Neural Computing and Applications
https://doi.org/10.1007/s00521-022-07019-5

ORIGINAL ARTICLE

Learning of physically significant features from earth observation data: an illustration for crop classification and irrigation scheme detection

Pattathal V. Arun[1] · Arnon Karnieli[2]

Received: 25 April 2021 / Accepted: 30 January 2022

Abstract
Earth observation data processing requires interpretable deep learning (DL) models that learn physically significant and meaningful features. The current study proposes approaches to make the network to learn meaningful features. In addition, a set of interpretability- and explanation-based evaluation strategies are proposed to evaluate the DL models. Adversarial variational encoding along with constraints to regulate latent representations and embed label information are employed to learn interpretable manifold. The proposed architecture, called interpretable adversarial encoding network (IAENet), significantly improves the results compared to other main existing DL models. The proposed IAENet learns the features which are essential in distinguishing the different classes thereby improving the interpretability of the model. The explanations for the different models are generated through analysis of the concepts learned by each model using activation maximization. Besides, the relevance assigned by the model to input features is also estimated using the layer-wise relevance propagation approach. Experiments on the phenological curve-based crop classification illustrate that IAENet learn relevant features (giving importance to the non-rainy season) to distinguish different irrigation schemes. The performance can be attributed to the learned interpretable manifold, and the refinement of architectural units and convolutions considering the point-nature and irregular sampling of the input data. Experiments on learning crop-specific features from multispectral images for crop-type classification indicate that IAENet learns red and green edge features crucial in distinguishing the studied crops. The improvement in interpretability of the DL models is found to reduce the sensitivity toward network parameters. The proposed evaluation measures facilitate ascertaining the physical significance of the learned manifold.

Keywords Interpretability · Deep learning · Crop-specific features · Phenological curves · VENμS · Classification

1 Introduction

Deep learning (DL) approaches, which learn abstract representations to transform inputs to intrinsic manifolds in an unsupervised manner, have reported better results than the conventional machine learning approaches for various Earth observation (EO) data applications [1–3]. Convolutional neural networks (CNNs) are supervised algorithms for DL and their numerous variants have been developed for processing and analysis of EO data [1, 3–5]. Although DL approaches result in state-of-the-art accuracies, it is essential to verify that the high measured accuracy results from the use of an appropriate latent representation and not from the exploitation of artifacts in the data [6–8]. Techniques for interpreting and understanding what the model has learned have become a vital ingredient of a robust validation procedure [9]. Some recent carefully designed interpretation techniques have shed light on the most complex and deepest machine learning models [7, 10, 11]. However, most of these approaches facilitate interpretation of the network but equally important is to make the

✉ Arnon Karnieli
karnieli@bgu.ac.il

[1] Swiss Institute for Dryland Environmental and Energy Research, Jacob Blaustein Institutes for Desert Research, Ben Gurion University of the Negev, Sede Boker Campus, 8499000 Beersheba, Israel

[2] The Remote Sensing Laboratory, French Associates Institute for Agriculture and Biotechnology of Dryland, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boker Campus, 8499000 Beersheba, Israel

Published online: 12 March 2022

network learn meaningful features. Learning of interpretable representations is critical in EO data-based analyses where the network needs to be constrained to learn physically significant features. The self-interpretability of deep networks can be explored to learn meaningful features tailored for specific tasks.

Generally, DL algorithms cannot learn meaningful representations without some implicit or explicit guidance in the pre-training phase [4, 12, 13]. Different autoencoder variants have been proposed to tackle this problem implicitly, including the sparse autoencoder (AE) [14–16], the denoising AE [14, 17, 18], and the contractive AE [8, 12, 19–23]. In attempting to learn meaningful features, the AEs mostly follow the principle of minimizing the reconstruction error and maximizing the robustness of the feature representations. However, learning the representations in the unsupervised pre-training phase means that the autoencoders do not know the particular supervised task in the fine-tuning phase [24]. The current study proposes DL architectures and approaches capable of learning meaningful features concerning the given objective in an end-to-end manner.

Evaluation of DL models is usually achieved through error computation on validation sets disjoint from the training data. However, the validation error is only a proxy for the true error as the validation set might differ statistically from the true distribution [25]. For EO data analysis, an inspection of the model rendered interpretable can thus be a good complement to the basic validation procedure [23, 26]. The explainable models give a collection of features that have contributed to the decision and the relevance scores indicating to what extent each feature contributes [18]. In the context of EO data analysis, the most relevant features that support the classification decision will be highlighted and given positive scores. The rich feedback provided by explanation allows in principle to explore the space of DL models in a more guided manner than a validation procedure based only on classification or mean squared errors. Consequently, interpretability and explainability mechanisms are proposed to effectively evaluate DL models in terms of the significance of the features learned.

In summary, the current study proposes generic architectures and regularizations to make convolutional models capable of learning meaningful features for a given task such as classification. The effectiveness of the proposed approaches is evaluated with reference to modeling features of vegetation index phenological curves and multispectral EO images. The physical significance of learned features and their effectiveness in distinguishing the irrigation schemes and crop types are illustrated. Learning meaningful features of Normalized Differential Vegetation Index (NDVI)-derived phenological curves that can differentiate irrigation schemes with a minimum number of training samples, effectively alleviating the noise and other irregularities, require an interpretable approach. Equally important is the proper modeling and learning of meaningful features for learning crop-specific features, with limited training samples and wavelength bands, to assign crop-labels to multispectral EO image pixels. This paper also proposes interpretation-based evaluation strategies suitable for producing explanations in the context of EO data-based DL models.

The main contributions of the current study are (1) development of a self-interpretable DL-based approach called interpretable adversarial encoding network (IAE-Net); (2) study of interpretability- and explanation-based evaluation measures for DL models; and (3) illustration of the proposed approaches for application in agriculture domain namely crop phenology-based classification and crop-specific feature learning.

2 Related studies

This section reviews the recent approaches related to the interpretability and explainability of DL models relevant to EO data analyses. Section 2.1 presents different prominent approaches that make the DL networks capable of learning meaningful features. Different approaches employed to interpret and explain DL models are reviewed in Sect. 2.2. The specific contributions of this research and the novelty of the proposed approaches are discussed in each subsection.

2.1 Learning meaningful representations

Recent advances for learning unsupervised representations are summarized in [17, 27, 28]. Although generative adversarial networks (GANs) are useful for learning optimal feature representations [29–32], they do not accurately model the spectral features and require a large number of training samples [33]. Discriminative and generative variants of autoencoders [28, 34–38] have also illustrated explicit modeling and leveraging of sample relations to encode real data manifold. Although these methods are much faster at inference time and leverage on large datasets, the importance of local features is not considered [24, 39–41]. Han et al. [39] illustrated a CNN-based task-specific feature generation. However, the architecture is not appropriate with limited training samples and for processing irregularly sampled point data. Long Short-Term Memory (LSTM) and variants also proved effective for learning latent manifolds of feature-rich spectral features such as phenological curves and pixel spectra [42, 43]. However, the LSTM-based approaches generally ignore

characteristic features of phenological curves and the Euclidean losses are prone to sampling biases. Most of the above discussed approaches focus on learning the intrinsic manifold but generally ignore the importance of learning physically significant features which is critical in EO data analyses.

Deep Belief Networks (DBNs) and stacked AEs trained with sparsity constraints have been proposed to penalize extreme neuron activations for improving the interpretability of the learned representations [15]. Although sparsity-constraint is perceived as prior knowledge of the input data, the difficulty in ascertaining the intrinsic sparsity of the input data makes the hyper-parameter selection difficult and the approach inefficient [24]. Similarly, selecting the level and type of correction in denoising AEs is not deterministic and affects the interpretability of the learned low-dimensional underlying manifolds [16]. It may be noted that the representations learned by contractive AEs [44], which use derivatives of latent features as a penalty term to reduce the sensitivity a priori, are robust but are not generally meaningful. An alternative approach is to combine the reconstruction error with possible embedding losses coming from Laplacian eigenmaps [45], multidimensional scaling [46], and margin-based labeling [47]. Zhuang et al. [48] attempted to model the features according to label information to reconstruct and produce labels simultaneously. Sun et al. [24] used classification error as a penalty term to the traditional reconstruction cost function to measure the benefit of the learned representations to the supervised task. Although disentanglement-based approaches [49–55] encourage independence in the latent space dimensions to improve interpretability, high values of penalty factor lead to poor reconstructions [56]. In addition, they are optimal for imparting feature disentanglement rather than for dynamically learning meaningful features.

It is hypothesized that the use of adversarial variational encoding along with approaches to refine the latent representations in accordance with the input data distribution can improve the significance of the learned features. In addition, classification-based constraints and losses, proposed in this study, refine the learned features in accordance to the classification objective. Besides, the embedding of label information in the latent space improves the interpretability of the features dynamically.

2.2 Interpretability and explainability of DL models

A recent survey of the different approaches to analyze the interpretability and explainability of DL models can be referred to [7, 12]. Most of these approaches employ different techniques such as attention-based models [9, 20], saliency map generation [57], variable effect obscuring [58], contribution score assignment [59], and layer-wise relevance propagation (LRP) [23, 60] for computing the relevance of input features [61, 62]. Another set of strategies such as model-specific local explanations [63], classification prototype estimations [51, 64], disentanglement [38, 54, 65] and local linear approximations [21, 66] attempt to explain the DL models based on the learned concepts or prototypes. Sensitivity analysis-based DL explanation approaches generally define a vector field where each vector indicates the direction of alternate classification [67]. Although tree-based explanations have been widely explored to explain DL models, most of them lacks smoothness and need of complex trees makes it difficult to interpret complex models. Model-agnostic methods separate explanation from a machine learning model, allowing the explanation method to be compatible with a variety of models [21, 68, 69]. Among the various DL-based explanation strategies, LRP method yield attribution scores that quantitatively better represent the importance of the input features [7, 12, 57]. LRP also benefits from straightforward implementation compared to the DeepLIFT method [59] that requires the determination of a reference input [70]. However, the modeling of LRP to consider the specific nature of the given models is least explored. Counterfactual-based explanation strategies involve detecting the smallest possible change in feature values that causes an alteration to the prediction of the model [71, 72]. Aravantinos and Diehl [73] attempted to trace each component of a final inference model to ensure that all choices, such as hyper-parameters and architecture, are well justified. In Al-Hmouz et al. [74], the operations of each neuron are restricted to limit the extracted features to a logical combination of the input features. However, the applicability of the approach is restricted only to simple models. Unlike the existing model-specific interpretability approaches, the current study explores a generic interpretability strategy to evaluate the physical significance of the learned features and network pipelines. The effectiveness of different interpretability and explainability strategies in the evaluation of DL models will be illustrated for the analysis of EO data.

