An Active Learning Framework for Computationally Efficient Resource Adequacy Analysis

Fengbin An, Nanpeng Yu
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, CA, USA
fan006@ucr.edu, nyu@ece.ucr.edu

Goutam Konapala, Osten Anderson
Cameron Bracken, Konstantinos Oikonomou
Pacific Northwest National Laboratory
{goutam.konapala, osten.anderson}@pnnl.gov
{cameron.bracken, konstantinos.oikonomou}@pnnl.gov

***Abstract*—Resource adequacy (RA) analysis is essential for assessing power system reliability but often requires computationally expensive simulations involving large-scale optimization problems. Machine learning models offer a promising approach to accelerating RA analysis; however, generating labeled samples still requires time-consuming RA simulations. To address this challenge, this paper proposes an active learning framework that selectively identifies informative and representative scenarios for simulation. The proposed framework combines density-based initialization with uncertainty- or diversity-based sampling to improve sampling efficiency and reduce computational burden. Numerical studies using real-world Western Electricity Coordinating Council data show that the proposed uncertainty-based method achieves an 18.9% RA run speedup while maintaining the expected unserved energy and loss of load expectation estimation errors below 0.08% and 0.23%, respectively.**

***Index Terms*—Resource adequacy analysis, active learning, power system planning, uncertainty-based sampling.**

I. Introduction

Resource adequacy (RA) analysis is essential for assessing whether a power system has sufficient capacity to reliably meet demand across a broad range of operating conditions throughout the planning horizon. For each time period, an RA scenario represents a specific realization of system conditions, including load, renewable generation, generator availability, weather, and hydro conditions [1]. By simulating these scenarios, planners can identify potential reliability violations and quantify unserved energy for each planning period. RA is evaluated by reliability metrics such as loss of load expectations (LOLE), loss of load probability (LOLP) [2], and expected unserved energy (EUE) [3]. Among these metrics, LOLE and LOLP describe the frequency and probability of shortage events, while EUE measures the amount of energy cannot be served during such events. Therefore, identifying the combinations of time periods and scenarios that result in unserved energy is a key step in quantifying power system reliability metrics for planning studies.

However, traditional RA analysis relies on detailed simulations over a large number of scenarios generated from different combinations of weather years, hydro years, planning years, and representative weeks. Each scenario simulation involves solving a large-scale optimization problem and can be computationally expensive and time consuming. For instance, Avila et al. [4] illustrated that modern RA assessments face substantial computational challenges due to large-scale systems, high-dimensional uncertainties, and the need to evaluate many weather, outage, and operational scenarios. Jonghwan et al. [5] formulated market-based RA and generation expansion planning as a bi-level optimization problem and solved as a mixed-integer linear programming (MILP) problem after linearization. Anderson et al. [6] showed that detailed power system planning with unit commitment and investment decisions can lead to extremely large-scale MILP problems, involving nearly 100 million variables and approximately 12 million binary variables, which required a surrogate level-based Lagrangian relaxation method to obtain a feasible solution within 48 hours. This creates a need for a data-driven acceleration framework that can efficiently identify high-risk week-scenario combinations and reduce the number of full RA simulations required.

A straightforward approach is to train a supervised machine learning (ML) model using input features and labeled simulation outcomes. Each week-scenario combination is represented by its features, and the model predicts whether it will result in unserved energy. By replacing a portion of detailed RA simulations with fast model inference, supervised ML can reduce computational burden. However, this approach requires a sufficiently large labeled dataset, whose generation still depends on time-consuming simulations across many scenarios.

Active learning (AL) addresses this limitation by avoiding the need to label a large amount of scenarios in advance. Instead, it iteratively selects a small subset of informative samples for simulation labeling and uses the newly labeled samples to update the supervised model accordingly [7], [8]. This strategy has been widely studied in rare-event detection and imbalanced classification problems, where informative or high-risk samples are difficult to identify through random sampling. For example, Pelleg and Moore [9] proposed a mixture-model-based AL method with an interleaving query strategy, where different mixture components take turns selecting samples for labeling, enabling faster discovery of rare categories in highly imbalanced datasets. Yu et al. [10] proposed a cost-sensitive AL method to improve learning performance under class imbalance. These studies show that active learning is particularly useful when informative or rare

samples are difficult to identify through random sampling. AL has been applied in several power system applications. For example, Malbasa et al. [11] used probabilistic ML classifiers with synchrophasor measurements as inputs to predict voltage-stability classes, and employed uncertainty- and margin-based sampling to selectively query operating points for detailed simulation labeling. These studies demonstrate that AL can reduce labeling costs by focusing simulations on informative samples while maintaining an effective training process.

