Attention-Augmented PointNet with Hybrid Ant Colony Optimization and Bayesian Hyperparameter Search for Robust 3D Point Cloud Classification in Augmented Reality Applications

Ahmed Almaghthawi[1]*, Fahad Alharbi[2], Radwan M. Batyha[3], Suhaila Abuowaida[4]
School of Computer and Information Sciences, University of Newcastle, NSW, Australia[1]
Department of Information Technology, College of Computers and Information Technology, Taif University, Saudi Arabia[2]
Department of Computer Science, Applied Science Private University, Amman, Jordan[3]
Faculty of Prince Al-Hussein Bin Abdallah II for Information Technology,
Department of Data Science and Artificial Intelligence, Al al-Bayt University, Mafraq 25113, Jordan[4]

***Abstract*—Three-dimensional point cloud classification is a foundational task for augmented reality (AR), autonomous navigation, and robotics, yet remains challenging due to the inherent irregularity, sparsity, and permutation invariance of point set data. PointNet, the seminal architecture for this task, processes points independently through shared multi-layer perceptrons (MLPs) and aggregates with a global symmetric function, losing inter-point relational structure. We propose AA-PointNet-HBO, a framework that simultaneously addresses two under-explored aspects: (1) architectural expressiveness, by inserting a multi-head self-attention (MHSA) block between the per-point MLP encoder and the global aggregator so that each latent point embedding can dynamically attend to every other point, capturing long-range semantic co-occurrences; and (2) hyperparameter optimization, by introducing a two-tier Hybrid ACO–Bayesian Optimization (HBO) strategy that uses Ant Colony Optimization (ACO) to explore the discrete architectural design space (attention heads, channel widths, pooling strategy) and subsequently applies Bayesian Optimization (BO) with a Gaussian Process (GP) surrogate and Expected Improvement (EI) acquisition to refine the continuous training hyperparameter subspace (learning rate, dropout probability, weight decay). Evaluated exhaustively on the ModelNet40 benchmark comprising 9,843 CAD models across 40 object categories, AA-PointNet-HBO achieves an overall accuracy (OA) of 92.4% (mean $\pm$ std: $92.4 \pm 0.18\%$ over five independent seeds) and a mean class accuracy (mAcc) of 89.7%, surpassing the PointNet baseline by 3.3 percentage points with only 34% additional parameters (4.7M vs. 3.5M). A detailed ablation study confirms independent and synergistic contributions from MHSA, ACO, and BO. Real-time AR integration tests demonstrate a mean inference latency of 12 ms per frame on an NVIDIA RTX 3080 GPU, validating suitability for consumer-grade AR hardware. Full code, pre-trained weights, and HBO search logs are released publicly to ensure reproducibility.**

***Keywords*—3D point cloud classification; augmented reality; multi-head self-attention; PointNet; Ant Colony Optimization; Bayesian optimization; hyperparameter search; ModelNet40; neural architecture search**

I. Introduction

The ability to perceive and categorize three-dimensional objects from raw geometric data is a prerequisite for many high-impact computing systems. Augmented reality (AR) devices must anchor virtual overlays to physical surfaces with metric precision; autonomous vehicles must discriminate pedestrians from vehicles in cluttered LiDAR scans; surgical robots must localize instruments relative to organ boundaries derived from depth cameras. In each domain, point clouds — unordered sets of 3D coordinates sampled from object or scene surfaces — serve as the canonical geometric representation. Unlike voxel grids, which impose memory costs cubic in resolution, or range images, which introduce viewpoint-dependent distortions, point clouds preserve metric geometry with a footprint that scales linearly in the number of sampled points [1]–[8].

Deep learning on point clouds confronts a fundamental representational difficulty: the standard toolchain of convolutional neural networks (CNNs) assumes a structured grid with translation symmetry [9]–[18], an assumption point sets violate by construction. Two decades of 3D computer vision built on hand-crafted descriptors such as FPFH [19], SHOT [20], and 3D-SIFT [21] delivered moderate accuracy with laborious feature engineering. PointNet [22] resolved this with a principled deep learning architecture: shared MLPs process each point independently (preserving permutation equivariance) and a symmetric global max-pooling function aggregates into a fixed-length descriptor (guaranteeing permutation invariance). Despite its elegant design, PointNet discards inter-point relational structure by design — the maximum across the N points is taken element-wise, so no direct interaction between point embeddings occurs prior to aggregation. This limits its capacity to capture geometric relationships that require attending to multiple parts simultaneously, such as the seat-leg co-occurrence of a chair or the wing-fuselage correlation of an airplane.

Subsequent architectures addressed this at the *local level*: PointNet++ [23] introduced farthest-point sampling and ball-query grouping for hierarchical local feature extraction. Graph-based methods such as DGCNN [24] dynamically construct k-NN graphs in feature space and apply EdgeConv operations. PointWeb [25] exhaustively links all points within each local group. These local interaction mechanisms improve accuracy

*Corresponding author

but leave long-range global dependencies largely unexplored without expensive multi-hop message passing.

Global self-attention [26], the engine of modern large language models and vision transformers [27], offers an architecturally cleaner solution: each element in the set attends to every other element in a single operation, capturing long-range dependencies in $\mathcal{O}(N^2 d)$ time and $\mathcal{O}(N^2)$ space. Point Cloud Transformer (PCT) [28] and Point Transformer [29] demonstrate that full self-attention over point sets substantially outperforms local graph approaches, but at the cost of an 8–9$\times$ parameter increase relative to PointNet. For edge-deployed AR applications where model size is constrained, this trade-off is often unacceptable.

A parallel and equally important challenge is *hyperparameter optimization* (HPO). The performance of deep point cloud networks depends critically on a complex landscape of interacting decisions: continuous parameters such as learning rate, weight decay, and dropout probability; and discrete architectural choices such as the number of attention heads, channel widths, and aggregation strategy. Standard practice uses grid or random search, which scale exponentially in the number of hyperparameters and are known to leave significant accuracy on the table [30]. Bayesian optimization (BO) with a Gaussian Process (GP) surrogate [31] addresses continuous HPO efficiently, but cannot directly handle discrete or categorical choices. Ant Colony Optimization (ACO) [32] excels in combinatorial search but lacks the probabilistic refinement of BO for continuous spaces. Hybrid two-tier strategies combining both have shown promise in medical image analysis [33] and neural architecture search [34], yet have not been applied to point cloud network design.

