av	TIMML	ESensI	OmniSat†	AnySat	DSensD	DSensD+
✓	✓	✓		62.7	68.5	68.7	62.4	62.5	**73.3**	67.5	68.4	70.5
-	✓	✓		60.8	66.1	**67.3**	61.6	65.8			65.5	65.9
✓	-	✓		62.4	67.0	69.1	61.9	65.0			69.2	**70.2**
✓	✓	-		57.7	32.3	61.5	56.3	55.5			64.2	**68.0**
✓	-	-	65.2	53.4	7.9	59.3	62.1	51.8	49.7	**75.0**	63.1	67.5
-	✓	-	57.3	51.8	13.2	54.4	55.1	52.3	55.9	51.9	55.0	**57.7**
-	-	✓	62.0	58.3	63.5	65.5	61.8	61.6	**71.0**	62.2	63.3	63.4
Average			61.5	58.2	45.5	63.7	60.2	59.2	62.5*	64.1*	64.1	**66.2**

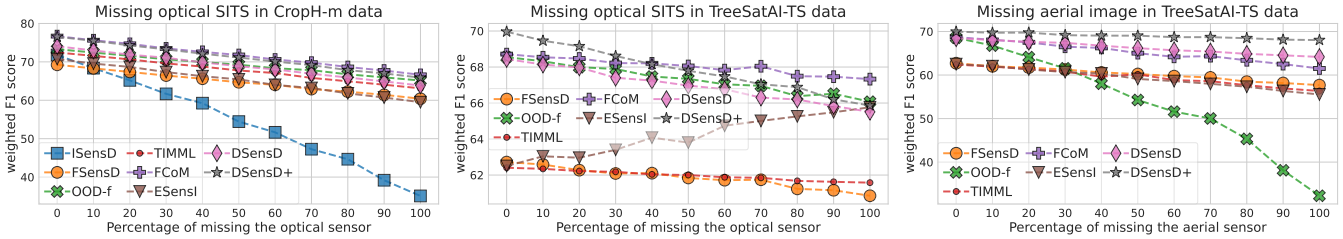


FIGURE 3. Robustness results when a specific sensor is missing for a given percentage of samples at inference.

single-sensors are available for inference, except for the topography. In this case, our DSensD+ obtains the second-best result, outperformed only by Galileo GeoFM. In the full-data prediction (no missing sensors), the DSensD+ model is the second-best after FCoM-av/TIMML. Nevertheless, with this model, we obtain the best overall score in the evaluated scenarios, and with DSensD, the third-best. This overall score considers the full-sensor data and different missing sensor cases, which is illustrated in the "average" row in the tables.

In the CropH-m data, we observe similar results to those in CropH-b. The difference is that the DSensD+ model is more competitive in the moderate missing sensor cases. Here, DSensD+ is the best in all missing sensor scenarios (moderate and extreme) except for the case in which optical SITS is missing, where it is the second-best after the FCoM-av model. Moreover, the simplified model version, DSensD, is among the top four in the overall score. In addition, our DSensD+ is the only one that outperforms all single-sensor baselines in both datasets CropH-m and CropH-b. This improvement over the single-sensor baselines is not straightforward, as several competitors fail to surpass them. Overall, the results clearly demonstrate the robustness of our method for the diverse moderate and extreme missing sensor scenarios that appear in both the binary and multi-class crop classification tasks.

We notice more variability in the results of the TreeSatAI-TS dataset. Likewise, in CropH-m and CropH-b datasets, our DSensD+ model obtains the second-best results when there is no missing data, only outperformed by the OmniSat model. Nevertheless, our model is the best in half of the missing sensor scenarios (moderate and extreme), being among the top three best in the other cases. Here, we observe that the best result depends on the specific missing sensors case. For instance, when optical SITS is missing, the best is FCoM-av, while Omnisat and Anysat are the best when optical SITS or aerial sensor is the only one available, respectively. Moreover, the DSensD+ achieves the best overall performance, illustrated in the average across the different missing sensors and full-data scenarios. Thus, the DSensD obtains the second-best results overall, tying with Anysat. These results show the effectiveness of both methods DSensD and DSensD+ for the missing sensor problem, not just in binary and multi-class tasks but also in multi-label classification.

D. RESULTS WITH PARTIAL MISSING SENSORS

For an in-depth analysis of the robustness to different scenarios, we consider partial missing sensors in the moderate case. To this end, we display the robustness of the models when a specific sensor is missing in just a subset of the inference samples, depicted in Fig. 3. These unavailable sensors represent the most significant variations in the results. These are optical SITS missing in CropH-m and TreeSatAI-TS datasets, as well as aerial missing in TreeSatAI-TS. The Anysat, Omnisat, and Galileo models are excluded from this analysis as they only operate in single-sensor fine-tuning.

