Paper No: 25PESGM2820

Advanced Semi-Supervised Learning With Uncertainty Estimation for Phase Identification in Distribution Systems

Kundan Kumar, Kumar Utkarsh, Jiyu Wang, Harsha Vardhana Padullaparti
National Renewable Energy Laboratory
{Kundan.Kumar, Utkarsh.Kumar, Jiyu.Wang, HarshaVardhana.Padullaparti}@nrel.gov

Why Phase Identification Needs a New Approach ?

- **Problem:** Utilities don’t know which phase customers are connected to — this affects voltage regulation, DER integration, and fault localization.

- **Challenge:** Ground truth phase data is **scarce**, **unreliable**, and **costly to collect**.

    - Supervised learning ML methods need lots of labeled data – often unavailable or unreliable.

- Motivation: How do we scale phase identification without needing tons of labeled data?

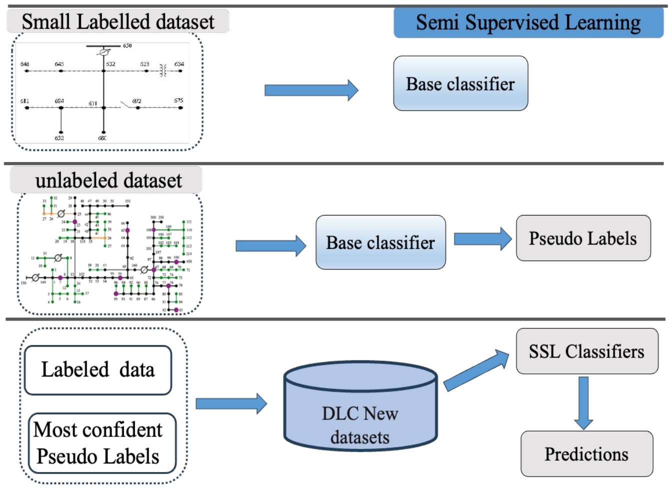


Fig. 1: Illustration of semi-supervised techniques

What is Semi-Supervised Learning (SSL) ?

SSL uses **a small amount of labeled data** + **a large pool of unlabeled data** to train better models.

What We Did: Hybrid SSL + BNN Framework

- In real-world grid data, most phase labels are missing.
- SSL learns from the small labeled set and improves by using patterns from the large unlabeled set.
- This helps models **scale** without needing manual labeling.

SSL Framework

- Self-training with ensemble MLP classifiers for pseudo-labeling.
- Label spreading to propagate labels through data similarity
- Bayesian Neural Networks (BNNs) to model prediction uncertainty.

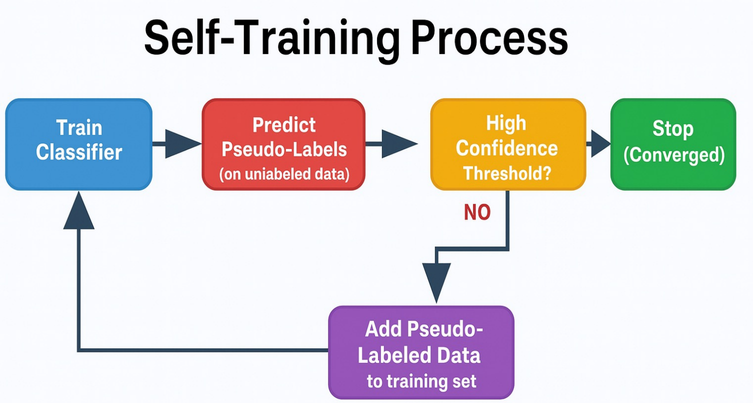
\+

- Designed custom BNN with epistemic and aleatoric uncertainty estimation.

- Evaluated model across varying percentages of labeled data (5% to 80%).

What are SSL Techniques used ?

