Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Limits of Semi-Supervised Learning for Modulation Classification in Software-Defined Radios

Youcef Mahdjoub [1, 2, *] Belkacem Benadda [3]

[1] *Dept. of Mathematics and Computer Science, University of Ghardaia, Algeria*

[2] *Laboratoire de Mathématiques et Sciences Appliquées (LMSA), University of Ghardaia, Algeria*

[3] *Dept. of Electrical and Electronics Engineering, Faculty of Technology, Aboubekr Belkaid University, Tlemcen, Algeria*

[*] *Corresponding Author:*

ARTICLE INFO

Received: 12 May 2026

Revised: 01 July 2026

Accepted: 14 July 2026

ABSTRACT

Automatic modulation classification (AMC) from raw I/Q samples is an enabling function of reconfigurable software-defined radio (SDR) receivers, spectrum monitoring, and cognitive radio: before a multi-standard platform can adapt its processing chain to a signal, it must identify what it is receiving. Labeling radio captures, however, requires expert annotation or cooperative transmitters, while unlabeled recordings accumulate nearly for free on any SDR platform. Semi-supervised learning (SSL) promises to exploit this asymmetry, and prior SSL-AMC studies report consistent gains, typically, however, against unaugmented supervised baselines, on inconsistent data splits, and with single-seed results. We re-examine the question under a rigorous protocol: one shared, SNR-stratified split for every method, a strong supervised reference using the full augmentation and regularization stack, and three seeds behind every headline number. We build a MixMatch-based recipe for raw I/Q signals (signal-domain strong augmentation, a per-(class, SNR) adaptive confidence threshold, and a Mean-Teacher consistency loss) and sweep the labeled budget on RML2016.10a. The recipe delivers large gains in the genuinely label-scarce regime: +7.8 accuracy points over the strong supervised baseline at 1% labels (200 per class), closing 44% of the gap to the fully supervised ceiling, and matching supervised training that uses five times as many labels. The advantage decays monotonically and vanishes between 5% and 10% labels, where every classical SSL baseline we evaluate (self-training, co-training, semi-supervised GAN) falls below the strong supervised reference. Ablations reveal that the gain is an interaction effect: neither strong augmentation nor Mean-Teacher helps alone; the augmentations act as the perturbation model that consistency training requires. We release code, per-SNR results, and the full protocol.

Keywords: *automatic modulation classification, semi-supervised learning, consistency regularization, software-defined radio, reconfigurable receivers, deep learning, RadioML.*

INTRODUCTION

Software-defined radio (SDR) has shifted transmission systems from fixed-function hardware toward software-based architectures in which a single reconfigurable platform supports multiple standards, waveforms, and services [20], [21]. The promise of such multiplatform telecommunication devices rests on reconfiguration: the same receiver adapts its transmission and reception parameters to the user's needs, the requested service, or the surrounding spectrum, communicating efficiently while avoiding interference with authorized and unauthorized users alike. Reconfiguration, however, presupposes awareness. Before a receiver can adapt its processing chain to a signal, it must identify what it is receiving, which makes automatic modulation classification (AMC), the identification of a signal's modulation scheme directly from baseband I/Q samples, a core enabling primitive of the SDR paradigm, alongside its established roles in spectrum monitoring, interference identification, cognitive radio [21], and signals intelligence.

Deep networks trained end-to-end on raw I/Q now dominate AMC [1], [2], but their appetite for labeled examples collides with an awkward property of the radio domain: producing a labeled capture requires either expert annotation of recorded spectrum or ground-truth access to the transmitter, while unlabeled I/Q recordings accumulate essentially for free on any SDR platform. This asymmetry is the textbook setting for semi-supervised learning.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A growing body of work applies SSL to AMC: GAN-based discriminative training, iterative pseudo-labeling, and more recently consistency- and contrastive-based methods [12]–[15], [22]–[28]. Reported gains are consistent, but three methodological weaknesses recur. First, the supervised baseline is typically trained without data augmentation or modern regularization, although these are free to apply and dramatically strengthen supervised performance at moderate label counts. Second, comparisons frequently mix data splits across methods, so test sets are not identical. Third, results are usually single-seed, while the effects claimed (1–3 points) are of the same order as seed variance. Oliver et al. [3] showed for image classification that SSL gains reported under such conditions often evaporate against properly tuned supervised baselines; whether this critique applies to radio data, and if so, where the genuine SSL advantage begins and ends, has not been quantified.

