

Kolmogorov–Arnold Network for Hyperspectral Change Detection

Seyd Teymoor Seydi¹, Member, IEEE, Mojtaba Sadegh¹, and Jocelyn Chanussot², Fellow, IEEE

Abstract—Hyperspectral change detection (HCD) techniques to monitor Earth’s surface processes advanced markedly in recent years. Seasonal variations and associated spectral signatures as well as nonlinear noise patterns emanating from sensors and atmospheric sources pose fundamental challenges in HCD. Advanced deep learning models, such as those that leverage convolutional neural networks (3D-Siamese) or transformers (MLP-Mixer), are increasingly employed to address these challenges. However, they often need substantial training data and computational resources. Here, we show that the Kolmogorov–Arnold network (KAN) can enhance HCD capabilities without the excessive training demand of deep networks. The Kolmogorov–Arnold theorem provides the theoretical foundation for our approach, which is particularly well-suited for hyperspectral data analysis by providing a rigorous basis for handling high-dimensional spectral signatures through dimensional reduction and feature extraction. Our architectural design employs this theoretical framework by incorporating specialized neural network layers that mirror the theorem’s compositional structure, thereby facilitating efficient processing of spectral bands. By replacing the linear weighting scheme with learnable nonlinear functions, the Kolmogorov–Arnold network (KAN) provides a unique capability to capture intricate patterns and irregularities in high-dimensional data. Here, we compare five KAN-based architectures and deep learning models such as the MLP-Mixer, 3D-Siamese, dual-branch Siamese spatial-spectral Transformer attention network (DBS³TAN), and the Swin Transformer for HCD and show that the Chebyshev-KAN model, with an average overall accuracy of 97.35% over four real-world benchmark cases, outperforms other models while having a marked lower complexity than the deep learning models. We also show that the choice of fit nonlinear function and model structure is more important than the number of parameters in KAN-based models.

Index Terms—Hyperspectral change detection (HCD), KAN, Kolmogorov–Arnold network (KAN), remote sensing.

I. INTRODUCTION

HYPERSPECTRAL sensors offer hundreds to thousands of spectral bands, producing images that are rich both

Received 27 August 2024; revised 14 November 2024 and 2 January 2025; accepted 30 January 2025. Date of publication 6 February 2025; date of current version 24 February 2025. The work of Mojtaba Sadegh was supported by the U.S. Joint Fire Science Program through the U.S. Department of the Interior/Bureau of Land Management under Grant L21AC10247. (*Corresponding author: Mojtaba Sadegh.*)

Seyd Teymoor Seydi and Mojtaba Sadegh are with the Department of Civil Engineering, Boise State University, Boise, ID 83706 USA (e-mail: seydsydi@boisestate.edu; mojtabasadegh@boisestate.edu).

Jocelyn Chanussot is with CNRS, Grenoble INP, GIPSA-Lab, University Grenoble Alpes, 38000 Grenoble, France, and also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Digital Object Identifier 10.1109/TGRS.2025.3539535

in spatial and spectral information and providing a wide range of discriminative information for a variety of applications, including change detection [1], [2], target detection [3], and image classification [4]. One of the most prominent applications of hyperspectral imagery is change detection, which aims to identify changes in specific geographical areas between two different times [5].

Change detection techniques are generally categorized into two types: traditional methods and machine/deep learning-based techniques [6], [7]. The application of traditional HCD techniques is often impeded by the high dimensionality and spectral redundancy that are intrinsic to hyperspectral data [8]. While methods such as change vector analysis (CVA) and iteratively reweighted multivariate alteration detection are computationally efficient, they face challenges in fully exploiting the rich spectral information available [8]. These techniques typically operate on a band-by-band basis or employ linear transformations, which may not fully capture the complex, nonlinear relationships inherent in hyperspectral data [9]. Additionally, they often rely on simplistic pixel-wise comparisons or transformations, making it challenging to distinguish subtle changes from noise or variations in imaging conditions [10]. Many traditional approaches also require manual parameter tuning or thresholding, which can be time-consuming and subjective, limiting their applicability across diverse datasets [11], [12]. In contrast, deep learning techniques have demonstrated the potential to address many of these limitations [13], [14], [15], [16]. These models offer significant advantages for HCD due to their advanced capabilities. They automate the process of feature extraction, learning hierarchical and abstract features from raw data to capture intricate and complex patterns, thereby surpassing traditional methods [17]. Their nonlinear modeling capacity enables the identification of complex changes that are not discernible by linear methods [18], [19]. These models readily adapt to diverse datasets and manage high-dimensional data with dimensionality reduction layers and attention mechanisms and are thus capable of handling complex data. Their end-to-end learning approach facilitates the streamlining of the detection pipeline, while appropriate training strategies enhance robustness to noise and varying conditions, rendering them more effective and reliable in comparison to conventional techniques [20], [21].

Deep learning methods have recently received much traction in the change detection community. Examples are numerous and we review some here. Wu et al. [22] proposed a multitask

framework for hyperspectral change detection (HCD) and band reweighting. Their approach integrates a differential band reweighting network with a Siamese detection network with multitemporal interaction and multidomain fusion modules. The framework also incorporates an unbalanced contrastive learning method to address sample imbalance and intraclass variation. Yang et al. [23] proposed a multiscale spatial–spectral features extraction framework based on multiscale pyramid convolution that uses a spatial–spectral residual attention module at each scale. Feng et al. [24] proposed an attention feature interaction change detection method based on detail enhancement for bi-temporal hyperspectral images. Their approach integrates three key components: a detail enhancement module to improve spatial resolution and capture subtle changes, an attention interaction module for temporal feature interaction, and a multiscale feature extraction module to detect changes of different sizes. Yu et al. [14] proposed a gated spectral–spatial–temporal attention network with spectral similarity filtering for HCD. Their approach integrates a spectral similarity filtering module (to reduce spectral redundancy using cosine similarity), a gated spectral–spatial attention module (to capture intraimage spatial features using single-head weak self-attention and gating mechanisms), and a gated spectral–spatial–temporal attention module (to extract interimage temporal changes). Ding et al. [25] proposed a novel graph representation learning-guided diffusion model for HCD, integrating graph convolutional networks with diffusion models to better capture the global context in complex scenes. Seydi et al. [26] proposed a double-stream HCD framework based on the differencing of features and a hybrid 3-D/2-D squeeze-and-excitation (SE) attention mechanism for informative deep feature generation. Zhang et al. [27] proposed a dual-branch Siamese spatial–spectral Transformer attention network (DBS³TAN) for hyperspectral image change detection. Their approach integrates convolutional neural networks and Transformers in a dual-branch architecture to extract both spatial and spectral features. Song et al. [28] proposed a few-shot hyperspectral image change detection method called 3DCNN-NF, which combines a 3-D convolutional neural network with normalization flows. Li et al. [29] proposed CBANet, a lightweight end-to-end deep learning network for HCD, integrating: 1) a cross-band feature extraction module using 1×1 convolutions to reduce spectral dimensionality while preserving characteristics; 2) a spectral–spatial feature extraction module employing 2-D convolution and pooling; and 3) a 2-D self-attention based deep feature extraction module to enhance spatial–spectral representation. Gong et al. [30] proposed attention-based methods to extract representative features through the channel attention mechanism. Finally, Shafique et al. [31] proposed an attention-based technique that can detect more discriminative features and long-range dependencies compared to traditional machine learning models. Huang et al. [32] proposed a parallel spectral–spatial attention network with feature redistribution loss to construct class-oriented feature distribution. This framework uses a parallel spectral–spatial attention module to improve relevant information.

A key challenge in HCD is that not all changes are derived from transitions between distinct object classes (e.g., forest to built-up areas). Seasonal variations and environmental conditions can lead to subtle changes that result in confusing change and no-change classes in different images. These classes are often only nonlinearly separable, requiring sophisticated modeling approaches. While deep learning models can capture such nonlinearities, they typically require an elevated number of parameters, leading to high computational costs and resource requirements.

Kolmogorov–Arnold networks (KANs) offer a promising solution to address these challenges in HCD. KANs are a class of neural networks inspired by the Kolmogorov–Arnold representation theorem, which states that any continuous function of n variables can be represented as a composition of univariate functions and addition operations. By replacing fixed activation functions on nodes with learnable activation functions on edges, KANs can capture highly nonlinear functional relationships between input and target data while maintaining a relatively shallow network structure [33], [34]. This characteristic makes KANs particularly suitable for modeling the complex, nonlinear relationships present in HCD tasks without the need for deep architectures and their associated computational overhead.

This article presents a pioneering work on the implementation and evaluation of KAN-based models for HCD. While traditional methods such as CVA rely on linear spectral analysis and recent deep learning approaches employ complex attention mechanisms, our KAN-based approach uniquely exploits the Kolmogorov–Arnold superposition theorem to decompose high-dimensional spectral changes into optimized univariate functions. This mathematical foundation provides several key advantages: 1) guaranteed universal approximation of any continuous multivariate function through optimal univariate compositions; 2) interpretable feature learning that reveals the importance of spectral bands; and 3) an adaptive architecture that adapts to data characteristics rather than using fixed convolution patterns or predefined attention mechanisms. Using four benchmark hyperspectral datasets, our study assesses the efficacy of five KAN-based models in comparison with state-of-the-art deep learning-based models such as DBS³TAN [27], 3D-Siamese [35], [36], [37], Shifted Window Transformer (Swin Transformer) [38], and MLP-Mixer [41]. This study aims to demonstrate the pathways that KANs can enhance HCD by effectively capturing the nonlinear separability of change and no-change classes while maintaining computational efficiency.