It is hypothesized that the modeling of activations, latent manifold, and layer-wise propagations can be employed to compute the representation of different concepts learned by the network. The inspection of the learned concepts is hypothesized to be an evaluation measure of the network's interpretability and learnability. Besides, the generative and discriminative priors are hypothesized to improve the LRP strategy to effectively explain the significance of the different input features concerning the concepts learned.

3 Materials and methods

This section discusses the datasets used, proposed approaches, and implementations. A brief discussion of the study area and datasets used are presented in Sect. 3.1. The proposed approaches are discussed in Sect. 3.2. The implementation details of the proposed approaches for two specific applications are presented in Sect. 3.3.

3.1 Datasets

The current study employs the Vegetation and Environment monitoring New Micro-Satellite (VENμS) data collected over two agricultural farms in Israel for phenology-based irrigation scheme classification and crop-specific feature learning. The VENμS sensor is characterized by a high spatial resolution of 5 m, a high spectral resolution of 12 narrow bands in the visible to near-infrared regions of the spectrum, and a high revisit time of 2 days at the same viewing and azimuth angles. For analyzing the proposed approaches for phenological curve-based irrigation scheme classification, 90 fields of wheat with two irrigation regimes (rainfed and irrigated) are considered. The NDVI of fields computed over three crop years 2018, 2019, and 2020 are used for the analysis. It may be noted that the temporal index curves, having a vector length of 27, are used as inputs for analyzing the index curve-based classification models. For analyzing the proposed approaches with regard to crop-specific feature learning, VENμS images covering 90 fields of wheat and 40 fields of potato are employed. It may be noted that the shapefiles of crop fields, along with the cropping, harvesting, and irrigation information obtained from framers, serve as ancillary data for labeling the phenological curves and image pixels. For sensor-specific feature learning model, VENμS image patches of size $5 \times 5 \times 12$ are fed to the model for learning spatial and spectral features. The spatial context of 5×5, considered in this study, is empirically found to give optimal results in terms of classification accuracy and computational performance.

3.2 Proposed approaches

The following subsections present the proposed approaches to develop interpretable DL frameworks. Section 3.2.1. presents a variational encoding strategy with adversarial training, called IAENet, that incorporate label information to learn discriminable representations. The constraints and losses proposed to make the network to learn meaningful features are presented in Sect. 3.2.2. Different explainability strategies to evaluate the proposed DL models are presented in Sect. 3.2.3.

3.2.1 Architecture for learning meaningful features

This study proposes variational autoencoder-based architecture, called IAENet, to learn meaningful representations for a given objective. The proposed architecture is presented in Fig. 1. The encoding function $q(z|x)$ defines an aggregated posterior distribution $q(z)$ of the hidden code vector of the autoencoder as follows:

$$q(z) = \int q(z|x)p(x)dx \tag{1}$$

where x is the input, z is the latent representation, $p(x)$ is the data distribution, and $q(z|x)$ is the encoding distribution. To facilitate the learning of meaningful disentangled features, the loss function of the proposed reconstruction stream constitutes of reconstruction and cross-entropy losses along with the disentanglement constraint as:

$$L_G = E_x\left[E_{q(z|x)}[-\log(p(x|z))] - \beta \mathrm{KL}(q(z|x)||q(z))\right] \tag{2}$$

where x is the input, z denotes the latent representation, $q(z)$ is the prior distribution imposed on z, $q(z|x)$ is the encoding distribution, $p(x|z)$ is the decoding distribution, β is the disentanglement penalty factor, and KL(.) is the Kullback–Leibler divergence. In order to consider the spectral fidelity and series nature of the remote sensing images and phenological curves, spectral dissimilarity and dynamic time wrapping (DTW)-based losses are also employed as follows:

$$L_s = \arccos\left(\frac{x.\tilde{x}}{|x||\tilde{x}|}\right) + \psi_{\alpha}A, \Delta(x, \tilde{x}) \tag{3}$$

where m is the length of the input vector x, $\tilde{x}$ is the reconstructed output, $\psi_\alpha(.)$ is the generalized minimizing function with a smoothing parameter α,$\Delta(.,.)$ denotes the cost matrix, and A is the alignment matrix.

The limitation of the standard variational AEs [35, 75] is that the learned latent code is not exclusive as it contains a stochastic variable that is randomly sampled from the prior distribution [24, 76, 77]. Also, the stochastically sampled latent code is unstable and will corrupt the features for classification. Furthermore, the approximation of the posterior distribution of z to the manually set prior distribution leads to information loss. To resolve these issues and facilitate the learning of features relevant to the given domain and task, in the proposed IAENet architecture (Fig. 1), instead of directly using the input x for sampling the latent code z, an encoded form of x (x_e) is used. Besides, the sampled latent code z is refined using an adversarial network (discriminator-generator network) that matches the aggregated posterior ($q(z)$) to $p(x)$. In other words, an additional adversarial loss is employed to increase the proportion of the inherited part and to decrease the proportion of stochastically sampled part in z, thereby

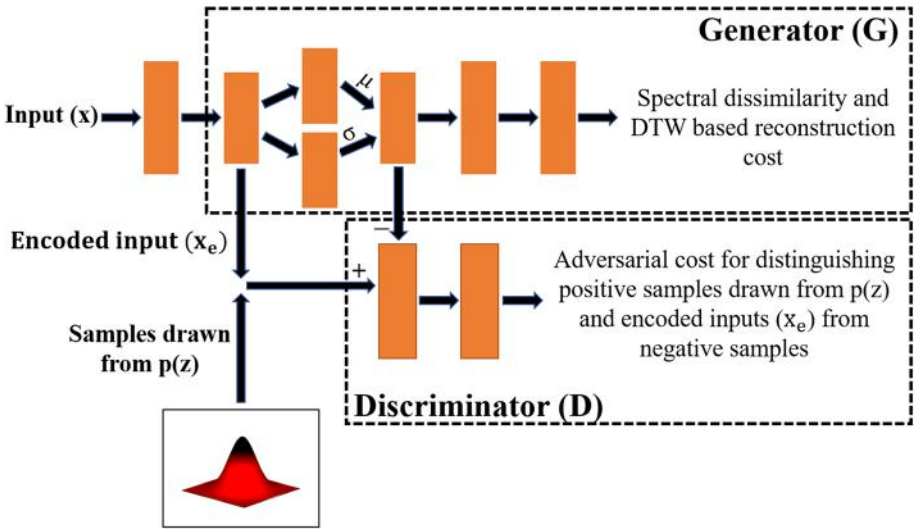

Fig. 1 Generic architecture of the proposed interpretable adversarial encoding network (IAENet) for learning meaningful features (μ and σ denote the mean and standard deviation of the latent representations)

increasing the interpretability of the learned manifold with respect to the given domain. The objective function of the proposed adversarial stream is formulated as follows:

$$L_{adv} = \min_G \max_D E_{x\sim p}[\log D(x)] + E_{z\sim q_{(z)}}[1 - \log D(G(z))] \tag{4}$$

where x is the input, z is the latent representation, p is the data distribution, and $q(z)$ denotes latent code distribution. The generator G(.) and the discriminator D(.) are alternatively trained. The generator attempts to fool the discriminator that tries to distinguish the encoded latent code from the code sampled from $q(z)$. The generator generates z based on the joint probability $q(\mu, \sigma, q(z))$ instead of the noise in standard GAN. The posterior $q(z|x)$ is no longer constrained to be Gaussian and the encoder can learn any arbitrary posterior distribution for a given input x. It may be noted that the approach is designed to consider both the labeled and unlabeled samples. Hence, in addition to classifying the latent codes as encoded or sampled from $q(z)$, the discriminator is modified to assign the latent codes to the correct class. This dual classification is achieved using the fake samples (sampled latent code z) as belonging to an additional class. Hence, the discriminator loss is formulated as follows:

$$\begin{aligned} L_D = &- \lambda_s * w_1\left(E_{z,y\sim q}[\log(D(y|z, y<k+1))]\right) \\ &- (1-\lambda_s)w_2\left(\left(E_{z\sim q(z|x)}[\log(D(y=k+1|z))]\right)\right. \\ &\left.- \left(E_{z\sim q(z|x)}[\log(1-D(y=k+1|z))]\right)\right) \end{aligned} \tag{5}$$

where x is the input, w_1 and w_2 are the weightage for the labeled and unlabeled losses, respectively, $q(z|x)$ is the decoding distribution, D(.) is the discriminator, z and y, respectively, refer the latent code and output of the discriminator, k is the number of classes, and λ_s symbolizes the flags for supervised training.

3.2.2 Constraints and losses for learning meaningful representations

This subsection proposes explicit and implicit guidance mechanisms to constraint the latent representations in IAENet for learning meaningful features. The constraints and losses proposed in this section can be extended to other architectures as well.

Let z be the latent code learned, the IAENet is constrained to fine-tune z for the classification task as follows:

$$\begin{aligned} L_f = \alpha \sum_{i=1}^{m} &-[y_i \log(\chi(w_c, z_i)) + (1-y_i)(1-\log(\chi(w_c, z_i)))] \\ &+ \lambda|z_i| \end{aligned} \tag{6}$$

where y_i is the hot encoded output of the discriminator for the ith sample, $\chi(.)$ is the SoftMax loss, w_c is the weight matrix of the encoding layers, α and λ are the scaling factors, z_i is the latent representation of the i^{th} sample, and m is the number of samples. An additional classification loss is employed to incorporate the label information of the source domain into the embedding space as follows:

$$L_c = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} 1\{y_i = j\} \log \frac{e^{l_i}}{\sum_{r=1}^{c} e^{l_{ir}}} \tag{7}$$

where $1\{\cdot\}$ is an indicator function, y_i is the discriminator output corresponding to the ith sample, l_i is the predicted label for the ith sample, n is the number of samples, and c denotes the total number of classes.

Although penalizing $KL(q(z)||p(z))$ term as in Eq. (2) facilitates disentanglement, it amounts to the loss of the information about x stored in z resulting in a poor reconstruction for high values of β. Hence, in this study, inspired

by Kim and Mnih [56], the formulation in Eq. (2) is modified as follows:

$$L_G = E_x\left[E_{q(z|x)}[-\log(p(x|z))]\right] - E_x[KL(q(z|x)||q(z))] - \beta KL\left(q(z)||\prod_{j=1}^{d} q(z_j)\right) \tag{8}$$

where x is the encoded input, z is the latent representation, $q(z)$ is the prior distribution imposed on z, $q(z|x)$ is the encoding distribution, $p(x|z)$ is the decoding distribution, d is the dimension of the latent space, β is the disentanglement penalty factor, and KL(.) is the Kullback–Leibler divergence. The last term in Eq. (8) measures the dependence for multiple variables. The additional constraint L_G is incorporated in the discriminator loss function discussed in Eq. (5).

3.2.3 Transparency and explainability

This subsection discusses the approaches used in this study to understand the concepts learned by the trained IAENet. The approaches to understand the way the input features have contributed to a given decision is also discussed. These approaches are used to evaluate and compare the DL models in terms of interpretability and explainability.