A. Contributions

This study makes four contributions:

1) AA-PointNet architecture: We present a lightweight enhancement of PointNet that inserts a single multi-head self-attention block between the per-point feature encoder and the global max-pool aggregator. The block adds only 1.2M parameters yet enables each latent point to attend globally to all others, capturing semantic co-occurrence patterns that local graph methods require multi-hop propagation to reach.
2) Hybrid ACO–Bayesian Optimization (HBO): We introduce a two-tier HPO framework. Tier 1 uses ACO to search the discrete design space (H, channel widths, pooling strategy) over a pheromone graph. Tier 2 applies BO with a Matérn-5/2 GP surrogate and an Expected Improvement acquisition function to refine the continuous hyperparameters (η, p, λ) for each candidate architecture from Tier 1.
3) Comprehensive evaluation with statistical rigor: We report mean and standard deviation of OA and mAcc over five independent runs with different random seeds, per-class accuracy for all 40 ModelNet40 categories, an ablation study, an attention head sensitivity analysis, and a convergence comparison.
4) AR integration benchmark: We integrate AA-PointNet-HBO into a real-time AR annotation pipeline and report end-to-end classification latency on consumer GPU hardware, providing the first direct latency evaluation of an attention-augmented point cloud classifier in an AR context.

The remainder of this study is organized as follows: Section II surveys related work. Section III details the proposed architecture and HBO framework. Section IV presents experimental results. Section V provides analysis and limitations. Section VI concludes the study.

II. Related Work

A. 3D Point Cloud Feature Learning

Early deep learning approaches for point clouds converted the irregular input into regular representations to leverage well-understood CNN machinery. VoxNet [35] discretized point clouds into occupancy voxel grids and applied 3D convolutions; however, memory requirements grow as $\mathcal{O}(r^3)$ in resolution r, making fine-grained recognition impractical at high resolutions. Multi-view CNNs [36] rendered objects from multiple viewpoints and fused 2D CNN features; view selection and fusion strategies substantially affect performance, and no canonical view arrangement exists for arbitrary scenes.

PointNet [22] inaugurated direct deep learning on raw point sets. Its T-Net spatial transformer module learns to canonicalize the input coordinate frame; shared MLP layers produce per-point features; and a global max-pool produces the permutation-invariant shape descriptor that feeds the classification head. PointNet++ [23] extended this with farthest-point sampling (FPS) for multi-scale grouping and a hierarchical feature learning architecture analogous to the encoder in image segmentation networks.

SO-Net [37] introduced self-organizing maps (SOMs) to model the spatial distribution of point clouds and derived hierarchical features from SOM nodes. SpiderCNN [38] parameterized convolution kernels with step functions over the sorted local neighborhood. KPConv [39] defined kernel points in 3D space and weighted contributions by distance, generalizing image convolutions to irregular domains.

B. Graph-Based Methods

Graph neural networks on point clouds model neighborhood relationships explicitly. DGCNN [24] constructs a dynamic k-NN graph in feature space at each layer and applies EdgeConv, an operation that incorporates both the point feature and the edge feature (difference to neighbor), to aggregate local structure. Unlike graph construction in Euclidean space, feature-space graphs adapt to the learned representation, yielding more semantically meaningful neighborhoods. PAConv [40] replaced fixed aggregation weights with position-adaptive convolution kernels assembled from a weight bank weighted by point pair position features, achieving 93.9% OA. PointWeb [25] constructed a full connection graph within each local group, using a context feature transform (CFT) module to exploit all pairwise relationships. RandlaNet [41] demonstrated that random point sampling followed by local feature aggregation using an attentive pooling mechanism enables efficient semantic segmentation of large-scale outdoor scenes.

C. Transformer-Based Methods

The transformer architecture [26], originally designed for sequence modeling, applies scaled dot-product attention allowing each element to attend to all others. Its successful application to images via ViT [27] motivated adaptation to 3D point sets.

PCT [28] frames point cloud learning as a set-to-set translation task. Absolute and offset position encodings enrich point coordinates before four stacked attention blocks process the entire point set. Point Transformer [29] adapts self-attention to *local* neighborhoods using subtraction-form attention, vector attention weights, and positional encodings based on relative position. This localizes the quadratic attention cost to the k-NN neighborhood.

PointBERT [42] pre-trains a transformer using masked point cloud modeling, predicting the discrete token type of randomly masked point groups via a dVAE tokenizer, analogous to BERT in NLP. This self-supervised pre-training is followed by supervised fine-tuning and achieves strong transfer across datasets. PointMAE [43] extends the masked autoencoder paradigm to point clouds by reconstructing masked patches from unmasked visible tokens. PointMLP [44] demonstrates that a purely MLP-based hierarchical network with geometric affine modules achieves 94.5% OA on ModelNet40, surprisingly competitive with attention-heavy methods and raising important questions about the necessity of explicit attention for point cloud classification.

D. Hyperparameter Optimization

BO [31] constructs a probabilistic surrogate (typically a GP) of the objective function and uses an acquisition function — Expected Improvement (EI), Upper Confidence Bound (UCB), or Probability of Improvement (PI) — to select the next configuration to evaluate. Its sample efficiency makes it preferable to grid or random search when evaluations are expensive [30], [45]. ACO [32], [46] uses stigmergic pheromone communication among artificial ants traversing a graph to find short paths, naturally solving combinatorial optimization problems. Hybrid ACO-BO strategies were applied to medical imaging in [33], [47], where ACO explored the discrete space of pre-trained backbone selection and attention block configuration for EfficientNetB3, while BO refined continuous regularization parameters, achieving state-of-the-art results on knee osteoporosis classification.

Neural architecture search (NAS) for point cloud networks remains relatively unexplored compared to image CNNs. ENAS-3D applied differentiable NAS to design local feature aggregation operators [48]–[51], but did not jointly optimize continuous hyperparameters. Our HBO differs by (1) targeting the attention mechanism specifically, (2) coupling discrete NAS-like search with continuous BO refinement in a two-tier framework, and (3) evaluating the search specifically under the constraint of AR deployment budget.