Results demonstrate the effectiveness of the Chebyshev-KAN model, achieving the highest overall accuracy while maintaining relatively low complexity. This analysis also provides insights into the impact of architectural choices and parameterization on the model performance, facilitating informed decision-making for model selection and design in HCD tasks. The remainder of this article is organized as follows. Section II describes case studies and data. Section III presents the methodology, including the KAN-based models and the comparison models. Section IV discusses the results

TABLE I
STATISTICAL DETAILS OF THE REFERENCE DATA
SAMPLES FOR JIANGSU FARMLAND

Class	All Samples	Training	Validation	Test
Unchanged	40,417	1,665	357	38,395
Changed	18,383	745	173	17,465
Total	58,800	2,410	530	55,860
Unchanged (%)	68.74%	69.09%	67.36%	68.73%
Changed (%)	31.26%	30.91%	32.64%	31.27%

TABLE II
STATISTICAL DETAILS OF THE REFERENCE DATA
SAMPLES FOR JIANGSU INUNDATION

Class	All Samples	Training	Validation	Test
Unchanged	101,885	4,170	939	96,776
Changed	9,698	404	66	9,228
Total	111,583	4,574	1,005	106,004
Unchanged (%)	91.31%	91.17%	93.43%	91.29%
Changed (%)	8.69%	8.83%	6.57%	8.71%

TABLE III
STATISTICAL DETAILS OF THE REFERENCE DATA
SAMPLES FOR OREGON AGRICULTURAL AREA

Class	All Samples	Training	Validation	Test
Unchanged	59,688	2,452	544	56,692
Changed	14,299	581	122	13,596
Total	73,987	3,033	666	70,288
Unchanged (%)	80.67%	80.84%	81.68%	80.66%
Changed (%)	19.33%	19.16%	18.32%	19.34%

TABLE IV
STATISTICAL DETAILS OF THE REFERENCE DATA
SAMPLES FOR KANSAS AGRICULTURAL AREA

Class	All Samples	Training	Validation	Test
Unchanged	79,074	3,227	717	75,130
Changed	13,462	566	116	12,780
Total	92,536	3,793	833	87,910
Unchanged (%)	85.45%	85.08%	86.07%	85.46%
Changed (%)	14.55%	14.92%	13.93%	14.54%

and analysis. Finally, Section V concludes the article and suggests future research directions.

II. STUDY AREAS AND DATA

We use four benchmark datasets for HCD in different areas with complex land cover. Our case studies involve the Hyperion and Prisma sensors with a 30-m spatial resolution.

Tables I–IV provide the statistical details of the reference data used in different study areas. It includes the total number of samples, as well as their distribution across training, validation, and test sets for each class (*Changed* and *Unchanged*).

- 1) *Jiangsu Farmland Area* [8]: Images were acquired on May 3, 2006 [Fig. 1(a)] and April 23, 2007 [Fig. 1(b)] from a farmland area near the city of Yuncheng in Jiangsu province, China, using the Hyperion sensor. Image size is 420 pixels × 140 pixels × 150 bands.

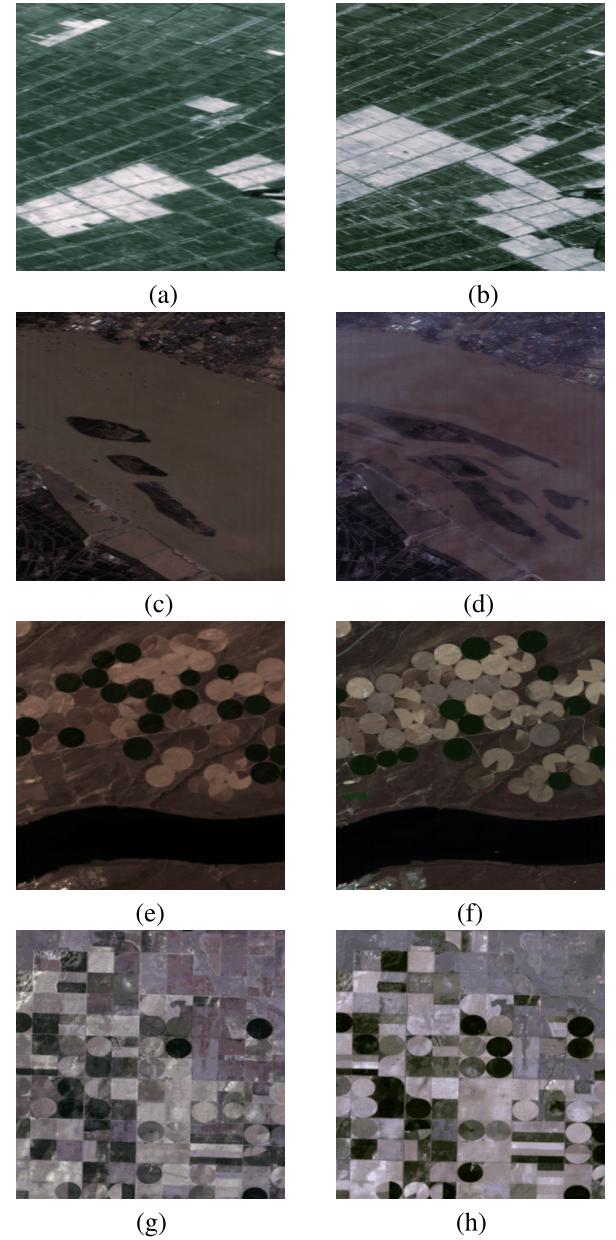


Fig. 1. True color representation of before (T1) and after (T2) images captured for four benchmark cases used in this study. Images have between 150 and 224 spectral bands, after removal of noisy bands from the hyperspectral data. (a) Jiangsu farmland T1. (b) Jiangsu farmland T2. (c) Jiangsu inundated T1. (d) Jiangsu inundated T2. (e) Oregon agricultural T1. (f) Oregon agricultural T2. (g) Kansas agricultural T1. (h) Kansas agricultural T2.

- 2) *Jiangsu Inundated Area* [39]: Images were captured on May 3, 2013 [Fig. 1(c)] and December 31, 2013 [Fig. 1(d)] in Jiangsu Province, China, using the Hyperion sensor. Image size is 436 pixels × 241 pixels × 198 bands.
- 3) *Oregon Agricultural Area* [8]: Images were captured on May 1, 2004 [Fig. 1(e)] and May 8, 2007 [Fig. 1(f)] from an irrigated agricultural field near Hermiston, Oregon, USA, using the Hyperion sensor. Image size is 307 pixels × 241 pixels × 154 bands.
- 4) *Kansas Agricultural Area* [31]: Images were acquired on March 18, 2021 [Fig. 1(g)] and May 9, 2021 [Fig. 1(h)] from a dry agricultural field near Manhattan, Kansas, USA, using the Hyperion sensor. Image size is 307 pixels × 241 pixels × 154 bands.

TABLE V
PSEUDOCODE FOR THE KAN

Input: Target function $f(x)$, input dimension n , number of hidden units $m = 2n + 1$, number of layers k , activation functions ϕ, ψ , learning rate η , epochs E .
Output: Trained KAN model approximating $f(x)$.
1. Initialization: Randomly initialize network weights and biases.
2. Forward Pass: 1. Compute first layer outputs using activation ϕ . 2. Compute subsequent layer outputs using activation ψ . 3. Sum outputs from all layers for final prediction.
3. Loss Function: Compute mean squared error between predictions and true values.
4. Backward Pass: 1. Compute gradients using backpropagation. 2. Update network parameters using gradient descent.
5. Training Loop: For each epoch: a. Forward pass b. Compute loss c. Update parameters d. Check convergence

[Fig. 1(h)] from agricultural regions near Kansas City, USA, using the PRISMA (PRecursore IperSpettrale della Missione Applicativa) sensor. Image size is 312 pixels \times 349 pixels \times 169 bands.

The source data for Jiangsu inundated and farmland areas, as well as the Oregon agricultural area, are available at <https://rslab.ut.ac.ir/data>, while the Kansas agricultural area dataset can be found at <https://citius.usc.es/investigacion/datasets/hyperspectral-change-detection-dataset>.

III. METHODOLOGY

KANs are based on the Kolmogorov–Arnold representation theorem [40], which states that any continuous multivariate function can be represented as a composition of continuous univariate functions. Using this theorem, a function $f(y_1, y_2, \dots, y_m)$ with m input variables can be expressed as

$$f(y) = f(y_1, y_2, \dots, y_m) = \sum_{r=1}^{2m+1} \Phi_r \left(\sum_{s=1}^m \varphi_{r,s}(y_s) \right) \quad (1)$$

where $\varphi_{r,s} : [0, 1] \rightarrow \mathbb{R}$ are inner univariate functions and $\Phi_r : \mathbb{R} \rightarrow \mathbb{R}$ are outer univariate functions. KANs are designed to approximate functions by parameterizing and learning these inner and outer univariate functions using a combination of spline coefficients and a set of simple functions. Unlike traditional neural networks with fixed activations, KANs can learn optimal activation functions during training, potentially improving model accuracy and interpretability. The spline parameterization allows visualizing and analyzing the learned activation functions $\phi_{l,j,i}$, providing insights into the functional relationships within the data. To evaluate the performance of KANs, we implemented the algorithm following the procedure outlined in Table V.

