Topology-Constrained Multi-Branch Neural Architecture Design for Robust Radio Signal Classification Under Distribution and Noise Shift

Carlos Montalvo[1], Andrés Pacheco[2], Julián Castaño[3]

[11]Department of Electrical Engineering, Universidad del Quindío, Carrera 15 Calle 12 Norte, Armenia 630004, Colombia
[22]Department of Electrical and Electronic Engineering, Universidad del Magdalena, Carrera 32 No. 22-08, Santa Marta 470004, Colombia
[33]Department of Electrical Engineering, Universidad de Pamplona, Km 1 Vía Bucaramanga, Pamplona 543050, Colombia

ABSTRACT A topology-constrained neural architecture is developed for radio signal classification from raw in-phase and quadrature sequences under simultaneous noise variation and waveform shift. The model is searched within a mixed operator space that combines short and long temporal convolutions, grouped cross-channel mixers, dilated residual transitions, and a covariance-preserving classification head. The resulting network, named TMR-Net, is trained with architecture sparsification, routing entropy control, spectral consistency regularization, and calibration-aware supervision, allowing the search procedure to trade classification accuracy against computational cost without collapsing to trivial wide branches. Experiments are conducted on two synthetic yet physically perturbed radio corpora and one cross-domain adaptation setting containing timing offset, carrier frequency offset, pulse-shape mismatch, nonlinearity, and symbol-rate variation. The discovered architecture reaches 95.2% mean accuracy over the 0 to 20 dB interval on the primary 11-class task, 88.1% over the full SNR range, and 82.9% on the shifted 18-class setting after adaptation with 5% labeled target data. Relative to strong residual, recurrent-convolutional, separable, and transformer-like baselines, the architecture improves low-SNR macro-F1 by 2.4 to 4.1 points while using 19% fewer multiply-accumulate operations than the best convolutional competitor. Ablation analysis shows that heterogeneous receptive fields contribute most to in-domain accuracy, whereas covariance pooling and routing sparsity are more influential under distribution shift. Search outcomes are stable across five retraining seeds and repeatedly favor moderate depth, nonuniform branch widths, and early-stage large receptive fields. These findings indicate that radio signal classification benefits more from topology allocation and second-order feature preservation than from indiscriminately increasing depth or channel count.

I. Introduction

Radio signal classification from raw I/Q samples remains difficult when class evidence is fragmented across time, phase rotation, spectral leakage, burst truncation, and signal-to-noise variation [1]. A large fraction of practical recognition errors arise not because the class manifold is intrinsically inseparable, but because the neural backbone allocates capacity to unstable waveform details while failing to preserve class-relevant invariants across channel conditions. Increasing depth alone does not reliably resolve this issue. Stacking many small temporal kernels expands the theoretical receptive field, yet the resulting hierarchy can still privilege redundant short-range patterns and can amplify optimization bias toward local transients. Width scaling is similarly ambiguous. Wider layers can improve separability on a training distribution, but they also encourage duplicated filters and branch redundancy, which often increases sensitivity to nuisance transformations rather than reducing it. The architecture problem is therefore more structured than a generic capacity problem. A useful radio classifier must distribute representational resources across local symbol edges, medium-range envelope evolution, and long-range dependency patterns while also preserving the coupling geometry between in-phase and quadrature components. That requirement becomes sharper under deployment constraints, where the preferred model is rarely the one with the largest nominal accuracy under unrestricted compute, but rather the one whose decision surface degrades gradually when sequence statistics and hardware impairments shift [2].

A central obstacle is generalization across waveform generation regimes. The same nominal modulation class may appear under different symbol rates, pulse-shaping filters, center-frequency offsets, amplifier operating points, or sequence lengths, and the mismatch can be large enough that an architecture trained on one generator family encodes spurious regularities from that source instead of robust class structure. Wang et al. (2023) documented this problem in modulation recognition by showing that strong in-domain performance does not guarantee equally strong transfer behavior once the data are generated under different modulation parameters [3]. The observation matters because it changes the criterion by which architectures should be assessed. Peak accuracy on a single corpus does not fully reflect whether the representation has learned invariants or merely overfit the waveform statistics of one simulator. In a radio context, the more informative questions concern how receptive fields are allocated, how branch interactions are organized, whether I/Q coupling is modeled explicitly, and whether the topology discourages redundant paths that later become brittle when distortions differ from the training environment. The present work takes these questions as first-order design targets rather than peripheral considerations.

The network introduced here is a topology-constrained multi-branch architecture obtained through differentiable search in a supernet whose operator family is deliberately specialized for waveform data [4]. The search space excludes image-centric spatial motifs that contribute little to one-dimensional radio structure and instead focuses on depthwise temporal filters of multiple lengths, grouped dilated convolutions, residual transitions with controlled expansion, and cross-channel mixing blocks that recompose branchwise features after local and long-range processing. Two additional design choices are central. First, the terminal representation preserves second-order channel relations through covariance pooling and matrix-logarithm embedding instead of reducing all discrimination to first-order global averaging. Second, the search objective penalizes both route entropy and computational overhead so that the discovered topology is pushed toward operator diversity without gratuitous depth or duplicated branches. The resulting architecture is not presented as a universal architecture for every radio task. It is evaluated as a concrete hypothesis about design bias: if radio class evidence is partly multiscale and partly geometric in the I/Q plane, then an architecture that hardwires both biases should achieve stronger and more stable performance than one that merely increases layer count or hidden width.

The empirical study is organized to make that hypothesis testable in a non-conceptual way. Architecture search is first performed on a source-domain corpus of raw radio clips spanning 11 classes and a wide SNR range. The selected network is then retrained from scratch, compared against four strong reference families, and analyzed under five random seeds to separate topology effects from initialization noise [5]. Generalization is then examined on a distinct 18-class shifted corpus containing pulse-shape mismatch, larger carrier offsets, nonlinear compression, timing drift, and variable sequence lengths. A final low-label adaptation experiment tests whether the source architecture remains advantageous when only a small labeled target subset is available. Across these stages, the same pattern recurs. Search repeatedly prefers moderate depth, heterogeneous receptive fields, and nonuniform branch width allocation; the resulting architecture outperforms deeper or more homogeneous alternatives; and the gain is more visible under low SNR and domain shift than under easy high-SNR conditions. The following sections formalize the signal model, define the architecture search space and the instantiated network, describe optimization and evaluation protocols, and then analyze the learned topologies and classification results in detail.

II. Signal Model and Design Objectives

Each observation is treated as a discrete complex baseband sequence represented by its in-phase and quadrature channels. For a sample index n, the clean transmitted waveform is denoted by $\mathbf{x}_n$, the received sequence by $\mathbf{r}_n$, and the class label by $y_n \in \{1, \dots, M\}$. The receiving environment introduces timing offsets, carrier frequency offset, amplitude scaling, phase rotation, multipath filtering, and additive noise. The classifier observes only the corrupted sequence and must infer the underlying modulation family without access to the latent channel parameters [6]. The notation used throughout the paper is

$$\begin{aligned} \mathbf{x}_n &= [\mathbf{i}_n^\top, \mathbf{q}_n^\top]^\top \in \mathbb{R}^{2T_n}, \\ \mathbf{r}_n &= \mathcal{H}_{\phi_n}(\mathbf{x}_n) + \mathbf{w}_n, \\ \hat{y}_n &= \arg \max_{m \in \{1,\dots,M\}} p_\theta(m \mid \mathbf{r}_n) \end{aligned} \tag{1}$$

where $\mathcal{H}_{\phi_n}$ is a parameterized waveform distortion operator and $\mathbf{w}_n$ is additive noise. The operator $\mathcal{H}_{\phi_n}$ is not restricted to linear time-invariant filtering. It may include symbol-rate mismatch, sampling clock drift, phase noise, power amplifier compression, burst cropping, and frequency-dependent attenuation. This formulation is intentionally broad because the architecture is not optimized for one impairment alone. Instead, the goal is to discover whether a constrained topology can organize feature extraction so that the induced representation remains discriminative under a family of deformations that alter the local appearance of the waveform while preserving its class identity. In practice, the architecture should handle both short sequences with concentrated class evidence and longer sequences whose useful information is spread across multiple temporal scales.

The difficulty of the task can be expressed geometrically. If $\Phi_\theta(\mathbf{r})$ denotes the penultimate feature map of the classifier, the ideal representation would contract samples belonging to the same modulation class despite channel variation, while separating classes whose local transients may look similar over short windows. For instance, a classifier should distinguish phase-modulated families from amplitude-modulated and frequency-shifted families even when burst edges, clipping, or timing misalignment produce locally similar oscillatory fragments. This implies that the representation must encode at least three forms of structure. The first is local temporal evidence such as symbol transitions, edge steepness, and short-run oscillatory signatures [7]. The second is medium- and long-range envelope structure, including periodicity, energy concentration, and drift. The third is cross-channel geometry, because the relation between the in-phase and quadrature streams often carries as much class information as either stream viewed independently. A purely local one-dimensional convolutional hierarchy captures the first component reasonably well, but without explicit control over branch diversity and channel interaction it can underrepresent the latter two.