Inspired by Montavon et al. [25], for interpreting the proposed IAENet, an approach based on activation mapping is employed. In this regard, the representative/prototype conditioned data density, $q\ (z)$ is the distribution of the learned latent space, and λ is the scaling factor.

Although interpreting the concepts/prototypes learned by DL models helps compare and contrast different models, learning the contribution of input features for each concept is also equally important. In this study, a modified version of LRP is employed to assign quantitative values to input features based on their relative significance in the output prediction. As shown in Bach et al. [60], the relevance can be distributed to input-layers using local redistribution rules as follows:

$$R_i^{(l)} = \sum_j \frac{a_j^{(l-1)} w_{ji}^{(l-1,l)}}{\sum_k a_k^{(l-1)} w_{ki}^{(l-1,l)}} R_i^{(l-1)} \tag{12}$$

where $a_j^{(l-1)}$ denotes the activation and j indexes all neurons of layer l-1 joined to the neuron i. Eq. (12) is applied in a backward pass through the network from the output layer to produce the relevance map. It may be noted that the summation of the relevance at any layer is conserved in the network. The introduction of a numerical stabilizer in Eq. (12), as discussed in Shrikumar et al. [59], removes noise elements in the explanations and restricts the number of features. However, the approach results in sparse explanations making them challenging to interpret. In order to improve the understandability of the explanations and to avoid unrelated concepts, Eq. (12) is modified for the Rectified Linear Units (ReLU) convolution layers as follows:

$$R_i^{(l)} = \left(\sum_j \left(\tau \frac{\left[a_j w_{ji}\right]^+}{\varepsilon.\psi\left(\sum_{i'} a_{i'} w_{i',i}\right) + \sum_{i'}\left[a_{i'} w_{i',i}\right]^+} + \delta \frac{\left[a_j w_{ji}\right]^-}{\varepsilon.\psi\left(\sum_{i'} a_{i'} w_{i',i}\right) + \sum_{i'}\left[a_{i'} w_{i',i}\right]^-}\right)\right) R_i^{(l-1)} \tag{13}$$

type of the class ω_c, which correspond to the most likely input x for class w_c, is found by optimizing:

$$\max_x (\log p(\omega_c|x) + \log p(x)) \tag{9}$$

where x is the input, and $p(w_c|x)$ and $p(x)$ are the class conditioned data density and data model, respectively. Adopting the generative model G(.) (discussed in Sect. 3.2.1) for modeling $p(x)$, Eq. (9) is reformulated as follows:

$$\max_z \left(\log p(\omega_c|G(z)) + \log q(z) - \lambda\|z\|^2\right) \tag{10}$$

$$x^* = G(z^*) \tag{11}$$

where z denotes the latent code, ω_c denotes the class, G(.) is the discriminator, $p(\omega_c|G(z))$ denotes the class

where R_i^l is the relevance score for neuron i in layer l, j indexes all neurons of layer l joined to the neuron i, ε is a numerical stabilizer, a_j is the activation of the jth neuron, w^{ji} is the weight between the ith and jth neuron, $[.]^+$ and $[.]^-$ are the positive and negative components, respectively, $f(x)$ is the total relevance at the output layer, and the function $\psi(.)$ yields the sign of the expression. The scaling factors τ and δ are constrained to $\tau + \delta = 1$ to ensure the conservation property. The reformulation facilitates to filter the spurious variations in convolution layers and is less sensitive to the entanglement in the upper layers. For the classification layer, the relevance is propagated from the outputs using the propagation rule as follows:

$$R_i^{(l-1)} = \left(\sum_j \left(\frac{a_i w_{ji}}{\varepsilon . \psi\left(\sum_{i'} a_{i'} w_{i',i}\right) + \sum_{i'} a_{i'} w_{i',i}} \right) \right) R_i^{(l)} \tag{14}$$

where R_i^l is the relevance score for neuron i in layer l, ε is a numerical stabilizer, a_i is the activation of the ith neuron, w^{ji} is the weight between the ith and jth neuron, and the function $\psi(.)$ yields the sign of the expression. By analyzing the relevance scores, regions, or patterns in the inputs with high positive relevance that mostly contribute to a classification decision are identified. As the probability of an input belonging to a certain class depends on the value at the output layer neuron, relevance scores can represent evidence for (positive values) and against (negative values) the classification decision.

The concepts learned by the model and the relevance assigned to the input features are analyzed to evaluate the physical significance of the learned features and interpretability of the models. Besides, the training data as well as parameter selection are also refined based on the working of the network pipeline interpreted from the analysis of learned prototype and relevance score assignments.

3.3 Implementation of the proposed IAENet

Modeling and designing the proposed IAENet model toward crop phenology-based water stress detection and crop-specific feature learning are discussed in the following subsections. The implementation details and optimal hyper-parameter settings for each of these applications are also discussed.

3.3.1 Classification of phenological curves based on irrigation scheme

This subsection discusses the implementation of IAENet for distinguishing the rainfed and irrigated wheat crops. The phenological curves based on the NDVI derived from multi-date VENμS images are used to train and validate the networks. The approach adopted is similar to the one discussed in Sect. 3.2. However, to resolve the effects of shifts, time series nature, and irregular sampling of the phenological curves, DTW-based nonlinear units are employed instead of the conventional neurons. The DTW units match similar features to the input and skip elements with a considerable distance to the weights and perform small translations. The activation of a given DTW node is computed as follows:

$$z^l = \phi\left(\sum_{(i,j)\in M_j} \left\| w_i^l - a_j^{l-1} \right\| \right) \tag{15}$$

where $a^l{}_j$ and $w^l{}_j$ are the jth activation vector and network weight vector, respectively, of the lth layer, $\phi(\cdot)$ is the activation function, and M_j is the set of matched indices corresponding to the index i of w^l and the index j of a^{l-1}, respectively. The set of matched indices M_j allows for duplicate and skipped values of w_i^l, and a_j^{l-1}.

The convolution operation is also modified to consider the specific nature of the phenological curves, such as irregularity and times series nature. The interpolated convolution of the vectorized phenological curve v with a kernel function κ(.), centered at a location $\tilde{x}$, is implemented as follows:

$$x * \kappa(\tilde{x}) = \sum_{x'} \frac{1}{N_{x'}} \sum_{k_\alpha} \varphi(\kappa_\alpha, x') v(\tilde{x} + x_\alpha).\kappa(x') \tag{16}$$

where φ(.,.) is an interpolation function that computes the weights based on a filter weight vector k_α and a given input point x', and $N_{x'}$ is the density normalization term to make the convolutions sparsity invariant. It may be noted that along with kernel size, kernel length $l \in R$ is another hyper-parameter that is defined as the distance between two adjacent weight vectors to control the receptive field. In addition to the reconstruction losses (for encoder-decoder stream) discussed in Sect. 3.2, to embed label information in the latent manifold learning, an additional loss (L_E) is employed as follows:

$$L_E = \sum_{v\in V^+} \left(v - D'_\theta(G_\theta(v))\right)^2 + \sum_{v \int V^-} \left(v - D'_\theta(G_\theta(v))\right)^2 \tag{17}$$

where V^+ and V^- are the sets of irrigated and rainfed samples, respectively, and G_Θ and D'_θ, respectively, are the generator and decoder networks. Further, to ensure the piece-wise similarity of the reconstructed and expected outputs, a multiscale version of the structural dissimilarity loss is also employed as follows:

$$L_{SD} = \sum_{p\in \mathbf{P}} 1 - \Omega(p)$$

where (18)

$$\Omega(p) = \frac{2\mu_p \mu_p^{'} + C_1}{\mu_p^2 + \mu_p^{'2} + C_1} . \frac{2\sigma_p \sigma_p^{'} + C_2}{\sigma_p^2 + \sigma_p^{'2} + C_2}$$

where $P \subseteq R$ is the set of all relative locations of the phenological curve, C_1 and C_2 are constants, μ_p and $\mu_{p'}$, respectively, represent the means of the patches of the reconstructed and ground truth phenological curves while

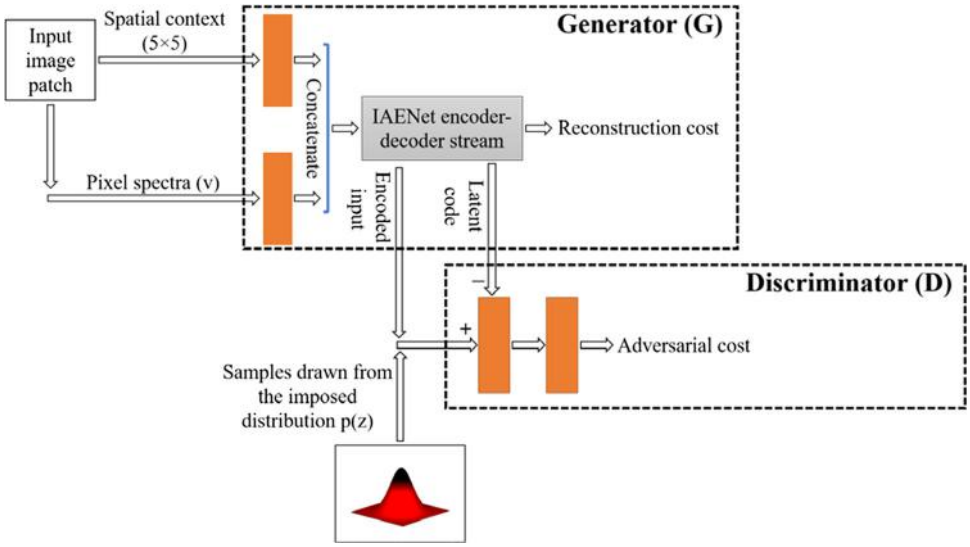

Fig. 2 Proposed implementation of interpretable adversarial encoding network (IAENet) for crop-specific feature learning

σ_p and $\sigma_p^{'}$, respectively, denote the corresponding standard deviations. The means and standard deviations are computed in neighborhoods (context) of varied extents to implement multiscale measurements of structural dissimilarity.

The implementation of IAENet, adopted in this study, for the phenological curve-based irrigation scheme classification uses multi-size kernels of sizes 1×2, 1×3, 1×5, 1×7, and 1×9 in the encoder and decoder streams. ReLU activations follow the padded convolutions in all layers, except the discriminator stream. The stride of pooling and unpooling layers are kept to two for, respectively, halving and doubling the resolution of the resulting feature maps. Convolutions following down-sampling steps double the number of feature maps while is halved by the convolutions following upsampling. The interpolated convolution kernels use Gaussian interpolation as the interpolation function, and the Gaussian bandwidth (3σ) is fixed to 0.1. The number of filters in the first encoding unit is empirically set to 64. It may be noted that for implementing multiscale structural dissimilarity measurements, the context extents are varied from 1, 3, 5, 7, and 9. The discriminator network consists of a fully connected layer having a depth of two. The network is trained for 300 epochs with an initial learning rate of 0.01 and a decay rate 0.5 every 100 epochs with a batch size 30. Hyper-parameter optimization, proposed in Bochinski et al. [78], is employed to optimize the parameters of the proposed network such as kernel size, number of filters, depth of the network, and number of epochs. The mean squared error (MSE)-based loss and cosine dissimilarity loss, along with the proposed piece-wise dissimilarity loss, are employed to learn the network weights.