A. Spline-KAN

In the Spline-KAN architecture, each activation function is parameterized as a curve with learnable coefficients of local B-spline basis functions. A spline activation function $\phi(y)$ is defined as the sum of a basis function $b(y)$ (similar to a residual connection) and the spline function

$$\phi(y) = w(b(y) + \text{spline}(y)) \quad (2)$$

where w is a learnable weight factor to control the overall magnitude of the activation function. The basis function $b(y)$ is often chosen to be the Sigmoid Linear Unit (SiLU) function:

$$b(y) = \text{SiLU}(y) = \frac{y}{1 + e^{-y}}. \quad (3)$$

The spline function is parameterized as a linear combination of B-spline basis functions

$$\text{spline}(y) = \sum_i c_i B_i(y) \quad (4)$$

where c_i are the learnable coefficients and $B_i(y)$ are the B-spline basis functions. The herein used code of Spline-KAN is adapted from the Spline-KAN repository: <https://github.com/KindXiaoming/pykan.git>

B. FCN-KAN

The fully connected network KAN (FCN-KAN) is a variant of the KAN architecture that integrates sinusoidal embeddings with MLP layers in the ϕ function defined as

$$\phi(y) = \text{Linear}(\text{SiLU}(\text{Linear}(\text{concat}(y, \sin(\omega y), \cos(\omega y))))) \quad (5)$$

where ω is a learnable parameter representing the frequencies of the sinusoidal functions and y is the input variable. The ϕ function first concatenates the input y with its sinusoidal embeddings $\sin(\omega y)$ and $\cos(\omega y)$ and then applies two linear transformations with a SiLU activation function in between. The herein used code of FCN-KAN is adapted from the FCN-KAN repository: <https://github.com/Zhangyanbo/FCN-KAN.git>

C. Chebyshev-KAN

The Chebyshev-KAN combines the theoretical foundations of the Kolmogorov–Arnold theorem with the powerful approximation capabilities of Chebyshev polynomials [42]. The Chebyshev-KAN model approximates the target multivariate function $f(x)$ using a single layer of Chebyshev interpolation

$$\tilde{f}(y) = \sum_{j=1}^{d_{\text{in}}} \sum_{k=0}^n \Theta_{j,k} T_k(\tilde{y}_j) \quad (6)$$

where $\tilde{y} = \tanh(y)$ is the normalized input tensor, $T_k(\tilde{y}_j)$ is the k th Chebyshev polynomial evaluated at \tilde{y}_j , n is the degree of the Chebyshev polynomials, and $\Theta \in \mathbb{R}^{d_{\text{in}} \times d_{\text{out}} \times (n+1)}$ are the learnable coefficients for the Chebyshev interpolation. The herein used code of Chebyshev-KAN is adapted from the Chebyshev-KAN repository: <https://github.com/SynodicMonth/ChebyKAN.git>

D. RBF-KAN

Radial Basis Function KAN (RBF-KAN) is a faster implementation of Spline-KAN that approximates the high-order B-spline basis using Gaussian RBFs defined as [43]

$$\phi(r) = \exp\left(-\left(\frac{r}{h}\right)^2\right) \quad (7)$$

where r is the radial distance between the input x and the grid centerpoints c_i calculated as $r = \|y - c_i\|$ and h is the width or spread of the Gaussian function determined by the range of the grid ($\text{grid_max} - \text{grid_min}$) divided by the number of grid intervals ($\text{num_grids} - 1$), that is, $h = ((\text{grid_max} - \text{grid_min}) / (\text{num_grids} - 1))$. The RBF-KAN layer computes the RBF basis using the above equation and multiplying it by the learnable spline weights w_i

$$f(y) = \sum_{i=1}^N w_i \phi(\|y - c_i\|) \quad (8)$$

where N is the total number of grid points, w_i are the adjustable weights, and c_i are the grid centerpoints. The herein used code of RBF-KAN is adapted from the RBF-KAN repository: <https://github.com/Sid2690/RBF-KAN.git>

E. Naive-Fourier-KAN

The Naive-Fourier-KAN uses 1-D Fourier coefficients instead of spline coefficients. The main idea behind this layer is to approximate the univariate functions in the KAN using the Fourier series. The 1-D Fourier equation used in this layer is defined as

$$f(y) = \sum_{k=1}^N a_k \cos(ky) + b_k \sin(ky) \quad (9)$$

where a_k and b_k are the Fourier coefficients, y is the input variable, and N is the number of Fourier terms (grid size). The code of Naive-Fourier-KAN is adapted from the Naive-Fourier-KAN repository: <https://github.com/GistNoesis/FourierKAN.git>

In this study, we compare the discussed KAN-based models against the state-of-the-art deep learning-based models such as DBS³TAN [27], 3D-Siamese [35], [36], [37], Swin Transformer [38], and MLP-Mixer [41] for HCD. All KAN models and MLP have three layers, they all have the same number of neurons per layer (32, 16, and 2 neurons, respectively), and they have similar learning rates, batch size, and number of epochs. Hyperparameters of all models tested in this study are listed in Table VI. We used the binary cross entropy as a loss function and the Adam optimizer [44] for tuning model parameters.

In this study, we conducted a comprehensive evaluation of our model's performance through both quantitative and visual analyses. The quantitative evaluation included nine different metrics: Accuracy, Precision, Recall, *F1* Score, Cohen's Kappa (KC), balanced accuracy (BA), Matthews correlation coefficient (MCC), and Jaccard Index. In addition, we performed visual comparisons of the results to validate our findings and provide qualitative insights into the performance characteristics of the model.

TABLE VI
HYPERPARAMETERS OF THE MODELS

Model	Hyperparameters
Spline-KAN	Size and number of hidden layers Number of grid intervals Number of piecewise polynomials
FCN-KAN	Size and number of hidden layers Number of sinusoidal embedding frequencies
Chebyshev-KAN	Size and number of hidden layers Degree of the Chebyshev polynomials
RBF-KAN	Size and number of hidden layers Number of grid intervals
Naive-Fourier-KAN	Size and number of hidden layers Number of grid intervals
MLP-Mixer	Number of blocks Window size Stem width MLP dimension Dropout Tokens MLP dimension
Swin-Transformer	Number of heads Window size Shift size Embedding dimension MLP dimension Dropout
3D-Siamese	Size and number of convolution layers Size and number of hidden layers for classification head
DBS ³ TAN	Maximum depth Number of heads Dropout rate Embedding dropout rate Size and number of hidden layers Window size

IV. EXPERIMENT AND RESULTS

A. Results of HCD for Jiangsu Farmland Case Study

Fig. 2 shows the HCD results for the Jiangsu farmland benchmark data. KAN-based models (e)–(i) generally outperform 3D-Siamese, Swin Transformer, DBS³TAN, and MLP-Mixer (a)–(d) in the change detection task. The KAN-based models show fewer false positives and false negatives, better edge preservation, and a more accurate representation of complex patterns. The non-KAN models, especially DBS³TAN, show more false negative (red) and false positive (light blue) pixels. Among all models, Chebyshev-KAN (i) performs very well, with its result closely resembling the ground truth (j). It shows minimal false positives and false negatives, and good preservation of details. However, the differences among the KAN-based models are subtle, and several KAN models perform similarly well.

Table VII presents a confusion matrix comparing the performance of various models for HCD on the Jiangsu farmland case study. The KAN-based models consistently outperform non-KAN models in HCD, with Chebyshev-KAN improving accuracy by 8% for *Unchanged* labels and 2% for *Changed* labels compared to the Swin Transformer. Among the KAN models, Fourier-KAN, RBF-KAN, and FCN-KAN show similarly strong performances, while Chebyshev-KAN slightly edges out others with a 97% accuracy for *Changed* and 98% accuracy for *Unchanged* labels.

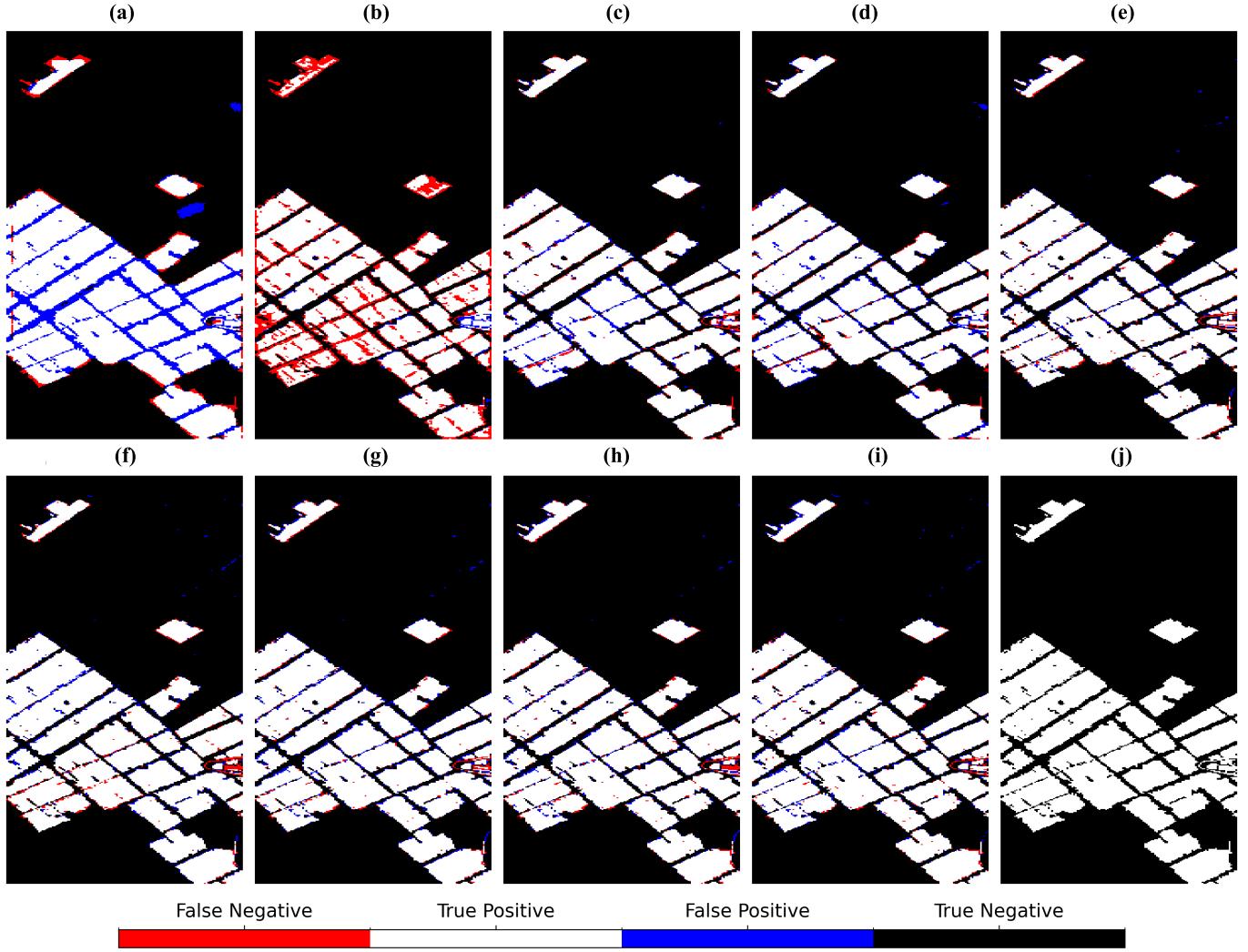


Fig. 2. Visual comparison of different models' predictions against the ground truth for the Jiangsu farmland data. (a) Swin Transformer. (b) DBS³TAN. (c) MLP-Mixer. (d) 3D-Siamese. (e) Spline-KAN. (f) Fourier-KAN. (g) RBF-KAN. (h) FCN-KAN. (i) Chebyshev-KAN. (j) Ground Truth.