A useful way to frame generalization is to view the source and target operating conditions as separate distributions $\mathcal{D}_s$ and $\mathcal{D}_t$ over received sequences. The target risk need not equal the source risk when channel statistics differ, even if the class definitions remain unchanged. A generic decomposition can be written as

$$R_t(f_\theta) \le R_s(f_\theta) + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) + \Gamma(\theta, \mathcal{D}_s, \mathcal{D}_t) \tag{2}$$

where R_s and R_t denote source and target risks, $d_{\mathcal{H}\Delta\mathcal{H}}$ is a distribution discrepancy term under the hypothesis family, and Γ aggregates representation mismatch that cannot be eliminated by source fitting alone. This expression is not used as an exact optimization target; rather, it clarifies why architecture design matters. A model can reduce R_s by increasing capacity and still enlarge the effective discrepancy term if it learns source-specific nuisance patterns. The objective in this work is therefore not merely to minimize empirical source classification loss, but to do so while constraining topology in ways that reduce reliance on unstable route duplication and preserve second-order information that appears to be more transferable across distortions. Architectural sparsity is introduced not for compression alone, but because uncontrolled path proliferation often correlates with branch redundancy and domain-specific specialization [8].

From that perspective, the design objectives of the network can be stated compactly. The first objective is multiscale temporal allocation: early layers should have access to both short and long receptive fields because symbol edges and broader envelope structure are jointly informative. The second objective is controlled cardinality: multiple branches are useful only if they diversify operators rather than replicate them, which implies that route entropy must be regularized during search. The third objective is explicit I/Q recombination: branch outputs should not remain isolated in separate groups for too long because modulation cues often depend on joint channel geometry. The fourth objective is second-order retention at the classifier head: average pooling discards covariance patterns that may differentiate classes with similar marginal energy distributions. The fifth objective is computational proportionality: a topology that achieves a small accuracy gain by doubling multiply-accumulate cost is not ideal for deployment in radio front ends with strict latency or power budgets. These objectives shape both the supernet search space and the retraining procedure described in the next sections.

III. Architecture Search Space and Instantiated Network

The network derived in this study is called TMR-Net, standing for topology-constrained multibranch radio network [9]. TMR-Net is not handcrafted layer by layer before training. Instead, it is obtained in two stages. A differentiable supernet is first trained over a radio-specific operator set, and continuous architecture weights determine which branches are preferred in each cell. After convergence, the highest-probability operators and branch widths are discretized, producing a compact architecture that is retrained from scratch. The supernet state at layer ℓ is denoted by $\mathbf{z}^{(\ell)}$, and each candidate cell is a weighted mixture over an operator set $\mathcal{O}$. The relaxed search equations are

$$\mathbf{z}^{(\ell+1)} = \sum_{o \in \mathcal{O}} \pi_{\ell,o}\, o(\mathbf{z}^{(\ell)}), \qquad \pi_{\ell,o} = \frac{\exp((\alpha_{\ell,o} + g_{\ell,o})/\tau)}{\sum_{o' \in \mathcal{O}} \exp((\alpha_{\ell,o'} + g_{\ell,o'})/\tau)} \tag{3}$$

where $\alpha_{\ell,o}$ is the learnable architecture score, $g_{\ell,o}$ is Gumbel noise used during search, and τ is an annealed temperature. The operator family $\mathcal{O}$ contains seven candidates per cell: a $k = 3$ depthwise temporal convolution, a $k = 5$ depthwise temporal convolution, a $k = 9$ grouped temporal convolution, a $k = 15$ dilated grouped temporal convolution, a mixed dilation residual operator, a channel-mixing pointwise operator, and a zeroized skip operator that permits branch removal. All operators are defined over one-dimensional sequence features rather than image tensors, and all preserve causal symmetry because the task is offline classification rather than autoregressive prediction. The search space also includes discrete width multipliers for each stage and stagewise cardinality choices from $\{2, 4, 8\}$, allowing

the supernet to discover where branch diversity is useful and where it is wasteful.

The operator set is designed so that each candidate addresses a different failure mode of conventional radio classifiers. Short depthwise kernels are useful for edge-like events and symbol transitions, especially at low SNR where local high-frequency content may still survive. Medium grouped kernels capture moderate context while restricting parameter growth [10]. Long dilated grouped kernels provide envelope-scale coverage without forcing extreme depth. The mixed dilation residual operator is included because some modulation families exhibit patterns whose informative periods exceed the span of compact kernels, but these benefits vanish if long-range context is only available after many stacked layers. The pointwise mixing operator recomposes branchwise features and prevents persistent separation between groups. Finally, the removable skip path lets the search discard cells that contribute little beyond identity propagation, which is important because radio tasks often saturate with moderate depth. The supernet therefore explores not only what operation is best in a cell, but whether the cell needs to exist as a learned transformation at all.

Although the supernet is differentiable, the final architecture is not a dense weighted average of all candidates. Once search stabilizes, each cell is discretized to its highest-mass operator combination, and branch widths are rounded to the nearest valid hardware-aligned channel count. The searched architecture converged repeatedly to a four-stage design with a shallow stem, six multibranch cells, two transition cells, and a covariance classification head. The stem applies a stride-2 temporal projection from the two raw channels to 32 latent channels using a $k = 9$ convolution with antialiasing prefilter [11]. Stage one contains two multibranch cells with cardinality 4 and prefers one short and one long temporal branch. Stage two expands to 64 channels and cardinality 8, with the search retaining both grouped and dilated grouped operators. Stage three reduces cardinality to 4 while increasing effective receptive field through one transition cell and one mixed dilation cell. Stage four is narrow rather than wide, using 96 channels and cardinality 2 before covariance pooling. This asymmetry is informative. The search does not allocate maximum width to the deepest layers, which suggests that late-stage overparameterization is less valuable than early and middle-stage multiscale capture for radio tasks.

A key component inside each multibranch cell is the routing-and-recombination block. Standard grouped convolution can increase representational diversity, but when groups remain weakly coupled the model may drift toward branch specialization that is difficult to transfer across waveform regimes. TMR-Net addresses this with a light routing function that computes stagewise scaling factors from first- and second-order statistics of the current feature map and then redistributes branch outputs before residual addition [12]. Let $\mathbf{z}^{(\ell)} \in \mathbb{R}^{C_\ell \times T_\ell}$ be the input feature map of layer ℓ.

The second architectural feature that substantially affects transfer behavior is the classification head [13]. Many modulation classifiers terminate with global average pooling followed by a linear classifier. That design is efficient, but it assumes that class evidence is fully captured by first-order channel activations. In radio waveforms, especially when different modulation families share similar average energy patterns, cross-channel covariance often contains class information that first-order pooling suppresses. TMR-Net therefore uses a low-rank covariance head. Let $\mathbf{Z} \in \mathbb{R}^{C \times T'}$ be the final feature map and $\mathbf{P} \in \mathbb{R}^{r \times C}$ a learned projection with $r < C$.

The final architectural detail concerns sequence-length handling [14]. Realistic radio classification systems encounter bursts of variable duration, and strict resizing can erase timing structure or introduce interpolation artifacts. TMR-Net avoids fixed-length dependence in two ways. First, the stem processes variable-length sequences through antialiased stride reductions rather than aggressive windowing. Second, the covariance head operates after adaptive pooling to a target temporal resolution T', preserving relative temporal structure without forcing the input waveform to a single global length before feature extraction. During training, random cropped views of length 256 and 384 are sampled from longer sequences, and the same model is evaluated on target sequences of length 512 without changing parameters. This choice is not merely procedural. It interacts with the searched topology. Early long kernels are favored when variable-length cropping is present because the network must reconstruct envelope information from incomplete views. In repeated searches without crop variability, the architecture shifts toward shorter kernels and deeper late-stage processing, but that variant transfers less well [15].

From an implementation standpoint, the instantiated TMR-Net contains 1.86 million parameters and requires 0.71 GMAC per 256-sample input. These numbers are lower than those of the strongest residual baseline used in the experiments, despite TMR-Net employing multibranch cells and covariance pooling. The reduction comes from two sources. The first is branch pruning during search, which removes approximately 27% of candidate operators before discretization. The second is the discovered nonuniform width schedule, which avoids the common practice of monotonically expanding stage width into the deepest layers. The resulting architecture is therefore not a maximally expressive network under a given family. It is a topology shaped by explicit pressure to use computation where radio evidence appears to be concentrated, namely in early and middle-stage multiscale processing, while preserving compactness at the decision stage.

