Kolmogorov–Arnold Networks for Multimodal Feature Fusion and mmWave Beam Prediction

Saba M. Hosseini, Amirhossein Safari, and Nazanin Rahnavard
Department of Electrical and Computer Engineering
University of Central Florida, USA
Email: {saba.mhosseini, amirhossein.safari, nazanin.rahnavard}@ucf.edu

MILCOM 2025 - 2025 IEEE Military Communications Conference (MILCOM) | 979-8-3315-0292-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/MILCOM64451.2025.11310028

***Abstract*—Optimal beam alignment in dynamic mmWave environments remains a key challenge. Integrated sensing and communication attempts to address this challenge leveraging heterogeneous sensor data. We propose a deep learning framework for the mmWave beam prediction (BP) based on Kolmogorov–Arnold Networks (KANs), a recently proposed alternative to Multi-Layer Perceptrons (MLPs) that replaces fixed node-wise activations with learnable spline-based functions on network edges. Our architecture fuses GPS, LiDAR, and vision modalities in a real-world vehicle-to-infrastructure scenario through KAN-based late fusion. We use convolutional neural networks and Gated Recurrent Units for spatial and temporal feature extraction before passing the features to KAN-based heads. We evaluate the models in regression and classification settings for BP. Extensive experiments demonstrate that KAN-based models outperform MLP-based models cross single- and multi-modal setups, yielding significant improvements in beam index prediction accuracy. These results highlight the effectiveness of KANs for robust BP in dynamic wireless environments.**

***Index Terms*—multi-modal deep learning, Kolmogorov-Arnold Networks, late fusion, mmWave beam predication**

I. Introduction

Millimeter-wave (mmWave) communication uses densely packed antenna arrays to enable high-capacity, directional MIMO transmission with precise beamforming and spatial multiplexing [1], [2]. Accurate beam prediction (BP) is crucial for directing energy to intended users, enhancing signal quality, reducing link setup time, and maximizing throughput [2]. Yet challenges such as path loss, penetration loss, and rain fading complicate high-frequency links, reinforcing the need for reliable BP [1].

Exhaustive beam alignment over large codebooks is a non-convex optimization that incurs high latency and computation [3], making it unsuitable for mobile or time-sensitive scenarios. To address this, recent work has adopted *learning-based* BP using classification frameworks [4]–[7]. Although Deep Learning (DL)-based BP is typically formulated as classification over a discrete codebook [8], the problem inherently lies in continuous angular space, where exploiting parameters such as direction and received power enables finer beam representations. Nevertheless, regression-based BP remains underexplored due to the reliance on discretized angular quantization in traditional systems [9].

Sensing-aided BP using multimodal data has emerged as a promising solution to challenges from environmental dynamics by leveraging heterogeneous sensor inputs to enhance situational awareness in BP. Early sensing-aided BP relied on GPS for user localization [10], while later work incorporated visual [4], [5] and LiDAR data [6], [7], providing complementary spatial and contextual cues that improved the accuracy and robustness of DL-based BP models.

In this paper, we propose a novel DL framework for BP in mmWave systems using Kolmogorov–Arnold Networks (KANs)—neural networks that model complex functions with spline-based transformations [11]–[13]. We adopt KAN as a *late-fusion* framework to integrate vision, LiDAR, and GPS data for improved BP. Unlike conventional multilayer perceptrons (MLPs), KAN replaces fixed activations with learnable spline-based parameterizations on the weights, yielding enhanced local adaptability and multi-resolution flexibility, while potentially leading to smaller computational graphs [14].

KANs have recently gained attention for capturing meaningful low-level features across diverse applications, including 5G communications [15]. They have been applied to both regression [16] and classification tasks [17]. Despite this progress, further study is needed to confirm their robustness and generalizability, particularly in communication systems where we aim to leverage their strengths. We summarize our main contributions as follows:

- **KAN-based BP:** We investigate the capability of standalone KAN and its integration with convolutional neural network (CNN) and Gated Recurrent Unit (GRU) feature extraction networks for BP in mmWave systems in regression and classification setups using vision, LiDAR, and GPS modalities.
- **KAN-based late-fusion framework.** We propose a novel and efficient KAN-based late fusion of various modalities. Our framework promotes modalities' distinct environmental cues.
- **Empirical evaluations.** Our extensive empirical analysis demonstrates that replacing MLP with KAN is the most effective approach to improve BP in single and multi-modal settings.

II. Related work

Since no single modality can fully capture the complexity of real-world propagation environments, researchers increasingly adopt multimodal fusion to integrate complementary data [18]. Among these strategies, late fusion, i.e., combining high-level features from independently processed modalities, has proven effective for integrating visual and positional data

[19]–[21]. For instance, Charan *et al.* [19] and Nie *et al.* [21] used late fusion of vision and GPS via MLPs in varied environments. Reus-Muns *et al.* [20] applied element-wise multiplication of convolutional features to preserve spatial structure in V2I mmWave links. For BP, TUNE [22] fuses GPS, vision, and LiDAR via penultimate-layer fusion in V2X, where Salehi *et al.* [23] concatenates latent embeddings from GPS, vision, and LiDAR through a CNN and softmax. Tian *et al.* [8] introduced transformer encoders after convolutional blocks to fuse image and LiDAR intermediate features, merging with GPS via MLP.