TABLE VII
COMPARISON OF THE CONFUSION MATRICES OF DIFFERENT MODELS ACROSS MULTIPLE DATASETS

Dataset	Metric	Swin-Transformer	DBS ³ TAN	3D-Siamese	MLP-Mixer	Spline-KAN	Fourier-KAN	RBF-KAN	FCN-KAN	Chebyshev-KAN
Jiangsu farmland	TP	0.95	0.80	0.96	0.97	0.96	0.97	0.97	0.97	0.97
	FP	0.10	0.01	0.03	0.03	0.02	0.02	0.03	0.02	0.02
	TN	0.90	0.99	0.97	0.97	0.98	0.98	0.97	0.98	0.98
	FN	0.05	0.20	0.04	0.03	0.04	0.03	0.03	0.03	0.03
Kansas agricultural	TP	0.56	0.38	0.59	0.76	0.75	0.76	0.68	0.75	0.76
	FP	0.02	0.0	0.01	0.03	0.03	0.03	0.02	0.03	0.02
	TN	0.98	1.0	0.99	0.97	0.97	0.97	0.98	0.97	0.98
	FN	0.44	0.62	0.41	0.24	0.25	0.24	0.32	0.25	0.24
Jiangsu inundated	TP	0.55	0.58	0.72	0.67	0.76	0.71	0.69	0.79	0.84
	FP	0.03	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	TN	0.97	0.99	0.28	0.98	0.99	0.99	0.99	0.99	0.99
	FN	0.45	0.42	0.28	0.33	0.24	0.29	0.31	0.21	0.16
Oregon agricultural	TP	0.68	0.82	0.92	0.92	0.96	0.95	0.92	0.97	0.97
	FP	0.07	0.01	0.03	0.03	0.01	0.00	0.00	0.00	0.00
	TN	0.93	0.99	0.97	0.97	0.99	1.00	1.00	1.00	1.00
	FN	0.32	0.18	0.08	0.08	0.04	0.05	0.08	0.03	0.03

Table VIII presents a numerical comparison of the performance of various models for HCD on the Jiangsu farmland

case study. KAN-based models offer a marked improvement in various performance metrics compared to other models. The

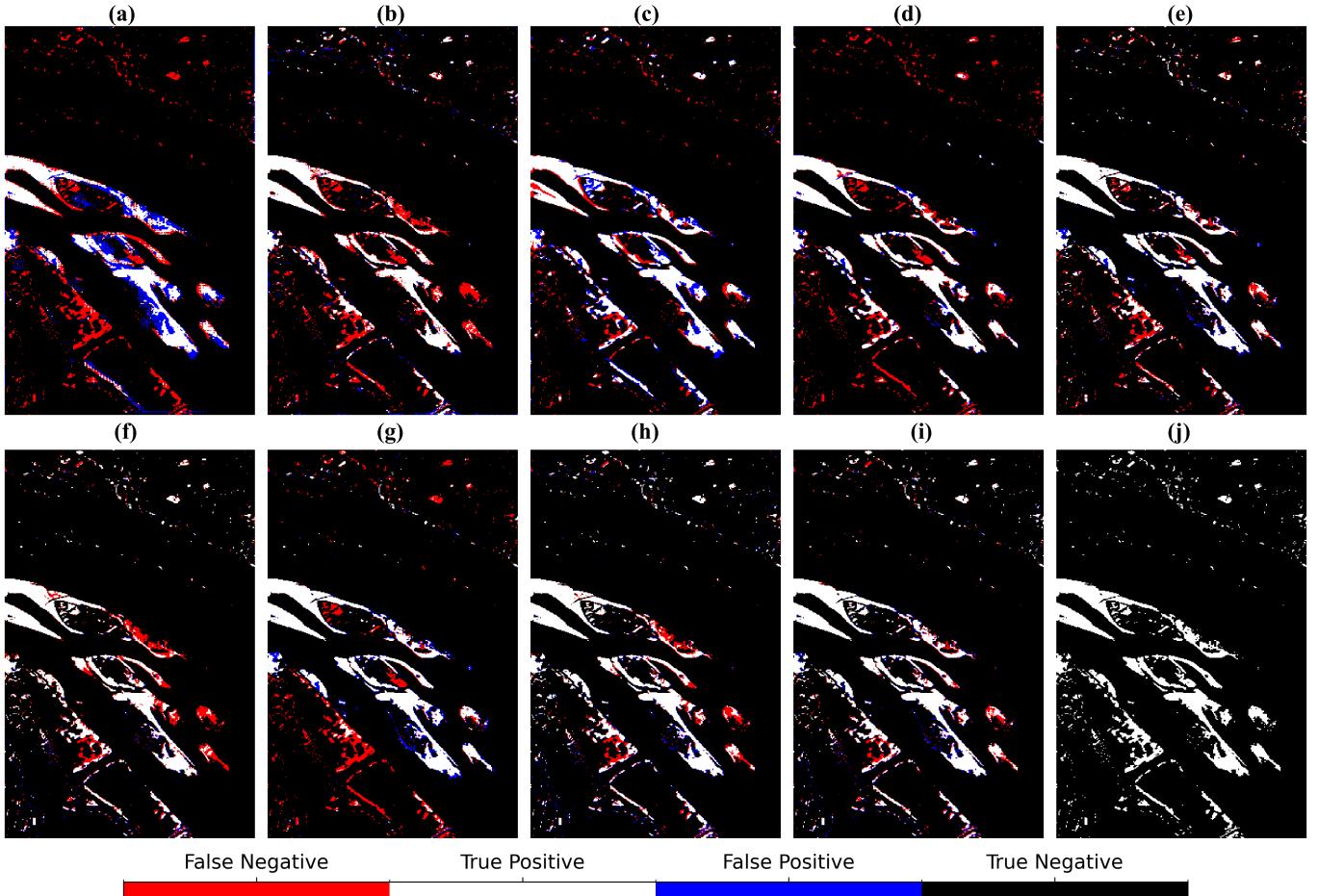


Fig. 3. Visual comparison of different models' predictions against the ground truth for the Jiangsu inundation case study. (a) Swin Transformer. (b) DBS³TAN. (c) 3D-Siamese. (d) MLP-Mixer. (e) Spline-KAN. (f) Fourier-KAN. (g) RBF-KAN. (h) FCN-KAN. (i) Chebyshev-KAN. (j) Ground Truth.

97% for *Unchanged* pixels but struggles with *Changed* pixels, correctly identifying only 55%. DBS³TAN performs slightly better, with 99% accuracy for *Unchanged* pixels and 58% for *Changed* pixels. The 3D-Siamese model improves upon this with 98% accuracy for *Unchanged* and 72% for *Changed* pixels, while the MLP-Mixer achieves 99% for *Unchanged* and 67% for *Changed* pixels. The KAN-based models demonstrate consistently superior performance, especially in detecting *Changed* pixels. Spline-KAN maintains a 99% accuracy for *Unchanged* pixels while dramatically improving *Changed* pixel detection to 76%. Fourier-KAN shows similar performance for the *Unchanged* class and achieves 71% accuracy for the *Changed* class. RBF-KAN slightly underperforms compared to other KAN models but still outperforms most deep learning models with 99% *Unchanged* and 69% *Changed* accuracy. FCN-KAN further improves the *Changed* pixel detection to 79% while maintaining 99% accuracy for *Unchanged* pixels. The standout performer is Chebyshev-KAN, which achieves the highest accuracy for both classes: 99% for *Unchanged* pixels and an impressive 84% for *Changed* pixels. This represents a significant improvement over all other models, particularly in the challenging task of identifying *Changed* pixels. Comparing the best KAN model (Chebyshev-KAN) to the best performing deep learning

model (3D-Siamese), we see a 1% improvement in *Unchanged* pixel detection and a substantial 12% improvement in *Changed* pixel detection.

Table VIII presents a numerical comparison of the performance of various models for HCD on the Jiangsu inundation case study. The Chebyshev-KAN model delivers the best performance among all models, with an accuracy of 97.74%, a precision of 97.69%, and an *F1* score of 97.71%. This represents an improvement of 4.56% in accuracy over the Swin Transformer, 2.14% over the DBS³TAN model, 2.11% over the 3D-Siamese model, and 1.35% over the MLP-Mixer. Other KAN-based models also outperform deep learning models across all metrics. In addition to accuracy, Chebyshev-KAN leads with a KC of 0.8538, a BA of 91.49%, a Jaccard Score of 95.72%, and a Matthew's correlation coefficient (MCC) of 0.8543.