Although AL has shown strong potential in reducing labeling costs, its application to RA analysis remains limited. One relevant study is Zhang et al. [12], which highlighted that AL can reduce the labeling burden of training surrogate models in multilevel Monte Carlo (MLMC)-based RA assessment to estimate reliability metrics. However, it mainly focuses on aggregate reliability metric estimation rather than identifying the specific week-scenario combination in which unserved energy occurs. This leaves a gap for a weekly scenario-level prediction framework that can directly detect unserved energy events. By selectively labeling informative and high-risk scenarios, the proposed framework improves the model's ability to identify critical RA outcomes under a limited simulation budget.

The contributions of this paper are summarized as follows:

- We propose an AL framework to accelerate RA analysis while maintaining high accuracy in estimating reliability metrics for power system planning studies.
- By combining density-based initialization with uncertainty-based sampling, the proposed framework selects informative samples, achieving an 18.9% RA run speedup while keeping the EUE and LOLE estimation errors below 0.08% and 0.23%, respectively.

The organization for the rest of the paper is as follows. Section II gives the problem formulation. Section III provides the overall framework and technical methods. Numerical study results are given in Section IV. Finally, the conclusions are presented in Section V.

II. Problem Formulation

In RA analysis, system reliability is evaluated for a target planning year under a large number of weather-driven operating conditions. In this study, the load profiles are associated with the planning year, while solar and wind generation profiles are generated from different historical weather years. Each weather year contains multiple weekly scenarios, and running full RA simulations for all candidate week-scenario combinations can be computationally expensive. However, only a subset of these cases may lead to unserved energy, making exhaustive simulation inefficient under limited computational resources.

To address this challenge, we formulate the RA analysis problem using the AL framework to accelerate RA analysis by identifying candidate week-scenario combinations that are likely to produce unserved energy. The selected cases are then passed to the RA simulation pipeline, and the resulting outcomes are used to estimate reliability metrics. Rather than exhaustively evaluating all week-scenario combinations, the proposed AL framework selectively simulates the most informative cases, thereby reducing computational burden while preserving the accuracy of reliability metric estimation.

Let the full RA analysis dataset be denoted as $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where N is the total number of week-scenario combinations. For each sample, the feature vector $\mathbf{x}_i$ contains hourly load, solar generation, and wind generation profiles across zones for a given weekly scenario, while the target y_i represents the total weekly aggregated unserved energy aggregated over all zones.

At the beginning of the AL framework, we strategically select an initial subset of samples $\mathcal{L}_0 = \{(\mathbf{x}_i, y_i)\}_{i\in\mathcal{I}_0}$ for labeling using only the feature vectors $\mathbf{x}_i$. The remaining data form the unlabeled data pool $\mathcal{U}_0 = \{\mathbf{x}_j\}_{j\in\mathcal{J}_0}$, where $\mathcal{I}_0$ and $\mathcal{J}_0$ denote the indices of labeled and unlabeled samples, respectively. At AL round t, an AL query model $f_{\theta_t}^{\mathrm{AL}}(\cdot)$ is first trained using the current labeled set $\mathcal{L}_{t-1}$. The trained model is then used by a query strategy, which selects a batch of informative samples $\mathcal{Q}_t \subset \mathcal{U}_{t-1}$ from the unlabeled pool for labeling. After obtaining their labels from the RA run, the labeled set and unlabeled pool are updated as $\mathcal{L}_t = \mathcal{L}_{t-1} \cup \{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathcal{Q}_t\}$ and $\mathcal{U}_t = \mathcal{U}_{t-1} \setminus \mathcal{Q}_t$. The final supervised regression-classification model is trained using the updated labeled set. The supervised model predicts the weekly unserved energy, and a threshold calibrated on the validation dataset is used to convert the regression output into a binary RA event indicator for each week-scenario pair. The next section introduces the initial sample selection method, AL query model, $f_{\theta_t}^{\mathrm{AL}}(\cdot)$, and query strategy designed to maximize the computational acceleration of RA analysis while minimizing the EUE estimation error.

