Spotting Virus From Satellites: Modeling the Circulation of West Nile Virus Through Graph Neural Networks

Lorenzo Bonicelli, Angelo Porrello, Stefano Vincenzi, Carla Ippoliti, Federica Iapaolo, Annamaria Conte, and Simone Calderara, *Member, IEEE*

Abstract—The occurrence of West Nile virus (WNV) represents one of the most common mosquito-borne zoonosis viral infections. Its circulation is usually associated with climatic and environmental conditions suitable for vector proliferation and virus replication. On top of that, several statistical models have been developed to shape and forecast WNV circulation: in particular, the recent massive availability of Earth observation (EO) data coupled with the continuous advances in the field of artificial intelligence offer valuable opportunities. In this article, we seek to predict WNV circulation by feeding deep neural networks (DNNs) with satellite images, which have been extensively shown to hold environmental and climatic features. Notably, while previous approaches analyze each geographical site independently, we propose a spatial-aware approach that considers also the characteristics of close sites. Specifically, we build upon graph neural networks (GNNs) to aggregate features from neighboring places and further extend these modules to consider multiple relations, such as the difference in temperature and soil moisture between two sites, as well as the geographical distance. Moreover, we inject time-related information directly into the model to take into account the seasonality of virus spread. We design an experimental setting that combines satellite images—from Landsat and Sentinel missions—with ground-truth observations of WNV circulation in Italy. We show that our proposed multiadjacency graph attention network (MAGAT) consistently leads to higher performance when paired with an appropriate pretraining stage. Finally, we assess the importance of each component of MAGAT in our ablation studies.

Index Terms—Deep learning (DL), graph neural network (GNN), Landsat, remote sensing, satellite imagery, self-supervised learning, Sentinel, West Nile virus (WNV).

I. INTRODUCTION

ZOONOSES[1] are diseases that are transmissible from animals to humans. Depending on the way of transmission, they can be foodborne, waterborne, vector-borne, and transmitted through direct contact with animals or indirectly by fomites or environmental contamination. These diseases represent a severe threat to worldwide public health by now, constituting approximately 60% of all the emerging infectious diseases reported globally [1]. For this reason, considerable efforts have recently been made to set up integrated surveillance plans [2], [3] paving the way toward early recognition and intervention of critical settings.

In such a scenario, West Nile virus (WNV) infection is one of the most widespread zoonoses in Eastern, Western, and Southern Europe. The disease is caused by the WNV—a positive-strand ribonucleic Acid (RNA) flavivirus— and is commonly transmitted by the bite of an infected mosquito. Incidentally, the transmission cycle can lead to human infections, where around one in five people present flu-like symptoms and 1 in 150 may develop a more serious— sometimes even fatal—illness. Several bird species are the main hosts of WNV [4], [5], [6], [7], but hundreds of cases each year concern infections in humans and other mammals (e.g., horses) considered dead-end hosts. Although most of these cases are asymptomatic, the evidence of viral circulation can be associated with clinical symptoms.

The persistence of the virus in nature is favored by various elements: the vector presence as mentioned, but also climatic and environmental factors play an important role [8], [9]. For example, the transmission of the pathogen is highly influenced by temperature, which determines both the survival conditions [10] and the intensity (i.e., the fractions of vectors that carry the disease) of virus spread [11]. Incidentally, climate change has been implicated as a contributing cause for the changing patterns of WNV transmission [12], [13], [14], [15], [16], [17], [18].

On top of this evidence, previous works [19], [20], [21], [22] attempted to model and track the spread of WNV infection by

---

[1]This work has been submitted to the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

examining climatic–environmental variables in an automated fashion. They typically rely on Earth observation (EO) to extract indicators that describe the land cover and are based on land surface temperature (LST), normalized difference vegetation index (NDVI, and modified normalized difference water index (MNDWI). Even if these derived indices have led to satisfactory results, the growth and massive availability of new satellite data from more recent space missions have enabled the exploitation of deep learning (DL) techniques for agriculture [23], or insect population models [24]. Concerning those problems dealing with animal and human health, the DL paradigm appears promising for modeling the environmental and climate factors, as these methodologies have the potential to learn the most appropriate high-level features directly from raw spectral band satellite images. Notably, these techniques took a few years to emerge as the state-of-the-art for diverse remote sensing tasks, including soil and crop classification [23], image fusion [25], and change detection [26], [27].

Following this trend, our research interest regards the analysis of the spread of the West Nile disease directly from remote sensing data. Specifically, we cast the problem as binary classification and propose a novel method that extracts both temporal and spatial cues from satellite imagery. In a slightly similar vein, recent works [24], [28] have tackled the issue of detecting not the circulation of the virus, but instead the presence/absence of the corresponding vector. These approaches usually examine the satellite imagery centered on a certain geographical site and attempt to infer the probability of vector presence. However, as a single point is considered for inference, the estimation provided by the model could be susceptible to outliers, obstructions (e.g., clouds, fog), noisy acquisitions, or other artifacts.

With the aim of obtaining robust predictions, our approach resorts to taking into account also the relationships between geographical sites: namely, areas with similar climatic and environmental conditions are potentially exposed to similar disease risks, even if they are geographically distant from each other [29]. On top of that relationship, we mainly contribute to the field by extending the input of the model and including a neighborhood of spatially close samples around the point of interest, thus studying the circulation at a wider geographical scale.

In methodological terms, we build upon the work of [28] and extract—for each site independently—a set of high-level features from raw multiband input. In addition, we aid the feature extractor to catch the seasonal patterns peculiar to the WNV circulation: we do so by conditioning [30] the computations of the internal representations on the month that spectral bands were captured by sensors. Then, to encompass the environmental and geographical relationships between nearby sites, we arrange their high-level representations as the nodes of a graph spanning multiple geographical locations. Thanks to such a formulation, we can model the edge between two sites by considering not only their geographical distance (e.g., using the Haversine formula) but also their affinity in terms of environmental and climatic variables, such as temperature and soil moisture. We then process the graph through a

modified version of graph attention networks (GATs) [31], revised to deal with multiple real-valued adjacency matrices. With the adoption of this architecture, named multiadjacency GAT (MAGAT), we can aggregate features of near sites by considering the broad spectrum of information.

For the task at hand, we collect satellite imagery coming from the Sentinel missions [32] and the Landsat-8 mission [33]. In particular, Sentinel-2A/2B and Landsat 8 satellites keep onboard multispectral devices (MSI) capable of acquiring 13 spectral bands for the Sentinel and nine for the Landsat. We then pair the remotely sensed data with the on-the-ground WNV circulation dataset described in [34], thus obtaining a dataset valid for supervised binary classification.

To prove the merits of our proposal, we conduct several experiments comparing the proposed graph aggregation with a baseline involving a single multiband image. We also perform several ablations studies to assess the merits of different inputs, pretraining stages, and neighborhood aggregation strategies.

II. RELATED WORKS

A. Vector-Borne Disease

In the past years, several studies applied machine learning methods in the field of vector-borne diseases and zoonoses. Candeloro et al. [34] used derived indices derived from remotely sensed data—i.e., LST, NDVI, and surface soil moisture (SSM)—to identify the areas at risk for WNV circulation in Italy. The idea was to use data collected during the 160 days before the infection date to estimate the potential circulation of the virus two weeks in advance. To this end, Chen and Guestrin [35] trained a model based on gradient boosting on epidemic data collected from 2017 to 2019.

In another line of work, Ippoliti et al. [29] identified climatic and environmental eco-regions, defined as areas within which there are associations of interacting biotic and abiotic features' [36]. The authors proposed to split the Italian territory into clusters (eco-climatic regions) according to seven variables, relevant to a broad set of human and animal vector-borne diseases: this way, they could highlight areas exposed to similar disease risks. Indeed, by relating the obtained results with ground-truth data, WN outbreak locations strongly end up being only a few—four out-of-the 22—eco-regions. Such a finding is valuable in practice, as it highlights areas where surveillance measures should be prioritized.