C. Results of HCD for the Oregon Agricultural Case Study

Fig. 4 shows the HCD results for the Oregon agricultural area dataset, providing important insights into the performance of different models. Here too, the KAN-based models (e)-(i) consistently show superior performance compared to the 3D-Siamese, Swin Transformer, DBS³TAN, and MLP-Mixer models (a)-(d). This superiority is evident in their

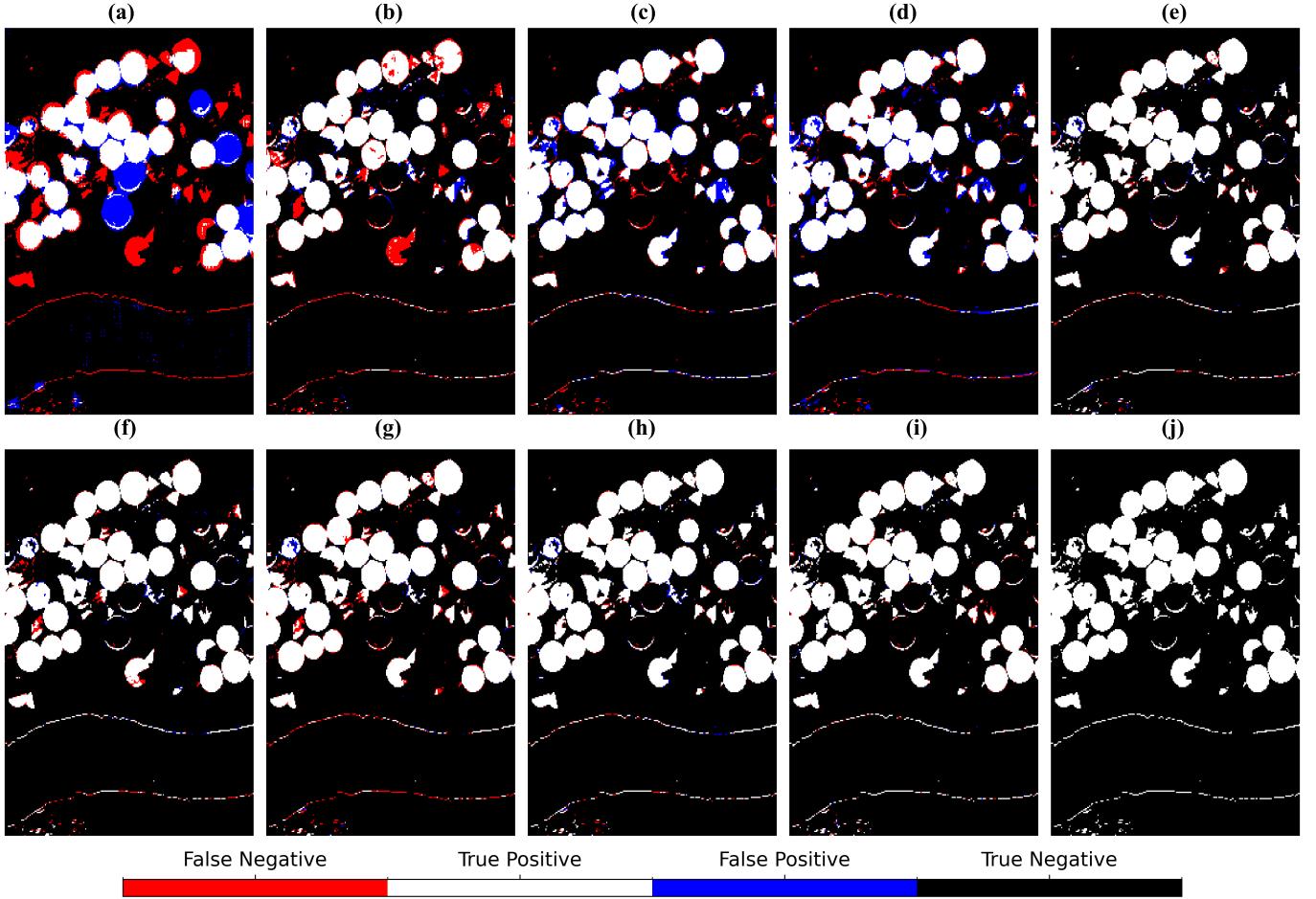


Fig. 4. Visual comparison of different models' predictions against the ground truth for the Oregon agricultural area case study. (a) Swin Transformer. (b) DBS³TAN. (c) 3D-Siamese. (d) MLP-Mixer. (e) Spline-KAN. (f) Fourier-KAN. (g) RBF-KAN. (h) FCN-KAN. (i) Chebyshev-KAN. (j) Ground Truth.

more accurate detection of modified areas with a noticeable reduction in false positives and false negatives. The non-KAN models, especially the Swin Transformer (a) and DBS³TAN (b), show a higher prevalence of false negatives (red pixels) and false positives (blue pixels), especially visible in the circular agricultural fields and along the edges of the changed areas. The 3D-Siamese model (c), while performing relatively well among the non-KAN models with fewer false positives and false negatives compared to the Swin Transformer and DBS³TAN, still lags behind the KAN-based models in overall accuracy. Among the KAN-based models, Chebyshev-KAN (i) stands out with exceptional performance. Its output is very close to the ground truth (j), demonstrating minimal false positives and negatives, excellent preservation of fine details and complex patterns, accurate detection of changes in the circular agricultural fields, and precise edge preservation, especially noticeable in the lower part of the image where other models struggle.

Table VII reveals the marked performance gap between KAN-based models and deep learning approaches in HCD. The deep learning models show varying levels of accuracy, with the Swin Transformer struggling the most, particularly in detecting changed pixels (68% accuracy). DBS³TAN shows improvement, achieving 82% accuracy for changed pixels and

99% for unchanged pixels. The 3D-Siamese and MLP-Mixer models perform identically and better than other deep learning models, with 92% accuracy for changed pixels and 97% for unchanged pixels. The KAN-based models show remarkable improvements. Spline-KAN shows significant improvement, correctly identifying 96% of changed pixels and 99% of unchanged pixels, a 4% increase in changed pixel detection over the best non-KAN models. The most impressive results come from Fourier-KAN, RBF-KAN, FCN-KAN, and Chebyshev-KAN, which achieve perfect accuracy (100%) in detecting unchanged pixels, completely eliminating false positives. These models show slight variations in detecting changed pixels, but all perform exceptionally well, with FCN-KAN and Chebyshev-KAN achieving 97% accuracy. Quantitatively, these top KAN-based models have improved the detection of changed pixels by 29% over the worst non-KAN model (Swin Transformer) and by 5% over the best non-KAN models. For unchanged pixels, they have achieved a perfect score, beating the worst non-KAN model by 7% and the best by 3%.

Table VIII presents a numerical comparison of the performance of various models for HCD on the Oregon agricultural case study. The provided metrics reveal a significant performance improvement of KAN-based models over deep learning

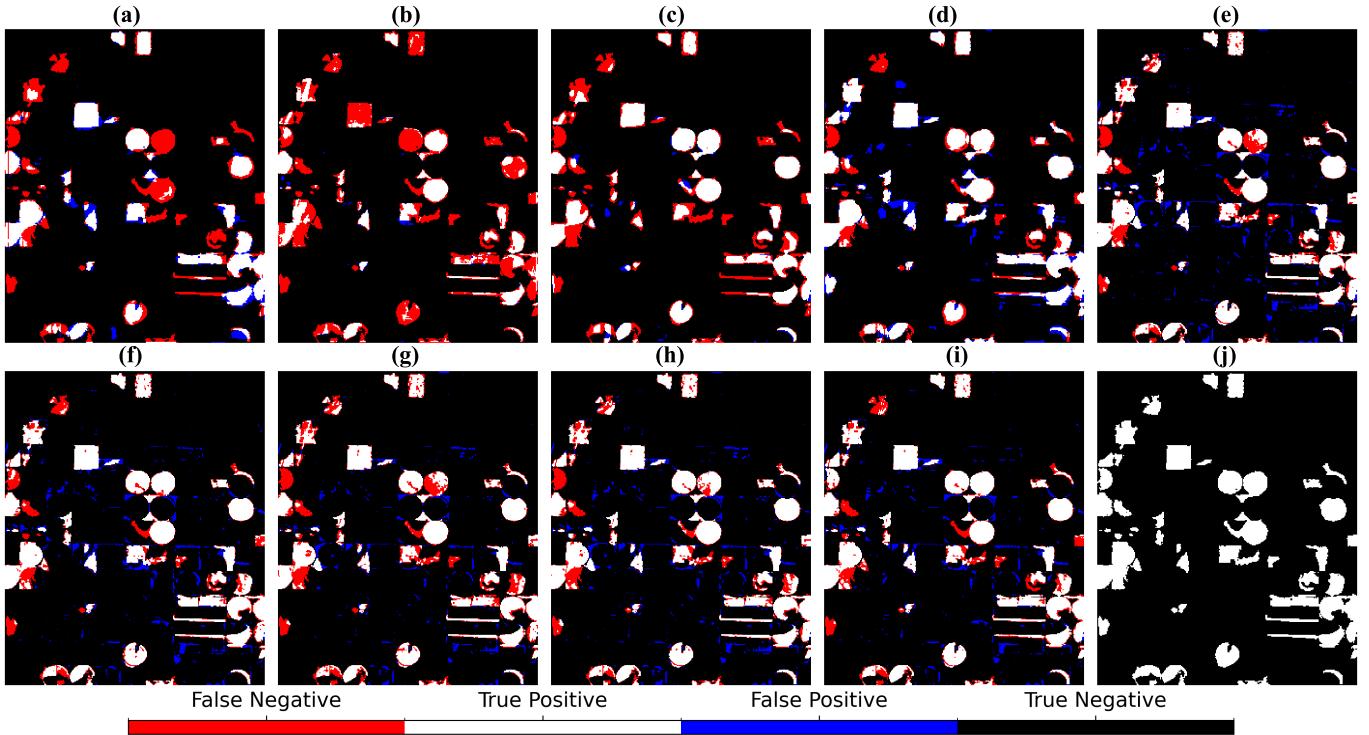


Fig. 5. Visual comparison of different models' predictions against the ground truth for the Kansas agricultural case study. (a) Swin Transformer. (b) DBS³TAN. (c) 3D-Siamese. (d) MLP-Mixer. (e) Spline-KAN. (f) Fourier-KAN. (g) RBF-KAN. (h) FCN-KAN. (i) Chebyshev-KAN. (j) Ground Truth.

approaches in HCD. Chebyshev-KAN once again emerges as the top performer, achieving an impressive minimum metric value of 99.36% across accuracy, precision, recall, and *F1* score. This represents a substantial improvement of 3.11% in accuracy and 3.09% in *F1* score compared to the best non-KAN model, 3D-Siamese. The KC score of 0.9793 for Chebyshev-KAN indicates near-perfect agreement between predictions and ground truth. Notably, all KAN-based models consistently outperform deep learning models across all metrics, with even the lowest-performing KAN model (RBF-KAN) surpassing the best non-KAN model in most categories. The consistency in high performance across different KAN implementations underscores the robustness of the KAN approach in HCD.