III. Technical Methods

A. Overall Framework

The proposed AL framework begins with density-based initialization rather than random sampling. The goal is to construct an initial labeled dataset that better represents different regions of the feature space and diverse power system operating conditions. These selected samples are then passed to the RA simulation pipeline for labeling.

Next, an AL query model based on supervised ML is trained using the selected features and corresponding targets. AL is then conducted on the remaining unlabeled data pool. Two query strategies are considered: uncertainty-based sampling and diversity-based sampling. Uncertainty-based sampling selects samples for which the model has the highest prediction uncertainty, while diversity-based sampling selects samples that are maximally different from each other to improve the representativeness of the training set.

Finally, the most informative batch of unlabeled samples is selected, labeled through RA simulation, and added to the training set. The regression model is then retrained using the updated labeled dataset. This process is repeated until a stopping criterion is met, such as reaching the labeling budget or achieving stable classification performance. Once the stopping condition is satisfied, the final regression-classification model

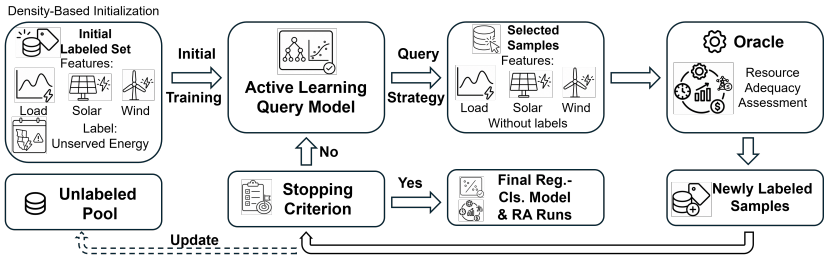

Fig. 1: Illustration of the AL framework for accelerating resource adequacy analysis.

is used to identify the remaining week-scenario combinations that are likely to produce unserved energy, and RA simulations are performed only for those selected cases. The unserved energy results from all labeled week-scenario combinations are then aggregated to compute reliability metrics such as EUE and LOLE.

B. Resource Adequacy Analysis

Labeled samples are obtained using Gridpath [13], an open source tool for capacity expansion and RA analysis. Gridpath simulates dispatch of a specified portfolio of generation, energy storage, and transmission resources to serve electricity demand over a given set of zones and hours. In this work, the RA tool uses a linear program optimization framework, without considering unit commitment or optimal power flow. The penalty for unserved energy is set sufficiently high (e.g. \$30,000 per MWh) to ensure load shedding only occurs when there are resource shortfalls. A small penalty on energy exported from a zone (e.g. \$10 per MWh) ensures each zone prioritizes its own energy needs first, such that load shedding in a given zone corresponds to capacity shortfall in that zone. The output of this model is a list of load shedding events, identifying the hours in which load shedding occurs, which zones shed load, and how much load is shed.

C. Density-Based Data Initialization

At the start of the AL process, a portion r_0 of the full dataset is strategically selected as the initial labeled set, and its targets, i.e., unserved energy statistics, are obtained through initial RA simulations. In addition, a fraction r_{val} of the full dataset is randomly selected as the validation set. The remaining samples form the unlabeled data pool, which is used for model evaluation before samples are queried for labeling.

For AL, the initial labeled set plays an important role in subsequent query decisions. Since RA samples are often imbalanced, random initialization may overlook sparse but potentially high-risk regions. Therefore, a density-based initialization strategy is used to construct a more representative initial labeled set across different data-density regions. First, principal component analysis (PCA) is applied to reduce the dimensionality of the original feature space of sample i: $\mathbf{z}_i = \text{PCA}(\mathbf{x}_i)$. The local sparsity of each sample is estimated using the average distance to its q nearest neighbors $\mathcal{N}_q(\mathbf{z}_i)$:

$$d_i = \frac{1}{q} \sum_{\mathbf{z}_j \in \mathcal{N}_q(\mathbf{z}_i)} \|\mathbf{z}_i - \mathbf{z}_j\|_2 . \tag{1}$$