Concerning the applications of DL approaches to vector-borne diseases, most of them adapt models conceived for computer vision tasks. The problem in doing so lies in the large domain shift: while these backbones leverage RGB inputs, remote sensing data usually come with multiple spectral bands; therefore, the naive application of popular techniques based on transfer learning could lead to unsatisfactory results. In this respect, Vincenzi et al. [28] recommended a colorization pretext task [37] to properly initialize a deep convolutional neural network (DCNN). The intuition was to exploit the high correlation between colors and the semantic characteristics of the environment (e.g., large bodies of water in blue, chlorophyll-rich leaves in green, and arable lands in warm tones).

Similar to our setting, Vincenzi et al. [24] pair satellite imagery from the Sentinel-2 mission with ground-truth information describing the presence of the vector—in their case, *Culicoles imicola*, a population of midges responsible for the transmission of the bluetongue and other orbivirus diseases in animals—on the ground. Different from our work, their model focused on a single site at once but spanned across several days, thus taking into account the temporal evolution of an area in terms of climatic and environmental variables. They argued that the temporal cues can help explain the spread of the vector: in fact, they observed that aggregating features from a recent temporal window leads to enhanced performance.

It is noted that our work differs from [24] in several aspects. First, we target virus circulation directly instead of vector presence (i.e., *Culicoides imicola*), which is just one of the necessary conditions influencing the viral cycle. Second, our analysis discards the temporal axis in favor of the spatial one: we indeed gather information from multiple nearby locations around the site of interest. However, we still maintain sensitivity to temporal cues by providing the model with the knowledge of the month the input acquisition was carried out within.

B. Graph Neural Networks

In the past decade, computer vision tasks have greatly profited from the rise of CNNs [38], [39]; however, these architectures cannot handle data structured in a nonregular and complex manner, which can be naturally described as graphs. For this reason, the generalization of standard CNNs has recently gained interest, leading to the introduction [40] of graph neural networks (GNNs). Several works recast convolution as a message-passing operation between neighboring nodes, which propagates features along the edges and then aggregates them to form the new representation for the pivot node. Remarkably, several efforts have also been spent on the design of tailored pooling layers [41], [42], which can be helpful for graph classification.

Defferrard et al. [43] defined graph convolution in the spectral domain by applying Chebyshev polynomials to the graph Laplacian. On top of that work, Kipf and Welling [44] introduce graph convolutional network (GCN), which approximates the polynomials with a truncated first-order expansion, computed on a renormalized adjacency matrix. Veličković et al. [31] proposed a more sophisticated approach called GAT, which recovers the attention mechanism to weigh the contribution of neighboring nodes. In their proposal, the topological information is injected by masking the attention coefficients based on a binary adjacency matrix.

Unfortunately, GAT can only handle a single binary adjacency matrix: as edges serve to mask the attention coefficients properly, GAT cannot profit from real-valued edges (such as similarities between nodes). GCN [44] can instead do it, but its application regards only graphs with a single adjacency matrix. As discussed in Section IV, our approach addresses these limitations; in a similar vein, Gong and Cheng [45]: 1) exploit graphs with multidimensional and real-valued edges and 2) update edges one layer after the other. To avoid exploding or vanishing values, the edge features are normalized by means of doubly stochastic normalization. Notably, their experiments on citation networks and molecular datasets show that the use of multidimensional edge features consistently outperforms the state-of-the-art competitors. Unlike their proposal, we incorporate an initial feature extraction phase, carried by a convolutional backbone. Our rationale for doing so is to simplify the graph-based processing using high-level features instead of raw high-resolution multiband images. In addition, we improve the learning process by including skip connections between each graph fusion layer.

Another line of work proposes to extend the message-passing operation of GCN and GAT to multiple edges by considering interlayer dependencies in a multigraph structure. Shanthamallu et al. [46] propose graph attention models for multi-layered embeddings (GrAMME), in which a specialized *fusion head* combines the representation produced by different GAT heads (one for each layer of a multigraph structure) aggregated by specialized fusion heads. Similarly, Zangari et al. [47] propose multilayer-GCN (ML-GCN) and ML-GAT to extend the graph convolution with intergraph connections. These solutions differ from our proposal, as they cannot take advantage of both GAT- and GCN-based processing simultaneously. Moreover, the aggregation mechanism of [47] is designed for multigraph structures in which both a separate adjacency matrix and a set of node features for each layer are defined; instead, our case study provides only a single set of node features. Nevertheless, for a fair comparison against our proposal, in Section V-G we introduce a multilayer GNN baseline inspired by GrAMME-Fusion [46] and ML-GCN [47].

Application of GNNs to Remote Sensing Data: In the EO field, GNNs have so far been used in few works. Wu et al. [48] propose a change detection model combining a multiscale segmentation technique and a GCN: the former extracts object-wise high-level features from the multiscale input images; then, for each scale, the features are arranged as the nodes of a graph, with the edges representing the spatial relationships between nearby objects. The graphs are then processed by GCNs and merged into a single one by a fusion model, producing the change detection result.

For the task of land cover classification, Censi et al. [49] propose an [attentive spatial-temporal graph convolutional Network (STEGON)] to model both the temporal and spatial aspects that characterize satellite data. Their method involves an initial 1-D CNN to extract relevant temporal patterns, followed by a graph aggregation technique via GAT and self-attention to incorporate the information about the spatial neighborhood of the target node. With their work, the authors emphasize the significance of considering both time and space dimensions when dealing with satellite data.

Ouyang and Li [50] point out that common semantic segmentation approaches overlook the strong correlation between different classes: for example, bridges tend to be closer to rivers and far from cultivated fields. On this basis, they propose to model such relationships through a dedicated graph, whose nodes are the clusters provided by the superpixel segmentation algorithm [51]; edge similarities are instead formulated in terms of average color similarity in input space. Finally, a GCN module performs the classification of each superpixel.
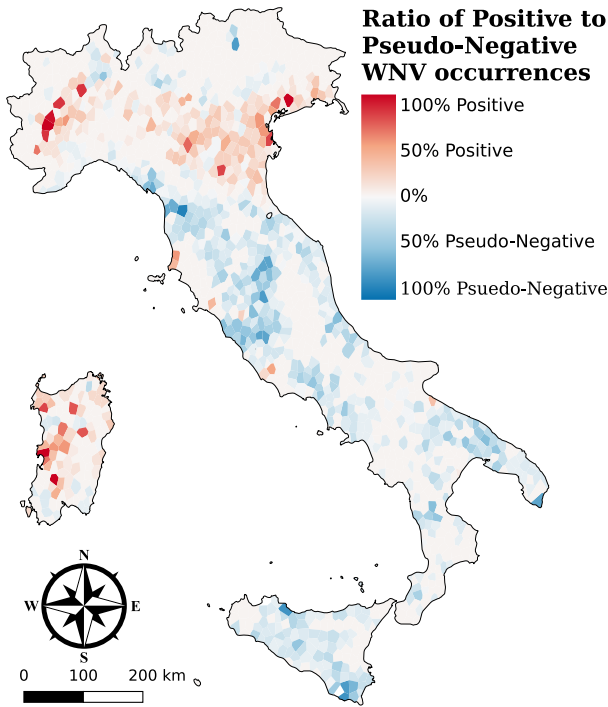
Fig. 1. **Distribution of ground-truth data** on the Italian peninsula. For better visualization, the individual occurrences are spatially clustered. We mark the prevalence of WNV occurrence (positive samples—red) over pseudoabsence (negative samples—blue). As can be seen, while the negative samples are spread across the territory, positive cases are more concentrated, indicating a strong correlation with the local environment (best seen in color).
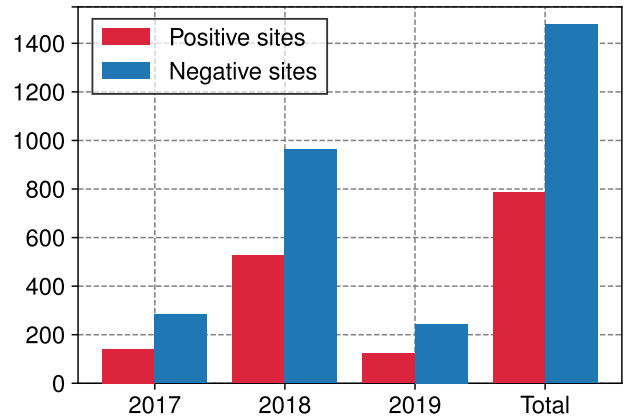

Fig. 2. **Number of recorded cases**—positive occurrences and pseudoabsence data—for each year.