D. Results of HCD for the Kansas Agricultural Case Study

Fig. 5 displays HCD results for the Kansas agricultural case study, indicating that performance varies considerably between models. Among non-KAN models, 3D-Siamese performs relatively well, showing better delineation of structural edges compared to the Swin Transformer and DBS³TAN. The MLP-Mixer shows a large number of false positives. The Swin Transformer shows a mix of false positives and false negatives, while DBS³TAN struggles with false negatives, especially in circular structures and edges. KAN-based models generally show improvement over non-KAN counterparts. Spline-KAN and Fourier-KAN reduce misclassifications, especially in background regions. RBF-KAN and FCN-KAN better preserve intricate details of circular and rectangular patterns, although some edge misclassifications persist. Chebyshev-KAN emerges as the top performer, with the fewest

false positives and false negatives and the most accurate edge definition across all structure types. Its change map is closest to the ground truth, accurately capturing various agricultural patterns and subtle landscape changes.

Among the deep learning models, DBS³TAN has the highest accuracy for *Unchanged* pixels at 100%, but struggles with *Changed* pixel detection, correctly identifying only 38% (Table VII). The Swin Transformer shows a slight drop in accuracy for *Unchanged* pixels at 98% but improves detection of *Changed* pixels to 56%. The 3D-Siamese model achieves 99% accuracy for *Unchanged* pixels and 59% for *Changed* pixels, showing a more balanced performance. The MLP-Mixer has a lower *Unchanged* pixel accuracy of 97%, but significantly improves *Changed* pixel detection to 76%, demonstrating the best performance among deep learning models in this respect. The KAN-based models consistently show strong performance in both true positive and true negative categories. Both Spline-KAN and Fourier-KAN achieve 97% accuracy for *Unchanged* pixels and 75%–76% for *Changed* pixels, matching or slightly outperforming the MLP-Mixer. RBF-KAN shows a slight improvement in the detection of *Unchanged* pixels at 98% but falls behind in the detection of *Changed* pixels at 68%. FCN-KAN maintains a high (97%) accuracy for *Unchanged* pixels while achieving 75% accuracy for *Changed* pixels. The Chebyshev-KAN model stands out with 98% accuracy for *Unchanged* pixels and 76% for *Changed* pixels, demonstrating the best overall balance between the two categories among all models. Chebyshev-KAN not only matches but often exceeds the performance of deep learning models in HCD tasks.

TABLE IX
COMPARISON OF MODEL COMPLEXITY VERSUS ACCURACY

Model	Mean Processing Time (s)	Trainable Parameters	Mean Overall Accuracy
Swin-Transformer	62.25	577,426	0.9138
DBS ³ TAN	453.13	688,593	0.9377
3D-Siamese	1052.72	88,047,362	0.9557
MLP-Mixer	68.28	1,642,757	0.9584
Spline-KAN	54.25	138,816	0.9671
FCN-KAN	24.08	77,120	0.9706
Chebyshev-KAN	17.98	77,248	0.9735
RBF-KAN	17.97	77,186	0.9638
Naive-Fourier-KAN	25.02	64,066	0.9684

Table VIII presents a numerical comparison of the performance of various models for HCD on the Kansas agricultural case study, confirming the superior performance of KAN-based models. The Chebyshev-KAN model achieves the highest accuracy of 94.46%, precision of 94.30%, recall of 94.46%, and *F*1 score of 94.34%, consistently outperforming all deep learning models. In terms of KC, Chebyshev-KAN scores 0.7677, demonstrating substantial improvements ranging from 1.16% over the MLP-Mixer to 26.79% over DBS³TAN. The BA of Chebyshev-KAN at 86.83% surpasses all other models. Chebyshev-KAN's Matthew's correlation coefficient of 0.7691 and Jaccard Score of 89.83% further underscore its superior performance. Other KAN-based models, particularly Fourier-KAN and FCN-KAN, also demonstrate strong performance, closely following Chebyshev-KAN across most metrics. Notably, even the lowest-performing KAN model (RBF-KAN) outperforms the Swin Transformer and DBS³TAN in all metrics and remains competitive with 3D-Siamese.

E. Model Complexity

Table IX compares the complexity and accuracy of various models for HCD. Chebyshev-KAN achieves the highest average overall accuracy of 0.9735 over the four benchmarks, while maintaining relatively low complexity, indicating a good balance. Spline-KAN is the most complex model but does not outperform FCN-KAN and Chebyshev-KAN in terms of accuracy, suggesting that increased complexity does not necessarily improve performance. FCN-KAN also performs well, with overall accuracy and complexity comparable to Chebyshev-KAN. RBF-KAN and Naive-Fourier-KAN have rather similar complexities to Chebyshev-KAN, but lower accuracies. The choice of model architecture played a crucial role in balancing complexity and accuracy in this experiment.

F. Sensitivity Analysis

We conducted a sensitivity analysis of the Chebyshev-KAN model over the Oregon Agricultural case study to explore the choices of model structure. As presented in Table X, Chebyshev-KAN has an optimal two-layer configuration that achieves an accuracy of 0.9934 with 54 848 parameters. This architectural approach leverages the expressive power of Chebyshev polynomials in activation functions, enabling efficient approximation of complex functions. The nonmonotonic trend in accuracy (0.9902, 0.9934, 0.9892, and 0.9895 for

TABLE X
COMPARISON OF MODEL ACCURACY AND TRAINABLE PARAMETERS VERSUS NUMBER OF LAYERS FOR THE CHEBYSHEV-KAN MODEL

Number of Layers	Overall Accuracy	Trainable Parameters
1	0.9902	49,664
2	0.9934	54,848
3	0.9892	60,032
4	0.9895	65,216

1–4 layers, respectively) suggests a delicate balance between model capacity and generalization. Deeper networks demonstrate the potential for overfitting, as evidenced by a decline in performance for test data despite an increase in parameters. The linear parameter growth, approximately 5184 per layer, implies a consistent layer structure. The efficacy of Chebyshev-KAN can be attributed to its capacity to approximate universal functions through the use of Chebyshev polynomial expansions, which may reduce the necessity for deeper architectures. The behavior of the model can be further elucidated in terms of the bias-variance tradeoff, wherein the two-layer configuration attains an optimal point within the model complexity spectrum. Furthermore, the diminishing returns observed in deeper networks can be attributed to the vanishing gradient problem, which is exacerbated by the unique properties of Chebyshev polynomial derivatives in deeper layers.

Further analysis of the Chebyshev-KAN model points to a relationship between the polynomial degree and the model performance, as shown in Table XI. The fourth-degree polynomial achieves optimal accuracy (0.9934) with 54 848 parameters. However, the accuracy of the model exhibits a nonmonotonic improvement with increasing polynomial degree. The first-degree polynomial achieved an accuracy of 0.9861, the second-degree polynomial achieved an accuracy of 0.9881, the third-degree polynomial achieved an accuracy of 0.9860, the fourth-degree polynomial achieved an accuracy of 0.9934, and the fifth-degree polynomial achieved an accuracy of 0.9887. The number of parameters increases linearly by approximately 10 944 with each additional degree of the polynomial. The fourth-degree configuration exhibits a 0.73% improvement in accuracy relative to the first-degree, associated with a 149% expansion in the number of parameters. The model exhibits a bias-variance tradeoff inflection point at degree 4, wherein the enhanced expressiveness (decreased bias) outweighs the potential for overfitting (increased variance). The decline in accuracy from the fourth to the fifth degree of 0.47% indicates the advent of overfitting despite the 20% increase in parameters. This phenomenon is consistent with the equioscillation property of Chebyshev polynomials, which suggests that higher degrees may introduce excessive oscillations, potentially impairing generalization performance.

G. Model Interpretability Analysis

To assess our model's interpretability, we employed SHAP analysis to visualize the contribution of spectral bands to change detection decisions (Fig. 6). The analysis reveals that our model primarily focuses on two critical spectral

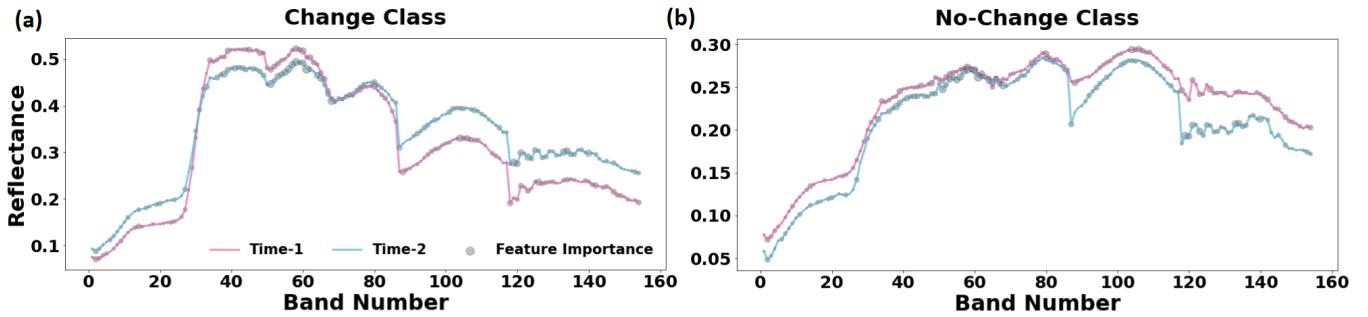


Fig. 6. Spectral signature analysis demonstrating the interpretability of the Chebyshev-KAN model using SHAP values. (a) Change class. (b) No-Change class.