The tendency in the behavior of the compared models is to have high robustness, depicted in a low F1 decay when a specific sensor is missing in more samples. However, a few models have a pronounced linear decay in the performance for specific missing sensors, i.e., low robustness. For instance, the ISensD model in CropH-b, FSensD model in CropH-m, and OOD-f model in TreeSatAI-TS datasets. Moreover, some models exhibited better behaviour in specific percentage intervals of missing sensors. For instance, when optical SITS is missing in the TreeSatAI-TS dataset. In this case, the best results are from the DSensD+ model when the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

TABLE 5. Ablation analysis using the weighted F1 scores in the CropH-m dataset with extreme missing evaluation (single-sensor predictions).

Variant	$\mathcal{L}_{\text{miss}}$	$\mathcal{L}_{\text{full}}$	$\mathcal{L}_s$	Full data	optical SITS	radar SITS	weather	topography	Average
DSensD+	✓	✓	✓	**77.2**	75.2	57.9	48.8	31.8	**58.2**
→ w/o norm	✓	✓	✓	75.6	74.2	56.7	47.8	31.2	57.1
→ w/o $\mathcal{L}_s$	✓	✓	-	74.4	74.4	**58.1**	**49.0**	**33.8**	57.9
→ w/o $\mathcal{L}_{\text{full}}$	✓	-	✓	75.0	**75.4**	**58.1**	**49.0**	32.0	57.9
→ w/o $\mathcal{L}_{\text{miss}}$	-	✓	✓	76.0	74.9	57.9	48.8	31.8	57.9
→ only $\mathcal{L}_s$	-	-	✓	68.5	74.4	57.7	48.8	33.0	56.5
DSensD	✓	-	-	73.9	72.4	53.8	46.0	25.9	54.4
Davg	-	✓	-	75.5	71.6	47.0	41.7	17.6	50.7

TABLE 6. Ablation analysis using the weighted F1 scores in the TreeSatAI-TS dataset with extreme missing evaluation (single-sensor prediction).

Variant	$\mathcal{L}_{\text{miss}}$	$\mathcal{L}_{\text{full}}$	$\mathcal{L}_s$	Full data	optical SITS	radar SITS	aerial	Average
DSensD+	✓	✓	✓	70.0	**67.9**	57.3	63.8	**64.7**
→ w/o norm	✓	✓	✓	67.4	61.2	56.0	**64.7**	62.3
→ w/o $\mathcal{L}_s$	✓	✓	-	**70.6**	64.1	54.8	63.0	63.1
→ w/o $\mathcal{L}_{\text{full}}$	✓	-	✓	69.0	67.8	58.0	63.1	64.5
→ w/o $\mathcal{L}_{\text{miss}}$	-	✓	✓	69.3	67.1	**58.5**	63.6	64.6
→ only $\mathcal{L}_s$	-	-	✓	67.4	67.6	58.4	63.9	64.3
DSensD	✓	-	-	68.0	63.0	55.3	63.4	62.4
Davg	-	✓	-	23.8	23.6	23.1	24.5	23.7

percentage is below 50%, after this, the FCoM-av model is the best (this aligns with the results shown in Table 4). The robustness curve of our DSensD is in the middle between the compared models, while DSensD+ has better behavior, especially when the aerial sensor is missing in the TreeSatAI-TS dataset. These results underline that our DSensD+ consistently exhibits a robust behavior across different missing sensor scenarios and conditions.

Overall, the better robustness of the decision-level SensD compared to input- and feature-level ones is aligned with the literature. For example, Hong et al. [29] and Mena et al. [11] show the effectiveness in achieving enhanced robustness of decision-level approaches in scenarios of missing sensor data. Thus, the evidence suggests that approaches performing fusion close to the output-level tend to be more robust to missing and noisy data [28], [24].

V. DISCUSSION

A. MODELING ALTERNATIVES

Here we study the effect of removing different modeling components in our method, providing an ablation analysis of these. This is shown in Table 5 for CropH-m dataset and in Table 6 for TreeSatAI-TS dataset. We notice that the normalization in Eq. (3) improves the modeling, as the results without this step decrease the performance by up to 3 points in some cases. Here, the Davg corresponds to the standard decision-level fusion model without SensD, while DSensD is the decision-level fusion model optimized only with the predictions after the SensD, i.e. $\mathcal{L}_{\text{miss}}$. When comparing Davg to DSensD model, we notice that the SensD brings benefits in the missing sensor scenarios. However, we observe a performance decline in the full-sensor data predictions for the CropH-m dataset. This phenomenon has been previously underlined in the literature when implementing SensD at the feature level [23], [14]. It can be attributed to the fact that full-sensor data represents just one scenario among numerous possible missing sensor combinations. Moreover, the DSensD+ model consistently outperforms both the DSensD and Davg models across all metrics. These results clearly emphasize the effectiveness of our proposed additional loss functions in addressing the limitations inherent to the standard SensD technique.