TABLE I

SPECTRAL BANDS SHARED BY SENTINEL-2 AND LANDSAT-8, WITH CORRESPONDING WAVELENGTH (μm) AND SPATIAL RESOLUTION (m)

	Bands	Wavelength	Res (Sentinel/Landsat)
B_1	Coastal Aerosol	0.443	60/30
$B_{2,3,4}$	BGR channels	0.490	10/30
B_{8A}	Vegetation Red Edge	0.865	20/30
B_{11}	SWIR	1.610	20/30
B_{12}	SWIR	2.190	20/30

III. PROBLEM SETTING

We aim to identify the conditions favoring WN virus circulation, which could enable the implementation of targeted surveillance plans. As mentioned, various works [20], [22] have already proven the connection between the WNV circulation and several environmental factors, such as vegetation, temperatures, and water coverage. We build upon that relationship and exploit herein a large amount of satellite data produced in the past years: different from previous studies based on classical statistical tools [34], we investigate whether the presence/absence of the WNV circulation in Italy can be inferred through DL-based approaches.

In the following, we describe the dataset we have forged to pursue our aim: basically, it pairs data describing the WNV circulation on the ground (see Section III-A) with satellite imagery (see Section III-B).

A. Data on West Nile Virus Circulation

We first gathered the veterinary dataset described in [34], collecting observations about the virus circulation in Italy from 2008 to 2019. Briefly, Colangeli et al. [52] collected ground-truth information regarding the geographical locations (coordinates) of the virus outbreaks, detected by the National Disease Notification System of the Italian Ministry of Health.[2] Specifically, we focus our analysis on the years 2017–2019 and consider *positive sites* all those places where the presence of the virus was observed in equids, birds, or mosquitoes.

[2]SIMAN, www.vetinfo.it

During those years, data have been recorded using the same methodology; any possible source of inconsistency (geographical location, sampling date, etc.) was verified and corrected.

This way, we could gather information only about confirmed positive instances. Since the surveillance activities were not performed in all Italian territories (but only in those areas with historical evidence of virus circulation or at higher risk of WNV introduction), the indication of which places can be considered as negative sites is not always available [34]. For this reason, pseudoabsence data were randomly generated and distributed in areas suitable for the virus presence—such as places characterized by an altitude under 600 m and located in medium inhabited centers—where the disease was never reported in the past.

To sum up, the dataset we collect comprises 2264 cases, split into 786 positives and 1478 negatives. To provide a visual description, we show the spatial distribution of positive and *pseudo-negative* sites in Fig. 1. Instead, we refer the reader to Fig. 2 for the temporal data distribution over the years.

B. Satellite Imagery

In the following paragraphs, we describe the main sources of remote sensing data, which we pair with the aforementioned information describing WNV circulation. These data are in the form of raw spectral bands [53], [54] and specific measurements (e.g., temperature and soil moisture) already acknowledged to partially explain WNV circulation.

First, we aim to represent each geographical site with a multiband satellite image depicting that site at a certain time. In this respect, we get the data captured by both Sentinel-2 and Landsat-8 satellites as follows.

1) *Sentinel-2:* The two twin satellites Sentinel-2A and 2B, launched in 2015 and 2017, respectively, in the frame of the European program Copernicus, acquire 13 spectral bands at different spatial resolutions (10, 20, and 60 m/pixel), with a revisit time equals to five days.

2) *Landsat-8:* Landsat-8 satellite, launched in 2013, carries onboard the Thermal Infrared Sensor (TIRS) and the Operational Land Imager (OLI), acquiring, respectively, two and nine spectral bands. The satellite flies in a near-polar orbit, at an altitude of 705 km, and acquires images of the same territory every 16 days. For this work, we consider the nine spectral bands captured by the OLI instruments, having a spatial resolution of 30 m/pixel.

We limit our analysis to the seven bands these two systems have in common (see Table I). As in [24], we adopt 20 m as the default spatial resolution and resize the available tiles to account for the different resolutions. We collect a total of 37 876 satellite images, which constitute the basis for both the train and test sets. Each of these is paired with the temporally closest ground-truth observation; therefore, the same observation can be associated with multiple satellite images.

1) Land Surface Temperature : It represents a fundamental determinant of the terrestrial thermal behavior, as it controls the effective radiating temperature of the Earth's surface. LST was captured separately during day and night from a dual-view scanning temperature radiometer, an instrument onboard the Copernicus Sentinel-3A and 3B satellites [55]. They fly in low Earth orbit at 800–830 km of altitude, feature a daily revisit time over the same place, and have a spatial resolution of 1 km.

2) Surface Soil Moisture : It represents the percentage of the relative water content of the top few centimeters of the soil, indicating how dry or wet it is. Moreover, it also provides insights into local precipitation impacts and soil conditions. The SSM data have been retrieved from the Copernicus Global Land Service [56], which includes two polar-orbiting satellites that operate day and night performing C-band synthetic aperture radar imaging; this ensures the acquisition regardless of the weather. Similar to Sentinel-3, Sentinel-1 has a daily revisit time and a spatial resolution of 1 km.

IV. MODEL

Overview: We look for a way to exploit neighbor sites while analyzing a given infection point. To this aim, we propose a model composed of three main blocks: a features' extractor (see Section IV-A), a graph aggregation module, and a final classifier (see Section IV-B). An overview of the architecture is shown in Fig. 3.

A. Features Extractor

The role of the feature extractor is to process the multiband and high-resolution satellite images to obtain a lower dimensional feature vector. During this step, redundant or noisy information is discarded, leaving only high-level details that better summarize the input patch.

1) Architecture: We adopt ResNet18 [39] as backbone, which comprises four blocks with two residual units each. The presence of residual connections prevents the representations from degrading toward the end of the network; moreover, as clarified in [57], ResNet-based networks are more suitable for self-supervised representation learning.

2) Pretraining: When dealing with a constrained number of images, common approaches involve transfer learning techniques to boost the quality of the representations extracted by the model. Among these, the leading approach usually involves an initial pretrain stage on a large dataset, followed by fine-tuning (knowledge transfer) on the target data [58], [59]. In the former phase, the dataset commonly adopted is ImageNet [60], a large image classification dataset. However, since it only features RGB images, its application appears less appropriate for the task at hand. Indeed, its use discards potentially meaningful information present in multispectral satellite imagery.

Instead, we rely on a tailored solution to exploit the additional information provided by satellite data. Following [28], we construct a colorization pretext task: we require the model to reconstruct the RGB spectral bands while taking as input the other ones. Starting from BigEarthNet [61], [62]—an open-access dataset containing 590 000 labeled (with land cover type) Sentinel-2 images—we train a ResNet-based autoencoder architecture on the colorization task. Once the model has reached good capabilities, we discard the decoder and keep the encoder parameters, which are used to initialize the feature extractor that will be trained to solve the WNV classification task.