Recent research shows that KANs significantly outperform traditional MLPs by replacing fixed activation functions with dynamically adjusted, spline-parametrized univariate functions. KANs enables adaptive modeling of complex nonlinear relationships and represents any ReLU^k-based MLP with comparable or fewer parameters—often outperforming MLPs in efficiency, interpretability, and adaptability [11], [12] over diverse datasets and modalities. K-DeepSC reduced parameters by 44% when replaced MLPs in a semantic encoder while preserving image reconstruction quality [24] for semantic communication. VLP-KAN [25] achieved sub-centimeter accuracy for visible light positioning using received signal strength, outperforming MLPs with lower complexity. In spectrum sensing and modulation classification, KANs showed improved noise robustness, reduced complexity, and competitive accuracy compared to conventional DNNs and CNNs [15], [26]. KANs excel at both vision and time-series tasks. On vision benchmarks like MNIST, KAN–MLP, and KAN–CNN hybrids perform strongly, though broader evaluations are needed [27]–[29]. For time-series, KANs capture temporal dependencies with fewer parameters than MLPs [17], [30]. This efficiency suggests that KANs can enhance or replace MLP/CNN layers in edge learning and communication systems.

III. System Model and Problem Formulation

A. System Model

Our communication system includes a base station (BS) with a mmWave antenna array of M elements in a Uniform Linear Array (ULA) configuration. The BS serves a single mobile user equipment (UE) with an omni-directional antenna. The system operates with Orthogonal Frequency Division Multiplexing (OFDM) using K subcarriers, and beamforming is performed at the BS using a predefined codebook $\mathcal{F} = \{\mathbf{f}_q\}_{q=1}^{Q}$, where each beamforming vector $\mathbf{f}_q \in \mathbb{C}^{M\times 1}$ and Q is the total number of available beams. The received signal at the UE on the k-th subcarrier at time t is $y_k[t] = \mathbf{h}_k^T[t]\mathbf{f}_{q[t]}x + v_k[t]$, where $\mathbf{h}_k[t] \in \mathbb{C}^{M\times 1}$ is the BS–UE channel vector, $q[t]$ is the selected beam index, $x \in \mathbb{C}$ is the transmitted symbol, and $v_k[t]$ is circularly symmetric complex Gaussian noise, $\mathcal{NC}(0, \sigma^2)$.

B. Problem Formulation

The BP's goal is to maximize beamforming gain by selecting the optimal beam index $q[t] \in \{1, \ldots, Q\}$ from a predefined codebook formulated as [31]:

$$q^*[t] = \underset{q\in\{1,\ldots,Q\}}{\operatorname{argmax}} \frac{1}{K}\sum_{k=1}^{K}\left|\mathbf{h}_k^T[t]\mathbf{f}_q\right|^2.$$

Traditional BP methods rely on exhaustive search or detailed channel knowledge, which are impractical in mmWave environments. While DL-based BP is commonly formulated as a classification task over a discrete codebook [8], the underlying problem inherently exists in continuous angular space. Exploiting continuous parameters such as direction and received power enables more refined representations for mmWave beamforming beyond discrete indices. We present our BP pipeline as both classification and regression tasks, applicable to both discrete and continuous beamforming scenarios. In cases with discrete target indices (as in this paper for the used dataset), we round continuous predictions of the regression model to the nearest available beam for evaluation. With a continuous dataset, regression could in principle yield finer, more physics-aligned predictions than classification restricted to discrete beams [9].

IV. KAN-Based Beam Prediction Framework

A. Preliminary

Kolmogorov–Arnold Theorem and Network: The Kolmogorov–Arnold representation theorem states that any continuous multivariate function on a bounded domain can be expressed as a composition and superposition of univariate functions, each depending on a subset of the input variables. Given a smooth function $f : [0,1]^{n_{\text{in}}} \to \mathbb{R}$, continuous univariate functions ψ_q and $\phi_{p,q}$ can be found such that:

$$f(\mathbf{x}) = \sum_{q=1}^{2n_{\text{in}}+1} \psi_q\left(\sum_{p=1}^{n_{\text{in}}}\phi_{p,q}(x_p)\right),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_{n_{\text{in}}})$ is the input vector consisting of n_{in} input variables. $\phi_{p,q} : [0,1] \to \mathbb{R}$ and $\psi_q : \mathbb{R} \to \mathbb{R}$ are univariate functions on the individual input vector components x_p and the summation of $\phi_{p,q}(x_p)$s, respectively.

The KAN architecture [11] is based on a generalization of the Kolmogorov–Arnold representation theorem and consists of L stacked layers defined as:

$$\text{KAN}(\mathbf{x}) = (\mathbf{\Phi}_{L-1}\circ\cdots\circ\mathbf{\Phi}_1\circ\mathbf{\Phi}_0)(\mathbf{x}), \quad x_{l+1,j} = \sum_{i=1}^{n_l}\phi_{l,j,i}(x_{l,i}),$$

where $\mathbf{x}_l \in \mathbb{R}^{n_l}$ is the input vector to layer $\mathbf{\Phi}_l$ and $\mathbf{x}_{l+1} \in \mathbb{R}^{n_{l+1}}$ is its output, for $l = 0, \ldots, L-1$. The element $x_{l,i}$ denotes the i-th component of $\mathbf{x}_l$ for $i = 1, \ldots, n_l$, while $j = 1, \ldots, n_{l+1}$ indexes the components of $\mathbf{x}_{l+1}$.

Each $\mathbf{\Phi}_l$ can be viewed as a transformation matrix where the element $\phi_{l,j,i}$ is a univariate nonlinear function applied to $x_{l,i}$ and connecting it to $x_{l+1,j}$. Each $\phi_{l,j,i}$ is parameterized by a combination of spline basis functions and a scaled sigmoid, given by $\phi(x) = \frac{w_b x}{1+e^{-x}} + \sum_{i=0}^{G+k-1} c_i B_i(x)$, where G is the spline grid size, B_i is the k-th order B-spline basis function defined on a uniform grid, and c_i, w_b are trainable parameters. KAN's unified use of linear transformations and spline-based nonlinearities improves approximation accuracy and flexibility over conventional neural networks [11].