All samples are then divided into density groups $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_M\}$ by applying K-means clustering to the local sparsity scores d_i, where $\mathcal{G}_1$ denotes the densest group and groups with larger indices correspond to sparser regions. The initial labeling budget is allocated among density groups in proportion to their respective sample size. Within each group $\mathcal{G}_m$, a local selected set $\mathcal{S}_m$ is initialized by choosing the sample closest to the group center in the reduced feature space. Then, k-center greedy selection is applied iteratively by choosing the sample $\mathbf{x}^*$ that has the largest minimum distance to the current selected set:

$$\mathbf{x}^* = \arg \max_{\mathbf{x}_i \in \mathcal{G}_m} \min_{\mathbf{x}_j \in \mathcal{S}_m} \|\mathbf{z}_i - \mathbf{z}_j\|_2 . \tag{2}$$

The selected samples are labeled and added to the initial training set $\mathcal{L}_0$. This initialization strategy improves the representativeness of the initial labeled data by covering both dense and sparse regions of the feature space.

D. Query Strategy

1) Uncertainty-based Query Strategy: The distribution of weekly unserved energy in RA analysis is often highly skewed, zero-inflated, and non-Gaussian, because most scenarios have no shortage while a few severe events produce large positive values. This motivates the use of a mixture density network (MDN) [14], which can model the full predictive non-Gaussian distribution rather than producing only a deterministic point estimate. For uncertainty-based AL, the MDN is used as the regression model to estimate the predictive distribution of weekly unserved energy. Given an input feature vector $\mathbf{x}_i$, the MDN outputs the parameters of a Gaussian mixture distribution $\mathcal{N}(\cdot)$, including the mixture weight $\pi_k(\mathbf{x}_i)$, mean $\mu_k(\mathbf{x}_i)$, and standard deviation $\sigma_k(\mathbf{x}_i)$ for each mixture component k:

$$p(y_i|\mathbf{x}_i) = \sum_{k=1}^{K} \pi_k(\mathbf{x}_i) \mathcal{N}\left(y_i \mid \mu_k(\mathbf{x}_i), \sigma_k^2(\mathbf{x}_i)\right), \tag{3}$$

where K is the number of mixture components, $\sum_{k=1}^{K} \pi_k(\mathbf{x}_i) = 1$, and $\mu_k(\mathbf{x}_i) \geq 0$.

The model is trained by minimizing the negative log-likelihood (NLL) loss:

$$\mathcal{L}_{\text{NLL}} = -\frac{1}{N_t} \sum_{i=1}^{N_t} \log \left[\sum_{k=1}^{K} \pi_k(\mathbf{x}_i) \mathcal{N}\left(y_i \mid \mu_k(\mathbf{x}_i), \sigma_k^2(\mathbf{x}_i)\right) \right], \tag{4}$$

where N_t is the number of labeled training samples.

The MDN predictive mean is computed as the weighted average of the component means: $\hat{\mu}(\mathbf{x}) = \mathbb{E}[y|\mathbf{x}] = \sum_{k=1}^{K} \pi_k(\mathbf{x})\mu_k(\mathbf{x})$.

The predictive uncertainty is decomposed into aleatoric and epistemic components. Aleatoric uncertainty represents the inherent noise or variability in the data and is estimated by the weighted variance of the mixture components: $U_{\text{alea}}(\mathbf{x}) = \sqrt{\sum_{k=1}^{K} \pi_k(\mathbf{x})\sigma_k^2(\mathbf{x})}$. Epistemic uncertainty reflects the disagreement among mixture components and is estimated by the weighted variance of the component means: $U_{\text{epi}}(\mathbf{x}) = \sqrt{\sum_{k=1}^{K} \pi_k(\mathbf{x}) \left(\mu_k(\mathbf{x}) - \hat{\mu}(\mathbf{x})\right)^2}$.

The total uncertainty score is defined as the sum of the two uncertainty terms: $U_{\text{total}}(\mathbf{x}) = U_{\text{alea}}(\mathbf{x}) + U_{\text{epi}}(\mathbf{x})$.

The top-B unlabeled samples with the largest total uncertainty are selected for labeling and then added to the labeled training set for model updating.

2) Diversity-based Query Strategy: For diversity-based AL, a supervised classification model is trained to estimate the probability that a weekly scenario contains unserved energy. For each unlabeled sample $\mathbf{x}_i$, the classifier outputs the probability of the positive class $p_i = P(y_i > 0 \mid \mathbf{x}_i)$. Each unlabeled sample is then represented as a binary probability vector $\mathbf{p}_i = [p_i,\ 1 - p_i]$.