This paper answers that question with a measurement campaign designed for auditability. Every experiment uses one shared 70/15/15 split of RML2016.10a, stratified jointly by modulation and SNR from a fixed seed, so all methods are scored on an identical 33,000-sample test set. The supervised reference is deliberately strong: random-crop/flip/noise augmentation, label smoothing, AdamW, learning-rate scheduling, and early stopping. Fully supervised training on 100% of labels provides the ceiling. Against these anchors we evaluate three classical SSL families (self-training, co-training, semi-supervised GAN) and a consistency-based recipe we construct from validated components: MixMatch [5] with distribution alignment [6], signal-domain strong augmentation, a FreeMatch-style [8] per-(class, SNR-bin) adaptive confidence threshold, and a Mean-Teacher [4] consistency loss (Fig. 1), then sweep the labeled budget over 1%, 2%, 5%, and 10% of the dataset with three seeds per configuration.

Our contributions are findings, not machinery:

1) A consistency-training recipe for raw I/Q whose value is an interaction effect. Signal-domain strong augmentation and Mean-Teacher consistency are individually neutral-to-harmful (−1.3 and −0.7 points versus the MixMatch baseline at 10% labels), yet jointly worth +2.5 points over that baseline and +3.3 points (≈4σ) over Mean-Teacher alone. The label-preserving I/Q transformations are best understood not as data augmentation but as the perturbation model that makes consistency regularization work for radio signals.

2) The first label-budget sweep for SSL-AMC against a strong supervised reference: +7.8 points at 1% labels (43.6% of the supervised gap closed), +5.2 at 2%, +2.7 at 5%, −0.2 at 10%. Unlabeled data acts as a ≈5× label multiplier: SSL with 200 labels per class matches supervised training with 1,000.

3) A domain-specific confirmation of the Oliver et al. critique: at 10% labels, every SSL method evaluated, including ours, ties or falls below the strong supervised reference; self-training and co-training trail it by 6.2 and 4.0 points.

4) Two analyzed negative results: two-view (I/Q + FFT) consistency through a shared backbone degrades accuracy by ≈9 points via BatchNorm-statistics corruption across input domains; class-debiased pseudo-label weighting is redundant once distribution alignment is applied.

We state the punchline where a reviewer will otherwise discover it: at 10% labels our method merely ties a strong supervised model, and that is the point. The study maps both sides of the crossover, which is what a practitioner deciding whether to deploy SSL actually needs. Section II reviews related work, Section III details the recipe, Section IV presents the campaign, Section V discusses implications, and Section VI concludes.

RELATED WORK

A. Deep Learning for AMC

O'Shea et al. introduced end-to-end convolutional classification of raw I/Q modulation data along with the RadioML datasets [1], later extended to over-the-air captures and deeper residual architectures [2]. Subsequent work explored recurrent, residual, and attention-augmented variants, with SE-enhanced residual networks on raw I/Q a common strong configuration [17], [18]; feature-domain alternatives (spectrograms, constellation images, cyclostationary features) trade end-to-end simplicity for inductive bias. RML2016.10a [1] remains the standard benchmark for controlled comparison: 220,000 length-128 I/Q frames across 11 modulations and 20 SNR levels. We adopt raw-I/Q SE-ResNet as the shared backbone for every method in this study.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

B. Semi-Supervised Learning

Modern SSL rests on two ideas. Pseudo-labeling trains on a model's own confident predictions [10], refined by confidence thresholds and, recently, self-adaptive per-class thresholds (FreeMatch [8]). Consistency regularization enforces stable predictions under input or model perturbation: the Π-model and temporal ensembling [9], Mean-Teacher's EMA-weight teacher [4], and the holistic combinations MixMatch [5], ReMixMatch (distribution alignment, augmentation anchoring) [6], and FixMatch (weak-to-strong consistency) [7]. MixUp [16] regularizes both objectives. Critically for this paper, Oliver et al. [3] demonstrated that SSL evaluations are frequently confounded by under-tuned supervised baselines and non-shared protocols, and recommended exactly the controls we adopt: identical splits, strong baselines, budget sweeps, and variance reporting.