KAN versus MLP:

KANs reduce to MLPs when edge splines are linear and node nonlinearities are absent. An MLP with depth L and width N has $\mathcal{O}(N^2L)$ parameters, whereas KANs replace weights with splines of $G+k$ coefficients, yielding $\mathcal{O}(N^2L(G+k))$ parameters. This added capacity enhances expressiveness, allowing KANs to match or exceed MLP accuracy with fewer layers or narrower widths [13].

KAN and BP:

Recent studies show that KANs reduce spectral bias compared to MLPs, enabling earlier capture of high-frequency features and fine-grained nonlinear patterns under rapid changes [11], [13]. These strengths make them well-suited for BP tasks with complex spatiotemporal dependencies in multimodal sensor data. Motivated by these advantages [11]–[13], we investigate KANs as decision heads atop CNN/GRU backbones, as fusion layers in multimodal setups, and as end-to-end models.

B. Proposed Network Structures

We propose KAN-based BP models using vision, LiDAR, and GPS inputs across classification and regression tasks in both single- and multi-modality settings. Each sample is mapped to a discrete beam index from the mmWave codebook in classification. In regression, the model predicts a continuous pseudo-angle that spans the beam range, capturing geometric relations between adjacent beams without requiring exact angular coverage.

Single-Modality Network: We apply KAN directly to raw GPS data, while for vision, applying KAN to extracted features yields significantly better performance. Detailed KAN-based single-modality architectures are provided in Table I, where FC(a, b) and KAN(a, b) denote FC and KAN layers with a inputs and b outputs, respectively. The proposed structures include a GRU layer with d hidden units, a 1D convolutional layer (Conv1D) with c channels and kernel size k, and DO(p) (Dropout layer with dropout probability p). Details of the CNN and CNN_GRU architectures are illustrated in Fig. 1.

Late-Fusion Network: We employ KAN-based late fusion to combine all modalities for BP, leveraging its expressive power to model nonlinear complex relationships. As shown in Fig. 1, features extracted via GRU for GPS and via CNN_GRU for the other modalities are concatenated and passed to a KAN layer for classification or regression. Compared to MLPs, KANs offer greater flexibility by learning dynamic, spline-based transformations, enabling more effective modeling of intricate dependencies in fused multimodal data.

V. Simulation Results

A. Dataset

DeepSense6G [32] is a real-world multimodal dataset collected in urban scenarios (32–34), where dynamic elements such as vehicles and pedestrians impact RF signals. The BS is equipped with mmWave, camera, LiDAR, and GPS, while the target vehicle carries GPS and a mmWave receiver. Each sample includes the last five camera and LiDAR frames, two GPS readings, and the corresponding optimal beam index.

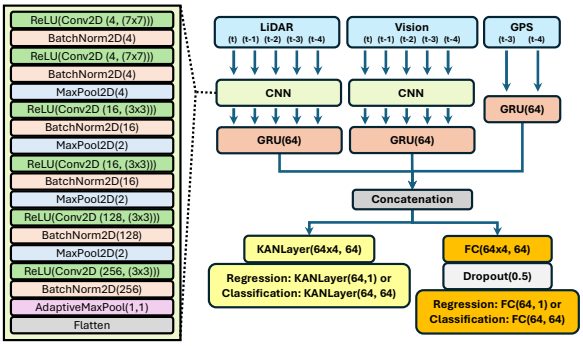

Fig. 1: CNN-GRU-KAN/MLP network structure for regression and classification tasks using late fusion of features from each modality.

TABLE I: Single-modality model architectures. Modalities are denoted as: G = GPS, L = LiDAR, V = Vision.

Input	Model	Architecture
G	GRU	[GRU(64), FC(64, n_{out})]
	MLP	[FC(2,5), DO(0.5), FC(5, n_{out})]
	CNN	[Conv1D(8,2), DO(0.5), Conv1D(16,1), DO(0.5), FC(16, n_{out})]
	KAN	[KAN(2,5), KAN(5, n_{out})]
	GRU_FC	[GRU(64), FC(64,64), DO(0.5), FC(64, n_{out})]
	GRU_KAN	[GRU(64), KAN(64,64), KAN(64, n_{out})]
L/V	CNN_GRU_FC	[CNN_GRU, FC(64,64), DO(0.5), FC(64, n_{out})]
	CNN_FC	[CNN FC(256×5, 64), DO(0.5), FC(64, n_{out})]
	CNN_KAN_FC	[CNN, KAN(256×5,64), KAN(64,64), DO(0.5), FC(64, n_{out})]
	CNN_GRU_KAN	[CNN_GRU, KAN(64,64), KAN(64, n_{out})]
	CNN_KAN	[CNN, KAN(256×5,64), KAN(64, n_{out})]

B. Data Preprocessing

LiDAR: The LiDAR preprocessing pipeline projects raw (x, y, z) point clouds and intensity values onto a 150×150 Bird's Eye View grid after two filters: an azimuth filter (110°–360°) to focus on the desired field of view and a spatial filter to retains points in a 100×100 m² area with ($x, y \in [-60, 40]$ m). It then generates an RGB BEV image where red (R) represents height, green (G) intensity, and blue (B) normalized point density per cell.

Camera: Vision samples are RGB images originally sized (540, 960, 3), resized to (150, 150, 3) to reduce computation and ensure modality consistency.