3.3.2 Learning crop-specific features for classification

This subsection discusses the implementation of IAENet for dynamically learning physically significant features for distinguishing crops from multispectral VENμS data. The labeled multispectral image patches are used to train and validate the network. The architecture and constraints discussed in Sect. 3.2 are adopted for the purpose. However, to consider the spatial and spectral contexts, an input stream consisting of 1D and 2D convolutions to, respectively, process the spectral and spatial features is embedded in IAENet framework as shown in Fig. 2.

In the current implementation of IAENet for crop-specific feature learning, the input stream employs 16 2D filters of kernel size 5×5 and 16 1D filters of size ranging from 1×3 to 1×7 to model the spatial and spectral features, respectively. The spectral and spatial features are concatenated and fed to IAENet framework that minimizes the losses (discussed in Sect. 3.2.) between the reconstructed (v') and original spectra (v)) to learn the network weights. The padded convolutions are followed by ReLU activations in all layers except the discriminator stream. The stride of pooling and unpooling layers are kept to two for, respectively, halving and doubling the resolution of the resulting feature maps. Convolutions following down-sampling steps double the number of feature maps while is halved by the convolutions following upsampling. The number of filters in the first encoding unit is empirically set to 64. The discriminator network consists of a fully connected layer having a depth of two. The network is trained for 200 epochs with an initial learning rate of 0.01 and a decay rate 0.5 every 100 epochs with a batch size 30. Hyper-parameter optimization, proposed in [79], is employed to optimize the parameters of the proposed

network such as kernel size, number of filters, depth of the network, and number of epochs.

4 Results

To verify the effectiveness of IAENet, extensive experiments were conducted using multi-date VENμS images for phenological curve-based irrigation scheme detection and crop-specific feature learning. The ablation analysis of IAENet for each of the applications is discussed in Sects. 4.1.1 and 4.2.1. Sections 4.1.2 and 4.2.2 present comparative analysis of IAENet with the benchmark approaches. Hyper-parameter optimization, proposed in Bochinski et al. [79], is employed to optimize the parameters of different models experimented in this study. It may be noted that an early stopping framework using k-fold validation forms the basis of the parameter selection. The confusion matrix-based Kappa statistics and overall accuracy are used for evaluating the classification results. High values of Kappa statistics and overall accuracy indicate high accuracy. An Z-score-based test statistics (discussed in [80]) is employed to analyze the significance of the results presented in this study. Along with confusion matrix-based measures, proposed interpretability techniques (Sect. 3.2.3) are used to evaluate the physical significance and interpretability of the models. For all the experiments adopted in this study, k-fold validation is adopted with k set to 10 for both the datasets.

4.1 Classification of phenological curves based on irrigation schemes

The pixel-level NDVI phenological curves derived from multi-date images are used to train IAENet to distinguish between irrigated and rainfed wheat crops. The ground truth ancillary data are used to generate labels for phenological curves and are used for training the network. A GAN-based augmentation, similar to the one adopted in [81], is used to increase the number of training samples. Besides, random Gaussian noise is added in irregular intervals to evaluate the effect of denoising. The approach is extensively analyzed over the data of wheat fields over three consecutive crop years. It may be noted that a total of 3600 samples are used for training and testing the model among which 800 are augmented patterns.

4.1.1 Ablation analysis of IAENet implementation for irrigation scheme detection

This subsection evaluates the effect of different regularizations and losses on the proposed architectures for index curve-based classification. In other words, experiments are conducted to analyze how the accuracy varies if the proposed architectures (Sect. 3.2.1) and constraints (Sect. 3.2.2) are altered or not applied. The results are summarized in Table 1. It is observed that the proposed strategies reduce the training sample requirement and significantly improve the results (in terms of Kappa and overall accuracy) as they facilitate learning interpretable manifold. Besides, the use of piece-wise loss (Sect. 3.3.1), interpolation-based convolution (Sect. 3.3.1), and DTW-based neural units (Sect. 3.3.1) also improves the classification accuracy. In addition to improved classification, the proposed architectures and constraints improve the concepts learned for both irrigated and rainfed crops. The relevance analysis of input features also indicates that the proposed constraints play a significant role in improving the interpretability of the learned latent space. It may be noted that the entanglement penalty (discussed in Sect. 3.3.2) is empirically set to 2 and successfully disentangles the latent codes with regard to the irrigated and rainfed classes.

Table 1 Analysis of the effect of proposed architectural variations and constraints

Architectural variations/losses	Kappa	Overall Accuracy	Z-score
Implementation without variational encoding constraint	0.79	84.34	2.65
Implementation without an adversarial loss for incorporating input prior to the Gaussian space	0.82	86.59	2.41
Implementation using classification loss instead of dual loss	0.90	93.68	2.15
Implementation without constraints for embedding classification prior to the adversarial loss	0.92	96.71	2.08
Implementation without label embedding constraint	0.93	96.08	1.98
Implementation without piece-wise loss	0.92	94.42	2.26
Implementation without cosine dissimilarity loss	0.93	96.80	1.99
Implementation without interpolated convolution	0.91	93.43	2.07
Proposed IAENet implementation	0.95	98.79	–

*Z-score > 1.96 shows a significant ($> 95\%$) difference between the confusion matrices of the existing approaches and interpretable adversarial encoding network (IAENet)

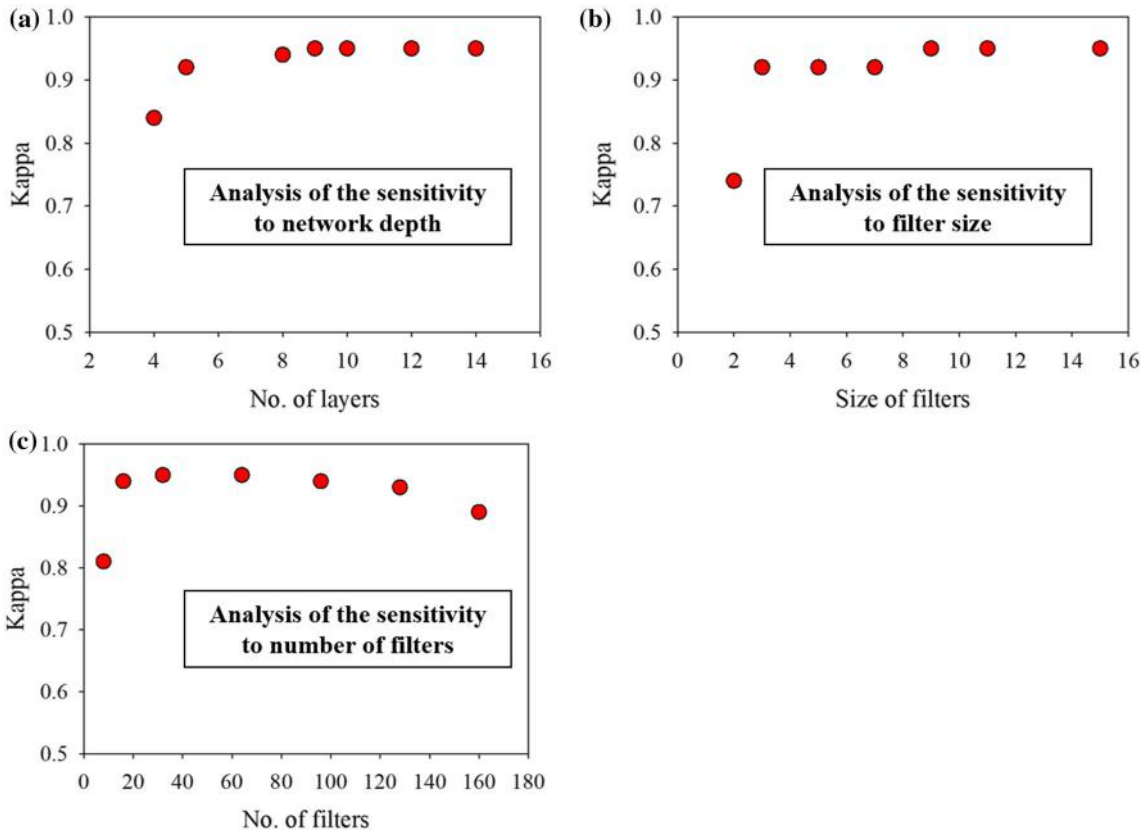

Fig. 3 Analysis of the sensitivity of interpretable adversarial encoding network (IAENet) toward **a** depth of the network layers; **b** size of filters; and **c** number of filters for phenological curve-based irrigation scheme detection

The sensitivity analysis of the proposed IAENet model (implemented for irrigation scheme classification) toward network parameters is presented in Fig. 3. An increase in network depth, number of filters, and filter-size are found to improve the accuracy to a limit beyond which it deteriorates or saturates due to over-/under-fitting. The reduction in sensitivity of IAENet to network parameters can be attributed to the improvement in interpretability of the learned manifold, achieved through adversarial encoding and data prior embedding. Empirically, for input curves having a length of 24–48, a 2–6-layered network yields the best results. Also, the approach is found to be less sensitive to slight changes in the depth of the reparameterization stream. The increase in the number of kernels improves the accuracy to a limit, but the trend saturates gradually. The size of filters is found to be a critical factor and needs to be tuned in accordance with the data. As the length of spectral features can vary from even one to a few pixels, too big sized kernels may sometimes ignore essential features. Also, very small-sized kernels may capture the noise instead of actual spectral features. In addition, the increase in size and number of filters exponentially increases the computational complexity of the network. Hence, a better trade-off needs to be adopted. The use of multi-sized kernels is found to be a viable alternative as it significantly improves the results without much affecting the execution time.