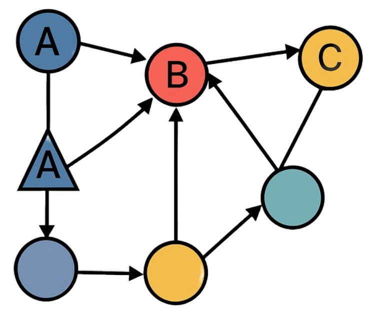
Self-Training



- Use an ensemble of MLPs.
- Add high-confidence pseudo-labels (probability) to labeled set.

Label Spreading

K-NN graph
Label Spreading Process

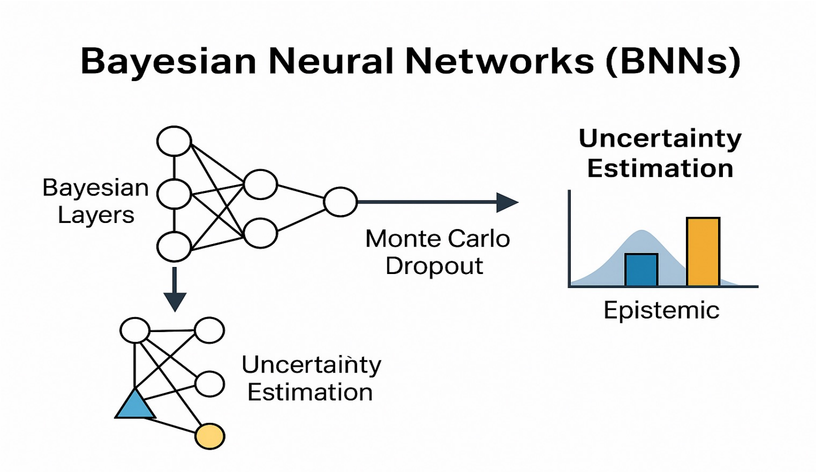


Labels spread on a graph based on feature similarity

- Build a kNN graph of the data points.
- Spread labels across nearby (similar) nodes.
- Captures structure in the data.

Bayesian Neural Networks (BNNs)

Neural Network (NNs) with probability distributions over weights



- Unlike standard NNs, BNNs assume each weight is not fixed but comes from a distribution.
- This gives **confidence** in every prediction.
- Handles uncertainty, robust with little data.

Uncertainty Estimation ?

In smart grids, wrong predictions can cause instability. So, we need to know not just what the model predicts — but how confident it is.

Aleatoric Uncertainty

- Comes from **data noise**.
- Can't be reduced even with more data.

Epistemic Uncertainty

- Comes from **lack of knowledge or data**.
- Can be reduced by giving the model more examples.

- **BNNs help quantify both**, giving utilities a "confidence score" along with each phase prediction.

How We Did It: Data, Training & Results

Labeled + Unlabeled Data → Feature Extraction → SSL Framework → Prediction + Uncertainty Estimation → Phase Assignment

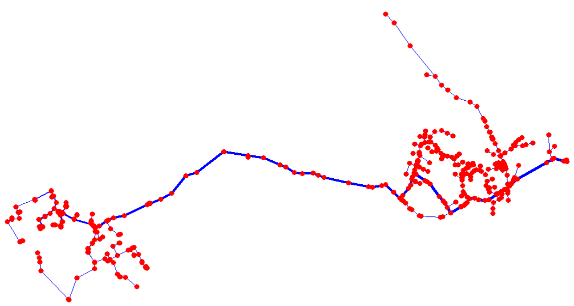


Fig. 2: Network topology of the selected distribution feeder.

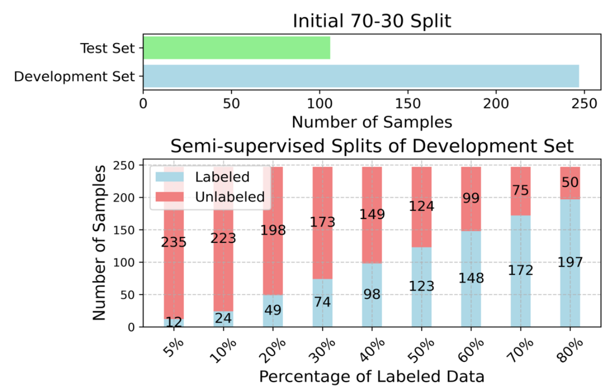


Fig. 3: Overview of different data partitions for training and testing.