GPS: GPS coordinates are projected to Cartesian-XY using UTM and normalized relative to the BS. To ensure spatial consistency across scenarios, calibration is then applied by aligning GPS angles with representative central-beam samples [8], [33].

C. Training

The dataset was partitioned into training and test sets (90:10) via stratified sampling to maintain class balance. From the training set, 20% is reserved for validation. Train/test sets include both Line of sight (LOS) and Non-LOS samples. Models are trained with the Adam optimizer (learning rate 10^{-3}) for up to 250 epochs with a batch size of 30. Early stopping is triggered if validation loss doesn't improve for 12 epochs, and the learning rate is reduced by a factor of 10 after 8 stagnant epochs. Results are averaged over five independent runs with distinct random seeds. We use the `KANLayer` from the KAN library (see https://github.com/

KindXiaoming/pykan) for all KAN layers in our models. As large grid sizes G can hinder KAN spline optimization and degrade performance [17], we found $G = 5$ and 3rd-order B-splines ($k = 3$) effective for our framework. KANs are typically trained with L-BFGS, a second-order optimizer that approximates the Hessian to improve convergence, but incurs high computational cost on large datasets. Thus, we use Adam—a faster, memory-efficient first-order optimizer—for end-to-end training. In our late-fusion models, we also explore two-step training: first using Adam, then fine-tuning only the KAN layers with L-BFGS, applied to concatenated features from fixed CNN and GRU components. For fine-tuning, we configure L-BFGS with a learning rate of 1, up to 20 iterations per step, and strict convergence criteria [12] for stability.

D. Performance Evaluation Metrics

1) Prediction Error: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are defined as $\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|\hat{q}_n - q_n|$ and $\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{q}_n - q_n)^2}$ for N samples, with $\hat{q}_n$ and q_n as the predicted and true beam indices.

2) Top-1 Accuracy: Top-1 accuracy is the fraction of predictions that, after rounding, exactly match the ground truth, applicable to both regression and classification models.

3) Top-K Beam: As discussed in [33], top-1 accuracy may be misleading in BP due to similar power among adjacent beams. Thus, we report the *top-K beam* metric (K=3), which verifies if the predicted beam lies among the K highest-power true beams, offering a more practical performance measure.

4) Power Ratio (PR): Following our earlier work [33], the $\text{PR} = P_{\text{predicted beam}}/P_{\text{max}}$ metric quantifies beam selection quality as the ratio of predicted to maximum available power, offering a practical measure of communication effectiveness.

5) Distance-Based Accuracy (DBA): The DBA metric [31] quantifies how close the top-K predicted beam indices $\hat{q}_{n,k'}$ ($k' \in \{1,\ldots,K\}$) are to the true index q_n in classification-based BP:

$$\text{DBA}_{\text{cls}} = \frac{1}{K}\sum_{k=1}^{K}\left[1 - \frac{1}{N}\sum_{n=1}^{N}\min_{1\le k'\le k}\left[\min\left(\frac{|\hat{q}_{n,k'} - q_n|}{\Delta}, 1\right)\right]\right],$$

where N is the number of test samples, K=3, and Δ=5 is a normalization constant. For regression-based BP, we propose DBA_{reg}, a simplified DBA_{cls} evaluated with K=1 using a single predicted index $\hat{q}_n$. This naturally fits scalar regression outputs and bounds the distance to q_n. Though conceptually similar, DBA_{reg} and DBA_{cls} are not directly comparable: the former reflects Top-1 accuracy and yields stricter scores, while the latter, based on Top-K, typically produces higher and more variable values. Unlike unbounded metrics like MAE or RMSE, DBA_{reg} provides a bounded, interpretable error measure aligned with scenarios where near-correct predictions suffice for reliable communication.

E. Results and Discussion

We compared the performance of our KAN-based BP architectures to conventional DL architectures through experimental evaluation in the classification and regression framework presented in Table II and Table III. Among single-modality models (Table II), GPS outperforms vision and LiDAR, in both regression and classification. Using raw GPS data, the standalone use of KAN cannot surpass GRU; however, it outperforms all other models. The best performance for BP using GPS data was achieved by applying KAN to extracted features from GRU, surpassing all models, including CNN, despite having fewer parameters (Table V). To further highlight KAN's superior expressiveness over MLP on low-dimensional GPS data, we analyzed model design choices by comparing network shape, parameter counts, and performance (Table IV). Results highligh that with identical [2,5,1] structures, KAN has slightly more parameters than MLP but far higher performance, while smaller KANs ([2,1,1]) nearly match MLP in parameter count yet still outperform it. For GPS data, $G = 2$ proved more effective than $G = 5$. Removing the hidden layer severely degraded KAN performance.

Feature extraction from vision yields compact inputs, enabling KAN and MLP to operate more efficiently. In contrast, applying them directly to raw images is inefficient, as flattening 2D inputs increases parameter count (Table V), computational cost, and disrupts spatial structure. To limit complexity, direct KAN application on raw vision uses a grayscale image from the last time step, while other models use RGB inputs. Both KAN and MLP performed poorly on raw images, so KAN was applied in this setting only as a baseline.

Our fusion model results in Table III show that KAN-based fusion of vision and GPS, in the classification framework, outperforms other combined modalities and regression frameworks. Adding LiDAR to vision and GPS yields similar or slightly improved performance in both frameworks. In comparing KAN-based and MLP-based fusion, the former indicates far more improvements than the latter in both frameworks. Moreover, two-stage training with Adam followed by L-BFGS optimizer significantly enhanced accuracy on modality fusion indicating the importance of fine-tuning KAN to achieve superior performance. The current CNN_GRU_MLP-based late-fusion models for classification simplifies our previous design [33] by removing MLPs from modality branches, retaining only CNN_GRU for feature extraction, and having MLP or KAN for fusion. In addition to reduced complexity, our new design is more effective, outperforming the earlier version in both MLP- and KAN-based fusion results presented in this study. After feature extraction, KAN acts as a flexible, locally adaptive decision head. BP involves irregular patterns arising from reflections or blockages. KAN models these piecewise relationships effectively, providing sharper decision boundaries and better local adaptation than traditional MLPs.