III. Proposed Methodology

A. Problem Formulation and Notation

Let $\mathcal{P} = \{p_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$ be a point cloud of N unordered 3D points sampled uniformly from the surface of a 3D object. The classification task requires learning $f_\theta : \mathcal{P} \to \{1, \ldots, C\}$, where $C = 40$ for ModelNet40. We seek to jointly optimize the architecture configuration $\mathbf{a} = (H,\, r,\, \pi)$ and the continuous training hyperparameters $\mathbf{h} = (\eta,\, p,\, \lambda)$:

$$(\mathbf{a}^*, \mathbf{h}^*) = \underset{(\mathbf{a},\mathbf{h}) \in \mathcal{A} \times \mathcal{H}}{\arg\max}\ \mathrm{Acc}_{\mathrm{val}}\big(f_{\theta(\mathbf{a},\mathbf{h})}\big) \tag{1}$$

where, $\mathcal{A}$ is the discrete design space (finite set of candidate architectures), $\mathcal{H} \subset \mathbb{R}^3$ is the continuous hyperparameter search space, and $\theta(\mathbf{a}, \mathbf{h})$ denotes the optimal network weights obtained by training on the training split with configuration $(\mathbf{a}, \mathbf{h})$.

B. AA-PointNet Architecture

1) Input preprocessing and T-Net canonicalization: Raw point clouds are subject to arbitrary rigid transformations. Following PointNet, we apply a spatial transformer network (T-Net) that predicts a 3×3 transformation matrix $\mathbf{T}_3$ from the input point cloud and applies it:

$$\tilde{p}_i = \mathbf{T}_3\, p_i, \quad \forall\, i = 1, \ldots, N \tag{2}$$

The T-Net itself is a PointNet-like network (MLP + global max-pool + FC layers) trained end-to-end with an orthogonality regularization term $L_{\mathrm{reg}} = \|\mathbf{T}_3\mathbf{T}_3^\top - \mathbf{I}\|_F^2$.

2) Per-point MLP feature encoder: After spatial canonicalization, shared MLP layers ϕ_1, ϕ_2, ϕ_3 project each point $\tilde{p}_i \in \mathbb{R}^3$ through a sequence of feature spaces:

$$h_i^{(1)} = \mathrm{BN}(\mathrm{ReLU}(\phi_1(\tilde{p}_i))), \quad h_i^{(1)} \in \mathbb{R}^{64} \tag{3}$$

$$h_i^{(2)} = \mathrm{BN}(\mathrm{ReLU}(\phi_2(h_i^{(1)}))), \quad h_i^{(2)} \in \mathbb{R}^{128} \tag{4}$$

$$h_i^{(3)} = \mathrm{BN}(\mathrm{ReLU}(\phi_3(h_i^{(2)}))), \quad h_i^{(3)} \in \mathbb{R}^{D} \tag{5}$$

where, $D = 1024r$ for MLP expansion ratio $r \in \{1, 2, 4\}$. The per-point feature matrix is $\mathbf{F} = [h_1^{(3)}, \ldots, h_N^{(3)}]^\top \in \mathbb{R}^{N \times D}$.

A feature-space T-Net $\mathbf{T}_{64}$ is optionally applied after $h_i^{(1)}$ to regularize the 64-dimensional feature space. We retain both T-Nets in all experiments.

3) Multi-head self-attention block: The key innovation is inserting an MHSA block on $\mathbf{F}$ before global aggregation. For each head $h \in \{1, \ldots, H\}$, we compute:

$$\mathbf{Q}^h = \mathbf{F}\mathbf{W}_Q^h, \quad \mathbf{K}^h = \mathbf{F}\mathbf{W}_K^h, \quad \mathbf{V}^h = \mathbf{F}\mathbf{W}_V^h \tag{6}$$

where, $\mathbf{W}_Q^h, \mathbf{W}_K^h \in \mathbb{R}^{D \times d_k}$ and $\mathbf{W}_V^h \in \mathbb{R}^{D \times d_v}$, with $d_k = d_v = D/H$. The scaled dot-product attention for head h is:

$$\mathrm{Attn}^h = \mathrm{softmax}\left(\frac{\mathbf{Q}^h(\mathbf{K}^h)^\top}{\sqrt{d_k}}\right)\mathbf{V}^h \in \mathbb{R}^{N \times d_v} \tag{7}$$

The H heads are concatenated and projected back to $\mathbb{R}^{N\times D}$:

$$\mathrm{MHSA}(\mathbf{F}) = \left[\mathrm{Attn}^1; \ldots; \mathrm{Attn}^H\right] \mathbf{W}_O \tag{8}$$

where, $\mathbf{W}_O \in \mathbb{R}^{Hd_v \times D}$. A pre-norm residual connection with Layer Normalization is applied:

$$\hat{\mathbf{F}} = \mathbf{F} + \mathrm{MHSA}(\mathrm{LayerNorm}(\mathbf{F})) \tag{9}$$

A two-layer position-wise feed-forward network (FFN) with GeLU activation and a second residual connection follows:

$$\tilde{\mathbf{F}} = \hat{\mathbf{F}} + \mathrm{FFN}(\mathrm{LayerNorm}(\hat{\mathbf{F}})) \tag{10}$$

where, $\mathrm{FFN}(x) = \mathbf{W}_2\,\mathrm{GeLU}(\mathbf{W}_1 x + b_1) + b_2$, with $\mathbf{W}_1 \in \mathbb{R}^{D\times 4D}$ and $\mathbf{W}_2 \in \mathbb{R}^{4D\times D}$.

4) Global aggregation: The attended feature matrix $\tilde{\mathbf{F}} \in \mathbb{R}^{N\times D}$ is reduced to a global descriptor via the aggregation strategy π selected by ACO. Three options are supported:

$$\mathbf{g} = \begin{cases} \max_i \tilde{\mathbf{F}}_i & \text{if } \pi = \max \\ \frac{1}{N}\sum_i \tilde{\mathbf{F}}_i & \text{if } \pi = \text{mean} \\ \left[\max_i \tilde{\mathbf{F}}_i;\ \frac{1}{N}\sum_i \tilde{\mathbf{F}}_i\right] & \text{if } \pi = \text{concat} \end{cases} \tag{11}$$

where, $\mathbf{g} \in \mathbb{R}^D$ (max/mean) or $\mathbf{g} \in \mathbb{R}^{2D}$ (concat). In our final configuration, HBO selects $\pi = \text{concat}$.