3) Conditional Batch Normalization: As discussed in [34], the yearly distribution of WNV infections exhibits a clear seasonal trend, highly affecting the distribution of positive occurrences. To capture such a pattern, we include the information of the relative timestamp as an additional covariate of our study. Specifically, we modulate the convolutional feature maps based on the month in which the input image was captured by satellites. In practice, we replace each batch normalization (BN) layer of the feature extractor with its counterpart conditional BN (CBN) [30].

In practice, while classical BN reckons on a single set of learnable affine parameters, the CBN layer we use provides 12 sets of parameters, one for each month. Therefore, given a batch $\mathcal{B} = \{\boldsymbol{H}^i\}_{i=1}^N$ of N samples, where $\boldsymbol{H} \in \mathbb{R}^{C \times W \times H}$ denotes a generic feature map produced from a convolutional layer, CBN on channel $c \in \{1, \ldots, C\}$ and location $(w, h) \in \{1, \ldots, W\} \times \{1, \ldots, H\}$ is defined as

$$\text{CBN}_c\big(H^i_{c,h,w}\big|\gamma_c, \beta_c\big) = \gamma_c \frac{H^i_{c,w,h} - \mathbb{E}_{\mathcal{B}}[\boldsymbol{H}_{c,\cdot,\cdot}]}{\sqrt{\text{Var}_{\mathcal{B}}[\boldsymbol{H}_{c,\cdot,\cdot}] + \epsilon}} + \beta_c \quad (1)$$

where ϵ is a constant that prevents numerical issues, and γ_c, β_c are the learnable parameters controlling the affine transformation the normalized feature maps are subject to.

B. Graph Aggregation

As previously mentioned, the density and spread of the vectors that carry WNV are closely related to climatic and
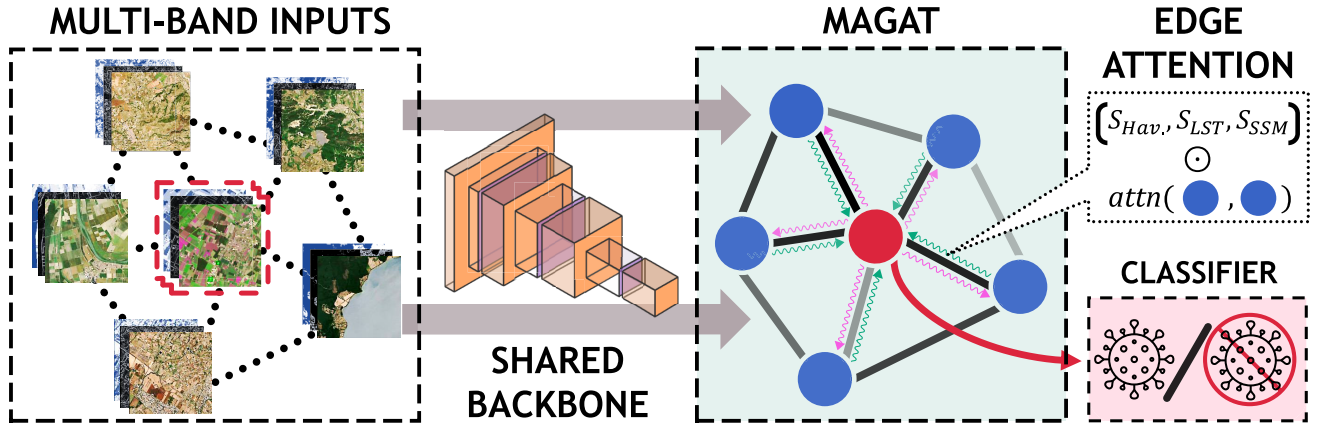
Fig. 3. **Overview of the proposed pipeline**. 1) Target multiband sample and its neighbors are independently passed through a shared convolutional network. 2) Obtained high-level features define the content of each node of a graph; the edges, instead, encode several distances. 3) Information is propagated by means of our MAGAT, which builds upon an attention function over the nearby nodes. 4) Final decision is obtained by a simple linear projection of the target's node features.

environmental characteristics. To model these relationships, we arrange the features produced by neighboring sites as nodes in a graph and apply a GNN.

1) Overview: Let N be the number of nodes in the graph, each with F^0-dimensional features. We then indicate with $X^0 \in \mathbb{R}^{N \times F^0}$ the input node features and with $S^0 \in \mathbb{R}^{N \times N \times G}$ its multivalued affinity matrix (G stands for the number of different input relationships). In our experiments, the following holds.

1) *Nodes:* We forge each input graph by sampling a random location—called **pivot** node—and then gathering the ten spatially closest neighbors' locations.
2) *Edges:* We provide a stack of three matrices: the geographical (Haversine) distance and the two other ones based on LST and SSM (namely, for each band we compute the mean absolute pairwise distance between the respective the nodes, thus outlining the average difference in temperature and soil moisture between the locations). All those distances are converted into similarities (through a Gaussian kernel with $\sigma = 1$) and then further preprocessed via double-stochastic normalization (discussed in the next paragraphs).

The building block of our proposed architecture consists of a multihead residual graph layer—termed MAGAT—which basically extends GAT [31] for leveraging multiple adjacency matrices. Briefly, each head operates on a separate adjacency matrix $S^0_{\cdot,\cdot,g}$ to process the input features X^0. The results are then aggregated to produce two results: a new $N \times F^1$ nodes' representation X^1, with $F^1 \triangleq G \cdot F^{\text{Int}}$, and $N \times N \times G$ edge feature matrices S^1. We stack multiple of these layers, each of which is fed with the affinity matrices produced by the previous layer.

After having applied these transformations, we discard the neighbors and focus only on the pivot node; this way a single vector embedding is fed to the final classification layer. The latter outputs the prediction for the presence of WNV circulation.

2) MAGAT Layer: In formal terms, let us first review the similarity between nodes i and j as originally conceived by

the authors of GAT

$$s(X_i^0, X_j^0) = \exp\{\text{LReLU}(p^T [V X_i^0 \| V X_j^0])\} \quad (2)$$

where $V \in \mathbb{R}^{F^{\text{Int}} \times F^0}$ linearly projects the input features to an intermediate dimension F^{Int}, $\|$ stands for the concatenation operator, and $p \in \mathbb{R}^{2F^{\text{Int}}}$ is a learnable vector that allows for a single scalar output. In this simple formulation, the only way to inject an external adjacency matrix is by zeroing out the contribution from unconnected nodes. Drawing inspiration from [45], MAGAT learns a separated similarity on top of each matrix $S^0_{\cdot,\cdot,g}$ provided as input. Formally, we modify (2) by introducing an attenuation factor that emphasizes the contributions from similar nodes in the neighborhood

$$\hat{\alpha}_{i,j,g} = S^0_{i,j,g} \cdot s(X_i^0, X_j^0) \quad (3)$$

where

$$S^0 = \begin{cases} S^0_{i,j}, & \text{if } j \in \text{Neighbors}(i) \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

In our approach, the resulting attention coefficients become the input affinity matrices for the subsequent layer. However, the repeated application of (3) may result in severe numerical instabilities. To avoid such an issue, we apply double-stochastic normalization on $\hat{\alpha}$ by means of the Sinkhorn–Knopp iterative algorithm [63], [64]. This way, we obtain the final normalized affinity matrices: for each $g = 1, 2, \ldots, G$ we have a square nonnegative real matrix with columns and rows summing to 1.