Furthermore, our analysis reveals that the mutual distillation loss (linked to the per-sensor predictions) contributes to model robustness, outweighing the impact of both the full-sensor and missing sensor losses. Among these components, the missing sensor loss demonstrates the least impact when removed from our formulation. Conversely, the full-sensor loss provides minimal contribution to robustness when applied alone. Notably, incorporating all loss functions does not consistently yield optimal results across all scenarios, particularly in full-sensor and single-sensor predictions. Nevertheless, this comprehensive formulation achieves an excellent balance between standard full-data performance and overall robustness, as evidenced by the average performances attained by DSensD+ across both datasets.

B. EFFECT OF DROPOUT RATIO

We assess the stability of our method to the choice of the dropout ratio α in the SensD technique ($d_s^{(i)}$ in Eq. (5)) by the performance with full-sensor data and extreme missing sensors. The results are depicted in Fig. 4 for the CropH-m dataset and in Fig. 5 for the TreeSatAI-TS dataset. We notice that the relative effect of the SensD ratio in the predictions may vary depending on the sensor and dataset under evaluation. For instance, in the CropH-m data, the variation in the single-sensor predictions is more evident for the radar and topography sensors, while for the optical sensor, the effect is negligible. Contrary, in the TreeSatAI-TS data, the most considerable variation in the single-sensor predictions

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

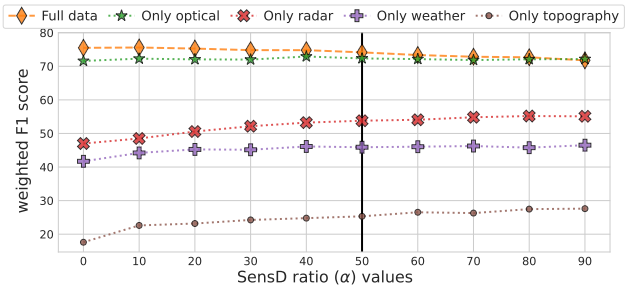


FIGURE 4. Predictive performance of the DSensD+ model using different ratios (α) in the SensD technique. Results are from the CropH-m data.

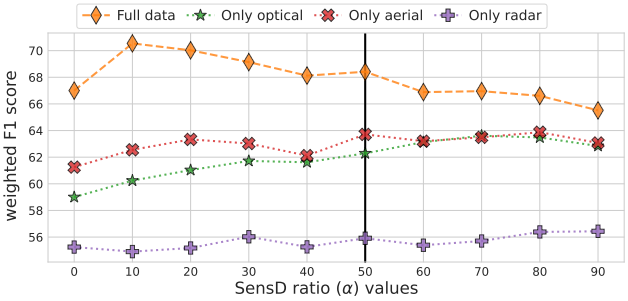


FIGURE 5. Predictive performance of the DSensD+ model using different ratios (α) in the SensD technique. Results are from the TreeSatAI-TS data.

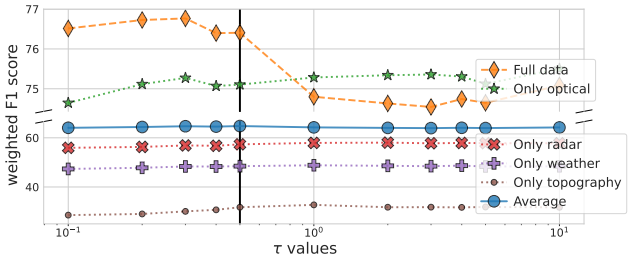


FIGURE 6. Predictive performance of the DSensD+ model using different temperature values in the distillation. Results are from the CropH-m data.

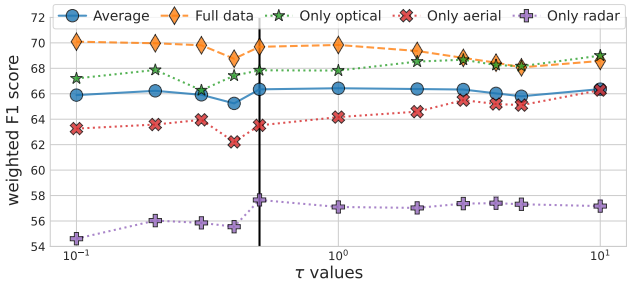


FIGURE 7. Predictive performance of the DSensD+ model using different temperature values in the distillation. Results are from the TreeSatAI-TS data.

is obtained with the optical sensor.