C. SSL for AMC

Applications of SSL to modulation classification have followed the evolution of the general SSL literature. Early work is GAN-based: a discriminator's class head is trained on the labeled subset while real/fake discrimination shapes features from unlabeled signals [12], [13], [22], [23]. A second family uses reconstruction or transfer: autoencoder features trained on unlabeled signals shared with a supervised classifier [24], and purpose-built loss designs such as SSRCNN [14]. More recent methods import contrastive pre-training (SemiAMC's self-contrastive framework [15], transformer-based contrastive SSL [25]) or consistency-and-pseudo-label training closest to ours: SemiAMR combines a Mean-Teacher framework with corrected pseudo-labels and mixed-sample augmentation on STFT inputs [26]; Ensemble SigMatch applies FixMatch-style pseudo-labeling with signal-domain augmentations and multi-view consistency [27]; and SSwsrNet couples a deep residual shrinkage network with a MixMatch module [28].

These works consistently report SSL gains, several on the same RML2016.10a/10b benchmarks we use [25]–[27]. Surveying their protocols, however, the supervised reference is typically trained without data augmentation or modern regularization, splits and preprocessing differ across compared methods, and single-seed gains of 1–3 points are common, precisely the conditions under which Oliver et al.'s critique [3] bites. None reports where its advantage vanishes as the label budget grows. Our contribution is therefore orthogonal: not a new loss, but a controlled map of when the existing machinery pays off (with shared splits, a strong augmented supervised anchor, seed variance, and a label-budget sweep), together with the identification of the augmentation×consistency interaction that determines whether it pays off at all.

METHODS

A. Problem Statement

Let $D = (x_i, y_i)$with $x_i \in R^{(2\times128)}$a baseband I/Q frame and $y_i \in 1..C, C = 11$. A fraction ρ of the dataset is labeled; the remaining training frames are available without labels. Each frame's SNR s_i is known (a mild assumption, since SNR is estimable from the signal itself) and is used only to bin the adaptive threshold (Sec. III-D). All frames are standardized per sample to zero mean and unit variance. The objective is accuracy on a held-out test partition, reported per SNR; we headline the mean over SNR >= 0 dB, the operationally relevant plateau.

B. Base Objective: MixMatch with Distribution Alignment and Confidence Masking

Our starting point (the baseline of every ablation) is MixMatch [5] with two established refinements. For each unlabeled batch, an EMA teacher [4] with decay 0.999 predicts on two weakly augmented views (additive Gaussian noise, σ = 0.03); the averaged prediction is corrected by distribution alignment [6] with a running marginal (momentum 0.99), sharpened at temperature T = 0.5, and used as the soft target for two strongly augmented student views. Labeled and unlabeled examples are jointly MixUp-ed (α = 0.75) [16]; the supervised term is soft cross-entropy on the labeled slice, and the unsupervised term is a Brier (squared-error) loss on the unlabeled slice, masked per sample by teacher confidence (τ = 0.90 on the unsharpened probability) and weighted by λ_u ramped linearly to 10 over 50 epochs. The EMA teacher is also the evaluation model, and model selection is by validation accuracy.

The full training objective combines the supervised loss, the masked unlabeled loss, and the Mean-Teacher consistency term of Sec. III-E:

$$L = L_x + \lambda_u(t)\, L_u + \lambda_{MT}\, L_{MT} \quad (1)$$

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

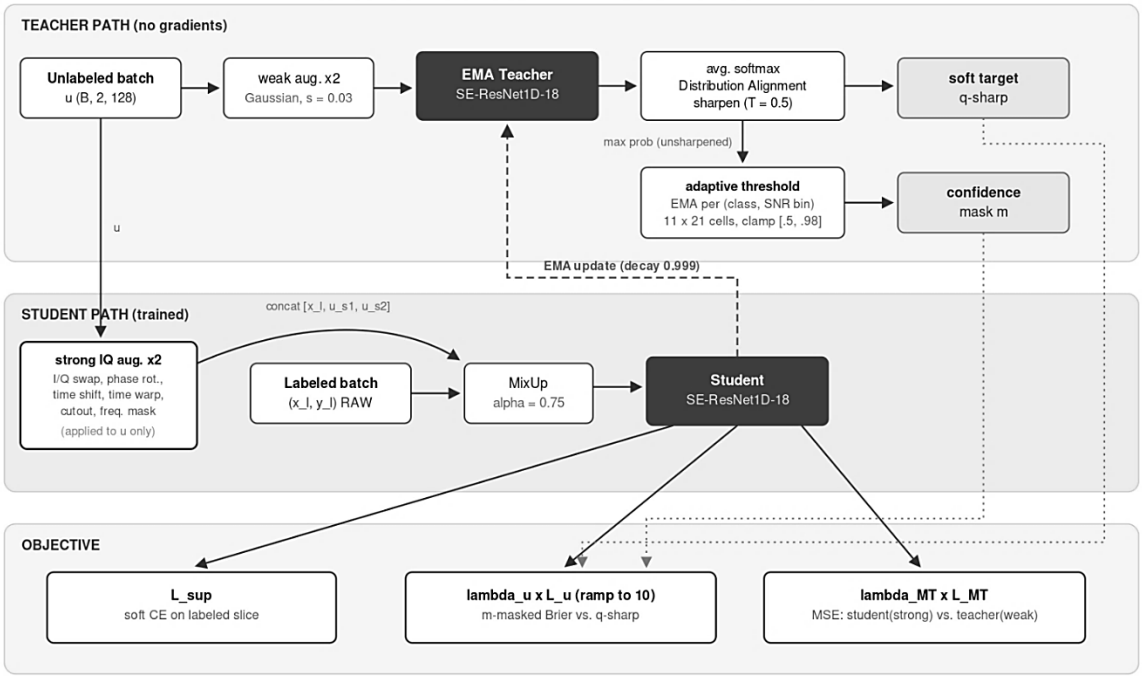

Figure 1. Overview of the proposed recipe. The EMA teacher pseudo-labels weakly augmented unlabeled views (with distribution alignment, sharpening, and a per-(class, SNR-bin) adaptive confidence mask); the student trains on strongly augmented views through MixUp, with an additional Mean-Teacher consistency term. Labeled samples are fed raw.

C. Component 1: I/Q-Aware Strong Augmentation

The strong student views are produced by a stochastic composition of six label-preserving signal transformations: I/Q channel swap (constellation symmetry), carrier-phase rotation $\varphi \sim U(-\pi/6, \pi/6)$, circular time shift up to ±10% (timing offset), mild non-linear time warp (clock jitter), time cutout up to 15% (burst interference), and frequency-band masking via FFT (narrowband interference). Each transform is applied independently with probability 0.4–0.7. Two design points matter. First, labeled samples are fed raw: these transformations exist solely to perturb the unlabeled student views. Second, each transformation corresponds to a physical channel or hardware impairment, so label preservation is an engineering argument, not an assumption imported from vision.

D. Component 2: Per-(Class, SNR-Bin) Adaptive Threshold

A single global confidence threshold either starves low-SNR samples (which rarely reach $\tau = 0.90$) or admits noise. Following the FreeMatch principle [8], we maintain a running EMA of teacher confidence per (predicted class, 2 dB SNR bin) cell, a table of 11 × 21 scalars, and use it as a per-sample threshold, clamped to [0.5, 0.98]. The SNR axis lets high-SNR and low-SNR populations settle at different operating points; the class axis absorbs per-class calibration differences.

Concretely, the threshold for cell (class c, SNR bin b) is updated as a running mean of teacher confidence:

$$\tau_{c,b} := \mu\,\tau_{c,b} + (1 - \mu)\,E[\,max_c\, q_i \mid c, b\,],\ \ \mu = 0.99 \qquad (2)$$

E. Component 3: Mean-Teacher Consistency

In addition to the MixMatch objective, we add an explicit consistency term: the squared distance between the student's softmax on a strongly augmented view and the EMA teacher's softmax on a weakly augmented view of the same unlabeled frame, weighted by $\lambda_{MT} = 1$. The term re-uses the teacher already maintained for pseudo-labeling, adding no parameters and one forward pass. As Section IV-D shows, this term is inert when student views are weak: its entire value emerges in combination with the strong augmentation of Sec. III-C. Two further candidates

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(consistency between I/Q and FFT-magnitude views through the shared backbone, and inverse-frequency pseudo-label re-weighting) were evaluated and rejected; they are described with the ablation in Section IV-D.

F. Implementation

SE-ResNet1D-18 backbone (Fig. 2); Adam, learning rate 1e-3, weight decay 1e-4, cosine annealing; batch 128; 150 epochs, where one epoch is defined as a fixed 171 optimization steps (the labeled-loader length at ρ = 10%) with loaders cycled, so every label budget receives the same optimization budget. Code and per-SNR results are released.