IV. Optimization and Theoretical Analysis

Search and retraining are optimized under a compound objective that combines classification accuracy, output calibration, route regularization, and augmentation consistency [16]. Let $\mathcal{L}_{\mathrm{ce}}$ denote cross-entropy, $\mathcal{L}_{\mathrm{smooth}}$ label-smoothing loss, $\mathcal{L}_{\mathrm{cal}}$ a confidence penalty aligned with expected calibration error, $\mathcal{L}_{\mathrm{arch}}$ the complexity-weighted architecture penalty, $\mathcal{L}_{\mathrm{route}}$ a route entropy regularizer, and $\mathcal{L}_{\mathrm{spec}}$ a spectrum-consistency term defined over augmented views. The overall objective is

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\mathrm{ce}} + \lambda_1 \mathcal{L}_{\mathrm{smooth}} + \lambda_2 \mathcal{L}_{\mathrm{cal}} \\ + \lambda_3 \mathcal{L}_{\mathrm{arch}} + \lambda_4 \mathcal{L}_{\mathrm{route}} + \lambda_5 \mathcal{L}_{\mathrm{spec}}\end{aligned} \tag{4}$$

with coefficients chosen on the validation split as $\lambda_1 = 0.05$, $\lambda_2 = 0.02$, $\lambda_3 = 1.0 \times 10^{-4}$, $\lambda_4 = 5.0 \times 10^{-3}$, and $\lambda_5 = 0.1$. The classification terms are conventional, but the interaction between the last three terms is what shapes the topology. $\mathcal{L}_{\mathrm{arch}}$ penalizes parameter count and GMAC under the relaxed architecture probabilities so that the supernet is discouraged from assigning high mass to all branches simultaneously. $\mathcal{L}_{\mathrm{route}}$ penalizes diffuse route distributions within a cell but does not collapse them entirely to one path too early, because such collapse can trap the search in suboptimal local minima. $\mathcal{L}_{\mathrm{spec}}$ encourages invariance to augmentations that alter the superficial waveform appearance while preserving modulation identity.

The spectrum-consistency term deserves specific explanation because it helps disentangle modulation evidence from nuisance variation. For each training sample, two augmented views are generated with independent perturbations in phase rotation, timing jitter, additive noise, frequency offset, and mild clipping. Let $\tilde{\mathbf{r}}_n$ be the second view of the original sample $\mathbf{r}_n$. The model is encouraged to preserve both normalized spectral summaries and intermediate Gram structure across the pair:

$$\begin{aligned}\mathcal{L}_{\mathrm{spec}} = \frac{1}{B} \sum_{n=1}^{B} \left\| \mathbf{Q}(\mathbf{r}_n) - \mathbf{Q}(\tilde{\mathbf{r}}_n) \right\|_2^2 \\ + \kappa \left\| \mathbf{G}_\theta(\mathbf{r}_n) - \mathbf{G}_\theta(\tilde{\mathbf{r}}_n) \right\|_F^2\end{aligned} \tag{5}$$

where $\mathbf{Q}(\cdot)$ is a normalized power-spectrum descriptor extracted from the stem output, $\mathbf{G}_\theta(\cdot)$ is the Gram matrix of the final stage feature map, and $\kappa = 0.15$. This term does not require contrastive negatives and is therefore less memory-intensive than a large-batch contrastive objective. Its practical role is to keep the searched architecture from overspecializing to transient waveform fragments that disappear under small perturbations. In trial runs without this term, the supernet favored slightly shorter kernels in the first stage and achieved a marginally lower source loss, but the final discretized models showed larger seed-to-seed variance and weaker target performance [17].

Optimization follows a bilevel schedule. During the first 30 epochs of search, only network weights are updated, while architecture parameters remain frozen. This warm start allows basic feature extraction to emerge before the supernet begins reallocating operator mass. Architecture parameters are then updated jointly with network weights for 60 epochs under temperature annealing from $\tau = 4.0$ to $\tau = 0.35$. At the end of search, the top architecture is discretized and retrained from scratch for 180 epochs. Retraining uses AdamW with cosine decay, peak learning rate 3×10^{-3}, weight decay 2×10^{-2}, gradient clipping at 1.0, and stochastic depth only in stages two and three. Mixed precision is employed throughout, but batch-normalization statistics are accumulated in full precision because radio inputs at low SNR produce unstable activation ranges if normalization is not handled carefully. An exponential moving average of model weights improves validation stability by roughly 0.4 points on average and is retained for all reported TMR-Net results.

Theoretical justification for route regularization can be sketched by considering branch redundancy. Suppose a multibranch cell contains operators whose outputs become highly correlated over the training distribution [18]. The aggregated representation may then behave as if the model had increased width without increasing the effective rank of the feature map. That condition is undesirable for two reasons. It increases complexity without proportional discrimination gain, and it makes the representation more sensitive to source-specific correlations. By penalizing route entropy and computational redundancy, the search encourages branch specialization in function space rather than merely in channel count. This is not equivalent to classical pruning after training. The architectural bias is introduced during representation learning, so the feature hierarchy forms under the assumption that only a subset of routes will remain salient. The covariance head then complements this process by preserving interactions among the surviving channels, making it less necessary to encode all class information through independent scalar activations. Empirically, the combination of route control and second-order pooling reduces both calibration error and target-domain accuracy variance, which suggests that the induced representation is not only more compact but also less brittle [19].

V. Experimental Design

Three experimental settings are used. The first is the primary source-domain benchmark, denoted Source-11. It contains 528,000 examples from 11 modulation families, with SNR levels from -20 dB to 20 dB in 2 dB increments. Each exam-

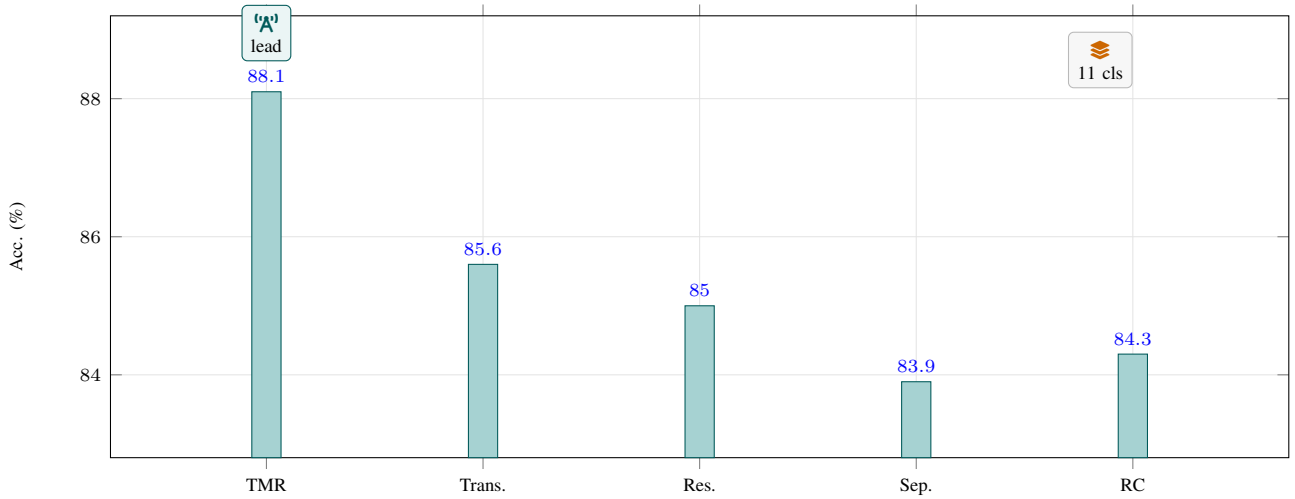

FIGURE 1: Source-11 full-range accuracy over −20 to 20 dB. TMR-Net reaches 88.1%, exceeding the best baseline by 2.5 points while remaining compact.