Once the final attention coefficients α have been computed, we perform the node feature aggregation on each affinity matrix. Specifically, we define a multihead structure in which the G heads perform a separate message-passing operation. This strategy ensures an efficient feature aggregation since each head can compute its output in parallel. In our design, each $g = 1, 2, \ldots, G$ sublayer computes an initial linear combination of the node features X^0, and then scales the result for $\alpha_{\cdot,\cdot,g}$. The independent results are then concatenated to

TABLE II
COMPARISON BETWEEN OUR GRAPH-BASED APPROACH AND A BASELINE FEATURING ONLY THE INFORMATION AVAILABLE FOR THE TARGET NODE. THE SUPERIOR PERFORMANCE OF MAGAT INDICATES THE ADVANTAGES OF ADDITIONAL INFORMATION FROM SURROUNDING ENVIRONMENT

	Domain	Pre-train.	Acc.	Pr.	Rc.	F1
ResNet18	RGB	ImageNet	.825	.722	.858	.780
	Spectral	Colorization	.840	.757	.861	.795
	Both	Ensemble	.855	.761	873	.813
MAGAT	RGB	ImageNet	.921	.834	.973	.897
	Spectral	Color.	.922	**.850**	.952	.898
	Both	Ensemble	**.926**	.844	**.977**	**.905**

obtain X^1. Formally, we compute

$$\hat{X} = \bigg\|_{g=1}^{G} \left(\boldsymbol{\alpha}_{\cdot,\cdot,g} \boldsymbol{W}_g \boldsymbol{X}^0 \right) \tag{5}$$

where $\boldsymbol{W}_g \in \mathbb{R}^{F^{\mathrm{Int}} \times F^0}$ is a learned parameter matrix, and $\|$ defines the concatenation operation.

Finally, we provide a residual connection between the input node features and the output of MAGAT, whose purpose is to prevent vanishing gradients thus easing the training of the model. In practice, if indicating with $u : \mathbb{R}^{F^0} \to \mathbb{R}^{F^1}$ a single linear projection layer, the MAGAT layer then computes

$$X^1 = \sigma(\hat{X}) + u(X^0) \tag{6}$$

where σ is the exponential linear unit (ELU) [65] activation function.

C. Model Ensemble

Following [28], we further take advantage of both the spectral bands (i.e., those being outside the visible spectrum) and the RGB ones. In particular, we set up an ensemble model that averages the predictions from two models: one is fed with RGB inputs and is pretrained on the ImageNet dataset; the other takes spectral bands as input and leverages colorization pretraining as was proposed in [28].

V. EXPERIMENTS

In this section, we present the results obtained in the scenario presented in Section III-A about WNV circulation. We first summarize the choices made to set up the experimental setting; then, several comparisons are proposed to highlight the contribution of each introduced component.

A. Experimental Details

1) Preprocessing: For each location included in the dataset, we sample a squared satellite patch centered around its spatial coordinates (latitude and longitude). We interpolate each band to obtain a resolution of 20 m, thus aligning different bands and obtaining a homogeneous format. All those pixels marked as invalid, saturated, or related to heavy clouds are set to a predefined "NoData" value: in this respect, we discard the examples presenting a ratio of NoData pixels higher than an acceptance threshold (set to 10% in our experiments).

2) Hyperparameters: We set the batch size to 16 and optimize for 20 epochs with plain stochastic gradient descent (SGD). To reduce the impact of overfitting, we use dropout before the final classification layer (with drop probability equal to 0.2). We minimize the binary cross-entropy loss during training.

3) Validation: Due to the class unbalance present in our dataset (65.3% of the samples are associated with a negative label), we resort not only to accuracy but also to proper metrics, such as *precision*, *recall*, and *F1-score*.

We split the dataset into two disjoint sets for training and inference: the former comprises data from 2017 and 2018, while the test phase performs on observations from 2019. As the split is made on the time dimension, a certain geographical location may appear in both the train and test sets due to the periodical nature of satellites' motion. Finally, all the results have been obtained by repeating each experiment five times and then reporting the average.

B. Value of Neighboring Nodes

We start the experimental investigation of our graph-expanded approach by comparing it with a very simple baseline, which does not involve graph learning at all and considers one location at once: this way, we can demonstrate the benefits of information from neighboring places. The baseline at stake is ResNet18, which we initialize according to the same pretrain strategies discussed in Section IV. We refer the reader to Table II, which reports the results of such a comparison. As can be observed, we report the performance for different input modalities.[3]

We can draw the following conclusions: first, the proposed approach is better at estimating the presence/absence of the WNV circulation, reaching a top accuracy of 92.6% and an F1-score of 90.5% for the ensemble strategy. Such evidence confirms our intuition, according to which the further consideration of the surrounding environment—embodied by neighboring nodes in our setting—could prove beneficial for shaping the trend of WNV circulation.

Second, while exploiting the spectral domain provides a consistent gain in terms of performance for the baseline approach, we note that this does not hold for MAGAT. In fact, when leveraging graph-learning, the results deriving from the two domains end up comparable. Namely, the extra information provided by the parts of the spectrum outside the visible appears less critical when considering multiple adjacent nodes. We conjecture that the message-passing operation could indeed mitigate the shortcomings of RGB, introducing novel sources of information.

Third, our results confirm what was said in [28]: as ensembling the two domains provides the best performance for both the approaches, the involved pretraining strategies seem to deliver different but valuable concepts.

C. Benefits of Deep Approaches

As deeply discussed, we rely on DL techniques and multiband satellite imagery to address the task at hand. However,

---

[3]For the rest of the article the term "*Spectral*" is meant to indicate the subset of the bands not including the visible part of the spectrum.

TABLE III

COMPARISON BETWEEN SHALLOW CLASSIFIERS
AND OUR DEEP PROPOSAL MAGAT

	Pre-train.	Acc.	Pr.	Rc.	F1
Logistic Regression	✗	.717	.571	.869	.689
Random Forest	✗	.811	.675	.917	.778
Gradient Boosting	✗	.837	.710	.926	.804
MAGAT	ImageNet	.921	.834	.973	.897
	Color.	.922	**.850**	.952	.898
	Ensemble	**.926**	.844	**.977**	**.905**

the actual benefits of such a demanding approach could be argued: could simpler solutions perform comparable or even better but at lower costs?

We herein attempt to answer these doubts by comparing MAGAT against several shallow classifiers that leverage hand-crafted features, such as the following.

1) **Logistic regression** [66] learns a linear decision boundary. It does so by maximizing—via gradient descent—the log-likelihood on the training set.
2) **Random forest** [67] is a *bagging* ensemble algorithm: a set of simple models are trained independently on subsets of the original data. During inference, the predictions of these models are combined via majority voting.
3) **Gradient boosting** [68] is based on *boosting*, an ensemble technique that trains a cascade of models, each of which focuses mainly on the examples misclassified by previous ones. The overall prediction is a weighted sum of the outputs of each model.

These methods cannot be fed with high-resolution multiband images; therefore, we provide them with the average values of the spectral bands introduced in Section III-B, along with the geographic coordinates and the NDVI. The latter is computed from the B_{8A} and B_4 bands (see Table I for reference).

We report the results for such a comparison in Table III. Among the shallow approaches, gradient boosting leads to higher performance. However, our approach leads by a margin of 10% on the F1 score. Such a gap is also clear when looking at the precision and recall metrics: all the shallow classifiers tend to overemphasize the positive class, thus presenting a high recall but at the cost of low precision.

While we focus on finding the best performer for the task, we still highlight that classical machine learning methods feature faster inference and less expensive training compared with their DL counterparts, which could favor them in scenarios with stricter constraints.