Overall, we observe that when the dropout ratio increases, the single-sensor predictions tend to improve, i.e., the robustness increases. However, the full-sensor performance tends to decrease. This behavior can be observed clearly in the TreeSatAI-TS dataset. This is expected, as there will be more samples with missing sensor data than full-sensor data when the SensD ratio increases. A good balance between the performance with and without missing sensors is reached when using $\alpha = 0.5$, which means that each sensor s will not be accessible for approximately half of the samples. We use this value for both methods DSensD and DSensD+.

C. EFFECT OF DISTILLATION TEMPERATURE

We analyze the effect of the distillation temperature τ (in $\mathcal{L}_{\text{kd}}$ from Eq. (8)) in the predictive performance with full-sensor data and extreme missing sensors. Following [27], we try both low and high values under a grid search exploration, as this temperature controls how soft the probability distributions are in the mutual distillation process. This is shown in Fig. 6 and Fig. 7 for the CropH-m and TreeSatAI-TS datasets, respectively. Here, the full-sensor data prediction tends to decrease when the temperature increases. This is expected as the loss weight λ in Eq. (8) is set as τ^2, i.e. the model tends to focus on the single-sensor predictions instead of the other losses. As noted in the literature [27], [46], the single-sensor predictions (derived from the mutual distillation process) tend to improve in the softer version (i.e., with a higher temperature).

We notice that there are significant differences in the predictive performance when the temperature changes slightly. This means that this is an essential hyper-parameter to tune in the mutual distillation process. Nevertheless, a value lower than 1 tends to provide a good balance in the results, especially in the CropH-m dataset. This balance corresponds to a good predictive performance with both full-sensor and missing sensor data. Based on this criterion, we set the temperature hyper-parameter to $\tau = 0.5$ for all our experiments involving the DSensD+ method.

D. EFFECT OF ENCODER ARCHITECTURE

Here we analyze the behaviour of the DSensD+ method when different temporal encoders are used to ingest time-series data, i.e. optical SITS, radar SITS, and weather. As encoders, we consider Multi-Layer Perceptron (MLP), Long Short Term Memory (LSTM), Transformer [52], ConvTran [53], and the one used in the main results, TempCNN. The results for the different datasets are shown in Fig. 8. We observe that the encoder that performs the worst depends on both the dataset and the sensor that is missing. For instance, the lowest scores are obtained with Transformer in the CropHarvest datasets, while in the TreeSatAI-TS dataset, the MLP achieves the lowest score. These results are aligned with the literature [8], noting that these fully connected and multi-head attention-based models are not naturally robust to missing data and require additional generalization components. Overall, the TempCNN architecture allows the DSensD+ model to exploit the missing patterns and improve its behaviour for different missing sensor scenarios, espe-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

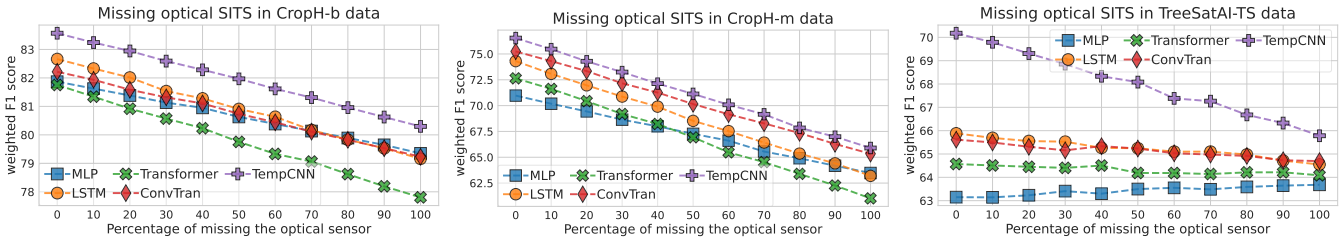


FIGURE 8. Robustness results for different encoder architectures used in the temporal sensor data in the DSensD+ model.

cially for the TreeSatAI-TS dataset. The importance of the encoder architecture is sometimes omitted in the literature. However, it allows extracting useful sensor-dedicated features that translate into better single-sensor predictions.

The robustness pattern of the DSensD+ model remains consistent across most tested temporal encoder architectures. However, superior performance is achieved in conjunction with the TempCNN encoder. This can likely be attributed to our hyper-parameter optimization process. Specifically, the method's key hyper-parameters ($\alpha = 0.5$ and $\tau = 0.5$) were selected on the validation set using this particular architecture. For optimal generalization across diverse encoder architectures, the DSensD+ method would benefit from architecture-specific hyper-parameter tuning.