TABLE XI

COMPARISON OF MODEL ACCURACY AND TRAINABLE PARAMETERS VERSUS POLYNOMIAL DEGREE FOR THE CHEBYSHEV-KAN MODEL

Degree of polynomial	Overall Accuracy	Trainable Parameters
1	0.9861	22,016
2	0.9881	32,960
3	0.9860	43,904
4	0.9934	54,848
5	0.9887	65,792

regions: the visible to near-infrared transition zone (bands 40–80) showing maximum reflectance differences ($\Delta\rho \approx 0.1$) with high SHAP values, and the near-infrared plateau region (bands 120–140) demonstrating significant temporal variations ($\Delta\rho \approx 0.05 - -0.08$). For the change class, these regions exhibit distinct spectral behavior between timestamps (reflectance range 0.4–0.5), while the no-change class shows minimal temporal variations ($\Delta\rho < 0.02$) and lower overall reflectance (0.05–0.30). The bimodal distribution of feature importance aligns with known spectral regions sensitive to vegetation and surface material changes, validating that our model has learned physically meaningful patterns. The learned importance patterns demonstrate that the model effectively captures both obvious spectral changes in primary vegetation-sensitive regions and subtle variations in secondary spectral regions, providing a clear interpretation of its decision-making process. These findings confirm that our model has learned interpretable and physically meaningful patterns in the hyperspectral data, focusing on spectral regions known to be sensitive to surface changes while appropriately down-weighting less informative bands.

V. DISCUSSION

Our comprehensive study of various machine/deep learning architectures for HCD reveals that KAN-based models have the potential to advance the field markedly. The Chebyshev-KAN model, in particular, exhibits superior performance, with an overall accuracy of 0.9735 across four diverse benchmark datasets (Jiangsu farmland, Jiangsu inundation, Oregon agricultural, and Kansas agricultural), consistently outperforming established deep learning models, including the MLP-Mixer, 3D-Siamese, DBS³TAN, and Swin Transformer. This remarkable degree of accuracy, coupled with relatively modest

computational demands, establishes KAN-based models as highly efficient solutions for HCD tasks.

The theoretical foundations of KAN architectures are anchored in the Universal Approximation Theorem and the Kolmogorov–Arnold representation theorem, offering a resilient framework for encapsulating intricate spectral-temporal interconnections in hyperspectral data. This is demonstrated by the capacity of KAN models to attain high accuracy in the detection of both *Changed* and *Unchanged* pixels, effectively addressing the prevalent challenge of class imbalance in change detection tasks. The efficiency of KAN-based models in approximating complex functions using simple building blocks translates to superior performance with fewer trainable parameters compared to more complex models, such as 3D-Siamese.

The superior performance of Chebyshev-KAN over other architectures can be attributed to its fundamental mathematical properties in spectral approximation and the equioscillation property of Chebyshev polynomials, which optimally minimize the maximum error across the entire spectral range. While Fourier-KAN employs sinusoidal basis functions that excel in periodic pattern detection, Chebyshev polynomials' ability to handle Runge phenomena and edge effects makes them particularly suitable for abrupt spectral changes characteristic of land-cover transitions. The improved BA and MCC of Chebyshev-KAN compared to RBF-KAN and Spline-KAN demonstrates its robustness in handling the bias-variance tradeoff, particularly evident in the model's fourth-degree polynomial configuration which achieves optimal expressiveness without overfitting. This theoretical advantage is particularly pronounced in agricultural monitoring scenarios, where the Chebyshev polynomial's ability to approximate complex spectral variations enables better discrimination between genuine land-cover changes and natural phenological variations. In addition, the advantages of Chebyshev-KAN go beyond the theoretical underpinnings through its unique architectural properties. The recursive formulation of Chebyshev polynomials facilitates stable gradient flow during backpropagation, addressing the critical challenge of gradient instability in deep networks when processing high-dimensional hyperspectral data. This mathematical property allows for efficient parameter utilization while maintaining robust approximation capabilities across the spectral domain. The model's hierarchical structure of orthogonal polynomials provides an optimal

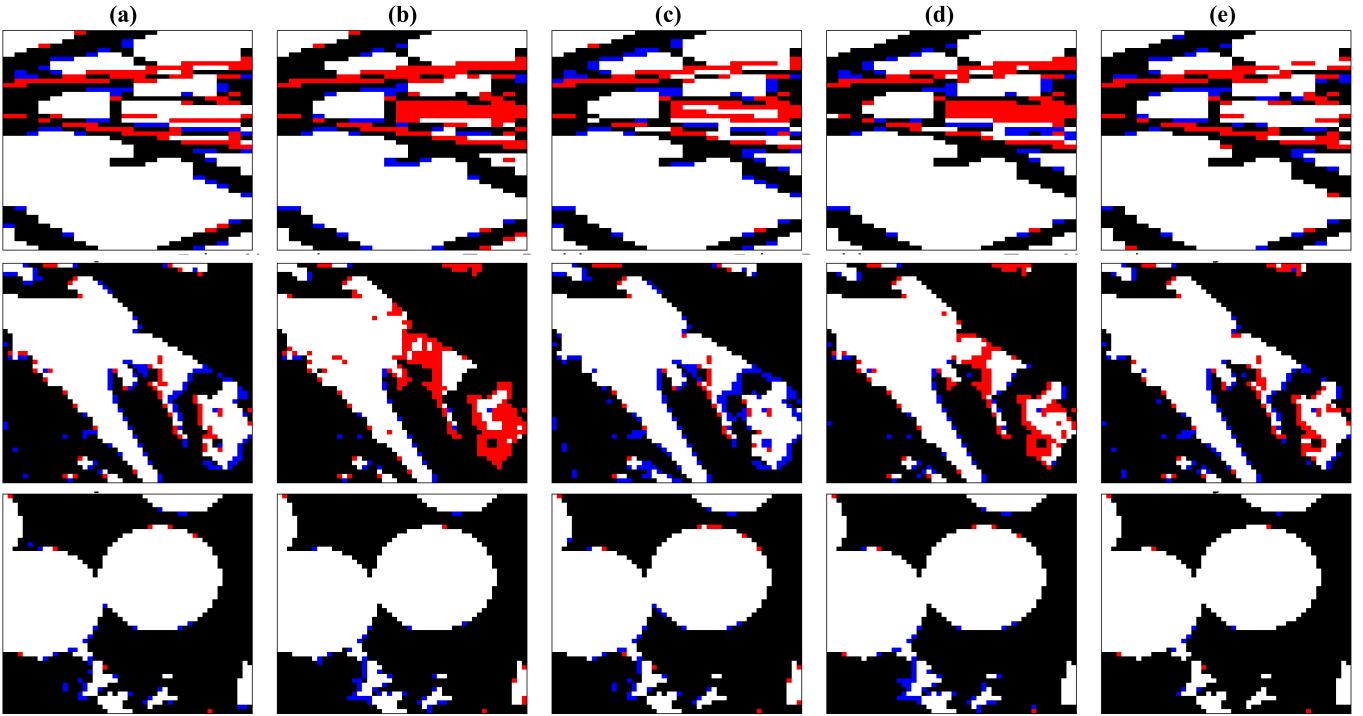


Fig. 7. Comparative visualization of the performance of the KAN models in zoomed areas. (a) Spline-KAN. (b) Fourier-KAN. (c) RBF-KAN. (d) FCN-KAN. (e) Chebyshev-KAN.

basis for representing both global spectral patterns and localized variations, which is particularly important in mixed-pixel scenarios where spectral signatures combine multiple land-cover transitions. The uniform approximation properties of the Chebyshev polynomials provide consistent performance across varying degrees of spectral complexity, while their inherent interpretability through polynomial coefficients provides valuable insights into the model's decision-making process, essential for operational remote sensing applications.

A comprehensive sensitivity analysis of the Chebyshev-KAN model indicates that optimal performance is achieved with a two-layer architectural configuration and a fourth-degree polynomial configuration. This finding elucidates the delicate balance in the bias-variance tradeoff, with the nonmonotonic trend in accuracy as a function of layers or polynomial degree indicating the existence of a threshold beyond which additional complexity yields diminishing returns. These insights are of paramount importance for future model design and hyperparameter tuning in KAN-based approaches to HCD, underscoring the necessity of prudent architectural selection over indiscriminate complexity increases.

A comparative analysis of model complexity versus accuracy offers insights into the efficiency of KAN-based models. Despite having a smaller number of trainable parameters, KAN models achieve superior accuracy in comparison to more complex architectures. This efficiency is particularly evident in the performance of Chebyshev-KAN, which maintains high accuracy while requiring substantially less processing time than its counterparts. These attributes render KAN-based models especially well-suited to real-world applications where

computational resources may be constrained, without any loss of performance.