To measure the diversity between two unlabeled samples $\mathbf{x}_i$ and $\mathbf{x}_j$, the symmetric Kullback–Leibler (KL) divergence is defined as $D_{\text{SKL}}(\mathbf{p}_i, \mathbf{p}_j) = \frac{1}{2}\left[D_{\text{KL}}(\mathbf{p}_i \| \mathbf{p}_j) + D_{\text{KL}}(\mathbf{p}_j \| \mathbf{p}_i)\right]$, where $D_{\text{KL}}(\mathbf{p}_i \| \mathbf{p}_j) = \sum_c p_{i,c} \log(p_{i,c}/p_{j,c})$, and $p_{i,c}$ denotes the predicted probability that $\mathbf{x}_i$ belongs to class $c \in \{0, 1\}$.

A pairwise diversity matrix $\mathbf{M}^{\text{div}}$ is constructed for the unlabeled pool $\mathcal{U}$, with $M_{ij}^{\text{div}} = D_{\text{SKL}}(\mathbf{p}_i, \mathbf{p}_j)$.

A diversity score is computed for each unlabeled sample by by summing the corresponding column of the matrix $\mathbf{M}^{\text{div}}$:

$S_{\text{div}}(\mathbf{x}_i) = H(\mathbf{p}_i) + \sum_{\mathbf{x}_j \in \mathcal{U}, j \neq i} D_{\text{SKL}}(\mathbf{p}_i, \mathbf{p}_j)$.

The query set is selected as the top-B samples with the largest diversity scores. These samples are considered to be the most diverse candidates in the unlabeled pool and are added to the labeled training set for the next AL iteration.

E. Stop Criterion and Final RA Run

In this study, we consider a practical setting in which the total RA analysis labeling budget is limited due to the high computational cost and long runtime of full RA simulations. Therefore, instead of exhaustively simulating all candidate week-scenario combinations, we assume that only a certain percentage of the samples are selected by the query algorithm and then labeled through RA simulations. This labeled subset is then used to train the supervised model, while the remaining candidate week-scenarios combinations are evaluated through model prediction.

Once the stopping criterion is satisfied, the selected samples are added to the labeled pool, and a final supervised regression-classification model is trained using the updated labeled dataset. The trained model is then applied to the remaining candidate week-scenario combinations to predict the weekly unserved energy amount. To convert the regression outputs into binary unserved-energy event labels, a threshold optimization is performed on the validation set. Specifically, a grid search is used to select the threshold that achieves the highest validation F1 score, and this threshold is then applied to classify whether each candidate week-scenario combination has unserved energy.

Based on the classification results, weeks with potential unserved energy are identified as high-risk weeks. Reliability metrics, such as LOLE and EUE, are then estimated from these selected weeks by passing them to the RA simulation pipeline for final evaluation.

IV. Numerical Study

A. Setup of Numerical Tests

RA analysis data is generated based on the WECC 2034 Anchor Data Set, an industry-accepted model of the Western Interconnection of the United States. This nodal model is aggregated to the balancing area-level to obtain a zonal model with 43 zones [15]. To capture the impact of weather variability on grid operations and reliability, 39 weather years are passed through the RA analysis tool, by way of 39 years of load, wind, and solar profiles. These hourly weather-driven time series are based on historical meteorology from years 1980 through 2018, following the methodology described in [16]. Rather than using Monte Carlo sampling and treating these profiles as independent, we synchronize each time series to preserve the impact of the underlying weather. The RA analysis tool is configured to simulate at an hourly resolution with weekly subproblems. Each weather year represented by 52 weekly samples. For each weekly sample, the input features include hourly load, solar and wind generation with 43 zones in the WECC region. The corresponding label is the weekly aggregated unserved energy across all zones in the WECC region. The target planning year for the RA analysis is 2034.

All experiments were conducted on a workstation equipped with an AMD Ryzen Threadripper 3970X 32-Core Processor, 251 GB RAM, and three NVIDIA GeForce RTX 2080 Ti GPUs with 11 GB memory each. All experiments are repeated over 30 random seeds. According to the RA simulation benchmark, one RA run for a full planning year with 52 weeks takes approximately one hour. In the AL setup, the initial training fraction is set to $r_0 = 10\%$, the validation data fraction is fixed at $r_{\text{val}} = 5\%$, and the query batch size is set to $B = 10$ samples per iteration. The PCA dimension is set to 32. The runtime of one AL round is approximately 6 s, which is negligible compared with the one-hour full-year RA simulation.