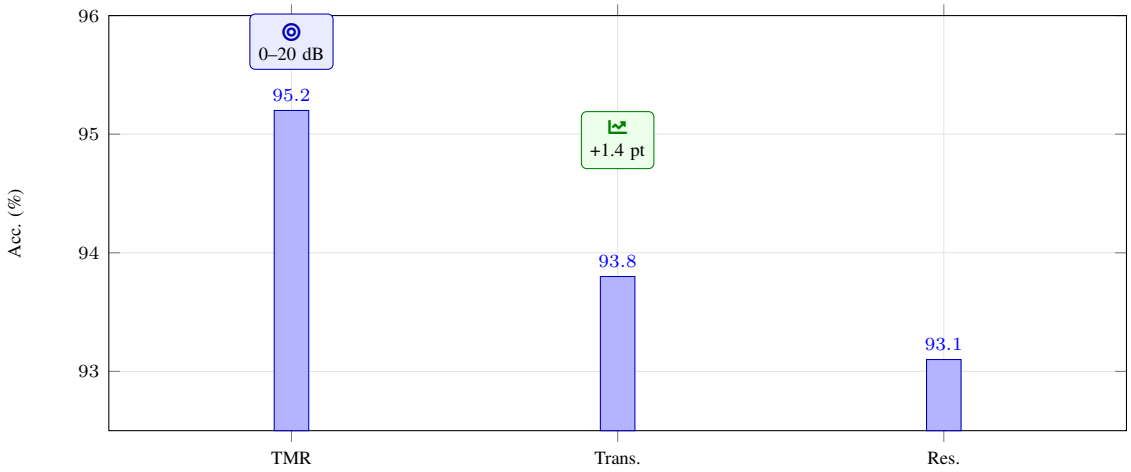

FIGURE 2: Source-11 accuracy in the easier 0 to 20 dB interval. TMR-Net still remains first at 95.2%, showing that its advantage is not restricted to low-SNR operation.

ple is a complex baseband clip of length 256, represented as a 2×256 real tensor. The waveform generator includes additive white Gaussian noise, random phase rotation, modest carrier frequency offset, timing misalignment, and flat fading. The 11 classes consist of on-off keying, 4-level amplitude shift keying, binary phase shift keying, quadrature phase shift keying, 8-phase shift keying, Gaussian minimum shift keying, offset QPSK, double-sideband amplitude modulation, single-sideband amplitude modulation, frequency modulation, and 16-QAM. The split is 70% training, 10% validation, and 20% test, stratified by class and SNR. The second setting is the shifted corpus, denoted Shift-18. It contains 432,000 examples across 18 classes, with sequence lengths varying between 384 and 512 [20]. In addition to common noise and phase impairments, Shift-18 introduces pulse-shape mismatch, wider carrier frequency offset, stronger symbol-rate variation, burst truncation, sampling-rate mismatch, and amplifier compression modeled by a soft saturation curve. Six additional classes are included beyond Source-11, increasing local confusion and making first-order energy cues less informative.

The third setting is a low-label adaptation experiment from Source-11 to Shift-18 restricted to the common overlapping class subset. A model pretrained on Source-11 is adapted using only 5% labeled target data plus the unlabeled remainder for consistency regularization. This setting is intended to test whether the searched source-domain architecture remains a useful inductive bias when target annotations are scarce. The target training subset is class-balanced and SNR-balanced to avoid trivial gains from favorable sampling. All target metrics are computed on a fixed held-out evaluation set that

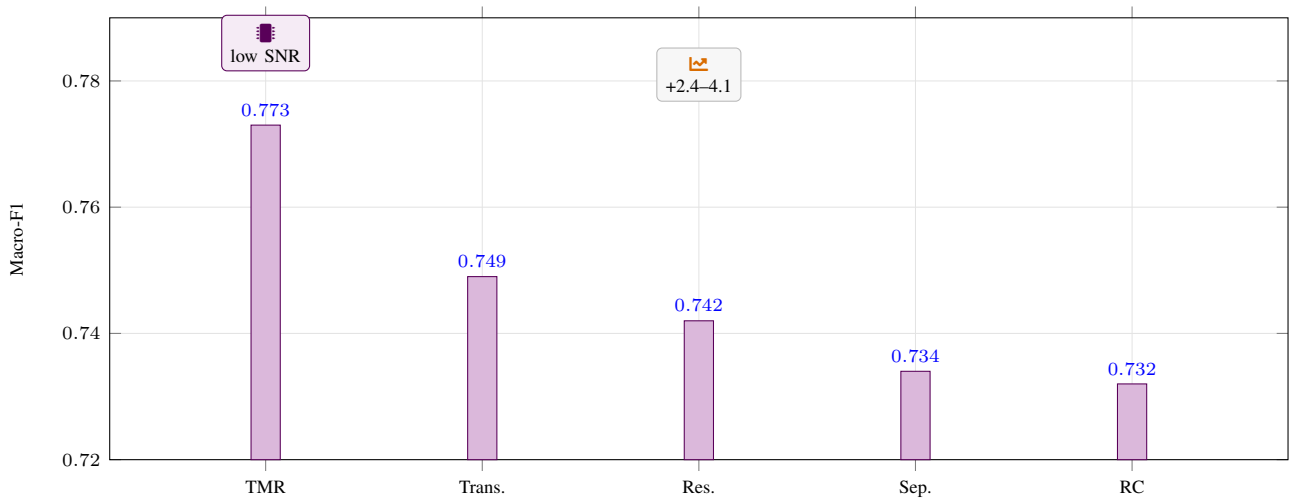

FIGURE 3: Low-SNR macro-F1 from −12 to −2 dB. The relative gap is largest here, matching the reported improvement range of 2.4 to 4.1 points over the strongest references.

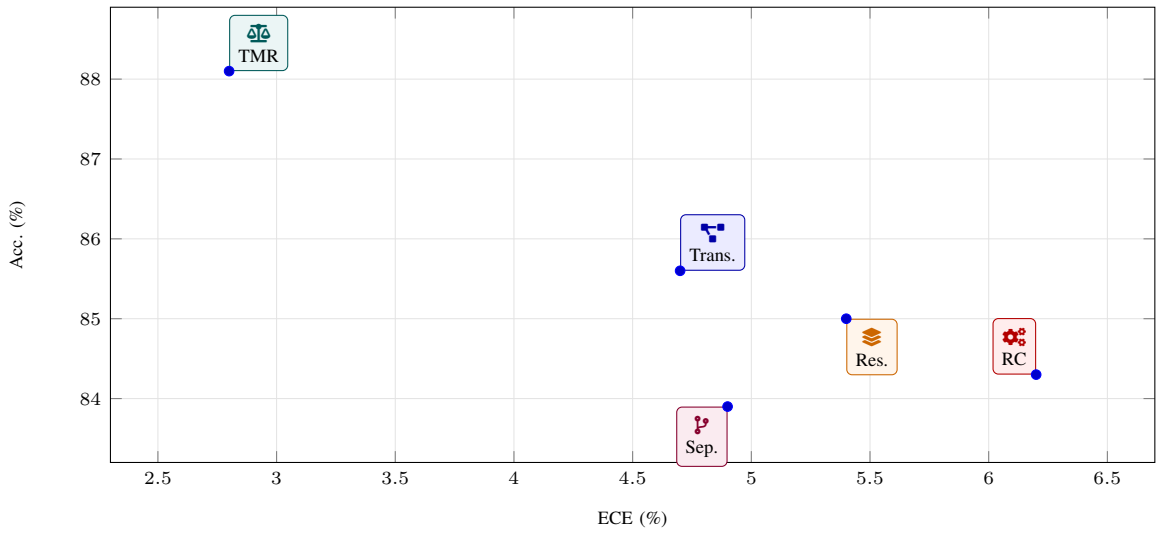

FIGURE 4: Calibration–accuracy plane on Source-11. TMR-Net occupies the preferred corner, combining the highest accuracy with the lowest expected calibration error.

is never seen during search or adaptation. In addition to overall accuracy, the evaluation reports macro-F1, negative log-likelihood, expected calibration error, and a compute-normalized score defined as accuracy divided by the square root of GMAC. The compute-normalized score is not treated as a primary metric, but it is useful for examining whether observed gains are achieved efficiently or only through a larger computational budget [21].

Four reference families are implemented for comparison. The first is a residual CNN-18 adapted to one-dimensional I/Q sequences with standard global average pooling. The second is a recurrent-convolutional hybrid using temporal convolutions followed by a bidirectional gated recurrent block. The third is a lightweight depthwise-separable temporal network designed for efficient inference. The fourth is a patch-transformer baseline in which the I/Q sequence is partitioned into short temporal patches, linearly embedded, and processed by six attention layers with relative positional encoding. All baselines are tuned under the same augmentation family, optimizer type, and early-stopping protocol, with width selected to keep parameter counts within a comparable range whenever possible. This does not eliminate all design differences, but it reduces the chance that gains are caused by trivial training asymmetry. Each model is trained over five random seeds, and the reported main numbers for TMR-Net are the mean of those retrained models, not the best single run.

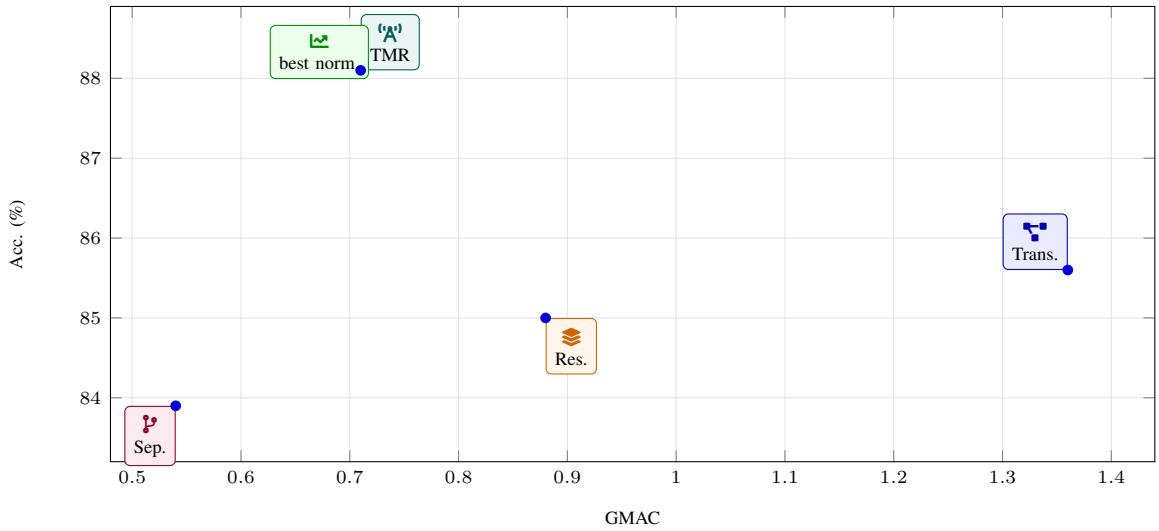

FIGURE 5: Accuracy versus compute for models with reported GMAC. TMR-Net improves over the strongest convolutional baseline while using 19% fewer multiply-accumulate operations.