VI. Conclusion

Accurate beam prediction (BP) reduces training overhead and ensures reliable connectivity in high-mobility mmWave systems. To model the piecewise and nonlinear patterns caused by beam switches, reflections, and blockages, we

TABLE II: Performance evaluation of *single-modality* BP models in both classification and regression settings. KAN and MLP architectures are specified as [input, hidden, output] neurons per layer. Top-1 A and Top-3 B refer to Top-1 accuracy and Top-3 beam selection, respectively. * Indicates grayscale vision input with a single time frame for reduced KAN complexity.

Task	Model	MAE (↓)	RMSE (↓)	Top-1 A (↑)	Top-3 B (↑)	DBA_{reg} (↑)	DBA_{cls}(↑)	PR (↑)
LiDAR Reg.	CNN_GRU_FC	2.95 ± 0.52	4.47 ± 0.92	0.14 ± 0.03	0.40 ± 0.06	0.53 ± 0.05	–	0.84 ± 0.03
	CNN_FC	9.01 ± 3.19	10.87 ± 3.74	0.03 ± 0.02	0.10 ± 0.06	0.17 ± 0.08	–	0.62 ± 0.06
	CNN_KAN_FC	6.76 ± 0.92	8.30 ± 1.13	0.03 ± 0.02	0.11 ± 0.06	0.19 ± 0.07	–	0.64 ± 0.05
	CNN_GRU_KAN	**2.93 ± 0.52**	**4.31 ± 0.65**	**0.14 ± 0.03**	**0.40 ± 0.06**	**0.53 ± 0.06**	–	**0.84 ± 0.03**
	CNN_KAN	6.59 ± 4.13	7.99 ± 4.24	0.07 ± 0.05	0.22 ± 0.15	0.31 ± 0.20	–	0.72 ± 0.11
Vision Reg.	CNN_GRU_FC	2.57 ± 0.18	3.90 ± 0.26	0.17 ± 0.03	0.43 ± 0.03	0.57 ± 0.02	–	0.86 ± 0.01
	CNN_FC	7.31 ± 4.85	8.65 ± 4.45	0.06 ± 0.06	0.18 ± 0.17	0.26 ± 0.22	–	0.68 ± 0.13
	CNN_KAN_FC	3.53 ± 1.83	4.78 ± 2.17	0.13 ± 0.07	0.36 ± 0.16	0.49 ± 0.14	–	0.82 ± 0.07
	CNN_GRU_KAN	**2.41 ± 0.53**	**3.63 ± 0.58**	**0.19 ± 0.05**	**0.48 ± 0.11**	**0.59 ± 0.08**	**–**	**0.87 ± 0.04**
	CNN_KAN	2.86 ± 1.18	3.93 ± 1.33	0.15 ± 0.06	0.41 ± 0.14	0.53 ± 0.13	–	0.84 ± 0.08
	MLP* [22500,64,1]	13.77 ± 5.74	15.97 ± 6.12	0.01 ± 0.01	0.04 ± 0.02	0.08 ± 0.03	–	0.56 ± 0.05
	KAN* [22500,64,1], G=5	12.23 ± 11.41	14.31 ± 12.01	0.08 ± 0.07	0.23 ± 0.19	0.32 ± 0.26	–	0.65 ± 0.24
GPS Reg.	MLP [2,5,1]	3.51 ± 0.20	4.47 ± 0.19	0.09 ± 0.01	0.23 ± 0.02	0.41 ± 0.03	–	0.77 ± 0.01
	KAN [2,5,1], G=5	2.54 ± 0.02	4.19 ± 0.01	0.20 ± 0.01	0.54 ± 0.00	0.61 ± 0.00	–	0.87 ± 0.00
	GRU	1.32 ± 0.01	2.47 ± 0.02	0.26 ± 0.00	0.74 ± 0.00	0.76 ± 0.00	–	0.96 ± 0.00
	CNN	3.82 ± 0.28	4.83 ± 0.28	0.06 ± 0.01	0.24 ± 0.03	0.38 ± 0.03	–	0.75 ± 0.02
	GRU_FC	1.44 ± 0.02	2.54 ± 0.01	0.24 ± 0.00	0.68 ± 0.01	0.74 ± 0.00	–	0.95 ± 0.00
	GRU_KAN	**1.24 ± 0.02**	**2.41 ± 0.02**	**0.29 ± 0.02**	**0.77 ± 0.01**	**0.78 ± 0.00**	**–**	**0.97 ± 0.00**
LiDAR Cls.	CNN_GRU_FC	2.48 ± 0.43	5.95 ± 0.73	0.32 ± 0.03	0.65 ± 0.04	–	0.78 ± 0.04	0.92 ± 0.02
	CNN_FC	7.46 ± 2.84	10.60 ± 2.52	0.14 ± 0.08	0.27 ± 0.15	–	0.42 ± 0.16	0.72 ± 0.09
	CNN_KAN_FC	2.53 ± 0.30	6.37 ± 1.16	0.33 ± 0.02	0.66 ± 0.03	–	0.79 ± 0.02	0.92 ± 0.00
	CNN_GRU_KAN	2.30 ± 0.22	5.38 ± 0.56	0.32 ± 0.01	0.65 ± 0.02	–	0.79 ± 0.01	0.93 ± 0.01
	CNN_KAN	**2.47 ± 0.46**	**6.09 ± 0.89**	**0.35 ± 0.03**	**0.69 ± 0.04**	**–**	**0.80 ± 0.03**	**0.93 ± 0.02**
Vision Cls.	CNN_GRU_FC	1.86 ± 0.29	4.64 ± 0.85	0.38 ± 0.02	0.73 ± 0.03	–	0.83 ± 0.02	0.94 ± 0.01
	CNN_FC	4.23 ± 3.35	6.84 ± 3.51	0.24 ± 0.11	0.47 ± 0.22	–	0.63 ± 0.21	0.84 ± 0.12
	CNN_KAN_FC	**1.40 ± 0.16**	**3.73 ± 0.39**	**0.41 ± 0.02**	**0.80 ± 0.02**	**–**	**0.87 ± 0.01**	**0.96 ± 0.01**
	CNN_GRU_KAN	1.59 ± 0.18	4.02 ± 0.30	0.38 ± 0.02	0.76 ± 0.03	–	0.85 ± 0.02	0.95 ± 0.01
	CNN_KAN	2.19 ± 0.46	5.68 ± 0.83	0.39 ± 0.02	0.75 ± 0.05	–	0.83 ± 0.03	0.94 ± 0.01
GPS Cls.	MLP [2,5,64]	2.75 ± 0.19	4.26 ± 0.22	0.22 ± 0.01	0.45 ± 0.03	–	0.69 ± 0.01	0.87 ± 0.01
	KAN [2,5,64], G=5	2.79 ± 0.09	4.83 ± 0.10	0.30 ± 0.00	0.54 ± 0.01	–	0.71 ± 0.01	0.87 ± 0.00
	GRU	1.30 ± 0.01	2.63 ± 0.02	0.39 ± 0.00	0.75 ± 0.00	–	0.86 ± 0.00	0.97 ± 0.00
	CNN	3.45 ± 0.46	4.85 ± 0.52	0.17 ± 0.02	0.35 ± 0.05	–	0.61 ± 0.04	0.81 ± 0.03
	GRU_FC	1.25 ± 0.01	2.58 ± 0.02	0.40 ± 0.01	0.76 ± 0.00	–	0.86 ± 0.00	0.97 ± 0.00
	GRU_KAN	**1.18 ± 0.01**	**2.53 ± 0.01**	**0.42 ± 0.00**	**0.79 ± 0.01**	**–**	**0.88 ± 0.00**	**0.97 ± 0.00**