E. COMPUTATIONAL USAGE & EXECUTION TIME

We depict the number of parameters and memory usage of all compared models in Table 7. We note that the input-level fusion model, ISensD, has the lowest number of parameters. This is expected based on the single encoder architecture that the model uses. On the other side, the models with the highest number of parameters are related to the GeoFMs, Anysat and Galileo, followed by the transformer-based ones, TIMML and then FSensD. A similar pattern is followed in the memory usage of the models, as this directly relates to the number of parameters the models have. Our DSensD and DSensD+ have the same number of parameters, while DSensD+ has extra memory usage due to the additional prediction from the full-sensor data compared to DSensD. Despite the prediction head per sensor used in our methods, the memory usage is similar or even more efficient compared to models with a single prediction head, e.g., FCoM-av, OOD-f, FSensD, and TIMML. This shows the ability of our methods to enhance robustness to missing sensors at the inference stage while matching or even lowering the computational requirements of state-of-the-art approaches.

TABLE 7. Learnable parameters and memory use of multi-sensor models calculated in the CropH-m dataset with a 128-batch-size. MB stands for megabytes.

Model	Parameters (millions)	Memory usage (MB)
ISensD	1.05	51.04
FSensD	3.70	155.22
OOD-f	3.15	103.12
FCoM-av	3.10	100.70
TIMML	4.38	196.17
ESensI	3.10	90.10
AnySat	252.01	1051.34
Galileo (nano)	7.65	341.86
DSensD	3.16	116.37
DSensD+	3.16	116.38

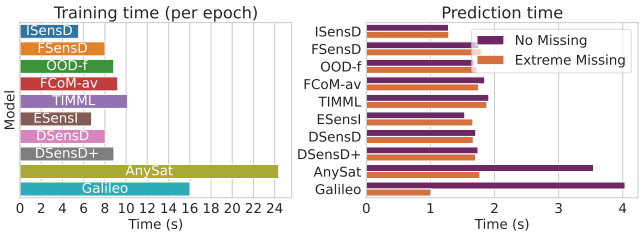


FIGURE 9. Execution time of different multi-sensor models calculated in the CropH-m data with a 128-batch-size. The prediction times are separated into no-missing and extreme missing sensor (only optical SITS available) cases.

We report the execution time of all compared models in Figure 9 using a GeForce RTX 3080 Mobile / Max-Q 8GB/16GB. We notice the high computational usage of the GeoFMs, Anysat followed by Galileo. For these models, the training time corresponds to the fine-tuning time. Surprisingly, Galileo (using the nano weights [49]) has the lowest prediction time in extreme missing sensor scenarios. However, considering the models trained from scratch, we found the following. During training, we notice that TIMML, FCoM-av, and OOD-f models have the highest execution time per epoch, while ISensD and ESensI are the most efficient ones. Our models are in the middle-lower part of the rank according to this analysis, with DSensD having a lower training time per epoch than DSensD+. These time differences between our models are due to the full-sensor data prediction and the computation of the additional loss functions in DSensD+. However, our models behave the same during inference, illustrated by the similar prediction time and third-best efficient time after ISensD and ESensI. The DSensD+ model, based on decision-level fusion, exhibits higher efficiency considering both training and prediction time (with and without missing data) than models based on feature-level fusion, such as TIMML and FCoM-av.

Overall, this analysis demonstrates that our approach offers greater practical applicability than existing models in the literature, offering advantages in runtime efficiency, computational resource usage, and predictive performance.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

F. LIMITATIONS

The results with our methods, DSensD and DSensD+, exhibit a slightly lower performance in the full-sensor data predictions regarding the competitors. This is observed in the main results (Tables 2, 3, and 4). However, the second-best results in this full-data prediction case are obtained with DSensD+, with a consistent gain in the missing data scenarios across the datasets. The advantage is greater in the extreme missing sensor scenarios for inference.

The additional loss terms and the SensD technique used in the DSensD+ method may increase the computational resources requirements. However, these components are only used during training. During inference, it acts as a simple decision-level fusion model, ignoring the predictions associated with the missing sensors from the aggregation. Moreover, our DSensD+ significantly outperforms the simple decision-level fusion model without additional components during training (i.e., Davg in Table 5 and 6). This happens in scenarios with and without missing sensor data.