D. Impact of Conditional Batch Normalization

As discussed in [22], in addition to being affected by environmental factors such as temperature, humidity, and soil moisture, the WNV circulation is also governed by seasonal patterns. In light of this, we introduced the CBN layer to provide the acquisition month to the feature extractor. Fig. 4 shows that such a design choice leads to a higher F1 score; moreover, the gain is consistent across different choices (i.e., backbone and input modality).
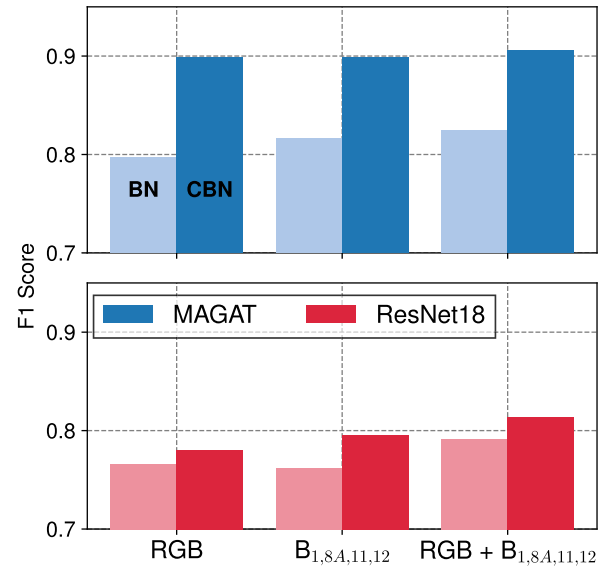

Fig. 4. Barplot showing the **impact** in terms of F1 score CBN over its naive counterpart (BN).

TABLE IV

IMPACT OF DIFFERENT EDGE CUES ON PERFORMANCE

Inputs	adjacency	Acc.	Pr.	Rc.	F1
RGB	*uniform*	.904	**.855**	.884	.869
	LST	.890	.818	.894	.854
	SSM	.900	.846	.885	.865
	LST + SSM	.889	.825	.879	.851
	all (ours)	**.920**	.834	**.973**	**.897**
Spectral	*uniform*	.889	.788	.948	.861
	LST	.880	.775	.942	.850
	SSM	.887	.786	.946	.859
	LST + SSM	.901	.805	**.960**	.876
	all (ours)	**.922**	**.861**	.952	**.898**
Both	*uniform*	.909	.830	.942	.883
	LST	.905	.830	.928	.876
	SSM	.904	.806	.968	.879
	LST + SSM	.907	.824	.944	.880
	all (ours)	**.926**	**.844**	**.977**	**.905**

E. Impact of Different Similarities in MAGAT

Here, we empirically review alternative strategies to build the input affinity matrices S. In particular, we compare ours (termed *all*, which comprises geographic, LST, and SSM distances) with using only LST and SSM similarities. In additionally, we also wonder what happens when providing no informative edges at all: in this respect, we set up a baseline called *uniform* that assigns the same importance to every edge [it ends up using node features solely, as held in (3)].

Table IV reports the results of this investigation for different modalities. Overall, using all the available information as advocated by our proposal consistently yields higher F1 scores. However, we also note that the *uniform* adjacency leads to competitive results: in two out of three cases, it even surpasses the combination of environmental bands (LST + SSM).

Furthermore, considering only the surface temperature or soil moisture seems misleading for the model. In fact, without providing any hints about the geographical distance between nodes, locations sharing similar environmental features may

TABLE V
IMPACT OF DIFFERENT PRETRAINING VERSUS TRAINING FROM SCRATCH

Inputs	Pre-training	ResNet18	MAGAT
RGB	✗	.704	.740
	ImageNet	.780	.897
Spectral	✗	.716	.867
	Color.	.795	.898
Both	✗	.747	.808
	ImageNet + Color.	.813	.905

TABLE VI
IMPACT OF MAGAT AGAINST OTHER GNNs, MEASURED AS F1 SCORE

Inputs	adjacency	RGB	Spectral	Ensemble
GCN	uniform	.767	.858	.821
	LST	.777	.864	.811
	SSM	.765	.855	.830
	Haversine	.729	.793	.777
GAT	-	.772	.795	.821
Fusion-GCN	all	.795	.849	.849
MAGAT (ours)	all	**.897**	**.898**	**.905**

be suspected of virus transmission even if they are too far apart for the virus to spread.

F. Relevance of Pretraining

Due to the limited size of the dataset used, our solution exploits different pretrain strategies (e.g., based on the ImageNet dataset [69]) to mitigate overfitting. We still question the relevance of that detail, hence evaluating the results of MAGAT with a randomly initialized feature extractor. Unsurprisingly, Table V highlights a clear advantage in using a pretrained network: when evaluating both colorization and RGB pretrain (last row), the ResNet18 and MAGAT ensemble models show an improvement of 6.6 and 9.7, respectively.

We also note that when pretraining is not involved, the performance of the MAGAT ensemble is inferior to the average performance of the models trained on the RGB and $B_{1,8A,11,12}$ bands separately. Such a result indicates that—without a proper initialization strategy—there are no advantages in combining the features learned by the two models.

G. Effectiveness of MAGAT Against GNN Baselines

Once we have discussed the importance of the various parts of MAGAT, we now wish to assess its performance under the light of other established GNN-based approaches. We consider both GAT and GCN as baselines due to their widespread use in literature. For the latter, we evaluate all the choices of single-dimensional adjacency matrix (*uniform*, LST, SSM, and geographic distance). Second, inspired by GrAMME-Fusion [46], we include a multiadjacency extension of GCN, which we call **Fusion-GCN**; herein, a *fusion* layer combines the contribution of multiple independent GCN layers as follows:

$$\mathbf{X} = \sum_{g=1}^{G} \beta_g \mathbf{z}_g \tag{7}$$

with $\mathbf{z}_g$ being the output of the GCN layer related to the gth adjacency matrix.

We first propose a quantitative comparison in Table VI: as can be seen, our proposal MAGAT reaches stronger results with respect to other GNN-based approaches under all the combinations of pretrain and adjacency matrix. This finding demonstrates the effectiveness of MAGAT in capturing the temporal and spatial dependencies of complex graph-structured data. Interestingly, the poor result of GAT indicates

a clear advantage to using the information regarding the edges, which is in line with our previous findings of Section V-E.

In addition to the quantitative analysis, in Fig. 5 we provide a qualitative assessment of the feature spaces generated by different GNN models. Inspired by other works [31], [44], [46], we use t-SNE to reduce and visualize the feature spaces generated by different models. Our observations reveal that MAGAT produces a feature representation that better separates the different classes, providing additional evidence for the effectiveness of our approach in capturing the underlying structures in the data. In addition, the representation produced by GAT suggests a poor separation between classes, which negatively affects the overall performance of the model as shown in Table VI. This effect is less pronounced in other GCN-based baselines, highlighting the importance of incorporating the environmental and climatic relationships between nearby locations.

H. Limitations and Future Work

Our evaluation focuses on the Italian peninsula, which occupies around 300.192 km^2 in the middle of Mediterranean Sea and it is characterized by high climatic variability and diversity. Concerning the WNV, since 2002 the Italian Ministry of Health has implemented a veterinary surveillance plan to monitor the viral introduction and circulation of WNV in the whole country. The virus circulation detected in animals (hereafter veterinary cases) is mandatory registered by the local veterinary authorities into the National Animal Disease Notification System (SIMAN) and forms the ground-truth database used in the analyses III-A. Thus, our analysis takes into account a wide geographical area, including many eco-climatic conditions over a long period of time.

Concerning other diseases, our model is based on the eco-climatic parameters which significantly influence the vectors' population density and the amplification of viral transmission. These modeled conditions can be used for the study of other vector-borne diseases (e.g., Usutu), but are not suitable for the evaluation of diseases that operate via a directly transmitted infection (e.g., Sars-Cov-2 infection).