Table 2 Comparison of interpretable adversarial encoding network (IAENet) with benchmark deep learning (DL) classifiers for 70% of training samples

Benchmark classifiers	Kappa statistics	Overall Accuracy
Karim et al. [43]	0.73	76.89
Han et al. [39]	0.76	80.10
Kang et al. [36]	0.80	84.43
Sun et al. [24]	0.84	87.91
Kim and Mnih [56]	0.76	80.08
Hang et al. [31]	0.81	84.43
Jiang et al. [32]	0.85	87.98
Mou and Zhu [42]	0.89	93.54
Honke et al. [38]	0.87	90.08
Proposed IAENet	0.96	98.27

Benchmark methods are implemented based on the available GitHub implementations and are fine-tuned based on the related publications

Table 3 Z-score-based significance analysis of interpretable adversarial encoding network (IAENet) in comparison with the benchmark deep learning (DL) classifiers

Benchmark smoothing approaches	Z-score of Kappa statistics as compared to IAENet	Z-score of Overall Accuracy as compared to IAENet
Karim et al. [43]	2.80	2.91
Han et al. [39]	2.42	1.98
Kang et al. [36]	1.99	2.08
Sun et al. [24]	2.52	2.16
Kim and Mnih [56]	2.79	2.35
Hang et al. [31]	1.92	2.19
Jiang et al. [32]	2.73	2.52
Mou and Zhu [42]	1.58	1.99
Honke et al. [38]	1.92	1.95

Z-score > 1.96 shows a significant (> 95%) difference between the confusion matrices of the existing approaches and interpretable adversarial encoding network (IAENet)

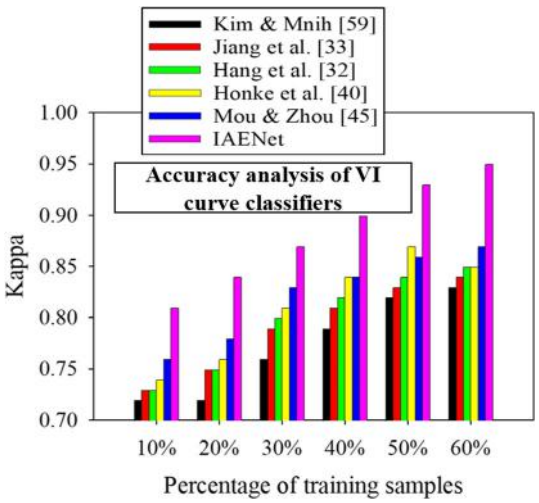

Fig. 4 Accuracy analysis of interpretable adversarial encoding network (IAENet) and deep learning (DL) classifiers with respect to the change in the percentage of training samples

4.1.2 Comparison of IAENet with the commonly used phenological curve classifiers

The commonly used classifiers, applicable to the phenological curve-based irrigation scheme classification, are compared with the proposed IAENet-based approach. The results are summarized in Table 2. The significance of the results of IAENet (at a confidence level of 95%) in comparison with the other approaches is analyzed in Table 3. Based on the discussions in [3, 38, 82, 83], some main existing classifiers are selected as the benchmark methods for comparison. It may be noted that some of the benchmark approaches are modified for the one-dimensional phenological curves. An analysis of the variation in the accuracy of different approaches according to the variation in the percentage of training samples is presented in Fig. 4. A total of 3600 samples are used in these experiments and tenfold validation is employed for each of the different sub-experiments (10%, 20%, 30%, etc.). As is evident from the results, IAENet better models phenological curves as compared to other prominent approaches. The proper modeling of features significantly improves the generalization capability of the network and results in improved classification accuracies even with a small number of training samples. The learning of physically significant features also resolves the issues of domain bias and inter-field variability of phenological curves. Besides, the DTW-based convolutional units and interpolation-based convolutions facilitate the effective transformation of vectorized phenological curves to a latent space that is more discriminative than the original space.

In addition to classification-based accuracy assessment, the models are also evaluated based on the significance of the features and concepts/prototypes learned. The concepts learned by different models for classifying the irrigation schemes are analyzed using the interpretability approach proposed in Sect. 3.2.3. Experiments indicate that the concepts learned by IAENet for distinguishing the irrigated and non-irrigated wheat crops highlight the features that correspond to the dry periods. The normalized difference between the non-rainy NDVI features for the prototype of irrigated and non-irrigated crops (in terms of area under the features) for different models is summarized in Table 4. The cosine dissimilarity of the concepts learned for both the classes is also presented in Table 4. The high feature differences and cosine dissimilarity values for IAENet indicate that the approach learns meaningful features to distinguish the irrigated and rainfed wheat crops. Besides, the prototype learned for both irrigated and non-irrigated crops can be compared with the meta data regarding the rainfall available from the metrological sources. Analysis

Table 4 Interpretability-based comparison of the different deep learning (DL) models based on the concepts learned

Benchmark classifiers	Normalized feature difference between the features of the concepts for irrigated and non-irrigated classes	Cosine dissimilarity between the concepts learned for irrigated and non-irrigated classes
Karim et al. [43]	0.09	0.895
Han et al. [39]	0.42	0.872
Kang et al. [36]	0.35	0.890
Sun et al. [24]	0.28	0.916
Kim and Mnih [56]	0.61	0.863
Hang et al. [31]	0.29	0.892
Jiang et al. [32]	0.30	0.932
Mou and Zhu [42]	0.58	0.916
Honke et al. [38]	0.49	0.902
Proposed IAENet	0.37	0.979

*Benchmark methods are implemented based on the available GitHub implementations and are fine-tuned based on the related publications

Table 5 Interpretability-based comparison of the different deep learning (DL) models

Benchmark classifiers	Normalized relevance assigned to the features of the non-rainy time slots
Karim et al. [43]	0.48
Han et al. [39]	0.39
Kang et al. [36]	0.56
Sun et al. [24]	0.63
Kim and Mnih [56]	0.60
Hang et al. [31]	0.52
Jiang et al. [32]	0.78
Mou and Zhu [42]	0.67
Honke et al. [38]	0.72
Proposed IAENet	0.96

of the learned concepts also indicates that the approach is less sensitive to the noise and irregularities in the phenological curves.

To further explain the network and analyze the contribution of input features, the modified LRP approach (Sect. 3.2.3) is adopted. The normalized relevance assigned by different approaches to the features of non-rainy time slots is presented in Table 5. The propagated relevance of IAENet, for distinguishing irrigated and rainfed wheat crops, indicates that the model gives importance to the NDVI features corresponding to the non-rainy time slots. An analysis of the learned features from bottom layers indicates that the non-rainy NDVI features are combined linearly and nonlinearly to guide the decision.

4.2 Learning crop-specific features for classification

The proposed implementation of IAENet, presented in Fig. 2, is analyzed for learning crop-specific features from VENμS satellite images. The ground truth ancillary data (Sect. 3.1) are used to label the pixels and are used for training the network. For the crops having a limited number of training samples, a GAN-based augmentation [84] is used to generate more training samples. A total of 4000 samples are used for training and testing the model among which 1600 are augmented patterns.

Table 6 Analysis of the alternative architectural choices of interpretable adversarial encoding network (IAENet) for crop-specific feature learning

Architectural variations/losses	Kappa statistics	Overall accuracy	Z-score
Implementation without variational encoding constraint	0.72	76.80	2.34
Implementation without an adversarial loss for incorporating input prior to the Gaussian space	0.75	79.15	2.06
Implementation using normal classification loss instead of dual loss	0.79	83.28	2.67
Implementation without constraint for embedding classification prior to the adversarial loss	0.77	84.15	2.81
Implementation without label embedding constraint	0.78	82.39	1.97
Implementation without disentanglement constraint	0.76	81.53	2.24
Implementation without cosine dissimilarity loss	0.80	84.28	2.15
Proposed IAENet implementation	0.82	89.05	–

*Z-score > 1.96 shows a significant (> 95%) difference between the confusion matrices of the existing approaches and interpretable adversarial encoding network (IAENet)

4.2.1 Ablation analysis of IAENet

An analysis of the proposed architectural variations and losses for crop-specific feature learning is presented in Table 6. The effect of each of the constraints and architectural modifications, used in the proposed architecture, is studied by evaluating the accuracy when each of them is not used. As is evident from the results, the use of adversarial encoding strategy and the incorporation of input prior to the Gaussian space improves the interpretability of the learned features resulting in improved Kappa and overall accuracy values. Besides, the use of cosine dissimilarity-

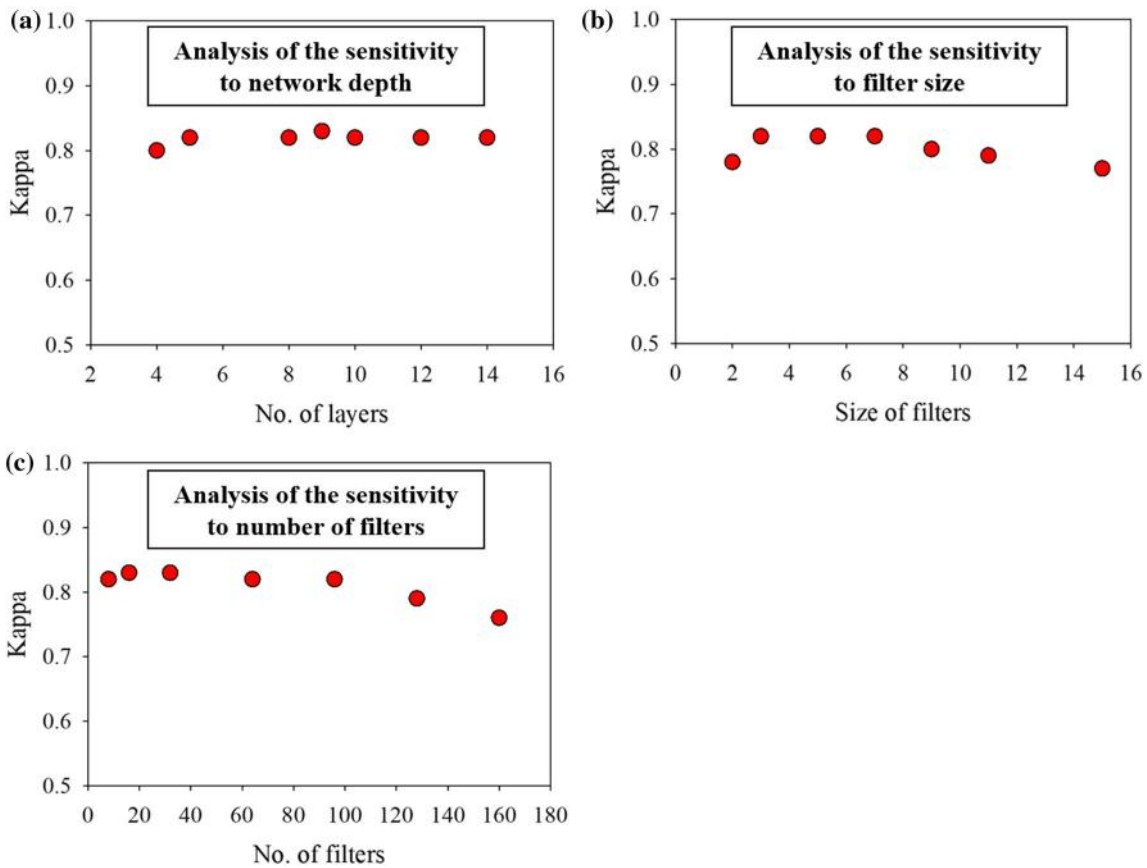

Fig. 5 Analysis of the sensitivity of interpretable adversarial encoding network (IAENet) toward **a** depth of the network layers; **b** size of filters; and **c** number of filters for crop-specific feature learning

based reconstruction loss along with the mean squared error (MSE) losses also improves the classification results (in terms of Kappa statistics and overall accuracy). It may be noted that the entanglement penalty (discussed in Sect. 3.3.2) is empirically set to 3 and successfully disentangles the latent codes with regard to the different crop classes.