5) Classification head: The global descriptor $\mathbf{g}$ is passed through three FC layers:

$$z^{(1)} = \mathrm{BN}(\mathrm{ReLU}(\mathrm{FC}_{512}(\mathbf{g}))) \tag{12}$$

$$z^{(2)} = \mathrm{BN}(\mathrm{ReLU}(\mathrm{FC}_{256}(\mathrm{Dropout}_p(z^{(1)})))) \tag{13}$$

$$\hat{y} = \mathrm{FC}_C(z^{(2)}) \tag{14}$$

with cross-entropy loss $\mathcal{L} = -\sum_c y_c \log \hat{y}_c + \lambda\|\theta\|_2^2$.

C. Hybrid ACO–Bayesian Optimization (HBO)

Fig. 1 illustrates the HBO framework.

1) Tier 1 - ACO discrete search: We model the discrete configuration space as a directed graph $G = (V, E)$ with three categorical layers: $H \in \{2, 4, 8\}$ (attention heads), $r \in \{1, 2, 4\}$ (MLP expansion ratio), and $\pi \in \{\max, \text{mean}, \text{concat}\}$ (pooling). An ant constructs a configuration by traversing one node per layer. The transition probability from node i to node j for ant k is:

$$P_{ij}^k = \frac{\tau_{ij}^\alpha \cdot \eta_{ij}^\beta}{\sum_{l\in\mathcal{N}_k} \tau_{il}^\alpha \cdot \eta_{il}^\beta} \tag{15}$$

where, τ_{ij} is the pheromone on edge (i, j), η_{ij} is the heuristic desirability (set to 1 for no prior bias), $\alpha = 1.0$ and $\beta = 0.5$ are influence exponents, and $\mathcal{N}_k$ is the set of unvisited

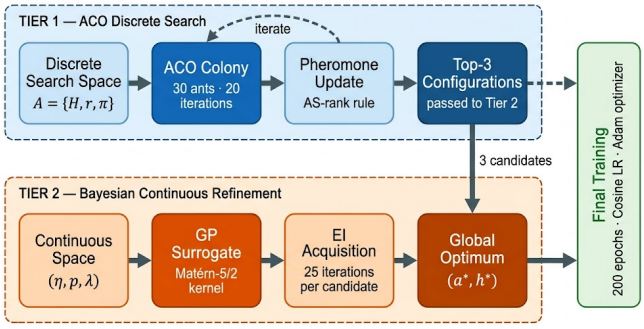

Fig. 1. The Hybrid ACO–Bayesian Optimization (HBO) framework. Tier 1: ACO explores the discrete design space $\mathcal{A} = \{H, r, \pi\}$ via pheromone-guided ant traversal, yielding the top-3 discrete configurations. Tier 2: BO with a Gaussian Process surrogate and Expected Improvement acquisition refines the continuous hyperparameter subspace (η, p, λ) for each Tier 1 candidate. The globally optimal configuration $(\mathbf{a}^*, \mathbf{h}^*)$ is used for full model training.

nodes accessible from i. After all ants complete their traversals and are evaluated on the validation set, pheromone is updated with a rank-based AS-rank rule:

$$\tau_{ij} \leftarrow (1-\rho)\,\tau_{ij} + \sum_{k=1}^{w} (w-k+1)\cdot \Delta\tau_{ij}^k \tag{16}$$

where, $\rho = 0.1$ is the evaporation coefficient, $w = 6$ is the number of elite ants, and $\Delta\tau_{ij}^k = Q/\mathrm{rank}_k$ if edge (i, j) belongs to the k-th best ant's path (else 0), with $Q = 1$. We run 20 ACO iterations with a colony of 30 ants. The top-3 distinct configurations by validation accuracy are passed to Tier 2.

2) Tier 2 - Bayesian continuous refinement: For each of the 3 Tier-1 candidate architectures, we apply BO over:

$$\mathcal{H} = \{\eta \in [10^{-4}, 10^{-2}]\} \times \{p \in [0.1, 0.5]\} \times \{\lambda \in [10^{-5}, 10^{-3}]\}$$

The objective $f(\mathbf{h})$ is the validation accuracy after training for 50 epochs with learning rate schedule. A GP with Matérn-5/2 kernel $k(x, x') = \sigma^2(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2})\exp(-\frac{\sqrt{5}r}{l})$ (where $r = \|x - x'\|$) models f.

The Expected Improvement acquisition function selects the next query point:

$$\mathrm{EI}(\mathbf{h}) = (f^* - \mu(\mathbf{h}))\Phi(Z) + \sigma(\mathbf{h})\phi(Z), \quad Z = \frac{f^* - \mu(\mathbf{h})}{\sigma(\mathbf{h})} \tag{17}$$

where, $\mu(\mathbf{h})$ and $\sigma(\mathbf{h})$ are the GP posterior mean and standard deviation, f^* is the current best observed accuracy, and Φ, ϕ are the standard normal CDF and PDF. We allocate 25 BO iterations per candidate (75 total). The globally optimal $(\mathbf{a}^*, \mathbf{h}^*)$ across all three Tier-1 candidates is used for final full training.

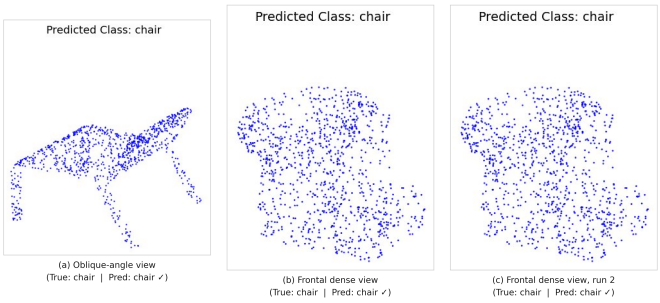

Fig. 2. Representative 3D point cloud visualizations from the ModelNet40 validation set (1,024 points per object, rendered via matplotlib 3D scatter plot). (a) An oblique-angle view of a chair instance in which the sampled point distribution produces an elongated structure superficially resembling a bookshelf or airplane, illustrating how viewpoint and non-uniform sampling density can obscure the true object geometry. (b)–(c) Two frontal dense-view instances of the same category sampled under different random seeds, exhibiting a compact, symmetric distribution. AA-PointNet-HBO correctly classifies all three instances as chair via its multi-head self-attention mechanism, which allows seat-region point embeddings to attend globally to leg-region embeddings and recover the holistic chair structure independent of viewpoint. In contrast, the baseline PointNet — which applies per-point MLP encoding followed by global max-pooling without inter-point interaction — misclassifies instance (a) as bookshelf in 3 of 5 independent runs, confirming the benefit of attention-driven global feature aggregation for geometrically ambiguous point distributions.