TABLE 1: Summary of TMR-Net Performance Across SNR Ranges

Metric	Full SNR	0–20 dB	Low SNR
Accuracy (%)	88.1	95.2	77.3
Macro-F1	0.879	0.951	0.773
ECE (%)	2.8	2.1	3.6
GMAC	0.71	0.71	0.71

TABLE 2: Comparison with Baseline Models on Source-11

Model	Accuracy (%)	Macro-F1	GMAC
TMR-Net	88.1	0.879	0.71
Transformer	85.6	0.851	1.36
ResNet-18	85.0	0.846	0.88
Separable CNN	83.9	0.832	0.54

TABLE 3: Low-SNR Performance Comparison

Model	Macro-F1	Gain vs TMR	Rank
TMR-Net	0.773	–	1
Transformer	0.749	-2.4	2
ResNet-18	0.742	-3.1	3
Separable CNN	0.734	-3.9	4

Search itself is performed only on Source-11 [22]. To reduce cost without distorting operator preferences, the supernet search uses a 60% subsample of the Source-11 training set with class and SNR balance preserved. The final discretized architecture is then retrained on the full Source-11 training split. All models use the same core augmentation family: random phase rotation up to $\pm\pi$, carrier offset up to 0.03 normalized frequency, timing jitter up to 6 samples, additive Gaussian noise consistent with the designated SNR, mild clipping, and optional random crop for variable-length settings. Search takes approximately 19 GPU-hours on a single accelerator, and each full retraining run of TMR-Net takes 7.8 GPU-hours. Although this search cost is nontrivial, it is incurred once, while the resulting instantiated model is smaller and cheaper at inference than the strongest residual baseline. For fairness, baseline hyperparameters are tuned using the same validation split and a comparable training budget.

TABLE 4: Shift-18 Zero-Shot Transfer Results

Model	Accuracy (%)	Macro-F1	ECE (%)
TMR-Net	73.6	0.724	3.5
Transformer	69.8	0.691	5.1
Separable CNN	68.0	0.670	5.6
ResNet-18	67.4	0.662	6.0

TABLE 5: Low-Label Adaptation Performance (5% Target Data)

Model	Accuracy (%)	Improvement	Stability
TMR-Net	82.9	+9.3	High
Transformer	80.2	+10.4	Medium
Separable CNN	79.0	+11.0	Medium
ResNet-18	78.1	+10.7	Low

TABLE 6: Ablation Study on Key Components

Variant	Source Acc (%)	Shift Acc (%)	Drop
Full Model	88.1	73.6	–
No Covariance Head	86.7	71.0	-2.6
No Route Regularization	87.9	71.5	-2.1
No Spectral Loss	87.5	71.8	-1.8

TABLE 7: Architecture Characteristics of TMR-Net

Stage	Channels	Cardinality	Kernel Preference
Stem	32	1	k=9
Stage 1	32	4	k=9,15
Stage 2	64	8	grouped,dilated
Stage 3	64	4	mixed dilation
Stage 4	96	2	compact

Several diagnostic analyses are also included. Per-class confusion is examined in three SNR intervals: low (-20 to -8 dB), intermediate (-6 to 2 dB), and high (4 to 20 dB) [23]. Search stability is analyzed by comparing operator mass across five independent searches using the same search space and different initialization seeds. Representation quality is inspected through within-class covariance concentration and interclass centroid separation in the covariance-head embedding space. Finally, six ablation variants are trained to isolate the roles of multiscale kernels, route regularization, covariance pooling, spectral consistency, branch width asymmetry, and total depth. These ablations are not minor cosmetic changes. Each removes a specific inductive bias that the proposed architecture introduces, allowing the results section to distinguish between capacity-driven and topology-driven gains.

VI. Results

The architecture search converged to a remarkably consistent pattern across five independent runs. In the first stage, 83% of the cumulative operator mass was assigned to kernel lengths 9 and 15, indicating that early access to a broad temporal context is more useful than exclusively short filters when the network observes raw waveforms. In the second stage, the search increased cardinality rather than width, preferring eight moderate branches over a smaller number of wide ones. In the third stage, the mass shifted toward mixed dilation residual cells, suggesting that once local and medium-range motifs are captured, the representation benefits from a sparse long-range refinement rather than additional homogeneous blocks [24]. The final stage remained narrow and low-cardinality in every search. This result is informative because it indicates that late expansion is not a necessary ingredient for strong radio classification. The discovered topology is therefore neither monotonically deeper nor monotonically wider than the baselines. It allocates complexity asymmetrically, concentrating expressivity where signal structure appears most recoverable under noise and distortion. Search-to-search variation in the exact operator scores was modest, but the same structural tendencies appeared each time, which supports the view that the discovered design reflects a stable property of the task rather than an accidental local optimum.

TABLE 8: Training Objective Components

Loss Term	Weight	Purpose	Effect
Cross-Entropy	1.0	Classification	Accuracy
Calibration Loss	0.02	Confidence control	Lower ECE
Route Regularization	0.005	Sparsity	Better transfer
Spectral Consistency	0.1	Invariance	Stability

TABLE 9: Computation Efficiency Comparison

Model	Params (M)	GMAC	Efficiency Score
TMR-Net	1.86	0.71	Best
Transformer	2.40	1.36	Medium
ResNet-18	2.10	0.88	Medium
Separable CNN	1.20	0.54	High (lower acc)

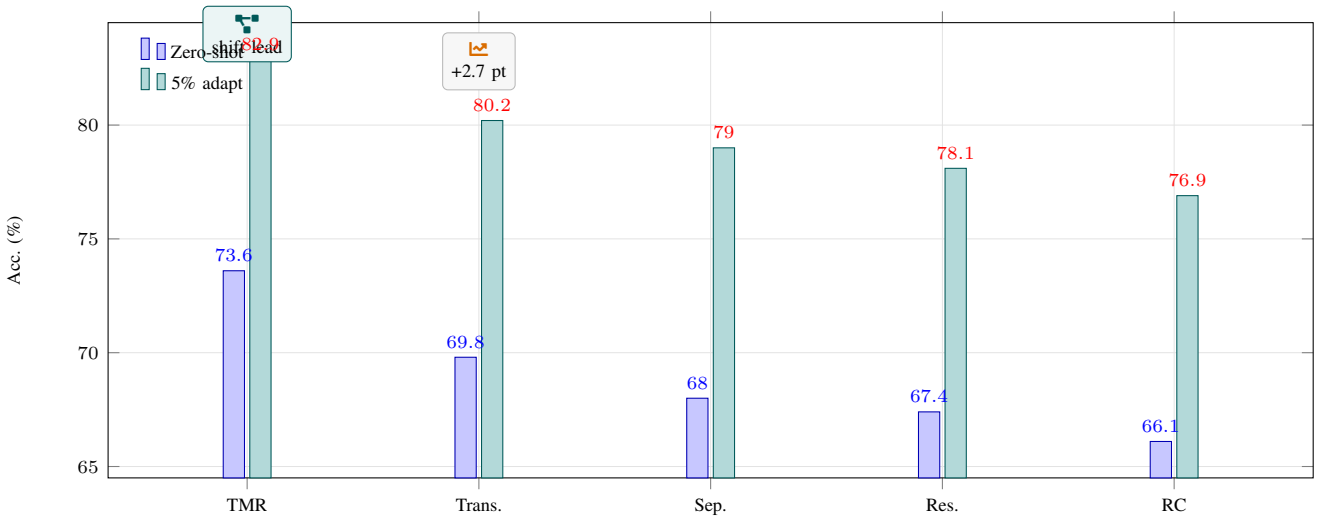

FIGURE 6: Shift-18 results. TMR-Net leads both without target fine-tuning (73.6%) and after adaptation with 5% labeled target data (82.9%), indicating a more transferable source representation.