Finally, our proposal could be applied to other geographical areas to model WNV circulation. However, it is necessary that these have the environmental and climatic conditions that allow the spread of the vector species and hosts.
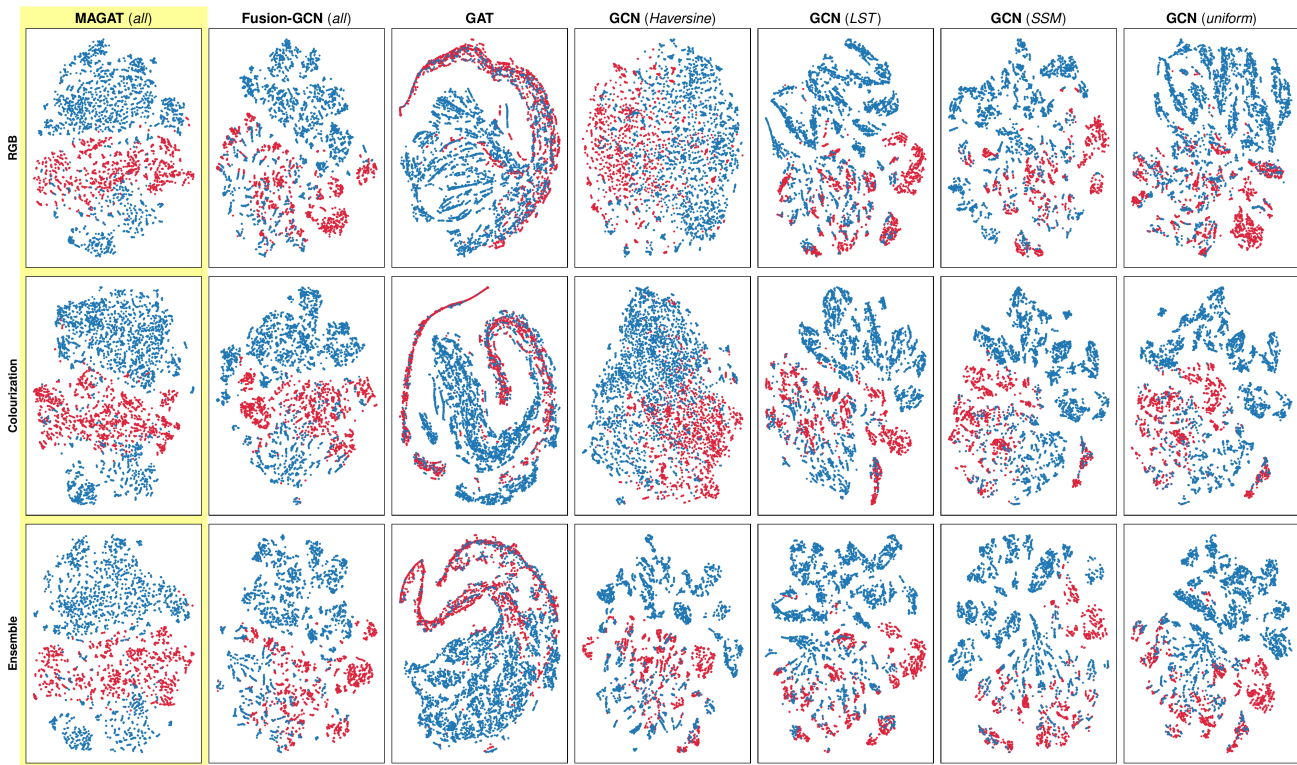
Fig. 5. Depiction of the **feature spaces spanned by different GNNs** after t-SNE [70] compression. As can be seen, the combined aggregation of all the edge features in MAGAT results in less overlapping between the positive samples (red) over the pseudonegatives (blue) (best seen in color).

VI. CONCLUSION

This work upholds the benefits of multiband satellite images for shaping the circulation of the WNV. We built upon the widely acknowledged relation between virus presence and the suitability of the surrounding environment to host specific vectors. Our inference schema goes beyond the only consideration of the climatic and ecological cues present in the geographic location of interest; we also attend to its close neighboring locations and model interactions through a GNN. The experimental results and several ablative studies reward our intuition, providing further insights regarding the relevance of circulation and seasonality modeling. In future works, we plan to extend our approach to embrace even multihorizon temporal dynamics.

ACKNOWLEDGMENT

REFERENCES

[1] L. H. Taylor, S. M. Latham, and M. E. J. Woolhouse, "Risk factors for human disease emergence," *Phil. Trans. Roy. Soc. London. Ser. B, Biol. Sci.*, vol. 356, no. 1411, pp. 983–989, Jul. 2001.

[2] Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, and World Health Organization. (2006). *Global Early Warning System for Major Animal Diseases Including Zoonosis (Glews)*. [Online]. Available: http://www.glews.net/

[3] *Taking a Multisectoral One Health Approach: A Tripartite Guide to Addressing Zoonotic Diseases in Countries*, FAO-OIE-OMS, Food & Agriculture Organization, Rome, Italy, 2019.

[4] N. Komar et al., "Experimental infection of North American birds with the New York 1999 strain of West Nile virus," *Emerg. Infectious Diseases*, vol. 9, no. 3, pp. 311–322, Mar. 2003.

[5] A. Rizzoli et al., "Understanding West Nile virus ecology in Europe: Culex pipiens host feeding preference in a hotspot of virus emergence," *Parasites Vectors*, vol. 8, no. 1, pp. 1–13, Dec. 2015.

[6] M. Spedicato et al., "Experimental infection of rock pigeons (Columba livia) with three West Nile virus lineage 1 strains isolated in Italy between 2009 and 2012," *Epidemiol. Infection*, vol. 144, no. 6, pp. 1301–1311, Apr. 2016.

[7] G. Mencattelli et al., "West Nile virus lineage 2 overwintering in Italy," *Tropical Med. Infectious Disease*, vol. 7, no. 8, p. 160, Jul. 2022.

[8] S. Paz and J. Semenza, "Environmental drivers of West Nile fever epidemiology in Europe and Western Asia—A review," *Int. J. Environ. Res. Public Health*, vol. 10, no. 8, pp. 3543–3562, Aug. 2013.

[9] A. M. Kilpatrick, M. A. Meola, R. M. Moudy, and L. D. Kramer, "Temperature, viral genetics, and the transmission of West Nile virus by *culex pipiens* mosquitoes," *PLOS Pathogens*, vol. 4, pp. 1–7, Jun. 2008.

[10] T. H. Jetten and D. A. Focks, "Potential changes in the distribution of dengue transmission under climate warming," *Amer. J. Tropical Med. Hygiene*, vol. 57, no. 3, pp. 285–297, Sep. 1997.

[11] D. J. Dohm, M. L. O'Guinn, and M. J. Turell, "Effect of environmental temperature on the ability of culex pipiens(diptera: Culicidae) to transmit West Nile virus," *J. Med. Entomology*, vol. 39, no. 1, pp. 221–225, Jan. 2002.

[12] S. Paz, "The West Nile virus outbreak in Israel (2000) from a new perspective: The regional impact of climate change," *Int. J. Environ. Health Res.*, vol. 16, no. 1, pp. 1–13, Feb. 2006.

[13] J. C. Semenza and B. Menne, "Climate change and infectious diseases in Europe," *Lancet Infectious Diseases*, vol. 9, no. 6, pp. 365–375, Jun. 2009.

[14] W. J. Tabachnick, "Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world," *J. Experim. Biol.*, vol. 213, no. 6, pp. 946–954, Mar. 2010.

[15] K. L. Ebi, E. Lindgren, J. E. Suk, and J. C. Semenza, "Adaptation to the infectious disease impacts of climate change," *Climatic Change*, vol. 118, no. 2, pp. 355–365, May 2013.

[16] C. B. Beard et al., *Vectorborne Diseases* (The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment). Washington, DC, USA: U.S. Global Change Research Program, 2016, Ch. 5, pp. 129–156.

[17] J. C. Semenza and J. E. Suk, "Vector-borne diseases and climate change: A European perspective," *FEMS Microbiology Lett.*, vol. 365, no. 2, p. 244, Jan. 2018.

[18] J. C. Semenza and S. Paz, "Climate change and infectious disease in Europe: Impact, projection and adaptation," *Lancet Regional Health Eur.*, vol. 9, Oct. 2021, Art. no. 100230.