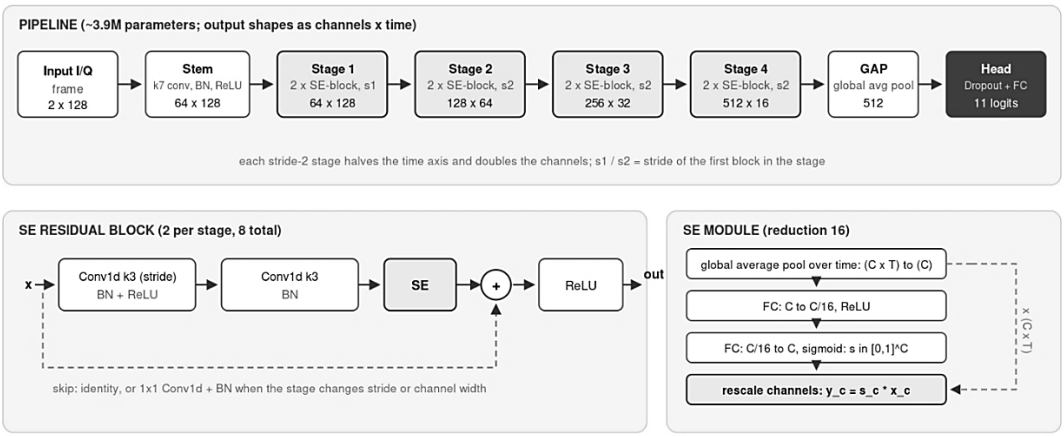

Figure 2. Structure of the SE-ResNet1D-18 backbone (~3.9M parameters) shared by all methods: the stage pipeline with output shapes, the SE residual block, and the squeeze-and-excitation module (reduction 16).

EXPERIMENTS AND RESULTS

A. Dataset and Experimental Setup

We evaluate on RadioML 2016.10a [1]: 220,000 complex baseband signals of length 128 spanning 11 modulation classes at 20 SNR levels from −20 dB to +18 dB (1,000 examples per class–SNR pair). Each example is standardized per sample (the SGAN baseline peak-normalizes to [−1, 1] to match its Tanh generator; the partition is unaffected). The data is partitioned 70/15/15 into train/validation/test, stratified jointly by modulation and SNR with a fixed split seed; every method, ablation, and label budget is evaluated on the identical 33,000-sample test set. Labeled subsets of 1%, 2%, 5%, or 10% of the dataset (200–2,000 per class) are drawn from the training partition by the same stratified procedure. All methods share the backbone of Sec. III-F. The supervised reference uses a strong recipe: crop/flip/noise augmentation, label smoothing 0.1, AdamW, ReduceLROnPlateau, early stopping (patience 15), up to 300 epochs at small budgets so it is never undertrained. Multi-seed experiments repeat training with three seeds controlling initialization and training stochasticity only (the split is never re-sampled) and report mean ± std.

B. Main Result: When Does Semi-Supervision Pay Off?

Table 1 and Fig. 3 sweep the labeled fraction with everything else fixed.

Table 1. Accuracy at SNR >= 0 dB vs. label budget (mean ± std, 3 seeds)

Gap closed is relative to the fully supervised ceiling of 90.19 ± 0.09%.

Labels (per class)	Supervised only	Proposed SSL	Δ	Gap closed
1% (200)	72.42 ± 0.74	80.16 ± 0.87	+7.75	43.6%
2% (400)	76.78 ± 0.24	82.00 ± 0.23	+5.23	39.0%
5% (1,000)	80.33 ± 1.00	83.01 ± 0.36	+2.68	27.1%
10% (2,000)	83.90 ± 0.48	83.74 ± 0.84	−0.16	≈0%

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

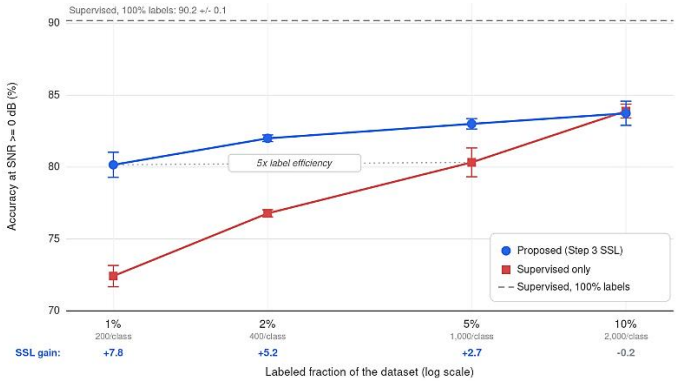

Figure 3. Accuracy at SNR >= 0 dB vs. labeled fraction (log scale, mean ± std over 3 seeds, identical test set). The proposed SSL method gains +7.8 points at 1% labels, matching supervised training with 5× more labels; a strong supervised baseline catches up by 10%.