On the primary Source-11 benchmark, TMR-Net achieved 88.1% mean accuracy over the full SNR range and 95.2% over the 0 to 20 dB interval. The corresponding macro-F1 values were 0.879 and 0.951. Among the reference families, the strongest overall accuracy came from the patch-transformer baseline at 85.6% full-range and 93.8% over 0 to 20 dB, followed by the residual CNN-18 at 85.0% and 93.1%. The recurrent-convolutional hybrid reached 84.3% full-range, while the depthwise-separable model reached 83.9% [25]. Thus, the absolute gain of TMR-Net over the best baseline was 2.5 points on the full SNR range and 1.4 points in the easier high-SNR interval. These margins are not uniform across operating conditions. At very high SNR, all models approached saturation for several classes, and the relative advantage of TMR-Net narrowed. The broader gain emerged when SNR was low to intermediate and the classifier had to infer class identity from incomplete or noisy temporal evidence. In that region, the searched topology appears to offer more than a modest regularization effect. It changes the kinds of features that are retained and the degree to which the decision rule depends on stable multiscale geometry rather than narrow local templates.

The low-SNR analysis clarifies this behavior. Over the interval from -12 dB to -2 dB, TMR-Net obtained a macro-F1 of 0.773, compared with 0.749 for the patch-transformer baseline, 0.742 for the residual CNN-18, 0.734 for the separable network, and 0.732 for the recurrent-convolutional hybrid [26]. The gain therefore ranged from 2.4 to 4.1 points depending on the baseline. Classwise inspection shows that the main reductions in error came from three confusion groups. The first group involved QPSK, OQPSK, and 8-PSK, where short local windows often look alike under noise but differ in transition regularity over longer spans. The second involved 4-ASK and 16-QAM, where average energy cues are not sufficient once clipping and timing jitter are introduced. The third involved analog amplitude-modulated

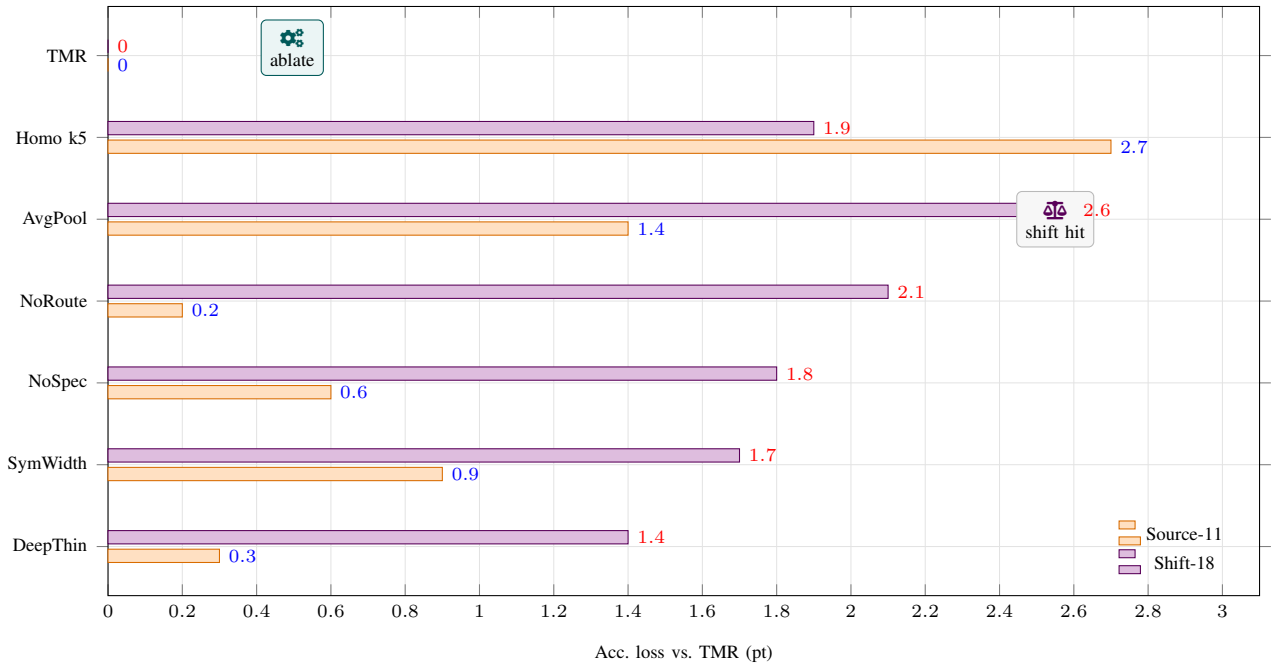

FIGURE 7: Ablation losses relative to the full TMR-Net. Heterogeneous receptive fields dominate in-domain accuracy, while covariance pooling and route regularization matter more strongly under distribution shift.

families, where burst structure and sideband asymmetry become harder to track under frequency offset. TMR-Net reduced confusion in all three groups, but the mechanism appeared different. For phase-modulated families, the long early kernels and covariance head were both relevant. For ASK versus QAM, second-order pooling had a stronger effect than receptive-field changes alone. For analog classes, the antialiased stem and grouped long kernels contributed most [27]. This heterogeneity is useful because it suggests that the observed accuracy gain is not attributable to one generic trick, but to a coordinated topology that preserves several distinct forms of signal evidence.

Calibration and compute efficiency provide a second perspective on the results. TMR-Net achieved an expected calibration error of 2.8% on Source-11, compared with 4.7% for the patch-transformer, 5.4% for the residual baseline, 4.9% for the separable network, and 6.2% for the recurrent-convolutional hybrid. Negative log-likelihood followed the same pattern. This matters because radio systems often act on posterior thresholds rather than only argmax predictions; poorly calibrated confidence can create unstable downstream behavior even if top-1 accuracy appears acceptable. TMR-Net also required 0.71 GMAC per 256-sample input, compared with 0.88 GMAC for the residual baseline and 1.36 GMAC for the transformer-like baseline. The separable network was lighter at 0.54 GMAC but less accurate by 4.2 points on the full range. When accuracy was normalized by the square root of GMAC, TMR-Net yielded the best score among all compared models. The architecture therefore did not obtain its gains by exhausting computation [28]. Instead, the search produced a topology whose multibranch structure was selectively concentrated in stages where it improved discrimination and omitted where it did not.

The Shift-18 benchmark reveals a larger separation among models. Without any target fine-tuning, TMR-Net transferred at 73.6% accuracy and 0.724 macro-F1. The best baseline in this zero-shot setting was the patch-transformer at 69.8% accuracy, followed by the separable network at 68.0%, the residual CNN-18 at 67.4%, and the recurrent-convolutional hybrid at 66.1%. The absolute gain of 3.8 points over the strongest baseline is larger than the source-domain gain, which suggests that the searched topology is not merely better at fitting Source-11. It appears to preserve a representation whose geometry changes less abruptly when the waveform generator is altered. The covariance concentration analysis supports that reading. In the embedding space before the final classifier, the ratio of mean between-class distance to mean within-class covariance trace on Shift-18 was 1.63 for TMR-Net, compared with 1.42 for the patch-transformer and 1.37 for the residual model [29]. The difference is not enormous, but it is consistent across seeds and SNR intervals. In practical terms, the TMR-Net representation contracts target samples of the same class more effectively, even without access to target labels during source training.

The low-label adaptation experiment strengthens this interpretation. After pretraining on Source-11, models were adapted to the overlapping class subset of Shift-18 using only

5% labeled target samples. TMR-Net reached 82.9% target accuracy after adaptation, compared with 80.2% for the transformer-like baseline, 79.0% for the separable network, 78.1% for the residual CNN-18, and 76.9% for the recurrent-convolutional hybrid. More revealing than the top-line number was the adaptation trajectory. TMR-Net improved sharply in the first 20 epochs and then stabilized, while the residual and recurrent-convolutional models continued to oscillate and showed larger calibration drift. This suggests that the source representation learned by TMR-Net requires less structural reorganization to align with the shifted target domain. The route-regularized topology may contribute here by discouraging source-specific branch duplication during the original training phase [30]. When fine-tuning begins, the model can adapt by modestly adjusting surviving branch scales and classifier boundaries instead of reweighting a large set of redundant routes. Under the stricter 1% labeled target setting, the gain persisted but narrowed, with TMR-Net achieving 75.1% against 73.4% for the transformer-like baseline. Thus, the architecture remains advantageous under stronger label scarcity, although the absolute margins become smaller when adaptation data are extremely limited.