B. Baseline Methods

The proposed framework involves two types of models: the AL model used during the sample selection process, and the final supervised learning model used for identifying week-scenario combinations with unserved energy events.

For the AL query model, different backbone models are used for different sampling strategies. The uncertainty-based AL strategy uses the proposed MDN, while the diversity-based

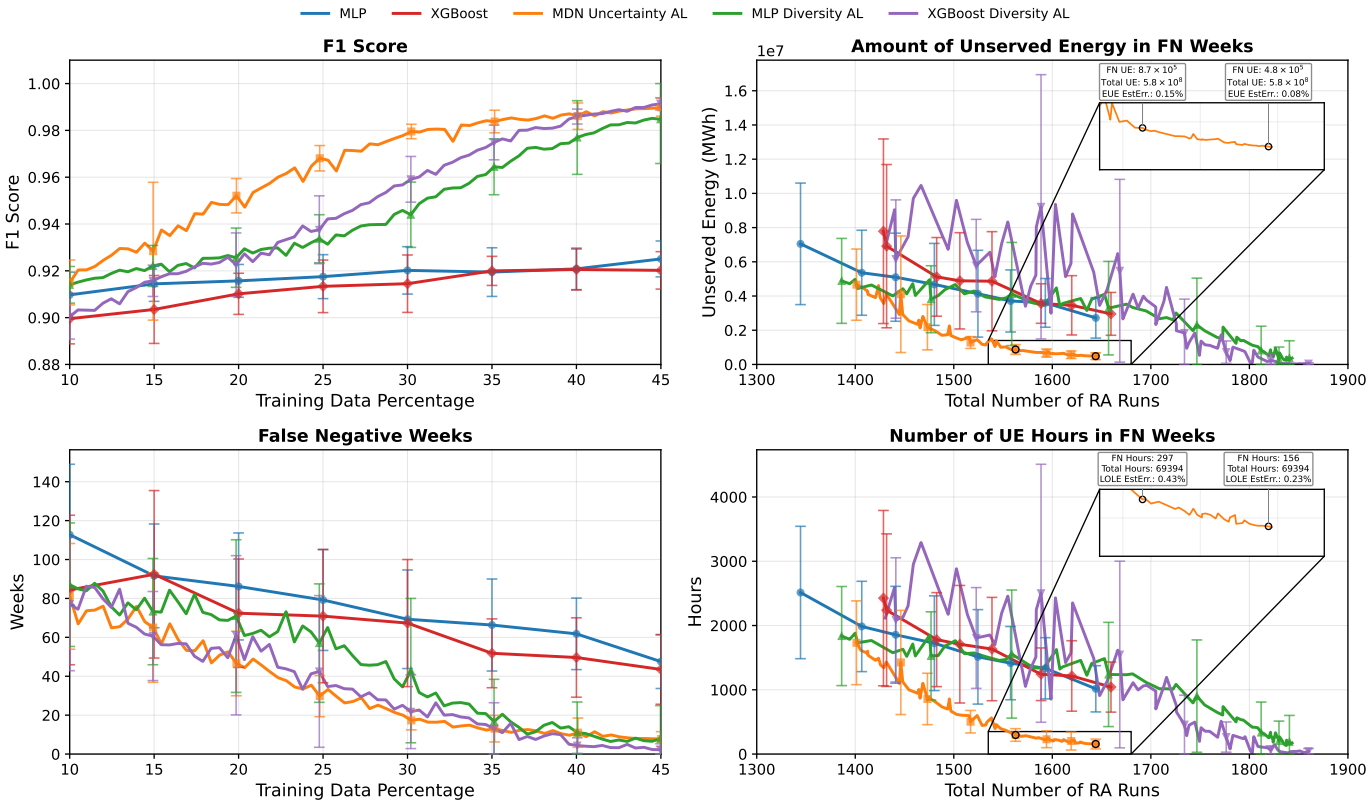

Fig. 2: Performance comparison of MLP, XGBoost, uncertainty-based AL, MLP-based diversity-based AL, and XGBoost-based diversity-based AL in terms of F1 score, FN weeks, unserved energy amount, and unserved energy hours in FN weeks.