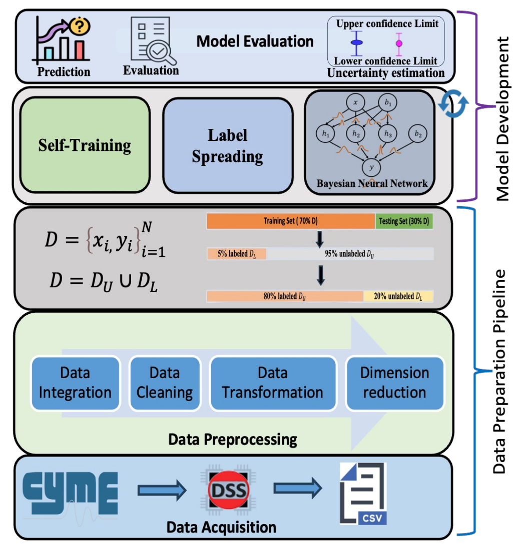


Fig. 4: Proposed SSL framework applied on utility AMI datasets.

- Dataset: Real utility data from Duquesne Light Company.
- Features: Max, Min, Avg Voltage; Power (P); Impedance (R0, X0, R1, X1).

Results

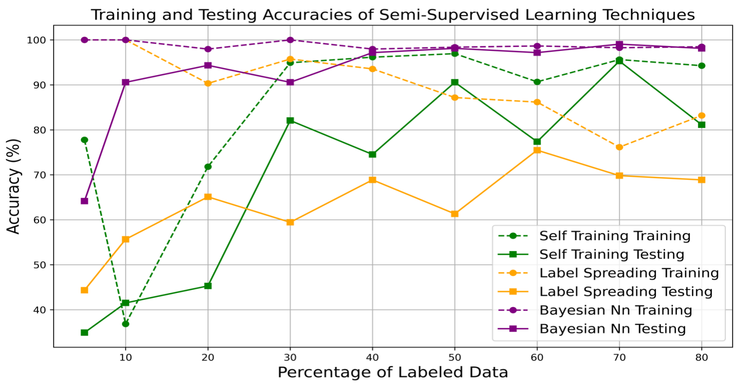


Fig. 5: Comparison of different SSL algorithms.

Ground Truth Percentage	Self Training (Accuracy)	Label Spreading (Accuracy)	BNNs (Accuracy)
5%	34.91 ± 0.11	44.34 ± 0.16	64.15 ± 0.14
10%	41.51 ± 0.12	55.66 ± 0.13	90.57 ± 0.11
20%	45.28 ± 0.11	65.09 ± 0.11	94.34 ± 0.10
30%	82.08 ± 0.12	59.43 ± 0.09	90.57 ± 0.09
40%	74.53 ± 0.11	68.87 ± 0.09	97.17 ± 0.07
50%	90.57 ± 0.13	61.32 ± 0.08	98.11 ± 0.06
60%	77.36 ± 0.12	75.47 ± 0.08	97.17 ± 0.06
70%	95.28 ± 0.10	69.81 ± 0.08	99.06 ± 0.06
80%	81.13 ± 0.10	68.87 ± 0.08	98.11 ± 0.07

TABLE I: Results of SSL Algorithms With Uncertainty Estimation.

Conclusion

- The semi-supervised learning framework combined with Bayesian Neural Networks enables accurate phase identification using AMI data.
- The proposed approach effectively utilizes limited and noisy labeled data, achieving up to 99% accuracy with just 50–70% labeled samples.
- By incorporating epistemic and aleatoric uncertainty estimation, the framework not only improves prediction performance but also offers confidence-aware decisions, which are critical in power system operations.