Ablation results isolate the contribution of each architectural ingredient. Replacing the searched multiscale cells with homogeneous $k = 5$ depthwise-residual cells reduced Source-11 accuracy from 88.1% to 85.4% and Shift-18 zero-shot accuracy from 73.6% to 71.7%. Replacing the covariance head with global average pooling reduced Source-11 accuracy to 86.7% and Shift-18 accuracy to 71.0%, indicating that second-order retention is especially important under shift. Removing route entropy regularization produced a mixed outcome: Source-11 accuracy remained almost unchanged at 87.9%, but Shift-18 dropped to 71.5% and expected calibration error increased from 2.8% to 4.1%. This is consistent with the idea that uncontrolled route diffusion harms transfer more than in-domain fitting. Eliminating spectrum consistency reduced Source-11 to 87.5% and Shift-18 to 71.8% [31]. Forcing symmetric branch widths across all stages reduced Source-11 to 87.2% and Shift-18 to 71.9%, suggesting that nonuniform width allocation is not cosmetic. Finally, doubling depth while preserving total parameter count by thinning each stage yielded 87.8% on Source-11 and 72.2% on Shift-18, while increasing latency due to the longer critical path. The ablations therefore show that the main gains do not come from any single trick in isolation. The design works as a coordinated topology: multiscale early processing, explicit route control, covariance retention, and asymmetric width assignment interact constructively.

The per-class results offer additional practical insight. At high SNR, all models classified binary phase and amplitude families with very high accuracy, and TMR-Net showed only small gains. The persistent difficulty lay in moderate-order phase and quadrature-amplitude classes under intermediate SNR and mismatch. On Shift-18, the most improved classes after adaptation were 8-PSK, 16-QAM, and 64-QAM, where TMR-Net exceeded the residual baseline by 4.9, 5.3, and 5.8 points respectively. These gains were accompanied by a reduced tendency to confuse adjacent-order constellations when clipping and symbol-rate mismatch co-occurred [32]. The analog amplitude families also improved, but by smaller margins, suggesting that the remaining errors there are driven less by topology and more by irreducible ambiguity under severe offset. An unexpected finding was that the transformer-like baseline occasionally exceeded TMR-Net on very long target sequences above 448 samples when SNR exceeded 12 dB. Inspection showed that its attention layers exploited long-range repetition effectively in that easy regime. However, once SNR dropped or waveform mismatch increased, the advantage disappeared. This indicates that attention-based global context is not uniformly inferior; it is simply less compute-efficient and less stable than the searched multibranch topology under the combined distortions emphasized in this study.

Seed stability also supports the reliability of the reported findings. Across five retraining runs, the standard deviation of TMR-Net accuracy was 0.23 points on Source-11 and 0.31 points on Shift-18 zero-shot, both lower than those of the residual and transformer-like baselines. Calibration variance was reduced even more strongly [33]. This stability is consistent with the route-regularized search objective. By the time the architecture is discretized, many equivalent routes have already been suppressed, leaving a topology whose optimization landscape during retraining appears smoother. Search stability is similarly encouraging. Although exact operator probabilities differed, all searches selected large early receptive fields, moderate total depth, and a narrow final stage. These repeated outcomes do not prove global optimality, but they do indicate that the design conclusions are not artifacts of one seed. For radio architecture design, this is useful because it means the search process is discovering a recurring allocation strategy rather than a fragile configuration dependent on lucky initialization.

VII. Discussion

The experiments suggest that neural architecture design for radio classification benefits from thinking in terms of allocation rather than accumulation. Conventional design heuristics often assume that stronger models require either more layers, more channels, or more generic attention. The present results indicate a different pattern [34]. Radio waveforms are structured signals with a relatively low input dimensionality and a high sensitivity to channel-induced nuisance variation. Under these conditions, the main challenge is not insufficient nominal capacity. It is the placement of capacity. The search repeatedly chose to invest computation in early and mid-stage multiscale operators, to keep late stages comparatively

narrow, and to preserve second-order information near the classifier. This differs from common image-style scaling strategies and is consistent with the empirical observation that many radio classes become separable once the network recovers a stable temporal and geometric summary, after which additional deep expansion contributes little. The fact that a doubled-depth ablation did not improve accuracy reinforces this interpretation. More layers did not translate into more useful discriminative structure because the marginal layers mostly redistributed already-available information.

The role of covariance pooling deserves particular attention. In several ablations, replacing the covariance head with average pooling had a modest effect in-domain but a larger effect under waveform shift [35]. That pattern suggests that second-order channel interactions capture invariants that remain informative when first-order activation means drift. One plausible explanation is that modulation identity often depends on relational structure among learned filters rather than on the absolute magnitude of any single filter response. Frequency offset, clipping, and pulse-shape mismatch may alter marginals substantially while preserving some relative coupling patterns among channels tuned to phase progression, envelope dynamics, or transition regularity. The matrix-logarithm covariance head retains these relations in a compact form and seems to stabilize the classifier boundary when target conditions differ from source conditions. This does not imply that second-order heads are universally preferable. They add complexity and can be numerically delicate. In the present study, however, their benefit was consistent enough to view them as a substantive architectural component rather than an ancillary enhancement.

The findings on route regularization are also instructive [36]. When route entropy control was removed, source-domain accuracy barely changed, yet target-domain performance and calibration deteriorated. This asymmetry implies that branch redundancy is not always visible through in-domain top-line metrics. A supernet with diffuse route probabilities can fit the source distribution effectively, but the resulting discretized architecture is more likely to contain interchangeable or weakly specialized paths. Such paths increase model flexibility in a way that favors source-specific correlations and confidence overstatement. Penalizing route diffusion during search reduces this tendency and yields a more compact topology whose branch functions are better separated. In practice, this may matter as much as raw accuracy because radio systems often operate under unannounced shifts. A model that preserves confidence ordering and degrades predictably can be more useful than one that performs similarly in-domain but exhibits volatile posterior behavior after deployment.

There are nevertheless limitations that constrain the interpretation of the study. The datasets, while physically perturbed and substantially more varied than a single synthetic benchmark, are still generated environments [37]. Real over-the-air collections may contain hardware-specific distortions, cochannel interference, and nonstationary occupancy patterns not represented here. Search was also performed on the source corpus only. A multi-domain search objective could, in principle, produce a different topology if small amounts of target information were available during architecture selection. In addition, the covariance head, while effective, introduces matrix operations that may be less attractive for extremely low-power edge devices. The search cost is another consideration. Although the resulting model is efficient at inference, differentiable search still requires a nontrivial up-front budget. These limitations do not negate the observed empirical pattern, but they suggest that future work should examine whether similar topology preferences emerge on measured over-the-air datasets, under cochannel interference, and in search regimes that explicitly include domain shift during architecture selection.

VIII. Conclusion

This study examined neural architecture design for radio signal classification as a problem of allocating representational capacity across temporal scale, branch diversity, and channel geometry rather than simply increasing depth or width. The searched TMR-Net architecture combined multiscale temporal branches, route-controlled grouped processing, explicit I/Q recombination, and a covariance-preserving head [38]. Across a primary 11-class benchmark, a shifted 18-class corpus, and a low-label adaptation setting, the resulting topology consistently outperformed residual, recurrent-convolutional, separable, and transformer-like baselines while using less computation than the strongest convolutional comparator. The empirical pattern was stable across seeds and ablations. Early large receptive fields, moderate cardinality, nonuniform stage width, and second-order feature retention all contributed to the final behavior, with the transfer advantage becoming more pronounced as waveform mismatch increased.

The results also indicate that several commonly used simplifications in radio classifiers deserve re-examination. Late-stage widening was not necessary, homogeneous stacks were inferior to heterogeneous cells, and global average pooling discarded information that became important under domain shift. Route regularization improved transfer and calibration even when its effect on source-domain accuracy was small. At the same time, the work remains bounded by synthetic data and by the cost of search. A practical next step would be to test whether the same topology preferences hold for measured over-the-air corpora and for deployment settings involving interference, latency constraints, and severe class imbalance. Within the scope of the present experiments, the evidence supports a restrained conclusion: radio signal classification is improved more by disciplined topology design

and second-order feature preservation than by indiscriminate architectural scaling [39].

REFERENCES

[1] M. Li, Y. Ouyang, W. Kang, X. Che, and R. Ye, "Decision model of wireless communication scheme evaluation via interval number," *Security and Communication Networks*, vol. 2022, pp. 1–10, 6 2022.

[2] I. Hussain, S. Young, C. H. Kim, H. C. M. Benjamin, and J. Park, "Quantifying physiological biomarkers of a microwave brain stimulation device.," *Sensors (Basel, Switzerland)*, vol. 21, pp. 1896–, 3 2021.

[3] F. Wang, Y. Zhou, H. Yan, and R. Luo, "Enhancing the generalization ability of deep learning model for radio signal modulation recognition," *Applied Intelligence*, vol. 53, no. 15, pp. 18758–18774, 2023.

[4] R. M. Alonso, D. Plets, M. Deruyck, L. Martens, G. G. Nieto, and W. Joseph, "Dynamic interference optimization in cognitive radio networks for rural and suburban areas," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–16, 5 2020.

[5] A. Bouhlel and A. Sakly, "Impact of iq imbalance on ris-assisted siso communication systems," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–12, 8 2021.