VI. CONCLUSION

The lack or unavailability of sensor data is inherent to the EO domain, as data collection is subject to real-world operational constraints that can, for example, limit the spatial coverage of certain sensors. Although various multi-sensor models have been proposed in the literature, they often fall short in terms of robustness under extreme missing sensor conditions during inference, such as scenarios where only a single sensor is available. To address this challenge, we propose a multi-sensor model that is robust by design not only to moderate sensor unavailability (i.e., a single missing sensor), but also to more severe, often overlooked, cases. Our method, termed Decision-level Sensor Dropout with mutual distillation (DSensD+), incorporates the SensD technique at the decision level, along with an additional mutual distillation loss. We evaluate model robustness across crop- and tree-mapping datasets, considering binary, multi-class, and multi-label classification tasks. The results demonstrate that our method achieves superior robustness across a variety of missing sensor scenarios, outperforming several recent competitors from the literature. Moreover, our model even surpasses competitors that have dedicated training for single-sensor predictions in some cases. Overall, the obtained findings position DSensD+ as a promising solution for applications involving missing sensor data, thus advancing the field of multi-sensor modeling in the EO domain.

REFERENCES

[1] G. Camps-Valls, D. Tuia, X. X. Zhu, and M. Reichstein, Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences. New York: John Wiley & Sons, 2021.

[2] R. Wu, H. Wang, H.-T. Chen, and G. Carneiro, "Deep multimodal learning with missing modality: A survey," arXiv preprint arXiv:2409.07825, 2024.

[3] F. Mena, D. Arenas, M. Nuske, and A. Dengel, "Common practices and taxonomy in deep multi-view fusion for remote sensing applications," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 4797 – 4818, 2024.

[4] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing information reconstruction of remote sensing data: A technical review," IEEE Geoscience and Remote Sensing Magazine, vol. 3, no. 3, pp. 61–85, 2015.

[5] V. Sainte Fare Garnot, L. Landrieu, and N. Chehata, "Multi-modal temporal attention models for crop mapping from satellite time series," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 187, pp. 294–305, 2022.

[6] B. L. Markham, J. C. Storey, D. L. Williams, and J. R. Irons, "Landsat sensor performance: History and current status," IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 12, pp. 2691–2694, 2004.

[7] P. Potin, O. Colin, M. Pinheiro, B. Rosich, A. O'Connell, T. Ormston, J.-B. Gratadour, and R. Torres, "Status and evolution of the Sentinel-1 mission," in IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022, pp. 4707–4710.

[8] M. Ma, J. Ren, L. Zhao, D. Testuggine, and X. Peng, "Are multimodal transformers robust to missing modality?" in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022, pp. 18 177–18 186.

[9] L. Fasnacht, P. Renard, and P. Brunner, "Robust input layer for neural networks for hyperspectral classification of data with missing bands," Applied Computing and Geosciences, vol. 8, 2020.

[10] S. Ofori-Ampofo, C. Pelletier, and S. Lang, "Crop type mapping from optical and radar time series using attention-based deep learning," Remote Sensing, vol. 13, no. 22, 2021.

[11] F. Mena, D. Arenas, M. Charfuelan, M. Nuske, and A. Dengel, "Impact assessment of missing data in model predictions for Earth observation applications," in IEEE International Geoscience and Remote Sensing Symposium, 2024, pp. 967–971.

[12] J. Kang, Z. Wang, R. Zhu, J. Xia, X. Sun, R. Fernandez-Beltran, and A. Plaza, "DisOptNet: Distilling semantic knowledge from optical images for weather-independent building segmentation," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022.

[13] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Urban land cover classification with missing data modalities using deep convolutional neural networks," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 6, pp. 1758–1768, 2018.

[14] F. Mena, D. Arenas, and A. Dengel, "Increasing the robustness of model predictions to missing sensors in Earth observation," arXiv preprint arXiv:2407.15512, 2024.

[15] G. Xu, X. Jiang, Y. Zhou, J. Fu, Z. Huang, and X. Liu, "Transformer-based incomplete multi-modal learning for land cover classification," in IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024, pp. 7276–7281.

[16] F. Mena, D. Arenas, and A. Dengel, "Missing data as augmentation in the Earth observation domain: A multi-view learning approach," arXiv preprint arXiv:2501.01132, 2025.

[17] Y. Cong, S. Khanna, C. Meng, P. Liu, E. Rozi, Y. He, M. Burke, D. Lobell, and S. Ermon, "SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery," Advances in Neural Information Processing Systems, vol. 35, pp. 197–211, 2022.

[18] G. Tseng, I. Zvonkov, M. Purohit, D. Rolnick, and H. Kerner, "Lightweight, pre-trained transformers for remote sensing timeseries," arXiv preprint arXiv:2304.14065, 2023.

[19] Y. Chen, M. Zhao, and L. Bruzzone, "A novel approach to incomplete multimodal learning for remote sensing data fusion," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[20] Y. Xie, J. Tian, and X. X. Zhu, "A co-learning method to utilize optical images and photogrammetric point clouds for building extraction," International Journal of Applied Earth Observation and Geoinformation, vol. 116, 2023.