Three observations. First, semi-supervision is highly effective in the genuinely label-scarce regime: at 1% labels the recipe gains 7.8 points (≈9σ), closing 44% of the gap to the ceiling. Second, unlabeled data acts as a label multiplier: SSL at 1% (80.16%) matches supervised at 5% (80.33%), a ≈5× efficiency gain, and at 2% surpasses it. The SSL curve is remarkably flat (80.2 to 83.7 across a 10× change in labels), indicating the method extracts most of the achievable decision structure from the unlabeled pool alone. Third, the advantage vanishes by 10% labels, where a well-regularized supervised model is statistically indistinguishable from the best SSL configuration, consistent with Oliver et al. [3], here quantified for radio: the crossover lies between 5% and 10% labels.

C. Comparison with SSL Baselines (10% Labels)

Table 2 and Fig. 4 compare classical SSL families under the identical split, backbone, and budget: iterative self-training (confidence 0.95, class-balanced selection, down-weighted pseudo-labels), co-training over I/Q and FFT-magnitude views with cross pseudo-labeling (averaged-softmax ensemble), and a semi-supervised GAN (feature-matching generator [12], shared trunk with real/fake and class heads).

Table 2. Overall and plateau accuracy at 10% labels

Method	Overall (%)	SNR >= 0 dB (%)	Seeds
Self-Training	51.45	77.73	1
Co-Training (I/Q+FFT ensemble)	52.72	79.87	1
SGAN	53.84	82.38	1
Proposed	**54.56**	**83.74 ± 0.84**	**3**
Supervised, same 10% labels	56.57	83.90 ± 0.48	3
Supervised, 100% labels	61.06	90.19 ± 0.09	3

The proposed method is the strongest SSL approach (+6.0 over self-training, +3.9 over co-training, +1.4 over SGAN at the plateau). More striking is that every classical SSL baseline falls below the strong supervised reference at this budget. Combined with Sec. IV-B, this yields the paper's practical guidance: at 2,000 labels per class, train a well-regularized supervised model; below ≈1,000, consistency-based SSL delivers large, reproducible gains.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

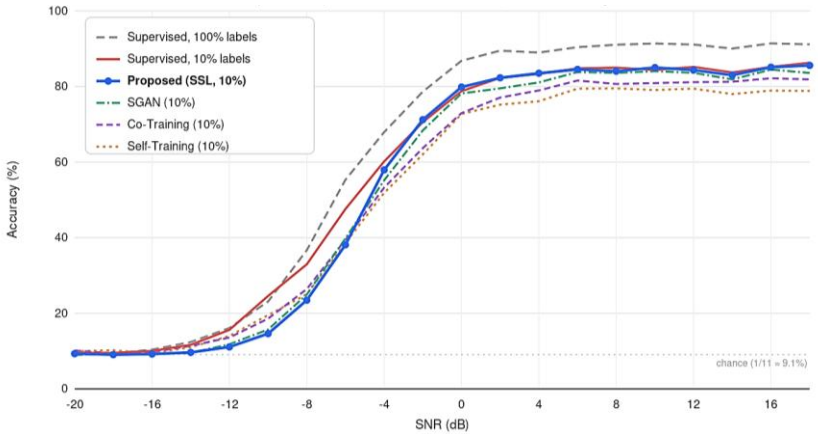

Figure 4. Accuracy vs. SNR at 10% labels for all methods on the shared split. Proposed and supervised anchors are means over 3 seeds.

D. Ablation and Component Analysis (10% Labels)

TABLE 3. Ablation at 10% labels

Cumulative rows add components in order; targeted rows isolate interactions.

Configuration	Overall (%)	SNR >= 0 dB (%)	Seeds
MixMatch baseline	53.69	81.21	1
+ I/Q-aware strong aug.	53.34	79.93	1
+ adaptive threshold	53.49	79.71	1
+ Mean-Teacher (= proposed)	**54.56**	**83.74 ± 0.84**	**3**
+ two-view (I/Q+FFT)	49.48	74.59	1
+ class-debiasing (cumulative)	49.23	74.26	1
Mean-Teacher only	52.32	80.47 ± 0.67	3
Proposed + class-debiasing	54.42	83.36 ± 0.84	3