TABLE III: Performance evaluation of *late-fusion* BP models, grouped by task (regression or classification) and fusion type (MLP, Adam-trained KAN, and LBFGS fine-tuned KAN*). Top-1 A and Top-3 B indicate Top-1 accuracy and Top-3 beam selection, respectively.

Task	Model	MAE (↓)	RMSE (↓)	Top-1 A (↑)	Top-3 B (↑)	DBA_{reg} (↑)	DBA_{cls} (↑)	PR (↑)
LV Reg.	CNN_GRU_MLP	3.00 ± 0.39	4.24 ± 0.42	0.11 ± 0.03	0.35 ± 0.07	0.50 ± 0.06	–	0.83 ± 0.03
	CNN_GRU_KAN	2.34 ± 0.49	3.68 ± 0.59	0.19 ± 0.04	0.51 ± 0.09	0.61 ± 0.07	–	0.88 ± 0.03
	CNN_GRU_KAN*	**1.93 ± 0.23**	**3.00 ± 0.26**	**0.19 ± 0.03**	**0.54 ± 0.05**	**0.66 ± 0.04**	**–**	**0.91 ± 0.02**
VG Reg.	CNN_GRU_MLP	1.41 ± 0.13	2.62 ± 0.18	0.28 ± 0.03	0.70 ± 0.05	0.75 ± 0.02	–	0.95 ± 0.01
	CNN_GRU_KAN	1.37 ± 0.15	2.61 ± 0.22	0.33 ± 0.02	0.75 ± 0.04	0.76 ± 0.02	–	0.95 ± 0.01
	CNN_GRU_KAN*	**1.13 ± 0.08**	**2.30 ± 0.04**	**0.36 ± 0.03**	**0.80 ± 0.03**	**0.80 ± 0.02**	**–**	**0.97 ± 0.00**
LVG Reg.	CNN_GRU_MLP	1.61 ± 0.15	2.76 ± 0.22	0.24 ± 0.03	0.62 ± 0.04	0.71 ± 0.02	–	0.93 ± 0.01
	CNN_GRU_KAN	1.55 ± 0.37	2.59 ± 0.29	0.26 ± 0.07	0.65 ± 0.14	0.72 ± 0.07	–	0.93 ± 0.04
	CNN_GRU_KAN*	**1.10 ± 0.04**	**2.23 ± 0.06**	**0.36 ± 0.02**	**0.81 ± 0.02**	**0.81 ± 0.01**	**–**	**0.97 ± 0.00**
LV Cls.	CNN_GRU_MLP	2.12 ± 0.30	5.57 ± 1.33	0.37 ± 0.01	0.73 ± 0.01	–	0.82 ± 0.01	0.94 ± 0.01
	CNN_GRU_KAN	1.75 ± 0.34	4.71 ± 0.64	0.39 ± 0.03	0.77 ± 0.04	–	0.85 ± 0.03	0.95 ± 0.01
	CNN_GRU_KAN*	**1.63 ± 0.80**	**4.24 ± 1.51**	**0.41 ± 0.05**	**0.77 ± 0.10**	**–**	**0.86 ± 0.06**	**0.95 ± 0.03**
VG Cls.	CNN_GRU_MLP	1.08 ± 0.04	2.49 ± 0.04	0.44 ± 0.01	0.82 ± 0.01	–	0.89 ± 0.01	0.97 ± 0.00
	CNN_GRU_KAN	1.06 ± 0.08	2.45 ± 0.09	0.43 ± 0.02	0.82 ± 0.02	–	0.89 ± 0.01	0.97 ± 0.00
	CNN_GRU_KAN*	**0.95 ± 0.03**	**2.40 ± 0.10**	**0.47 ± 0.01**	**0.87 ± 0.01**	**–**	**0.91 ± 0.00**	**0.98 ± 0.00**
LVG Cls.	CNN_GRU_MLP	1.13 ± 0.07	2.53 ± 0.07	0.43 ± 0.02	0.80 ± 0.02	–	0.88 ± 0.01	0.97 ± 0.00
	CNN_GRU_KAN	**0.94 ± 0.03**	**2.28 ± 0.12**	**0.46 ± 0.02**	**0.86 ± 0.01**	**–**	**0.91 ± 0.00**	**0.98 ± 0.00**
	CNN_GRU_KAN*	**0.95 ± 0.03**	**2.30 ± 0.13**	**0.46 ± 0.01**	**0.86 ± 0.01**	**–**	**0.91 ± 0.00**	**0.98 ± 0.00**