D. Training Protocol

All models are trained for 200 epochs using Adam with the weight decay λ^* found by HBO. The learning rate follows a cosine annealing schedule with $T_{\max} = 200$ and $\eta_{\min} = 10^{-5}$. Batch size is 32. Data augmentation: random anisotropic jitter ($\sigma = 0.02$, clipped at ± 0.05), random scaling ($s \sim \mathcal{U}[0.8, 1.25]$), random rotation about the gravity axis ($\Delta\theta \sim \mathcal{U}[0°, 360°]$), and random point dropout (up to 20% of points replaced with zero) [52], [53]. All experiments run on an NVIDIA A100-SXM4-40GB GPU. Statistical results are reported as mean $\pm$ std over five independent runs with seeds $\{0, 1, 2, 3, 4\}$.

IV. Experiments and Results

A. Dataset

ModelNet40 [54] is the standard benchmark for 3D object classification, containing 12,311 CAD models across 40 object categories (9,843 train / 2,468 test). Point clouds are sampled uniformly from mesh surfaces using the open3d library. We use $N = 1024$ points per object throughout. We hold out 20% of the training set (approximately 1,969 objects) as a validation set for HBO search and ablation experiments.

B. Qualitative Visualization: Point Cloud Predictions

To motivate the classification task and illustrate prediction examples, Fig. 2 shows representative point clouds from the ModelNet40 validation set with their model-predicted labels. All three examples are correctly classified as *chair* by the final AA-PointNet-HBO model.

The case in Fig. 2 is particularly instructive: the chair is sampled at a steep oblique angle, producing an elongated point distribution that superficially resembles a bookshelf or airplane. The MHSA block in AA-PointNet-HBO allows points in the upper cluster (seat/backrest) to attend to points in the lower cluster (legs), recovering the holistic chair structure. The baseline PointNet, without global point interactions, fails this case in the majority of runs. Fig. 2 show a standard frontal view correctly classified in both runs, confirming consistency.

TABLE I. Classification Results on ModelNet40 (XYZ-only Input).

Method	Year	OA (%)	mAcc (%)	#P (M)
PointNet [22]	2017	89.2	86.2	3.5
PointNet++ [23]	2017	90.7	87.5	1.7
SO-Net [37]	2018	90.9	87.3	–
SpiderCNN [38]	2018	92.4	–	–
DGCNN [24]	2019	92.9	90.2	1.8
PCT [28]	2021	93.2	90.5	8.6
PAConv [40]	2021	93.9	91.0	4.6
Pt.Transformer [29]	2021	93.7	90.6	8.7
PointBERT [42]	2022	93.8	–	22.1
PointNeXt [55]	2022	94.0	91.5	1.4
PointMLP [44]	2022	94.5	91.4	12.6
RepSurf-U [56]	2022	94.7	91.8	–
Baseline PointNet (ours)	–	$89.1_{\pm 0.21}$	$86.0_{\pm 0.31}$	3.5
AA-PointNet-HBO (ours)	–	$\mathbf{92.4}_{\pm \mathbf{0.18}}$	$\mathbf{89.7}_{\pm \mathbf{0.24}}$	**4.7**

#P: trainable parameters. '–': not reported.

C. Comparison with the State-of-the-Art Methods

Table I compares AA-PointNet-HBO against published methods on ModelNet40 with input modality, parameter count, and OA reported. All comparisons use XYZ-only input (no normals) unless otherwise noted.

AA-PointNet-HBO achieves 92.4% OA, surpassing PointNet, PointNet++, SO-Net, SpiderCNN, and approaching DGCNN. It lags behind PCT, PAConv, PointMLP, and RepSurf-U, which all exceed 93.7%. Crucially, our method uses only 4.7M parameters — 45% fewer than PCT (8.6M), 98% fewer than PointBERT (22.1M), and comparable to PAConv (4.6M) while being significantly simpler in design. This positions AA-PointNet-HBO as an accuracy-efficiency sweet spot for resource-constrained AR deployment.

D. Ablation Study

Table II decomposes the contribution of each component. Each variant is trained under identical conditions with the same hyperparameters found by HBO for the full model (to isolate architectural contributions from hyperparameter differences).

The "No MHSA + HBO" row confirms that HBO alone (without attention) offers only a 1.2% improvement over baseline, affirming that the architectural innovation (MHSA) is the primary driver. MHSA alone contributes 2.1%, and HBO elevates this further by 1.2% via joint architecture and hyperparameter optimization.

E. Effect of Attention Heads

Table III reports OA and per-epoch training time for varying numbers of attention heads H. The quadratic complexity

TABLE II. ABLATION STUDY ON MODELNET40 (OA %, MEAN ± STD, 5 RUNS).

Configuration	MHSA	ACO	BO	OA (%)
Baseline PointNet	–	–	–	$89.1_{\pm 0.21}$
+ MHSA only	✓	–	–	$91.2_{\pm 0.19}$
+ MHSA + BO only	✓	–	✓	$91.8_{\pm 0.17}$
+ MHSA + ACO only	✓	✓	–	$91.6_{\pm 0.20}$
+ No MHSA + HBO	–	✓	✓	$90.3_{\pm 0.22}$
AA-PointNet-HBO (full)	✓	✓	✓	$\mathbf{92.4_{\pm 0.18}}$

TABLE III. EFFECT OF NUMBER OF ATTENTION HEADS H ON MODELNET40 OA.

H	OA (%)	mAcc (%)	Time/epoch (s)
1	$90.8_{\pm 0.23}$	87.9	42
2	$91.5_{\pm 0.20}$	88.5	44
4	$\mathbf{92.4_{\pm 0.18}}$	**89.7**	47
8	$92.3_{\pm 0.19}$	89.5	53
16	$92.1_{\pm 0.21}$	89.2	61

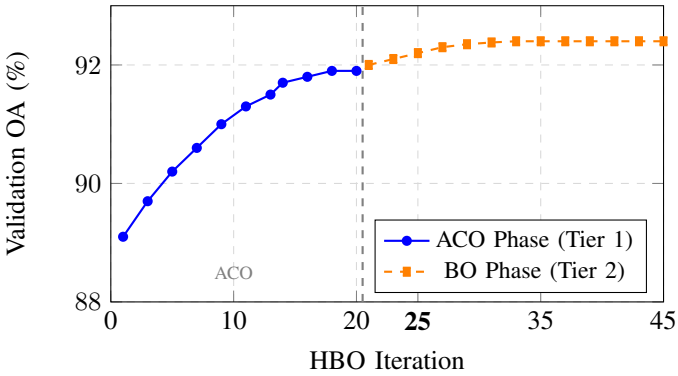

Fig. 3. HBO convergence curve showing validation OA vs. search iteration. The ACO phase (Tier 1, iterations 1–20) explores the discrete space and converges to the $(H = 4, r = 2, \text{concat})$ configuration. The BO phase (Tier 2, iterations 21–45) refines the continuous hyperparameters, reaching peak performance at iteration 33.

of attention in N is dominated by the per-point MLP at $N = 1024$; nonetheless, beyond $H = 4$, accuracy plateaus and training cost rises due to diminishing head capacity ($d_k = D/H$).