Analysis of the sensitivity of DL models for learning crop-specific features shows that an increase in network depth and kernel sizes improves the accuracy (in terms of Kappa statistics and overall accuracy) to a limit beyond which it deteriorates. The use of multi-sized kernels improves the results without much affecting the execution time. The sensitivity analysis of the IAENet, implemented for feature learning, toward network parameters is presented in Fig. 5. Although an increase in the number and size of filters and depth of the network, without increasing the training data, deteriorates the classification accuracy of the existing DL models, the proposed approach is comparatively stable. The less sensitivity of IAENet can be attributed to the improved interpretability achieved through proposed architectural variations, loss functions, and constraints.

4.2.2 Comparison of IAENet with the commonly used DL approaches for crop-specific feature learning

A comparative analysis of IAENet with recently published approaches, relevant to latent feature learning, is summarized in Table 7. The significance of the results of IAENet (at a confidence level of 95%) in comparison with the benchmark methods is illustrated in Table 8. The selected benchmark approaches are improved versions of the ones that reported the state-of-the-art results [1, 2, 4, 12, 17, 28, 33, 34, 42, 48, 65, 83, 85]. An analysis of the variation in the accuracy (in terms of Kappa statistics and overall accuracy) of different approaches with respect to the variation in the percentage of training samples is presented in Fig. 6. A total of 4000 samples are used in these experiments and tenfold validation is employed for each of the different sub-experiments (10%, 20%, 30%, etc.). As is evident from the results, IAENet better models the manifold as compared to other prominent DL approaches. The proper learning of meaningful features significantly improves the generalization capability of the network and results in better classification accuracies, even with a small number of training samples.

Table 7 Comparison of interpretable adversarial encoding network (IAENet) with the benchmark deep learning (DL)-based representation learning approaches for 70% of the training samples

Benchmark classifiers	Kappa statistics	Overall Accuracy
Hoshen [35]	0.61	66.92
Zhuang et al. [48]	0.60	65.40
Sun et al. [24]	0.63	67.12
Subramanian et al. [15]	0.61	65.58
Kang et al. [40]	0.64	68.92
Kang et al. [36]	0.65	70.35
Anirudh et al. [29]	0.67	71.80
Pfau et al. [65]	0.71	75.96
Emami et al. [30]	0.74	78.29
Zhong and Deng [47]	0.78	82.26
Proposed IAENet	0.86	90.89

Benchmark methods are implemented based on the available GitHub implementations and are fine-tuned based on the related publications

Feature learning benchmark methods are followed by fully connected network

The comparison of the concepts learned by the networks (derived using the approach discussed in Sect. 3.2.3) with respect to the reference spectra is presented in Table 9. The interpretability-based analysis (Sect. 3.2.3) of IAENet for learning crop-specific features indicates that the concepts learned for each of the crops align with the corresponding crop's spectral characteristics. Analysis indicates that the proposed model emphasizes spectral bands ranging from 555 to 620 nm and 702 nm to 910 nm, specifically to the red edge bands. Also, the variational encoding strategy alleviates the noise effects, even with a minimum number of training samples.

The modified LRP-based explanation strategy (Sect. 3.2.3) is employed to analyze the relevance of different input features in accordance with the decision of the IAENet. The comparison of normalized relevance scores assigned to the red edge, green edge and near infrared features by different approaches is presented in Table 10. The relevance scores indicate that the red edge and green edge features as well as the spectral bands ranging from 555 to 910 nm contribute significantly to the decisions of IAENet. The linear and nonlinear features synthesized in deeper layers in IAENet are found to be combinations of the near infrared bands.

5 Discussion

Experiments on IAENet for different applications (discussed in Sect. 3.3) illustrate that the proposed approaches improve the interpretability of the models, yielding better results as compared to the main existing approaches. A detailed analysis of the results of each of the proposed approaches is presented in the following subsections.

Table 8 Z-score-based significance analysis of interpretable adversarial encoding network (IAENet) in comparison with deep learning (DL)-based representation learning approaches

Benchmark classifiers	Z-score of Kappa statistics in comparison with IAENet	Z-score of overall accuracy in comparison with IAENet
Hoshen [35]	2.43	2.53
Zhuang et al. [48]	1.89	2.16
Sun et al. [24]	2.15	1.92
Subramanian et al. [15]	2.09	2.33
Kang et al. [40]	2.67	2.50
Kang et al. [36]	2.23	2.69
Anirudh et al. [29]	2.71	2.15
Pfau et al. [65]	2.42	2.08
Emami et al. [30]	2.05	1.99
Zhong and Deng [47]	1.98	2.25

Z-score > 1.96 shows a significant (> 95%) difference between the confusion matrices of the existing approaches and interpretable adversarial encoding network (IAENet)

Benchmark methods for feature learning are followed by fully connected network

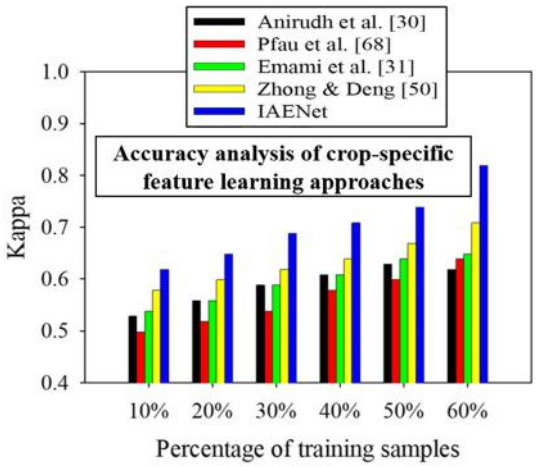

Fig. 6 Accuracy analysis of interpretable adversarial encoding network (IAENet) and the benchmark representation learning approaches with respect to the change in the percentage of training samples

5.1 Learning meaningful representations

The architecture of DL models is found to influence the capability to learn meaningful data manifolds. The use of the variational AE model, instead of the model adopted in IAENet, can constraint the latent features to a normally distributed space. However, additional sampling layers in variational AE are found to adversely affect the modeling of EO data, especially when training samples are limited. As discussed in Sect. 3.2.1 and illustrated in Sect. 4.1., the proposed IAENet uses variational encoding to project the latent representation to a normal distribution. Besides, the proposed adversarial constraints and losses use the input prior for improving the projected space to have meaningful

Table 9 Comparison of the concepts learned for each crop with the corresponding reference spectra

Benchmark classifiers	Cosine similarity
Karim et al. [43]	0.78
Han et al. [39]	0.86
Kang et al. [36]	0.95
Sun et al. [24]	0.90
Kim and Mnih [56]	0.92
Hang et al. [31]	0.94
Jiang et al. [32]	0.89
Mou and Zhu [42]	0.92
Honke et al. [38]	0.90
Proposed IAENet	0.99

Table 10 Comparison of the normalized relevance scores assigned to the red edge, green edge and near infrared features

Benchmark classifiers	Normalized Relevance Score
Karim et al. [43]	0.49
Han et al. [39]	0.52
Kang et al. [36]	0.58
Sun et al. [24]	0.61
Kim and Mnih [56]	0.81
Hang et al. [31]	0.76
Jiang et al. [32]	0.88
Mou and Zhu [42]	0.85
Honke et al. [38]	0.86
Proposed IAENet	0.98

and interpretable representations. The multiple kernels and multi-layer abstractions facilitate the modeling of different characteristic features at different resolutions. The depth of the network determines the level of abstraction, whereas the number of kernels determines the latent space dimension. Different experiments in this study illustrate that using multi-size kernels facilitates the modeling of features more effectively rather than using single-size kernels. As is evident from Sects. 4.1.2 and 4.2.2, the improvement in interpretability of learned manifold reduces the model's sensitivity toward network parameters.

Lack of generalizability usually makes the DL networks trained on one type of data less effective for another. The data and domain bias and the limited availability of training samples also affect the effectiveness of DL approaches. As illustrated in Sects. 4.1 and 4.2., the proposed strategy of using adversarial variational encoding and data prior constraints have significantly resolved these issues (Sect. 4.1). For the data having time series nature and are irregularly sampled, such as phenological curves, DTW-based convolutional units and interpolation-based convolution are found to be useful compared to the normal ones. For learning the features from image patches (Sect. 4.2), the use of a separate stream for initial learning of spatial and spectral features improves the results compared to the use of 3D filters. The differences in results are specifically significant when the number of training samples is limited.

The experiments on both the applications, namely irrigation scheme detection and learning of crop-specific features, indicate that the evaluation matrices based on classification or reconstruction accuracy are insufficient for DL models. Interpretability analyses based on the prototype learned by the models (as explained in Sect. 3.2) facilitate explanations to the models as presented in Tables 4 and 9. The concept-based interpretability analysis (Tables 4 and 9) indicates the features to which the network is giving importance. Similarly, the relevance propagation-based analyses (explained in Sect. 3.2 and results presented in Tables 5 and 10) attempt to explain the models based on the relevance assigned to the input features. The proposed IAENet fares well in terms of interpretability and explainability compared to the other DL models considered in this study. Besides, rather than using the interpretability evaluation for mere quantitative comparison, this study shed light on the use of the same for understanding the network through concepts it learned and features it is giving importance to. As illustrated in Sect. 4, the proposed IAENet is generic and can be extended to different applications.

5.2 Transparency and explainability

For EO data-based analyses, the spectral and spatial features and their characteristics, such as depth, width, and position, are found to be crucial. As is evident from Sect. 4, the conventional evaluation matrices do not give an idea about the learning mechanisms and capabilities of the models. The interpretability-based evaluation strategy (Sect. 3.2.2) is found to be useful in understanding the concepts learned by the models. The proposed approaches also demonstrate the modeling of interpretability analysis in accordance with the specific models.

The concepts learned for each of the classes provide an understanding of the learning capability of the DL networks. The prototype's analysis gives an idea of the characteristic features learned by the model for distinguishing different classes. For instance, the NDVI corresponding to non-rainy time slots are important in distinguishing the irrigated and non-irrigated crops. Hence, a good model is expected to give importance to the features corresponding to the non-rainy time slots and the model can be explained by analyzing the same. As discussed in Sect. 4.1.2, the features relevant for irrigation schemes can be verified based on the prototype learned by the network, and the training data can be refined accordingly. Similarly, the concepts learned for classifying crop types by IAENet (Sect. 4.2.2) can also be verified to refine the training data.