proposed a Kolmogorov–Arnold Network (KAN)-based BP framework. In our experiments, KAN offered greater expressiveness than MLP and consistently outperformed it in single-modality and late-fusion settings with similar or lower structural complexity. KAN-based fusion achieved the highest classification accuracy by effectively capturing spatial and temporal features from heterogeneous sensor data. Future work will investigate convolutional KANs to enhance scalability for high-dimensional, image-based inputs.

REFERENCES

[1] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[2] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[3] K. Vuckovic, M. B. Mashhadi, F. Hejazi, N. Rahnavard, and A. Alkhateeb, "Paramount: Towards generalizable deep learning for mmwave beam selection using sub-6ghz channel measurements," *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 5187–5202, 2024.

[4] T. Nishio, Y. Koda, J. Park, M. Bennis, and K. Doppler, "When wireless communications meet computer vision in beyond 5G," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 76–83, 2021.

[5] H. Ahn, I. Orikumhi, J. Kang, H. Park, H. Jwa, J. Na, and S. Kim, "Machine learning-based vision-aided beam selection for mmWave multi-user miso system," *IEEE Wireless Communications Letters*, 2022.

[6] M. B. Mashhadi, M. Jankowski, T.-Y. Tung, S. Kobus, and D. Gündüz,

TABLE IV: Comparison of MLP and KAN models with different architectures for GPS-based regression. Top-1 A and Top-3 B denote Top-1 accuracy and Top-3 beam selection, respectively.

Task	Model	Parameters (↓)	MAE (↓)	RMSE (↓)	Top-1 A (↑)	Top-3 B (↑)	DBA_{reg} (↑)	PR (↑)
MLP	[2,5,1]	21	3.51 ± 0.20	4.47 ± 0.19	0.09 ± 0.01	0.23 ± 0.02	0.41 ± 0.03	0.77 ± 0.01
KAN	[2,5,1], G=5	150	2.54 ± 0.02	4.19 ± 0.01	0.20 ± 0.01	0.54 ± 0.00	0.61 ± 0.00	0.87 ± 0.00
KAN	[2,1,1], G=5	30	3.26 ± 0.04	5.07 ± 0.02	0.13 ± 0.00	0.44 ± 0.01	0.54 ± 0.01	0.84 ± 0.00
KAN	[2,1,1], G=2	21	3.11 ± 0.02	4.79 ± 0.01	0.13 ± 0.00	0.42 ± 0.01	0.54 ± 0.00	0.84 ± 0.00
KAN	[2,1], G=2	14	20.61 ± 0.01	26.04 ± 0.00	0.02 ± 0.00	0.05 ± 0.01	0.09 ± 0.00	0.49 ± 0.00

TABLE V: Parameter count across model architectures.

Modality	Model	Reg.	Cls.
LiDAR/Vision (RGB)	CNN-GRU-FC	384,917	389,012
	CNN-FC	400,917	405,012
	CNN-KAN-FC	1,183,253	1,187,348
	CNN-GRU-KAN	422,292	462,612
	CNN-KAN	1,138,708	1,179,028
Vision (Gray)	KAN	14,400,640	–
	MLP	1,440,129	–
GPS	GRU	12,929	17,024
	MLP	21	399
	CNN	185	1,256
	KAN	150	3,300
	GRU-FC	17,089	21,184
	GRU-KAN	54,464	94,784
LV Late Fusion	CNN-GRU-MLP	769,705	773,800
	CNN-GRU-KAN	843,944	884,264
VG Late Fusion	CNN-GRU-MLP	401,877	405,972
	CNN-GRU-KAN	476,116	516,436
LVG Late Fusion	CNN-GRU-MLP	786,665	790,760
	CNN-GRU-KAN	897,768	938,088

"Federated mmWave beam selection utilizing LIDAR data," *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2269–2273, 2021.

[7] M. Zecchin, M. B. Mashhadi, M. Jankowski, D. Gündüz, M. Kountouris, and D. Gesbert, "LiDAR and position-aided mmWave beam selection with non-local CNNs and curriculum training," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 2979–2990, 2022.

[8] Y. Tian, Q. Zhao, F. Boukhalfa, K. Wu, F. Bader *et al.*, "Multimodal transformers for wireless communications: A case study in beam prediction," *ITU Journal on Future and Evolving Technologies*, vol. 4, no. 3, pp. 461–471, 2023.