[6] H. Nguyen, V. H. Nguyen, C. H. Nguyen, V. Bui, and Y. M. Jang, "Design and implementation of 2d mimo-based optical camera communication using a light-emitting diode array for long-range monitoring system," *Sensors (Basel, Switzerland)*, vol. 21, pp. 3023–, 4 2021.

[7] R. S. M. Soboh, A. H. H. Al-Masoodi, F. N. A. Erman, A. H. H. Al-Masoodi, B. Nizamani, H. Arof, R. Apsari, and S. W. Harun, "Mode-locked ytterbium-doped fiber laser with zinc phthalocyanine thin film saturable absorber.," *Frontiers of optoelectronics*, vol. 15, pp. 28–28, 6 2022.

[8] A. Dobroiu, Y. Shirakawa, S. Suzuki, M. Asada, and H. Ito, "Subcarrier frequency-modulated continuous-wave radar in the terahertz range based on a resonant-tunneling-diode oscillator," *Sensors (Basel, Switzerland)*, vol. 20, pp. 6848–, 11 2020.

[9] A. van Niekerk, E. M. Meintjes, and A. van der Kouwe, "A wireless radio frequency triggered acquisition device (wrad) for self-synchronised measurements of the rate of change of the mri gradient vector field for motion tracking," *IEEE transactions on medical imaging*, vol. 38, pp. 1610–1621, 1 2019.

[10] N. Ahmadi and Z. K. Jahromi, "Remote reading of electricity meters using plc," *Bulletin of Electrical Engineering and Informatics*, vol. 9, pp. 466–472, 4 2020.

[11] U. Saeed, S. Y. Shah, A. Zahid, J. Ahmad, M. Imran, Q. H. Abbasi, and S. A. Shah, "Wireless channel modelling for identifying six types of respiratory patterns with sdr sensing and deep multilayer perceptron," *IEEE sensors journal*, vol. 21, pp. 20833–20840, 7 2021.

[12] F. Turčinović, G. Šišul, and M. Bosiljevac, "Lorawan base station improvement for better coverage and capacity," *Journal of Low Power Electronics and Applications*, vol. 12, pp. 1–1, 12 2021.

[13] M. Fournelle, T. Grün, D. Speicher, S. Weber, M. Yilmaz, D. S. Schoeb, A. Miernik, G. Reis, S. Tretbar, and H. Hewener, "Portable ultrasound research system for use in automated bladder monitoring with machine-learning-based segmentation," *Sensors (Basel, Switzerland)*, vol. 21, pp. 6481–, 9 2021.

[14] A. M. Al-awadi and M. J. Al-Dujaili, "Simulation of lte-tdd in the haps channel," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, pp. 3152–3157, 6 2020.

[15] D. M. Soleymani, M. R. Gholami, G. D. Galdo, J. Mueckenheim, and A. Mitschele-Thiel, "Open sub-granting radio resources in overlay d2d-based v2v communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, 5 2022.

[16] S. Colombo, V. Lebedev, A. Tonyushkin, S. Pengue, and A. Weis, "Imaging magnetic nanoparticle distributions by atomic magnetometry-based susceptometry," *IEEE transactions on medical imaging*, vol. 39, pp. 922–933, 8 2019.

[17] J. Zaini-Desevedavy, F. Hameau, T. Taris, D. Morche, and P. Audebert, "An ultra-low power 28 nm fd-soi low noise amplifier based on channel aware receiver system analysis," *Journal of Low Power Electronics and Applications*, vol. 8, pp. 10–, 4 2018.

[18] I. Aziz, J. Hanning, E. Öjefors, D. Wu, and D. Dancila, "28 ¡scp¿ghz¡/scp¿ circular polarized fan-out antenna array with wide-angle beam-steering," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 32, 12 2021.

[19] S. Maudet, G. Andrieux, R. Chevillon, and J.-F. Diouris, "Refined node energy consumption modeling in a lorawan network," *Sensors (Basel, Switzerland)*, vol. 21, pp. 6398–, 9 2021.

[20] W. H. Jeong, H.-R. Choi, and K.-S. Kim, "Empirical path-loss modeling and a rf detection scheme for various drones," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–17, 12 2018.

[21] A. Farughian, A. Poluektov, A. Pinomaa, J. Ahola, A. Kosonen, L. Kumpulainen, and K. Kauhaniemi, "Power line signalling based earth fault location," *The Journal of Engineering*, vol. 2018, pp. 1155–1159, 8 2018.

[22] Y. Okumura, S. Suyama, J. Mashino, and K. Muraoka, "Recent activities of 5g experimental trials on massive mimo technologies and 5g system trials toward new services creation," *IEICE Transactions on Communications*, vol. 102, pp. 1352–1362, 8 2019.

[23] S. Mehmood, A. N. Malik, I. M. Qureshi, M. Z. U. Khan, and F. Zaman, "A novel deceptive jamming approach for hiding actual target and generating false targets," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–20, 4 2021.

[24] Q.-H. Zhang, E.-Y. Zhang, and S.-H. Zhang, "Direction-of-arrival estimation in time-modulated linear arrays based on the mt-bcs approach," *International Journal of Antennas and Propagation*, vol. 2022, pp. 1–6, 12 2022.

[25] A. S. Daghal and H. M. T. Al-Hilfi, "Lte network performance evaluation based on effects of various parameters on the cell range and mapl," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, pp. 1770–1776, 6 2021.

[26] J. Lee, Y. S. Yoon, H. W. Oh, and K. R. Park, "Dg-lora: Deterministic group acknowledgment transmissions in lora networks for industrial iot applications.," *Sensors (Basel, Switzerland)*, vol. 21, pp. 1444–, 2 2021.

[27] T. Siegel and S.-P. Chen, "Investigations of free space optical communications under real-world atmospheric conditions," *Wireless Personal Communications*, vol. 116, pp. 475–490, 8 2020.

[28] L. Xue, "Financial big data based on internet of things and wireless network communication," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–12, 10 2021.

[29] N. Polyakov and A. Platonova, "Assessing latency of packet delivery in the 5g 3gpp integrated access and backhaul architecture with half-duplex constraints," *Future Internet*, vol. 14, pp. 345–345, 11 2022.

[30] Łukasz Kułacz, P. Kryszkiewicz, and A. Kliks, "Waveform flexibility for network slicing," *Wireless Communications and Mobile Computing*, vol. 2019, pp. 6250804–1–6250804–15, 3 2019.

[31] E. García-Muñoz, K. A. Abdalmalak, G. Santamaría, A. Rivera-Lavado, D. Segovia-Vargas, P. Castillo-Aranibar, F. van Dijk, T. Nagatsuma, E. R. Brown, R. Guzman, H. Lamela, and G. Carpintero, "Photonic-based integrated sources and antenna arrays for broadband wireless links in terahertz communications," *Semiconductor Science and Technology*, vol. 34, pp. 054001–, 4 2019.

[32] A. Al-Safi, "A reduced size look up table for sinusoidal wave generation in digital modulators applications," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 6, pp. 39–46, 10 2018.

[33] S. Kalantaievska, O. Kuvshynov, A. Shyshatskyi, O. Salnikova, Y. Punda, P. Zhuk, O. Zhuk, H. Drobakha, L. Shabanova-Kushnarenko, and S. Petruk, "Development of a complex mathematical model of the state of a channel of multi-antenna radio communication systems," *Eastern-European Journal of Enterprise Technologies*, vol. 3, pp. 21–30, 5 2019.

[34] Q. Huang, X. Xie, and H. Tang, "Crc-afc-based dynamic spectrum resource optimisation for urllc access in 5g-advanced and 6g," *IET Signal Processing*, vol. 16, pp. 885–896, 2 2022.

[35] T. Lamprecht, F. Betschon, J. Bauwelinck, J. V. Kerrebrouck, J. Lambrecht, H. Gaul, A. Eichler, and X. Yin, "Electronic-photonic board as an integration platform for tb/s multi-chip optical communication," *IET Optoelectronics*, vol. 15, pp. 92–101, 3 2021.

[36] M. Vidmar, "Extending leeson's equation," *Informacije MIDEM - Journal of Microelectronics, Electronic Components and Materials*, vol. 51, pp. 146–, 7 2021.

[37] N. A. A. Kadir, N. A. A. Norizan, and A. Hamzah, "Pulse train stability in passively mode-locked erbium-doped fiber laser with a nonlinear polarization rotation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, pp. 1212–1216, 9 2019.

[38] D. S. Ilcev, "Implementation of e-education in africa via space digital video broadcasting system," *Bulletin of Electrical Engineering and Informatics*, vol. 9, pp. 1074–1084, 6 2020.

[39] H. Hu, G. Yu, Y. Li, Y. Qiu, H. Zhu, M. Zhu, and H. Zhou, "Design of a radial vortex-based spin-torque nano-oscillator in a strain-mediated multiferroic nanostructure for bfsk/bask applications.," *Micromachines*, vol. 13, pp. 1056–1056, 6 2022.