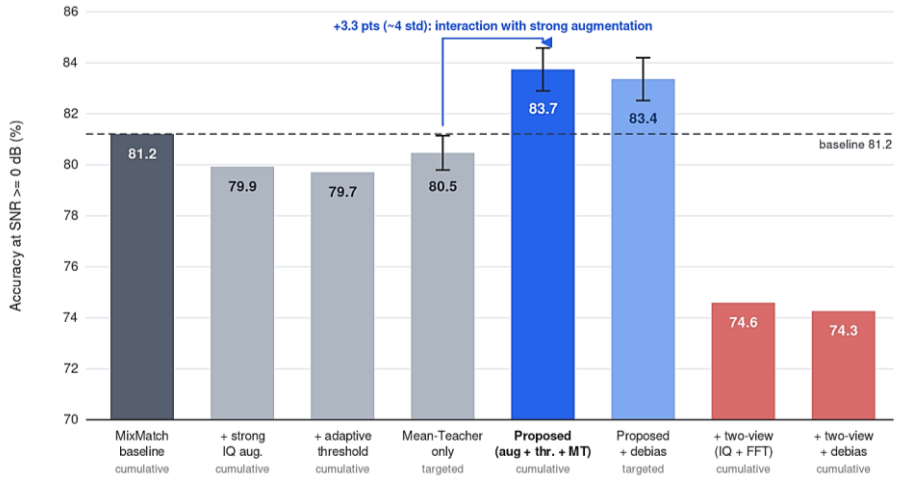

Figure 5. Ablation at 10% labels. Neither strong augmentation nor Mean-Teacher helps alone; their combination produces the entire gain. Two-view consistency through a shared backbone is harmful.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The gain is an interaction, not a sum (Fig. 5). Strong augmentation and adaptive thresholding reduce accuracy when added alone. Mean-Teacher on the bare baseline, where students see only light Gaussian noise, is indistinguishable from the baseline (80.47 ± 0.67). The same Mean-Teacher loss consuming strongly augmented student views reaches 83.74 ± 0.84, a 3.3-point (≈4σ) jump. When teacher and student see nearly identical inputs, the consistency loss carries no learning signal; the strong, label-preserving I/Q transformations create precisely the prediction disagreement the teacher–student objective needs.

Class-debiasing is neutral (−0.4, within one std): distribution alignment already corrects most pseudo-label imbalance. **Two-view consistency is harmful in a shared backbone** (−9.2 points) even with the FFT view log-compressed, standardized identically to the I/Q input, trained against detached pseudo-labels, confidence-masked, and weight-ramped. We attribute this to a single set of BatchNorm statistics [19] serving two input distributions; in a preliminary experiment feeding the FFT view unnormalized, training collapsed to chance entirely, confirming statistics corruption as the mechanism. Domain-specific normalization is a plausible remedy left to future work.

E. Threats to Validity

Results are from a single dataset; validating the crossover on RML2016.10b or 2018.01A is a natural extension. Single-seed rows in Tables 2–3 carry the ±0.2–1.0-point seed variability observed in multi-seed runs; all headline comparisons rest on three seeds. The supervised reference and the SSL recipe differ in supervised ingredients (the reference uses labeled-data augmentation and label smoothing; the recipe feeds labeled samples raw), which if anything biases the 10% comparison against SSL. The 15% validation split used for model selection exceeds the labeled budget at every fraction; genuinely label-scarce deployments need smaller validation protocols.

DISCUSSION

Why the SSL curve is flat. Between 1% and 10% labels (a 10× change), the recipe moves only from 80.2% to 83.7%, while the supervised model moves from 72.4% to 83.9%. The unlabeled pool (132,000–152,000 frames) evidently determines most of the achievable decision structure; labels mainly calibrate it. This reframes the method as unlabeled-data-dominant: its accuracy is set by the corpus, and the labeled set buys the last few points.

Practical guidance. For an 11-class deployment on RML2016.10a-like data: below ≈1,000 labels per class, the recipe converts free unlabeled recordings into the equivalent of roughly 4× additional annotations; above ≈1,000–2,000 per class, skip SSL and spend the effort on augmentation and regularization of a supervised model. The crossover, not any single accuracy number, is the study's transferable output.

Deployment context: reconfigurable SDR receivers. The recipe fits the SDR workflow that motivated this work. An SDR receiver captures unlabeled I/Q on-platform at negligible cost; the classifier, a software component of the receiver stack, is retrained with a few hundred labels per class and redeployed as a software update, with no change to the underlying platform. In a multi-standard device, the same mechanism extends naturally to recognizing the waveforms of the successive standards the receiver is reconfigured to support, which is precisely the label-scarce, unlabeled-data-rich regime where Section IV-B shows SSL pays off.