[9] T. Yassine, B. Chatelier, V. Corlay, M. Crussière, S. Paquelet, O. Tirkkonen, and L. Le Magoarou, "Model-based deep learning for beam prediction based on a channel chart," in *2023 57th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2023, pp. 1636–1640.

[10] J. Morais, A. Bchboodi, H. Pezeshki, and A. Alkhateeb, "Position-aided beam prediction in the real world: How useful gps locations actually are?" in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 1824–1829.

[11] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov-Arnold Networks," Feb. 2025, arXiv:2404.19756 [cs]. [Online]. Available: http://arxiv.org/abs/2404.19756

[12] Z. Liu, P. Ma, Y. Wang, W. Matusik, and M. Tegmark, "KAN 2.0: Kolmogorov-Arnold Networks Meet Science," Aug. 2024, arXiv:2408.10205 [cs]. [Online]. Available: http://arxiv.org/abs/2408.10205

[13] Y. Wang, J. W. Siegel, Z. Liu, and T. Y. Hou, "On the expressiveness and spectral bias of KANs," Feb. 2025, arXiv:2410.01803 [cs]. [Online]. Available: http://arxiv.org/abs/2410.01803

[14] T. Ji, Y. Hou, and D. Zhang, "A Comprehensive Survey on Kolmogorov Arnold Networks (KAN)," Jan. 2025, arXiv:2407.11075 [cs]. [Online]. Available: http://arxiv.org/abs/2407.11075

[15] P. Patel and D. K. Patel, "Smart sensing framework using kolmogorov-arnold networks for enhanced detection in 5g environments," in *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, 2025, pp. 332–339.

[16] N. R. Panczyk, O. F. Erdem, and M. I. Radaideh, "Opening the black-box: Symbolic regression with kolmogorov-arnold networks for energy applications," 2025. [Online]. Available: https://arxiv.org/abs/2504.03913

[17] C. Dong, L. Zheng, and W. Chen, "Kolmogorov-Arnold Networks (KAN) for Time Series Classification and Robust Analysis," Sep. 2024, arXiv:2408.07314 [cs]. [Online]. Available: http://arxiv.org/abs/2408.07314

[18] J. Summaira, X. Li, A. M. Shoib, S. Li, and J. Abdul, "Recent advances and trends in multimodal deep learning: A review," 2021.

[19] G. Charan, T. Osman, A. Hredzak, N. Thawdar, and A. Alkhateeb, "Vision-position multi-modal beam prediction using real millimeter wave datasets," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2727–2731.

[20] G. Reus-Muns, B. Salehi, D. Roy, T. Jian, Z. Wang, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on visual and location data for v2i mmwave beamforming," in *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*, 2021, pp. 559–566.

[21] J. Nie, Y. Cui, T. Yu, J. Mu, W. Yuan, and X. Jing, "An efficient nocturnal scenarios beamforming based on multi-modal enhanced by object detection," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 515–520.

[22] J. Gu, B. Salehi, S. Pimple, D. Roy, and K. R. Chowdhury, "Tune: Transfer learning in unseen environments for v2x mmwave beam selection," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 1658–1663.

[23] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7639–7655, 2022.

[24] W. Liu, Q. Zeng, L. Lu, and W. Abdul, "Intelligent semantic communication system based on kolmogorov–arnold networks driven by dynamic terminal-side computing power network," *Electronics*, vol. 13, no. 20, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/20/4076

[25] F. Seguel, D. Salihu, S. Hägele, and E. Steinbach, "Vlp-kan: Low-complexity and interpretable rss-based vlp using kolmogorov-arnold networks," in *2025 20th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2025, pp. 1–8.

[26] J. Pan, "Kolmogorov-arnold networks for spectrum sensing and modulation classification," Ph.D. dissertation, University of British Columbia, 2024. [Online]. Available: https://open.library.ubc.ca/collections/ubctheses/24/items/1.0447238

[27] M. M. Ferdaus, M. Abdelguerfi, E. Ioup, D. Dobson, K. N. Niles, K. Pathak, and S. Sloan, "Kanice: Kolmogorov-arnold networks with interactive convolutional elements," 2024. [Online]. Available: https://arxiv.org/abs/2410.17172

[28] M. Liu, D. Geißler, D. Nshimyimana, S. Bian, B. Zhou, and P. Lukowicz, "Initial Investigation of Kolmogorov-Arnold Networks (KANs) as Feature Extractors for IMU Based Human Activity Recognition," Jun. 2024, arXiv:2406.11914 [cs]. [Online]. Available: http://arxiv.org/abs/2406.11914

[29] M. Cheon, "Kolmogorov-arnold network for satellite image classification in remote sensing," 2024. [Online]. Available: https://arxiv.org/abs/2406.00600

[30] C. J. Vaca-Rubio, L. Blanco, R. Pereira, and M. Caus, "Kolmogorov-arnold networks (kans) for time series analysis," 2024. [Online]. Available: https://arxiv.org/abs/2405.08790

[31] G. Charan, U. Demirhan, J. Morais, A. Behboodi, H. Pezeshki, and A. Alkhateeb, "Multi-modal beam prediction challenge 2022: Towards generalization," 2022.

[32] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, J. Morais, U. Demirhan, and N. Srinivas, "Deepsense 6G: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 122–128, 2023.

[33] K. Vuckovic, S. M. Hosseini, and N. Rahnavard, "Revisiting performance metrics for multimodal mmwave beam prediction using deep learning," *Proc. IEEE Military Communications Conference (MILCOM)*, 2024, to be published.