[21] G. Astruc, N. Gonthier, C. Mallet, and L. Landrieu, "OmniSat: Self-supervised modality fusion for Earth observation," in European Conference on Computer Vision. Springer, 2025, pp. 409–427.

[22] ——, "AnySat: An Earth observation model for any resolutions, scales, and modalities," in *Accepted* at the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025.

[23] Y.-J. Li, J. Park, M. O'Toole, and K. Kitani, "Modality-agnostic learning for radar-lidar fusion in vehicle detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022, pp. 918–927.

[24] F. Mena, D. Arenas, M. Miranda, and A. Dengel, "On what depends the robustness of multi-source models to missing data in Earth observation?"

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

in *Accepted* at the IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2025.

[25] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, "Deep mutual learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018, pp. 4320–4328.

[26] Y. J. E. Gbodjo, O. Montet, D. Ienco, R. Gaetano, and S. Dupuy, "Multisensor land cover classification with sparsely annotated data based on convolutional neural networks and self-distillation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 11 485–11 499, 2021.

[27] G. Hinton, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[28] J. Gawlikowski, S. Saha, J. Niebling, and X. X. Zhu, "Handling unexpected inputs: Incorporating source-wise out-of-distribution detection into SAR-optical data fusion for scene classification," EURASIP Journal on Advances in Signal Processing, vol. 2023, no. 1, 2023.

[29] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 5, pp. 4340–4354, 2021.

[30] S. Woo, S. Lee, Y. Park, M. A. Nugroho, and C. Kim, "Towards good practices for missing modality robust action recognition," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 3. AAAI, 2023, pp. 2776–2784.

[31] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778–782, 2017.

[32] N. Audebert, B. Le Saux, and S. Lefevre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 140, pp. 20–32, 2018.

[33] F. Mena, D. Arenas, and A. Dengel, "In the search for optimal multi-view learning models for crop classification with global remote sensing data," arXiv preprint arXiv:2403.16582, 2024.

[34] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022, pp. 16 000–16 009.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[36] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2018, pp. 4171–4186.

[37] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "BRITS: Bidirectional recurrent imputation for time series," Advances in Neural Information Processing Systems, vol. 31, 2018.

[38] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Hyperspectral image classification using random occlusion data augmentation," IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 11, pp. 1751–1755, 2019.

[39] Y. Wang, C. M. Albrecht, and X. X. Zhu, "Self-supervised vision transformers for joint SAR-optical representation learning," in IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022, pp. 139–142.

[40] B. Ekim and M. Schmitt, "Deep occlusion framework for multimodal Earth observation data," IEEE Geoscience and Remote Sensing Letters, 2024.

[41] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," Advances in Neural Information Processing Systems, vol. 30, 2017.

[42] S. Wang, Z. Yan, D. Zhang, H. Wei, Z. Li, and R. Li, "Prototype knowledge distillation for medical segmentation with missing modality," in IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2023.

[43] D. Ienco and C. F. Dantas, "DisCoM-KD: Cross-modal knowledge distillation via disentanglement representation and adversarial learning," in British Machine Vision Conference, 2024.

[44] L. Zhang, C. Bao, and K. Ma, "Self-distillation: Towards efficient and compact neural networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 8, pp. 4388–4403, 2021.

[45] S. Black and R. Souvenir, "Multi-view classification using hybrid fusion and mutual distillation," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2024, pp. 270–280.

[46] N. Bakalos, S. Sykiotis, A. Temenos, I. Rallis, A. Doulamis, and N. Doulamis, "Segmentation of remote sensing data with missing modalities through prototype knowledge distillation," in IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024, pp. 10 015–10 018.

[47] S. Sun, W. Ren, J. Li, R. Wang, and X. Cao, "Logit standardization in knowledge distillation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2024, pp. 15 731–15 740.

[48] G. Tseng, I. Zvonkov, C. L. Nakalembe, and H. Kerner, "CropHarvest: A global dataset for crop-type classification," Proceedings of NIPS Datasets and Benchmarks Track, 2021.

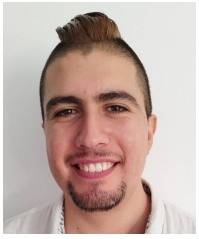
[49] G. Tseng, A. Fuller, M. Reil, H. Herzog, P. Beukema, F. Bastani, J. R. Green, E. Shelhamer, H. Kerner, and D. Rolnick, "Galileo: Learning global and local features in pretrained remote sensing models," arXiv preprint arXiv:2502.09356, 2025.

[50] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," Remote Sensing, vol. 11, no. 5, 2019.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016, pp. 770–778.

[52] M. Rußwurm and M. Körner, "Self-attention for raw optical satellite time series classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 169, pp. 421–435, 2020.