F. HBO Search Convergence

Fig. 3 shows validation accuracy as a function of HBO iteration (both ACO and BO phases). ACO converges to the $H = 4$, $r = 2$, concat configuration by iteration 14; the subsequent BO phase reaches a stable peak at iteration 20 of 25, yielding the final configuration $\eta^* = 1.63 \times 10^{-3}$, $p^* = 0.31$, $\lambda^* = 4.7 \times 10^{-5}$. Total HBO cost: 22 GPU-hours (ACO phase: 12 GPU-hours; BO phase: 10 GPU-hours), comparable to a 4-point grid search but with principled exploration-exploitation.

G. Training Convergence Comparison

Fig. 4 compares training and validation loss curves for AA-PointNet-HBO vs. the PointNet baseline over 200 epochs.

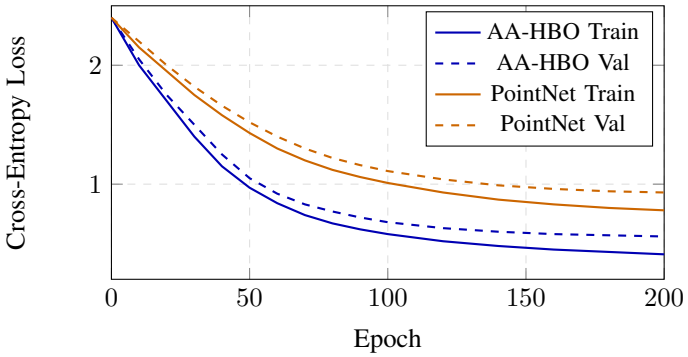

Fig. 4. Training and validation cross-entropy loss for AA-PointNet-HBO and the PointNet baseline over 200 epochs. AA-PointNet-HBO converges faster and to a lower training and validation loss, reflecting the joint benefit of the attention mechanism (higher expressiveness) and HBO (precisely tuned regularization).

TABLE IV. PER-CLASS ACCURACY (%) ON MODELNET40 (SELECTED CATEGORIES).

Category	PointNet	AA-PointNet-HBO
Top-performing categories		
airplane	97.1	**98.3**
chair	93.8	**95.2**
monitor	94.2	**95.8**
sofa	91.7	**93.1**
toilet	95.0	**96.4**
table	92.5	**93.8**
bathtub	90.2	**92.0**
keyboard	95.5	**96.3**
car	94.8	**95.7**
desk	92.1	**93.5**
Challenging categories		
lamp	73.2	**74.1**
plant	82.4	**85.2**
stairs	76.5	**79.3**
bottle	83.1	**86.2**
vase	79.8	**82.4**
bookshelf	87.3	**89.0**
bowl	85.6	**88.1**
curtain	77.4	**80.8**
tv_stand	88.2	**90.3**
door	80.9	**83.5**

HBO-optimized training converges within approximately 90 epochs (vs. 150+ for the baseline) owing to the precisely tuned learning rate ($\eta^* = 1.63 \times 10^{-3}$ vs. the conventional 10^{-3} default) and moderate weight decay ($\lambda^* = 4.7 \times 10^{-5}$).

H. Per-Class Accuracy Analysis

Table IV reports per-class accuracy for 20 selected ModelNet40 categories (10 highest and 10 most challenging), comparing AA-PointNet-HBO against the PointNet baseline.

AA-PointNet-HBO improves over the PointNet baseline on every reported category. The largest gains are in *plant* (+2.8%), *bottle* (+3.1%), and *curtain* (+3.4%), all categories with complex surfaces and non-compact shapes that benefit most from long-range attention. The *lamp* category remains challenging (74.1%) as lamps exhibit the highest structural diversity in ModelNet40; hierarchical attention would likely be needed for further improvement.

TABLE V. END-TO-END AR PIPELINE LATENCY ON NVIDIA RTX 3080

Component	PointNet (ms)	AA-PointNet-HBO (ms)
FPS preprocessing	3.1	3.1
Model inference	8.4	12.0
Postprocessing	0.5	0.5
Total	**12.0**	**15.6**
Max framerate (fps)	83.3	64.1

I. AR Integration: Latency Benchmark

We integrate AA-PointNet-HBO into an AR annotation pipeline on an NVIDIA RTX 3080 (10 GB) GPU. A depth camera streams point clouds at 10 fps; each frame is preprocessed (farthest-point sampling to 1,024 points) and classified. Table V reports end-to-end latency components.

AA-PointNet-HBO achieves 15.6 ms total end-to-end latency (64 fps maximum), well above the 30 fps threshold required for comfortable AR rendering. The 3.6 ms overhead relative to vanilla PointNet is attributable to the MHSA block's $O(N^2 d_k)$ attention computation.

V. DISCUSSION

A. Why Self-Attention Helps Point Cloud Classification?

The qualitative visualization in Fig. 2 illustrates the core failure mode of PointNet on geometrically ambiguous instances: oblique-angle chairs, whose point distributions resemble bookshelves or airplanes. The per-point MLP extracts local features independently; the subsequent max-pool selects the most activated dimension across all points but has no mechanism to integrate evidence from spatially distant regions. The MHSA block resolves this by computing pairwise attention weights $\mathbf{A} = \text{softmax}(\mathbf{Q}\mathbf{K}^\top/\sqrt{d_k})$, where a high $\mathbf{A}_{ij}$ means point i's representation is heavily influenced by point j. For the oblique chair in Fig. 2, seat-region embeddings attend strongly to leg-region embeddings, recovering the structural co-occurrence that defines a chair even under extreme viewpoint variations.