Why the classical baselines underperform. Self-training and co-training accumulate pseudo-label errors: ≈40% of the unlabeled pool sits below −8 dB where predictions approach chance, and hard selection followed by retraining compounds early mistakes. The SGAN's discriminative features are shaped by a real/fake task only loosely aligned with class structure, and its preprocessing differs by necessity (Tanh generator), a documented confound. Consistency-based training sidesteps both failure modes: soft targets, per-sample masking, and no irreversible selection.

Negative results as design lessons. The two-view failure is a caution for any multi-domain input scheme sharing BatchNorm; the debiasing redundancy suggests distribution alignment already occupies that niche in MixMatch-family objectives.

CONCLUSION

We asked when semi-supervised learning pays off for modulation classification and answered with a measured map rather than a single claim: consistency-based SSL is a ≈5× label multiplier below roughly 1,000 labels per class (+7.8

points at 200 per class against a strong supervised baseline), the advantage decays monotonically, and it vanishes between 5% and 10% of RML2016.10a, a budget at which every classical SSL baseline also fails to beat strong supervised training. The mechanism behind the low-label gains is an interaction: Mean-Teacher consistency is inert without signal-domain strong augmentation, which supplies the perturbation model the objective requires. Future work: validating the crossover on RML2016.10b/2018.01A and over-the-air captures; domain-specific normalization to rehabilitate multi-view consistency; harmonizing supervised ingredients into the SSL recipe; sub-1% budgets; and label-scarce validation protocols.

CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

REFRENCES

[1] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in Proc. EANN, 2016.
[2] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learningbased radio signal classification," IEEE J. Sel. Topics Signal Process., vol. 12, no. 1, 2018.
[3] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in Proc. NeurIPS, 2018.
[4] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in Proc. NeurIPS, 2017.
[5] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in Proc. NeurIPS, 2019.
[6] D. Berthelot et al., "ReMixMatch: Semi-supervised learning with distribution alignment and augmentation anchoring," in Proc. ICLR, 2020.
[7] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in Proc. NeurIPS, 2020.
[8] Y. Wang et al., "FreeMatch: Self-adaptive thresholding for semi-supervised learning," in Proc. ICLR, 2023.
[9] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in Proc. ICLR, 2017.
[10] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in ICML Workshop, 2013.
[11] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in Proc. COLT, 1998.
[12] T. Salimans et al., "Improved techniques for training GANs," in Proc. NeurIPS, 2016.
[13] A. Odena, "Semi-supervised learning with generative adversarial networks," arXiv:1606.01583, 2016.
[14] Y. Dong, X. Jiang, L. Cheng, and Q. Shi, "SSRCNN: A semi-supervised learning framework for signal recognition," IEEE Trans. Cogn. Commun. Netw., 2021.
[15] D. Liu, P. Wang, T. Wang, and T. Abdelzaher, "Self-contrastive learning based semi-supervised radio modulation classification," in Proc. IEEE MILCOM, 2021.
[16] H. Zhang, M. Cisse, Y. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in Proc. ICLR, 2018.
[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. CVPR, 2018.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. CVPR, 2016.
[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. ICML, 2015.
[20] J. Mitola, "The software radio architecture," IEEE Commun. Mag., vol. 33, no. 5, 1995.
[21] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," IEEE Pers. Commun., vol. 6, no. 4, 1999.
[22] Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," Comput. Mater. Continua, 2018.
[23] M. Li, O. Li, G. Liu, and C. Zhang, "Generative adversarial networks-based semi-supervised automatic modulation recognition for cognitive radio networks," Sensors, vol. 18, no. 11, 2018.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[24] Y. Wang et al., "Transfer learning for semi-supervised automatic modulation classification in ZF-MIMO systems," IEEE J. Emerg. Sel. Topics Circuits Syst., 2020.
[25] W. Kong et al., "A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition," IEEE Trans. Cogn. Commun. Netw., 2023.
[26] Y. Guo et al., "SemiAMR: Semi-supervised automatic modulation recognition with corrected pseudo-label and consistency regularization," IEEE Trans. Cogn. Commun. Netw., 2024.
[27] H. Wang et al., "Semi-supervised modulation classification via an ensemble SigMatch method," IEEE Internet Things J., 2024.
[28] H. Zhang et al., "SSwsrNet: A semi-supervised few-shot learning framework for wireless signal recognition," IEEE Trans. Commun., 2024.