Although analysis of the prototypes learned by DL models gives an idea about the learning capability of the network, it is further required to analyze the relevance of the different input features and also to interpret the manifold accordingly. As is evident from Sects. 4.1.2 to 4.2.2, the LRP approaches facilitate identifying the contribution of the input features in terms of the relevance scores. It is observed that the irrigation scheme detection using IAENet gives importance to NDVI features during the non-rainy season of the year. This can be attributed to the fact that NDVI responses are almost indistinguishable in the rainy dates of the season and hence may not be relevant in distinguishing the schemes. The high relevance scores of red edge features in the classification of crop types (Sect. 4.2.2) indicate the ability of the interpretability analysis in verifying the significance of feature learning. Besides, the analysis of the features at a deeper level indicates the significant contribution from these relevant spectral bands.

6 Conclusion

DL-based EO data analyses require unsupervised learning of hierarchical representations of the input to find the intrinsic data manifold. This research proposed the use of

adversarial and variational encoding strategies and data prior embedding to develop a generic interpretable DL framework. The projection of latent representations to normally distributed space, as in variational encoders, affects the physical significance of the learned manifold. Although the use of entanglement penalty is found to improve the independence of the latent dimensions and enhance sparsity, they do not improve the interpretability of the network for the given task or data. In this regard, the proposed IAENet combines variational and adversarial encoding schemes along with constraints to fine-tune the manifold according to a given objective, such as classification. The proposed strategies enforce the network to learn meaningful features and facilitate the use of the information prior for improving physical significance. The improved interpretability and physical significance of the learned representations, along with the transparency of the proposed pipeline, significantly improve the results. The current study also proposed interpretability-based evaluation measures to compare different DL models. The predictor conditioned distribution of input is modeled to understand the most likely input of the model for a given output. The comparison of the learned concepts facilitates to understand the physical significance of the features. To further evaluate the interpretability of the DL models, the contribution of each input feature to the nonlinear hidden layer abstractions and their relevance to the network's decision are analyzed using a modified LRP technique. The proposed relevance propagation strategy explores the availability of generative and adversarial priors to understand the relevance better. It may be noted that the activation maximization- and LRP-based approaches provide explanations of IAENet based on the concept learned and the importance given to the input features.

The constraints and encoding schemes, along with the DTW-based similarity measures, achieve effective classification with a minimal number of training samples. Unlike the conventional convolution, a point-based convolution is proposed to process the irregularly sampled phenological curves. The concepts learned by the network and the relevance given to the characteristic features serve as a way to explain the model in conjunction with expert consultation. The modeling of IAENet for distinguishing irrigation schemes illustrated that the approach learns meaningful features (over the dry period) and the specific nature of the phenological curves. The learning of crop-specific features from multis-spectral images using IAENet also illustrated the effectiveness of the proposed strategies. The learned features are found to be a linear and nonlinear combination of red edge and green edge features, confirming the capability of the approach in learning relevant features. It is also observed that the improvement in interpretability achieved in IAENet significantly reduces the sensitivity of the network toward hyper-parameters. In addition, the requirement of training samples has also been reduced to a large extent.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F, Fraundorfer F (2017) Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci Remote Sens Mag 5:8–36. https://doi.org/10.1109/MGRS.2017.2762307
2. Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA (2019) Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J Photogramm Remote Sens 152:166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015
3. Li S, Song W, Fang L, Chen Y, Ghamisi P, Benediktsson JA (2019) Deep learning for hyperspectral image classification: an overview. IEEE Trans Geosci Remote Sens 57:6690–6709. https://doi.org/10.1109/TGRS.2019.2907932
4. Yuan Q, Shen H, Li T, Li Z, Li S, Jiang Y, Xu H, Tan W, Yang Q, Wang J, Gao J, Zhang L (2020) Deep learning in environmental remote sensing: achievements and challenges. Remote Sens Environ 241:111716. https://doi.org/10.1016/j.rse.2020.111716
5. Wang L, Shi C, Diao C, Ji W, Yin D (2016) A survey of methods incorporating spatial information in image classification and spectral unmixing. Int J Remote Sens 37:3870–3910. https://doi.org/10.1080/01431161.2016.1204032
6. Soneson C, Gerster S, Delorenzi M (2014) Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. PLoS ONE 9:e100335. https://doi.org/10.1371/journal.pone.0100335
7. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012
8. Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller KR (2020) Toward interpretable machine learning: transparent deep neural networks and beyond, ArXiv. http://arxiv.org/abs/2003.07631
9. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128:336–359. https://doi.org/10.1007/s11263-019-01228-7
10. Vikranth Jeyakumar J, Noor J, Cheng YH, Garcia L, Srivastava M (2020) How can I explain this to you? An empirical study of deep neural network explanation methods. https://github.com/nesl/Explainability-Study (accessed November 14, 2020)
11. Tjoa E, Guan C (2020) A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Networks Learn Syst. https://doi.org/10.1109/tnnls.2020.3027314
12. Fan F, Xiong J, Wang G (2020) On Interpretability of Artificial Neural Networks, ArXiv. http://arxiv.org/abs/2001.02522

13. Chai X, Gu H, Li F, Duan H, Hu X, Lin K (2020) Deep learning for irregularly and regularly missing data reconstruction. Sci Rep 10:1–18. https://doi.org/10.1038/s41598-020-59801-x
14. Trifonov V, Ganea OE, Potapenko A, Hofmann T (2018) Learning and evaluating sparse interpretable sentence embeddings, ArXiv. 200–210. https://doi.org/10.18653/v1/w18-5422
15. Subramanian A, Pruthi D, Jhamtani H, Berg-Kirkpatrick T, Hovy E (2018) SpinE: Sparse interpretable neural embeddings, In: 32nd AAAI Conf Artif Intell AAAI 2018, pp 4921–4928. http://arxiv.org/abs/1711.08792 (accessed November 14, 2020)
16. Liu D, Sun K, Wang Z, Liu R, Zha ZJ (2020) Frank-wolfe network: an interpretable deep structure for non-sparse coding. IEEE Trans Circuits Syst Video Technol 30:3068–3080. https://doi.org/10.1109/TCSVT.2019.2936135
17. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35:1798–1828. https://doi.org/10.1109/TPAMI.2013.50
18. Spinner T, Körner J, Görtler J, Deussen O (2018) Towards an interpretable latent space: an intuitive comparison of autoencoders with variational autoencoders, IEEE VIS 2018. https://kops.uni-konstanz.de/handle/123456789/43657 (accessed November 5, 2020)
19. Kovalev MS, Utkin LV, Kasimov EM (2020) SurvLIME: A method for explaining machine learning survival models, ArXiv. http://arxiv.org/abs/2003.08371 (accessed November 14, 2020)
20. Serrano S, Smith NA (2019) Is Attention Interpretable?, ACL 2019 - 57th Annu Meet Assoc Comput Linguist Proc Conf 2931–2951. http://arxiv.org/abs/1906.03731 (accessed November 14, 2020)
21. Shankaranarayana SM, Runje D (2019) ALIME: Autoencoder based approach for local interpretability, In: Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), Springer, pp 454–463. https://doi.org/10.1007/978-3-030-33607-3_49
22. Fan F, Li M, Teng Y, Wang G (2020) Soft autoencoder and its wavelet adaptation interpretation. IEEE Trans Comput Imaging 6:1245–1257. https://doi.org/10.1109/TCI.2020.3013796
23. Montavon G, Lapuschkin S, Binder A, Samek W, Müller KR (2017) Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognit 65:211–222. https://doi.org/10.1016/j.patcog.2016.11.008
24. Sun Y, Mao H, Sang Y, Yi Z (2017) Explicit guiding auto-encoders for learning meaningful representation. Neural Comput Appl 28:429–436. https://doi.org/10.1007/s00521-015-2082-x
25. Montavon G, Samek W, Müller KR (2018) Methods for interpreting and understanding deep neural networks. Digit Signal Process A Rev J 73:1–15. https://doi.org/10.1016/j.dsp.2017.10.011
26. Lapuschkin S, Binder A, Montavon G, Muller KR, Samek W (2016) Analyzing classifiers: fisher vectors and deep neural networks, In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, IEEE Computer Society pp 2912–2920. https://doi.org/10.1109/CVPR.2016.318
27. Cheriyadat AM (2014) Unsupervised feature learning for aerial scene classification. IEEE Trans Geosci Remote Sens 52:439–451. https://doi.org/10.1109/TGRS.2013.2241444
28. Girin L, Leglaive S, Bie X, Diard J, Hueber T, Alameda-Pineda X (2020) Dynamical variational autoencoders: a comprehensive review. http://arxiv.org/abs/2008.12595 (accessed October 25, 2020)
29. Anirudh R, Thiagarajan JJ, Kailkhura B, Bremer PT (2020) MimicGAN: robust projection onto image manifolds with corruption mimicking. Int J Comput Vis 128:2459–2477. https://doi.org/10.1007/s11263-020-01310-5
30. Emami H, Aliabadi MM, Dong M, Chinnam RB (2019) SPA-GAN: spatial attention GAN for image-to-image translation, IEEE Trans Multimed, 1–1. http://arxiv.org/abs/1908.06616 (accessed October 25, 2020)
31. Hang R, Zhou F, Liu Q, Ghamisi P (2020) Classification of hyperspectral images via multitask generative adversarial networks. IEEE Trans Geosci Remote Sens. https://doi.org/10.1109/tgrs.2020.3003341
32. Jiang T, Li Y, Xie W, Du Q (2020) Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. IEEE Trans Geosci Remote Sens 58:4666–4679. https://doi.org/10.1109/TGRS.2020.2965961
33. Gui J, Sun Z, Wen Y, Tao D, Ye J (2020) A review on generative adversarial networks: algorithms, theory, and applications, http://arxiv.org/abs/2001.06937 (accessed October 25, 2020)
34. Tschannen M, Bachem O, Lucic M (2018) Recent advances in autoencoder-based representation learning. http://arxiv.org/abs/1812.05069 (accessed October 26, 2020)
35. Hoshen Y (2018) Non-adversarial mapping with VAES, In: Adv Neural Inf Process Syst pp 7528–7537
36. Kang J, Fernandez-Beltran R, Duan P, Liu S, Plaza AJ (2020) Deep unsupervised embedding for remotely sensed images based on spatially augmented momentum contrast. IEEE Trans Geosci Remote Sens. https://doi.org/10.1109/tgrs.2020.3007029
37. Peng X, Zhu H, Feng J, Shen C, Zhang H, Zhou JT (2019) Deep Clustering With Sample-Assignment Invariance Prior. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/tnnls.2019.2958324
38. Honke G, Higgins I, Thigpen N, Miskovic V, Link K, Duan S, Gupta P, Klawohn J, Hajcak G (2020) Representation learning for improved interpretability and classification accuracy of clinical factors from EEG. http://arxiv.org/abs/2010.15274 (accessed November 25, 2020)
39. Han P, Li G, Skulstad R, Skjong S, Zhang H (2020) A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data. IEEE Trans Instrum Meas. https://doi.org/10.1109/tim.2020.3016413
40. Kang Z, Lu X, Liang J, Bai K, Xu Z (2020) Relation-guided representation learning, Neural Netw. 131: 93–102. http://arxiv.org/abs/2007.05742 (accessed October 26, 2020).
41. Charte D, Charte F, del Jesus MJ, Herrera F (2020) An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges. Neurocomputing 404:93–107. https://doi.org/10.1016/j.neucom.2020.04.057
42. Mou L, Zhu XX (2020) Learning to pay attention on spectral domain: a spectral attention module-based convolutional network for hyperspectral image classification. IEEE Trans Geosci Remote Sens 58:110–122. https://doi.org/10.1109/TGRS.2019.2933609
43. Karim F, Majumdar S, Darabi H, Harford S (2018) Multivariate LSTM-FCNs for time series classification. Neural Netw 116:237–245. https://doi.org/10.1016/j.neunet.2019.04.014
44. Rifai S, Vincent P, Muller X, Glorot X, Bengio Y (2011) Contractive auto-encoders: Explicit invariance during feature extraction, In: Proc 28th Int Conf Mach Learn ICML 2011 pp 833–840
45. Hamdi SM, Angryk R (2020) Interpretable feature learning of graphs using tensor decomposition, In: Institute of Electrical and Electronics Engineers (IEEE), pp 270–279. https://doi.org/10.1109/icdm.2019.00037
46. Rudolph M, Wandt B, Rosenhahn B (2019) Structuring autoencoders, In: Proc 2019 Int Conf Comput Vis Work ICCVW 2019, Institute of Electrical and Electronics Engineers Inc., pp 615–623. https://doi.org/10.1109/ICCVW.2019.00075