[19] V. Chevalier, A. Tran, and B. Durand, "Predictive modeling of West Nile virus transmission risk in the Mediterranean basin: How far from landing?" *Int. J. Environ. Res. Public Health*, vol. 11, no. 1, pp. 67–90, Dec. 2013.

[20] A. Tran et al., "Environmental predictors of West Nile fever risk in Europe," *Int. J. Health Geographics*, vol. 13, no. 1, p. 26, Dec. 2014.

[21] M. Marcantonio et al., "Identifying the environmental conditions favouring West Nile virus outbreaks in Europe," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0121158.

[22] A. Conte et al., "Spatio-temporal identification of areas suitable for West Nile disease in the Mediterranean basin and central Europe," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0146024.

[23] T. Zhang, J. Su, C. Liu, W.-H. Chen, H. Liu, and G. Liu, "Band selection in Sentinel-2 satellite for agriculture applications," in *Proc. 23rd Int. Conf. Autom. Comput. (ICAC)*, 2017, pp. 1–6.

[24] A. Porrello et al., "Spotting insects from satellites: Modeling the presence of culicoides imicola through deep CNNs," in *Proc. 15th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2019, pp. 159–166.

[25] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[26] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[27] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, p. 871, Feb. 2022.

[28] S. Vincenzi et al., "The color out of space: Learning self-supervised representations for Earth observation imagery," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 3034–3041.

[29] C. Ippoliti et al., "Defining ecological regions in Italy based on a multivariate clustering approach: A first step towards a targeted vector borne disease surveillance," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0219072.

[30] A. C. Courville, "Modulating early visual processing by language," in *Proc. NIPS*, 2017, pp. 1–11.

[31] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lió, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[32] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "ESA's sentinel missions in support of Earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, May 2012.

[33] D. P. Roy et al., "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, Apr. 2014.

[34] L. Candeloro et al., "Predicting WNV circulation in Italy using Earth observation data and extreme gradient boosting model," *Remote Sens.*, vol. 12, no. 18, p. 3064, Sep. 2020.

[35] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794.

[36] R. G. Bailey, "Identifying ecoregion boundaries," *Environ. Manage.*, vol. 34, no. S1, pp. 14–26, Apr. 2004.

[37] G. Larsson, M. Maire, and G. Shakhnarovich, "Colorization as a proxy task for visual understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 840–849.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[40] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[41] A. Porrello, D. Abati, S. Calderara, and R. Cucchiara, "Classifying signals on irregular domains via convolutional cluster pooling," in *Proc. 22nd Int. Conf. Artif. Intell. Statist.*, 2019, pp. 1388–1397.

[42] D. Mesquita, A. Souza, and S. Kaski, "Rethinking pooling in graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 2220–2231.

[43] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.

[44] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–14.

[45] L. Gong and Q. Cheng, "Exploiting edge features for graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9203–9211.

[46] U. S. Shanthamallu, J. J. Thiagarajan, H. Song, and A. Spanias, "GrAMME: Semisupervised learning using multilayered graph attention models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3977–3988, Oct. 2020.

[47] L. Zangari, R. Interdonato, A. Calió, and A. Tagarelli, "Graph convolutional and attention models for entity classification in multilayer networks," *Appl. Netw. Sci.*, vol. 6, no. 1, p. 87, Dec. 2021.

[48] J. Wu et al., "A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102615.

[49] A. M. Censi, D. Ienco, Y. J. E. Gbodjo, R. G. Pensa, R. Interdonato, and R. Gaetano, "Attentive spatial temporal graph CNN for land cover mapping from multi temporal remote sensing data," *IEEE Access*, vol. 9, pp. 23070–23082, 2021.

[50] S. Ouyang and Y. Li, "Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery," *Remote Sens.*, vol. 13, no. 1, p. 119, Dec. 2020.

[51] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1356–1363.

[52] P. Colangeli, S. Iannetti, A. Cerella, C. Ippoliti, and A. Di, "Sistema nazionale di notifica delle malattie degli animali," *Veterinaria Italiana*, vol. 47, no. 3, pp. 291–301, 2011.

[53] M. Calzolari et al., "West Nile virus surveillance in 2013 via mosquito screening in northern Italy and the influence of weather on virus circulation," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0140915.

[54] S. Paz, "Climate change impacts on West Nile virus transmission in a global context," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 370, no. 1665, Apr. 2015, Art. no. 20130561.

[55] K. Fletcher, *Sentinel-3: ESA's Global Land and Ocean Mission for GMES Operational Services*. Oakville, ON, Canada: ESA Communications, 2012.

[56] B. Bauer-Marschallinger et al., "Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 520–539, Jan. 2019.

[57] A. Kolesnikov, X. Zhai, and L. Beyer, "Revisiting self-supervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1920–1929.

[58] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[59] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[60] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[61] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "BigearthNet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 5901–5904.

[62] G. Sumbul et al., "BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]," *IEEE Geosci. Remote Sens. Mag. Replaces Newsletter*, vol. 9, no. 3, pp. 174–180, Sep. 2021.

[63] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pacific J. Math.*, vol. 21, no. 2, pp. 343–348, May 1967.

[64] R. P. Adams and R. S. Zemel, "Ranking via sinkhorn propagation," 2011, *arXiv:1106.1925*.

[65] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," *in Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–14.

[66] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc., Ser. B, Methodol.*, vol. 20, no. 2, pp. 215–232, Jul. 1958.

[67] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[68] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[69] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[70] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

Lorenzo Bonicelli received the bachelor's and master's degrees from the University of Modena and Reggio Emilia, Modena, Italy, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include machine learning, deep learning, and the recent advances in continual learning and geometric deep learning.

Angelo Porrello received the master's degree in computer engineering and the Ph.D. degree in ICT from the University of Modena and Reggio Emilia, Modena, Italy, in 2017 and 2021, respectively.

He is currently a Research Fellow with the AImageLab Group, Department of Engineering "Enzo Ferrari," University of Modena and Reggio Emilia. His research interests include deep learning techniques: more precisely on continual learning, re-identification, and anomaly detection.

Stefano Vincenzi received the master's degree in computer engineering from the University of Modena and Reggio Emilia, Modena, Italy, in 2018.

He was a Research Fellow with the AImageLab Group, University of Modena and Reggio Emilia. He is currently a Machine Learning Engineer at Ammagamma, Modena. His research interests include machine learning and deep learning, with a focus on image segmentation, text recognition, and remote sensing.

Carla Ippoliti is a Mathematician at the National Reference Centre for Epidemiology, Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale," Teramo, Italy. Her main research interests include spatial analysis of vectors and vector-borne diseases and the related driving factors. She has a great experience in EO data and their use in veterinary epidemiology.

Federica Iapaolo is a Veterinarian at the National Reference Centre for Foreign Animal Diseases (CESME), Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale," Teramo, Italy. She has experience in WNV surveillance and control activities in Italy. She is involved in the management of data flow collected within WNV and USUV and SIMAN (National Information System on animal disease) implementation to ensure the information system arrangement to the information debt toward WOAH, European Commission, and EFSA. Her main research interests include West Nile virus (WNV) and Usutu virus (USUV) epidemiology.

Annamaria Conte is the Head of the Statistics and GIS Unit, National Reference Centre for Epidemiology, Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale," Teramo, Italy. Her research interests include epidemiological analytical methods and spatial epidemiology of major animal infectious diseases, including zoonoses, and the identification of factors influencing the spread and persistence of vector-borne diseases.

Simone Calderara (Member, IEEE) received the master's and Ph.D. degrees in computer engineering from the University of Modena and Reggio Emilia, Modena, Italy, in 2005 and 2009, respectively.

He is currently an Assistant Professor with the AImageLab Group, University of Modena and Reggio Emilia. His research interests include computer vision and machine learning applied to human behavior analysis, visual tracking in crowded scenarios, and time series analysis for forensic applications.