A comprehensive analysis of model complexity versus accuracy reveals the remarkable efficiency-performance balance achieved by KAN-based architectures. The Chebyshev-KAN achieves superior accuracy with only parameters and minimal processing time. This efficiency stems from the theoretical advantages of Chebyshev polynomials in spectral approximation, where the orthogonality properties enable more effective feature extraction with fewer parameters. The complexity analysis shows diminishing returns in deeper architectures, with the optimal two-layer configuration outperforming deeper variants, suggesting that the polynomial basis functions effectively capture spectral-temporal relationships without requiring additional complexity. Notably, while 3D-Siamese employs a high number of parameters with high processing time, its accuracy falls short of Chebyshev-KAN, highlighting that architectural efficiency rather than model size drives performance. The FCN-KAN and Naive-Fourier-KAN demonstrate similar efficiency patterns, though their slightly lower accuracies suggest that Chebyshev polynomial basis functions provide optimal spectral approximation capabilities. Finally, the increased complexity of Spline-KAN due to its architecture poses significant challenges for HCD compared to other KAN-based models. The interaction between spline knots and coefficients creates parameter redundancy in the spectral-domain representation, leading to inefficient feature learning despite the larger parameter space. In contrast, the orthogonal polynomial basis of Chebyshev-KAN provides a more structured and efficient parameter space through its mathematically based basis functions.

The consistent performance improvements observed across various evaluation metrics, including KC, BA, and Jaccard Score, serve to underscore the robustness of KAN-based models. The capacity to maintain high performance across diverse landscapes indicates the potential for generalization to a wide range of HCD scenarios. This versatility, coupled with computational efficiency, establishes KAN-based models as promising candidates for large-scale, operational HCD applications in fields such as precision agriculture, environmental monitoring, and disaster response.

Our findings indicate that an increase in model complexity does not necessarily result in enhanced performance in HCD tasks. This is illustrated by the Spline-KAN model not outperforming FCN-KAN and Chebyshev-KAN, despite being the most complex among the shallow networks evaluated. This observation serves to reinforce the principle that the design of an efficient architecture and the appropriate parameterization are of greater consequence than mere increases in model complexity to capture the intricate patterns present in HCD tasks.

Our research demonstrates that KAN-based models, particularly Chebyshev-KAN, offer a powerful and efficient framework for addressing the challenges of HCD. Furthermore, it is more efficient in the detection of subtle changes (Fig. 7). The superior performance across diverse datasets, coupled with lower computational requirements, establishes these models as a leading technology in the field of HCD. As the field of remote sensing continues to evolve, with increasing demands for accurate and efficient change detection in various environmental and agricultural applications, KAN-based approaches present a promising avenue for future research and practical implementation. Their ability to balance high accuracy with computational efficiency makes them particularly well-suited for the development of next-generation HCD systems capable of processing vast amounts of hyperspectral data in real-time scenarios.

Beyond its utility in HCD, the KAN architecture exhibits great potential for various other remote sensing applications, including the fusion of multimodal data from different sensors. The remarkable efficiency of Chebyshev-KAN in handling high-dimensional spectral information highlights its suitability for real-time implementations in domains such as precision agriculture, environmental monitoring, and urban change tracking, where computational resources may be constrained but maintaining high spectral fidelity is of utmost importance.

VI. CONCLUSION

This study compares the performance of various KAN-based architectures against leading deep learning-based models for HCD using four benchmark datasets. The results indicate that the Chebyshev-KAN model outperformed others, achieving the highest overall accuracy while maintaining a relatively low complexity. The choice of model architecture plays a crucial role in balancing complexity and accuracy for change detection, and an increased model complexity does not guarantee improved performance.

Our findings suggest that KAN-based models, particularly the Chebyshev-KAN architecture, have the potential

to effectively capture the intricate patterns and irregularities present in HCD problems. By exploiting the theoretical foundations of the Kolmogorov–Arnold theorem, these models can approximate complex nonlinear functions and model high-dimensional relationships in hyperspectral data.

Future research could explore the integration of KAN-based architectures with deep learning techniques, such as attention mechanisms or Transformer models, to further enhance the performance of HCD systems, among other geophysical problems. Investigating potential improvements in model optimization or exploring hybrid architectures combining KAN with attention mechanisms might offer new insights and possibilities for further work.

REFERENCES

- [1] B. Yang, W. Sun, and J. Peng, "SAGN: Sharpening-aware graph network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5518812.
- [2] P. Jian, Y. Ou, and K. Chen, "Uncertainty-aware graph self-supervised learning for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5509019.
- [3] C.-I. Chang, "Constrained energy minimization (CEM) for hyperspectral target detection: Theory and generalizations," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5522921.
- [4] F. Ullah, I. Ullah, R. U. Khan, S. Khan, K. Khan, and G. Pau, "Conventional to deep ensemble methods for hyperspectral image classification: A comprehensive survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3878–3916, 2024.
- [5] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [6] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2006.
- [7] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [8] M. Hasanlou and S. T. Seydi, "Hyperspectral change detection: An experimental comparative study," *Int. J. Remote Sens.*, vol. 39, no. 20, pp. 7029–7083, Oct. 2018.
- [9] W. Dong, Y. Yang, J. Qu, S. Xiao, and Y. Li, "Local information-enhanced graph-transformer for hyperspectral image change detection with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509814.
- [10] Z. Lv, Z. Lei, L. Xie, N. Falco, C. Shi, and Z. You, "Novel distribution distance based on inconsistent adaptive region for change detection using hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4404912.
- [11] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.
- [12] K. Meshkini, Y. T. Solano-Correia, F. Bovolo, and L. Bruzzone, "Multianual change detection in long and dense satellite image time series based on dynamic time warping," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 441152.
- [13] Y. Huang, L. Zhang, W. Qi, R. Song, C. Huang, and Y. Cen, "Semi-supervised hypermatch-driven cross temporal and spatial interaction transformer for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 6426–6443, 2024.
- [14] H. Yu, H. Yang, L. Gao, J. Hu, A. Plaza, and B. Zhang, "Hyperspectral image change detection based on gated spectral-spatial-temporal attention network with spectral similarity filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511313.
- [15] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and X. Gao, "DCENet: Diff-feature contrast enhancement network for semi-supervised hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511514.
- [16] L. Wang, J. Zhang, and L. Bruzzone, "MixCDNet: A lightweight change detection network mixing features across CNN and transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4411915.

- [17] B. Yang, X. Cheng, W. Chen, and X. Ye, "A graph-based hyperspectral change detection framework using difference augmentation and progressive reconstruction with limited labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5518914.
- [18] S. Dara and P. Tumma, "Feature extraction by using deep learning: A survey," in *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2018, pp. 1795–1801.
- [19] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [20] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets," *Vis. Comput.*, vol. 38, no. 8, pp. 2939–2970, 2022.
- [21] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," *Image Vis. Comput.*, vol. 106, Feb. 2021, Art. no. 104078.
- [22] X. Wu, P. Gamba, J. Feng, R. Shang, X. Zhang, and L. Jiao, "A multitask framework for hyperspectral change detection and band reweighting with unbalanced contrastive learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530213.
- [23] Y. Yang, J. Qu, S. Xiao, W. Dong, Y. Li, and Q. Du, "A deep multiscale pyramid network enhanced with spatial-spectral residual attention for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5525513.
- [24] S. Feng, J. Zhang, F. Huang, R. Peng, and C. Zhao, "An attention feature interaction change detection method based on detail enhancement for dual-temporal hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [25] X. Ding, J. Qu, W. Dong, T. Zhang, N. Li, and Y. Yang, "Graph representation learning-guided diffusion model for hyperspectral change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [26] S. T. Seydi, M. Boueshagh, F. Namjoo, S. M. Minouei, Z. Nikraftar, and M. Amani, "A hyperspectral change detection (HCD-Net) framework based on double stream convolutional neural networks and an attention module," *Remote Sens.*, vol. 16, no. 5, p. 827, Feb. 2024.
- [27] Y. Zhang, T. Wang, C. Zhang, S. Xu, H. Gao, and C. Li, "A dual-branch Siamese spatial-spectral transformer attention network for hyperspectral image change detection," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122125.
- [28] B. Song et al., "3DCNN-NF: Few-shot hyperspectral image change detection based on 3-D convolution neural network and normalizing flow," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530615.
- [29] Y. Li et al., "CBA-Net: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513011.
- [30] M. Gong et al., "A spectral and spatial attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521614.
- [31] A. Shafique, S. T. Seydi, T. Alipour-Fard, G. Cao, and D. Yang, "SSViT-HCD: A spatial-spectral convolutional vision transformer for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6487–6504, 2023.
- [32] Y. Huang, L. Zhang, C. Huang, W. Qi, and R. Song, "Parallel spectral-spatial attention network with feature redistribution loss for hyperspectral change detection," *Remote Sens.*, vol. 15, no. 1, p. 246, Dec. 2022.
- [33] A. N. Kolmogorov, "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition," *Doklady Akademii Nauk SSSR*, vol. 114, no. 5, pp. 953–956, 1961.
- [34] A. Afzal Aghaei, "RKAN: Rational Kolmogorov-arnold networks," 2024, *arXiv:2406.14495*.
- [35] L. Wang, L. Wang, Q. Wang, and P. M. Atkinson, "SSA-SiamNet: Spectral-spatial-wise attention-based Siamese network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2021.
- [36] A. Song, J. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3D fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, p. 1827, Nov. 2018.
- [37] S. T. Seydi, R. Shah-Hosseini, and M. Amani, "A multi-dimensional deep Siamese network for land cover change detection in bi-temporal hyperspectral imagery," *Sustainability*, vol. 14, no. 19, p. 12597, Oct. 2022.
- [38] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
- [39] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2018.
- [40] Z. Liu et al., "KAN: Kolmogorov-arnold networks," 2024, *arXiv:2404.19756*.
- [41] I. O. Tolstikhin et al., "MLP-mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24261–24272.
- [42] S. S. Sidharth, A. R. Keerthana, R. Gokul, and K. P. Anas, "Chebyshev polynomial-based Kolmogorov-Arnold networks: An efficient architecture for nonlinear function approximation," 2024, *arXiv:2405.07200*.
- [43] Z. Li, "Kolmogorov-Arnold networks are radial basis function networks," 2024, *arXiv:2405.06721*.
- [44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.