[53] N. M. Foumani, C. W. Tan, G. I. Webb, and M. Salehi, "Improving position encoding of transformers for multivariate time series classification," Data Mining and Knowledge Discovery, vol. 38, no. 1, pp. 22–48, 2024.



FRANCISCO MENA (Graduate Student Member, IEEE) received the master's and bachelor's degree in computer engineering from the Federico Santa Maria Technical University (UTFSM), Valparaíso, Chile, in 2020. He is currently working toward the Ph.D. degree in computer science with the University of Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany. During 2020 and 2021, he gave some lectures on computational statistics and artificial neural networks in the computer engineering program with the UTFSM, and in 2024 on applications of machine learning and data science with the RPTU. He is currently researching with the German Research Center for Artificial Intelligence (DFKI GmbH), Kaiserslautern, under the supervision of Prof. A. Dengel. His research interests include deep neural networks, dimensionality reduction, multiview or multimodal learning, data fusion, and Earth observation applications. Mr. Mena is involved as a Member of the Geoscience and Remote Sensing Society (GRSS) and a Reviewer for IEEE Transactions on Geoscience and Remote Sensing.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3568706

DINO IENCO (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Torino, Turin, Italy, in 2006 and 2010, respectively. In 2011, he joined the TETIS, National Research Institute for Agriculture, Food and the Environment (INRAE), University of Montpellier, Montpellier, France, as a Junior Researcher. His main research interests include machine learning, data science, graph databases, social media analysis, information retrieval, and spatiotemporal data analysis with a particular emphasis on remote sensing data and Earth observation data fusion. Dr. Ienco served on the program committee of many international conferences on data mining, machine learning, and database, including the IEEE International Conference on Data Mining (ICDM), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Asian Conference on Machine Learning (ACML), and International Joint Conference on Artificial Intelligence (IJCAI). He served as a reviewer for many international journals in the general field of data science and remote sensing.

CÁSSIO F. DANTAS received the Ph.D. degree in signal, image, and vision from Inria, Rennes, France, in 2019. He is a Research Scientist at INRAE, TETIS, Montpellier, France. Between 2020 and 2022, he was a Postdoctoral Researcher with IRIT (computer science laboratory of Toulouse) and IMAG (mathematics laboratory of the University of Montpellier). His current research interests include interpretable artificial intelligence, optimization algorithms, and machine learning for remote sensing data with applications to agriculture, ecosystems, and the environment.

ROBERTO INTERDONATO received the Ph.D. degree in computer engineering from the University of Calabria, Arcavacata, Italy, in 2015. His Ph.D. dissertation was titled "novel ranking problems in information networks". He is currently a Research Scientist with CIRAD, UMR TETIS, Montpellier, France. He was previously a Postdoctoral Researcher with the University of La Rochelle, La Rochelle, France, Uppsala University, Uppsala, Sweden, and University of Calabria, Arcavacata, Italy, His research interests include the design of data science techniques applied to the analysis of complex networks (e.g., social media networks, trust networks, semantic networks, bibliographic networks) and the extraction of information from remote sensing data. His most recent contributions concern the implementation of deep learning methods for land use classification based on the analysis of time series of multisensor satellite images (optical, radar, high/very high spatial resolution), the application of complex network analysis techniques for the extraction of spatialized indicators (landscape, socio-economic) from multisource data (remote sensing, survey data, statistics, social networks, etc.). His thematic interests mainly concern the characterization of tropical agricultural landscapes, the production of spatial information for food security, and the analysis of the transnational land trade market. On these topics, he has coauthored journal articles and conference papers, organized workshops, presented tutorials at international conferences, and developed practical software tools

ANDREAS DENGEL received the diploma degree in computer science from the University of Kaiserslautern and the Ph.D. degree from the University of Stuttgart. He is a Scientific Director with the German Research Center for Artificial Intelligence (DFKI GmbH), Kaiserslautern. In 1993, he became a Professor at the Computer Science Department, University of Kaiserslautern, where he holds the chair Knowledge-Based Systems. Since 2009, he has been a Professor (Kyakuin) at the Department of Computer Science and Information Systems, Osaka Prefecture University. He was also with IBM, Siemens, and Xerox Parc. Moreover, he has co-edited international computer science journals and has written or edited 12 books. He has authored more than 300 peer-reviewed scientific publications and has supervised more than 170 Ph.D. and master's theses. He is a member of several international advisory boards, has chaired major international conferences, and founded several successful start-up companies. He is an IAPR Fellow and has received prominent international awards. His main scientific emphasis is in the areas of pattern recognition, document understanding, information retrieval, multimedia mining, semantic technologies, and social media.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/