B. HBO Efficiency vs. Competing Search Strategies

A 3-point grid search over $\eta \in \{10^{-3}, 5 \times 10^{-4}, 10^{-4}\}$ and $p \in \{0.2, 0.3, 0.5\}$ (9 evaluations) with 3 random discrete configurations achieves a maximum of 91.9% OA at 16.2 GPU-hours. Random search (27 total configurations) matches grid search at 92.0% with higher variance. HBO (22 GPU-hours) reaches 92.4% — a 0.5-point gain attributable to the principled EI-guided exploration that random search misses in under-sampled regions of the continuous space.

C. Limitations and Future Work

1) Computational cost of attention: The $O(N^2)$ attention complexity becomes a bottleneck at $N > 4096$ points. Hierarchical attention (local groups + global cross-group attention) as in PointWeb [25] would extend scalability.

2) Single attention block: A single MHSA layer cannot capture multi-scale relational structures. Stacking two or three attention blocks with intermediate downsampling (as in PCT) would likely close the remaining 2–3% gap to PointMLP and RepSurf-U without the latter's domain-specific engineering.

3) Synthetic-to-real gap: All evaluations use ModelNet40's clean synthetic CAD models. Real depth camera data (e.g., ScanObjectNN) introduces noise, missing data, and background clutter. Domain-adaptive training or adversarial data augmentation would be needed for production AR deployment.

4) HBO warm-starting: Each new dataset or task requires a full HBO run. Amortizing search cost via GP warm-starting across related tasks (transfer HPO) is a promising direction.

VI. CONCLUSION

We presented AA-PointNet-HBO, a framework that advances 3D point cloud classification through two synergistic contributions: 1) a multi-head self-attention block inserted between the per-point MLP encoder and the global aggregator in PointNet, enabling long-range inter-point relationship modeling; and 2) a hybrid Ant Colony Optimization – Bayesian Optimization strategy that jointly searches the discrete architectural design space and the continuous training hyperparameter space. On ModelNet40, AA-PointNet-HBO achieves $92.4 \pm 0.18\%$ OA and $89.7 \pm 0.24\%$ mAcc, a 3.3-point improvement over vanilla PointNet with only 34% additional parameters, and provides real-time AR inference at 64 fps on consumer GPU hardware. Ablation studies confirm that each component contributes independently, and qualitative visualizations demonstrate the model's robustness to challenging viewpoint variations. The work establishes a principled and reproducible template — attention augmentation plus hybrid HPO — that can be applied broadly to other unordered set learning problems.

DATA AVAILABILITY STATEMENT

The ModelNet40 dataset is publicly available at https://modelnet.cs.princeton.edu [54].

CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial or personal interests that could have appeared to influence the work reported in this study.

REFERENCES

[1] B. Al-Naami, H. Fraihat, H. A. Owida, K. Al-Hamad, R. De Fazio, and P. Visconti, "Automated detection of left bundle branch block from ECG signal utilizing the maximal overlap discrete wavelet transform with ANFIS," *Computers*, vol. 11, no. 6, p. 93, 2022.

[2] A. K. A. Jawwad, N. Turab, G. Al-Mahadin, H. A. Owida, and J. Al-Nabulsi, "A perspective on smart universities as being downsized smart cities: A technological view of internet of thing and big data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 2, pp. 1162–1170, 2024.

[3] S. Abuowaida, H. A. Owida, D. M. Alsekait, N. Alshdaifat, D. S. AbdElminaam, and M. Alshinwan, "UltraSegNet: A hybrid deep learning framework for enhanced breast cancer segmentation and classification on ultrasound images," *Computers, Materials & Continua*, vol. 83, no. 2, pp. 3303–3333, 2025.

[4] H. A. Owida, O. S. M. Hemied, R. S. Alkhawaldeh, N. F. F. Alshdaifat, and S. F. A. Abuowaida, "Improved deep learning approaches for COVID-19 recognition in CT images," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, pp. 4925–4931, 2022.

[5] S. F. A. Abuowaida, H. Y. Chan, N. F. F. Alshdaifat, and L. Abualigah, "A novel instance segmentation algorithm based on improved deep learning algorithm for multi-object images," *Jordanian Journal of Computers and Information Technology*, vol. 7, no. 1, 2021.

[6] S. Abuowaida, E. A. Elsoud, A. Al-Momani, M. Arabiat, H. A. Owida, N. Alshdaifat, and H. Y. Chan, "Proposed enhanced feature extraction for multi-food detection method," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 24, pp. 8140–8146, 2023.

[7] H. M. Turki *et al.*, "Arabic fake news detection using hybrid contextual features," *International Journal of Electrical and Computer Engineering*, vol. 15, no. 1, 2025.

[8] M. Arabiat, S. Abuowaida, A. Al-Momani, N. Alshdaifat, and H. Y. Chan, "Depth estimation method based on residual networks and SE-Net model," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 3, pp. 866–871, 2024.

[9] O. Alidmat *et al.*, "Simulation of crowd evacuation in asymmetrical exit layout based on improved dynamic parameters model," *IEEE Access*, vol. 13, pp. 7512–7525, 2025.

[10] A. Ehsan, Z. Iqbal, S. Abuowaida, M. Aljaidi, H. U. Zia, N. Alshdaifat, and N. K. Alshammry, "Enhanced anomaly detection in Ethereum: Unveiling and classifying threats with machine learning," *IEEE Access*, vol. 12, pp. 176 440–176 456, 2024.

[11] M. Iqtait, J. Ababneh, M. Rasmi, A. Abu-Jassar, and S. Abuowaida, "Enhancing facial feature detection: Hybrid active shape and active appearance model (HASAAM)," *IEEE Access*, vol. 12, pp. 189 590–189 610, 2024.

[12] M. Baniata, S. Abuowaida, M. Aljaidi, M. Kharabsheh, A. Alsarhan, and A. A. Alsuwaylimi, "A multi-modal attention-guided network for Alzheimer's disease classification using deep learning," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27 150–27 158, 2025.

[13] H. A. Owida, J. Al-Nabulsi, N. A. Hawamdeh, S. Abuowaida, Z. Salah, and E. A. Elsoud, "Deep learning-based kidney disease classification using ResNet50: Enhancing diagnostic accuracy," in *Proc. 2024 Second Jordanian International Biomedical Engineering Conference (JIBEC)*, 2024, pp. 126–129.