47. Zhong Y, Deng W (2019) Adversarial learning with margin-based triplet embedding regularization, In: Proc IEEE Int Conf Comput Vis Institute of Electrical and Electronics Engineers Inc., pp 6548–6557. https://doi.org/10.1109/ICCV.2019.00665
48. Zhuang F, Cheng X, Luo P, Pan SJ, He Q (2017) Supervised representation learning with double encoding-layer autoencoder for transfer learning. ACM Trans Intell Syst Technol 9:1–17. https://doi.org/10.1145/3108257
49. Burgess CP, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, Lerchner A (2018) Understanding disentangling in β-VAE. http://arxiv.org/abs/1804.03599 (accessed November 25, 2020)
50. Chen TQ, Li X, Grosse R, Duvenaud D (2018) Isolating sources of disentanglement in variational autoencoders, In: 6th Int Conf Learn Represent. ICLR 2018 Work Track Proc
51. Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P (2016) InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, In: Adv Neural Inf Process Syst, pp 2180–2188. https://doi.org/10.5555/3157096.3157340
52. Gaujac B, Feige I, Barber D (2020) Learning disentangled representations with the Wasserstein Autoencoder. http://arxiv.org/abs/2010.03459 (accessed November 25, 2020)
53. Higgins I, Chang L, Langston V, Hassabis D, Summerfield C, Tsao D, Botvinick M (2020) Unsupervised deep learning identifies semantic disentanglement in single inferotemporal neurons, ArXiv. http://arxiv.org/abs/2006.14304 (accessed November 25, 2020)
54. Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A (2017) B-VAE: Learning basic visual concepts with a constrained variational framework, In: 5th Int Conf Learn Represent. ICLR 2017 Conf Track Proc
55. Higgins I, Amos D, Pfau D, Racaniere S, Matthey L, Rezende D, Lerchner A (2018) Towards a definition of disentangled representations, ArXiv. http://arxiv.org/abs/1812.02230 (accessed November 25, 2020)
56. Kim H, Mnih A (2018) Disentangling by Factorising, In: 35th Int Conf Mach Learn ICML 2018, International Machine Learning Society (IMLS), pp 4153–4171. http://arxiv.org/abs/1802.05983 (accessed November 25, 2020)
57. Simonyan K, Vedaldi A, Zisserman A (2014) Deep inside convolutional networks: Visualising image classification models and saliency maps, In: 2nd Int Conf Learn Represent ICLR 2014 Work. Track Proc
58. Adler P, Falk C, Friedler SA, Nix T, Rybeck G, Scheidegger C, Smith B, Venkatasubramanian S (2018) Auditing black-box models for indirect influence. Knowl Inf Syst 54:95–122. https://doi.org/10.1007/s10115-017-1116-3
59. Shrikumar A, Greenside P, Kundaje A (2017) Learning important features through propagating activation differences, In: 34th Int Conf Mach Learn ICML 2017. 7: 4844–4866. http://arxiv.org/abs/1704.02685 (accessed November 20, 2020)
60. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W (2015) On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS ONE 10:e0130140. https://doi.org/10.1371/journal.pone.0130140
61. Datta A, Sen S, Zick Y (2016) Algorithmic transparency via quantitative input influence: theory and experiments with learning systems, In: Proc 2016 IEEE Symp Secur Privacy SP 2016, Institute of Electrical and Electronics Engineers Inc., pp 598–617. https://doi.org/10.1109/SP.2016.42
62. Henelius A, Puolamäki K, Boström H, Asker L, Papapetrou P (2014) A peek into the black box: exploring classifiers by randomization. Data Min Knowl Discov 28:1503–1529. https://doi.org/10.1007/s10618-014-0368-8
63. Lundberg SM, Allen PG, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. https://github.com/slundberg/shap (accessed November 20, 2020)
64. Chen C, Li O, Tao C, Barnett AJ, Su J, Rudin C (2019) This looks like that: deep learning for interpretable image recognition
65. Pfau D, Higgins I, Botev A, Racanière S (2020) Disentangling by Subspace Diffusion, ArXiv
66. Ribeiro MT, Singh S, Guestrin C (2016) "Why should i trust you?" Explaining the predictions of any classifier, In: Proc ACM SIGKDD Int Conf Knowl Discov Data Min, Association for Computing Machinery, New York, NY, USA, pp 1135–1144. https://doi.org/10.1145/2939672.2939778
67. Baehrens D, Harmeling S, Kawanabe M, Hansen Khansen K, Edward Rasmussen C (2010) How to explain individual classification decisions timon Schroeter * Klaus-Robert M¨ uller, https://doi.org/10.5555/1756006.1859912
68. Zhou Z, Sun M, Chen J (2019) A model-agnostic approach for explaining the predictions on clustered data, in: Proc. - IEEE Int Conf Data Mining ICDM, Institute of Electrical and Electronics Engineers Inc., pp 1528–1533. https://doi.org/10.1109/ICDM.2019.00202
69. Jiarpakdee J, Tantithamthavorn C, Dam HK, Grundy J (2020) An empirical study of model-agnostic techniques for defect prediction models. IEEE Trans Softw Eng. https://doi.org/10.1109/tse.2020.2982385
70. Grezmak J, Zhang J, Wang P, Loparo KA, Gao RX (2020) Interpretable convolutional neural network through layer-wise relevance propagation for machine fault diagnosis. IEEE Sens J 20:3172–3181. https://doi.org/10.1109/JSEN.2019.2958787
71. Wachter S, Mittelstadt B, Russell C (2017) Counterfactual explanations without opening the black box: automated decisions and the GDPR, SSRN Electron J. http://arxiv.org/abs/1711.00399 (accessed November 20, 2020)
72. Barredo-Arrieta A, Del Ser J (2020) Plausible counterfactuals: auditing deep learning classifiers with realistic adversarial examples, In: Proc Int Jt Conf Neural Networks. http://arxiv.org/abs/2003.11323 (accessed November 20, 2020)
73. Aravantinos V, Diehl F (2018) Traceability of Deep Neural Networks, ArXiv. http://arxiv.org/abs/1812.06744 (accessed November 20, 2020)
74. Al-Hmouz R, Pedrycz W, Balamash A, Morfeq A (2019) Logic-driven autoencoders. Knowl Based Syst 183:104874. https://doi.org/10.1016/j.knosys.2019.104874
75. Ghosh P, Sajjadi MSM, Vergari A, Black M, Schölkopf B (2019) From Variational to Deterministic Autoencoders. http://arxiv.org/abs/1903.12436 (accessed October 25, 2020)
76. Zhang X, Yao L, Yuan F (2019) Adversarial Variational Embedding for Robust Semi-supervised Learning, In: Proc ACM SIGKDD Int Conf Knowl Discov Data Min, 139–147. https://doi.org/10.1145/3292500.3330966
77. Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B (2015) Adversarial Autoencoders, http://arxiv.org/abs/1511.05644 (accessed November 23, 2020).
78. Arun PV, Buddhiraju KM, Porwal A, Chanussot J (2020) CNN-based super-resolution of hyperspectral images. IEEE Trans Geosci Remote Sens 58:6106–6121. https://doi.org/10.1109/tgrs.2020.2973370
79. Bochinski E, Senst T, Sikora T (2018) Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms, In: Proc Int Conf Image Process. ICIP, IEEE Computer Society, pp 3924–3928. https://doi.org/10.1109/ICIP.2017.8297018.
80. Herrmann I, Shapira U, Kinast S, Karnieli A, Bonfil DJ (2013) Ground-level hyperspectral imagery for detecting weeds in wheat fields. Precis Agric 14:637–659. https://doi.org/10.1007/s11119-013-9321-x

81. Zhang Z, Duan F, Sole-Casals J, Dinares-Ferran J, Cichocki A, Yang Z, Sun Z (2019) A novel deep learning approach with data augmentation to classify motor imagery signals. IEEE Access 7:15945–15954. https://doi.org/10.1109/ACCESS.2019.2895133
82. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P-A (2018) Deep learning for time series classification: a review. Data Min Knowl Discov 33:917–963. https://doi.org/10.1007/s10618-019-00619-1
83. Imani M, Ghassemian H (2020) An overview on spectral and spatial information fusion for hyperspectral image classification: current trends and challenges. Inf Fusion 59:59–83. https://doi.org/10.1016/j.inffus.2020.01.007
84. Cubuk ED, Zoph B, Shlens J, Le QV (2019) RandAugment: Practical automated data augmentation with a reduced search space, In: IEEE Comput Soc Conf Comput Vis Pattern Recognit Work. 2020-June 3008–3017. http://arxiv.org/abs/1909.13719 (accessed October 25, 2020)
85. Ghamisi P, Yokoya N, Li J, Liao W, Liu S, Plaza J, Rasti B, Plaza A (2017) Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art. IEEE Geosci Remote Sens Mag 5:37–78. https://doi.org/10.1109/MGRS.2017.2762087

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com