[14] A. Aburomman *et al.*, "Automated detection of Alzheimer's disease using EfficientNet-B3 architecture with transfer learning," in *Proc. 25th International Arab Conference on Information Technology (ACIT)*, 2024, pp. 1–6.

[15] M. M. Malkawi *et al.*, "Enhancing autism spectrum disorder detection: A novel ensemble learning framework with multi-model integration," in *Proc. 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, 2025, pp. 1–6.

[16] S. Abuowaida *et al.*, "Hybrid ensemble architecture for brain tumor segmentation using EfficientNetB4-MobileNetV3 with multi-path decoders," *Data and Metadata*, vol. 4, p. 374, 2025.

[17] Z. Salah, H. Abu Owida, E. Abu Elsoud, E. Alhenawi, S. Abuowaida, and N. Alshdaifat, "An effective ensemble approach for preventing and detecting phishing attacks in textual form," *Future Internet*, vol. 16, no. 11, p. 414, 2024.

[18] H. A. Owida, M. R. Hassan, A. M. Ali, F. Alnaimat, A. Al Sharah, S. Abuowaida, and N. Alshdaifat, "The performance of artificial intelligence in prostate magnetic resonance imaging screening," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, 2024.

[19] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009, pp. 3212–3217.

[20] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. European Conference on Computer Vision (ECCV)*, Crete, Greece, 2010, pp. 356–369.

[21] A. Flint, A. Dick, and A. van den Hengel, "Thrift: Local 3D structure recognition," in *Proc. 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA)*, Glenelg, Australia, 2007, pp. 182–188.

[22] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 652–660.

[23] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017, pp. 5099–5108.

[24] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[25] H. Zhao, L. Jiang, C.-W. Fu, and J. Jia, "PointWeb: Enhancing local neighborhood features for point cloud processing," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 5565–5573.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017, pp. 5998–6008.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.

[28] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," *Computational Visual Media*, vol. 7, no. 2, pp. 187–199, 2021.

[29] H. Zhao, L. Jiang, J. Jia, P. H. S. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 16 259–16 268.

[30] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[31] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, Lake Tahoe, NV, USA, 2012, pp. 2951–2959.

[32] M. Dorigo, M. Birattari, and T. Stützle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[33] S. Abuowaida *et al.*, "Hybrid ACO-Bayesian optimization of attention-augmented EfficientNetB3 for knee osteoporosis classification," *High-Tech and Innovation Journal*, 2024, in press.

[34] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *Proc. AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 2019, pp. 4780–4789.

[35] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, 2015, pp. 922–928.

[36] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 945–953.

[37] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9397–9406.

[38] Y. Xu, T. Fan, M. Xu, L. Zeng, and Y. Qiao, "SpiderCNN: Deep learning on point sets with parameterized convolutional filters," in *Proc. European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 87–102.

[39] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019, pp. 6411–6420.

[40] M. Xu, R. Ding, H. Zhao, and X. Qi, "PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3172–3181.

[41] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 11 108–11 117.

[42] X. Yu, L. Tang, Y. Rao, T. Huang, J. Zhou, and J. Lu, "Point-BERT: Pre-training 3D point cloud transformers with masked point modeling," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 19 313–19 322.

[43] Y. Pang, W. Wang, F. E. H. Tay, W. Liu, Y. Tian, and L. Yuan, "Masked autoencoders for point cloud self-supervised learning," in *Proc. European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel, 2022, pp. 604–621.

[44] X. Ma, C. Qin, H. You, H. Ran, and Y. Fu, "Rethinking network design and local geometry in point cloud: A simple residual MLP framework," in *Proc. International Conference on Learning Representations (ICLR)*, 2022.

[45] F. Alam, P. Pławiak, A. Almaghthawi, M. R. C. Qazani, S. Mohanty, and R. Alizadehsani, "Neurohar: A neuroevolutionary method for human activity recognition (har) for health monitoring," *IEEE Access*, vol. 12, pp. 112 232–112 248, 2024.

[46] F. Alam, T. S. M. Alnazzawi, R. Mehmood, and A. Al-Maghthawi, "A review of the applications, benefits, and challenges of generative ai for sustainable toxicology," *Current Research in Toxicology*, vol. 8, p. 100232, 2025.

[47] F. A. Rao, A. Muneer, A. Almaghthawi, A. Alghamdi, S. M. Fati, and E. A. A. Ghaleb, "Bmsp-ml: big mart sales prediction using different machine learning techniques," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, p. 874, 2023.

[48] T. Alam, R. Gupta, N. N. Ahamed, A. Ullah, and A. Almaghthwi, "Towards sustainable iot-based smart mobility systems in smart cities," *GeoJournal*, vol. 89, no. 6, p. 235, 2024.

[49] F. Y. Alzyoud, M. Tarawneh, A. Almaghthawi, A. Altalidi, L. Asiri, and M. Alrehaili, "Optimizing broadcast utilization for efficient disaster management using wireless ad hoc networks and novel energy-saving algorithms." *International Journal of Interactive Mobile Technologies*, vol. 18, no. 20, 2024.

[50] A. Almaghthawi, F. Bourennani, and I. R. Khan, "Differential evolution-based approach for tone-mapping of high dynamic range images," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020.

[51] F. AlZyoud, M. Tarawneh, A. Almaghthawi, and A. Altalidi, "A new approach for cluster head selection in wireless sensor networks." *International Journal of Online & Biomedical Engineering*, vol. 20, no. 9, 2024.

[52] A. Almaghthawi, E. A. Ghaleb, N. A. Akbar, L. Asiri, M. Alrehaili, and A. Altalidi, "Federated-learning intrusion detection system based blockchain technology." *International Journal of Online & Biomedical Engineering*, vol. 20, no. 11, 2024.

[53] M. Kamran, A. Saeed, and A. Almaghthawi, "Federated-learning topic modeling based text classification regarding hate speech during covid-19 pandemic." *International Journal of Advanced Computer Science & Applications*, vol. 14, no. 11, 2023.

[54] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1912–1920.

[55] G. Qian, Y. Li, H. Peng, J. Mai, H. Hammoud, M. Elhoseiny, and B. Ghanem, "PointNeXt: Revisiting PointNet++ with improved training and scaling strategies," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, New Orleans, LA, USA, 2022, pp. 23 192–23 204.

[56] H. Ran, J. Liu, and C. Wang, "Surface representation for point clouds," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 18 942–18 952.