AL strategy uses either an MLP or XGBoost model. The hyper-parameters of these neural networks are set as follows:

1) MDN: A three-layer fully connected network with 512, 256, and 128 hidden neurons. The model uses three output heads to estimate the mixture weights, means, and standard deviations, with softmax activation for the weights and softplus activation for the means and standard deviations.

2) MLP: A three-layer fully connected network with 512, 256, and 128 hidden neurons.

3) XGBoost: An XGBoost regression model with squared-error loss. The model uses 500 estimators, a maximum tree depth of 6, a learning rate of 0.05, and subsample and column subsample ratios of 0.8.

For consistency, the final supervised learning model uses the same backbone and hyper-parameter settings as its corresponding AL query model. The predicted weekly EUE values are then converted into binary event labels using a validation-based threshold selected by maximizing the F1 score.

C. Performance Evaluation

To evaluate the effectiveness of the proposed AL framework under limited simulation budgets, Fig. 2 compares different methods in terms of unserved energy event detection performance and the number of false-negative (FN) weeks as the training data fraction increases from 10% to 45%. FN weeks refer to weeks that are predicted by the ML model to be reliable but actually contain unserved energy. As the training data fraction increases, the F1 scores of the supervised ML baselines, MLP and XGBoost, gradually improve from approximately 0.91 and 0.90 to about 0.93 and 0.92, while the number of FN weeks decreases from 113 and 84 to around 48 and 43. In comparison, the proposed AL framework generally achieves substantially better performance, with the uncertainty-based AL method providing the best overall classification results under limited simulation budgets. Its F1 score improves from approximately 0.92 at the initial training stage to nearly 0.99, while the number of FN weeks decreases from around 80 to fewer than 10, indicating that uncertainty-based sampling effectively improves the identification of high-risk weeks. The diversity-based AL approaches achieve slightly lower performance, particularly when the training data fraction is small. Among them, the XGBoost-based diversity sampling method attains a slightly higher F1 score but also results in more FN weeks than its MLP-based counterpart.

The right panel of Fig. 2 further evaluates the reliability impact of FN weeks by comparing the amount and hours of unserved energy in FN weeks. Unlike the left panel, which uses training percentage as the horizontal axis, the right panels use the total number of RA runs, including labeled training samples, fixed validation samples, and test weeks predicted to have unserved energy. Therefore, the effective computational cost depends on both the labeled-data percentage and the number of predicted positive test weeks passed to the

RA simulation pipeline. Note that the eight points on each curve in the right panel correspond to the eight training data percentages shown in the left panel. With 1644 total RA runs, corresponding to an actual speedup of 18.9%, the uncertainty-based AL method reduces the unserved energy amount and hours in FN weeks to 4.8×10^5 MWh and 156 hours, respectively. The corresponding EUE and LOLE estimation errors are reduced to 0.08% and 0.23%. It is also noticeable that, with only 1563 RA runs, corresponding to a higher speedup of 22.9%, the same method still keeps the FN unserved energy amount and hours at 8.7×10^5 MWh and 297 hours, with EUE and LOLE estimation errors of 0.15% and 0.43%, respectively. Compared with the other methods, the uncertainty-based AL method achieves lower reliability metrics estimation errors with fewer RA runs, indicating better computational efficiency. These results show that the proposed AL framework can reduce both the number and severity of missed reliability-risk events, while maintaining a favorable balance between accuracy and computation efficiency.

D. Ablation Study

To evaluate the contribution of the training data initialization strategy, an ablation study is conducted by comparing AL methods with and without data initialization.

TABLE I: Performance comparison of AL methods with and without data initialization at the stopping point.

Method	FN Weeks	UE in FN Weeks (MWh)	UE Hours in FN Weeks
U AL	9 ± 2	$(5.1 \pm 1.8) \times 10^5$	173 ± 48
U AL + DI	8 ± 4	$(4.8 \pm 2) \times 10^5$	156 ± 76
D AL	13 ± 4	$(7.1 \pm 0.3) \times 10^5$	277 ± 54
D AL + DI	7 ± 2	$(3.7 \pm 0.1) \times 10^5$	172 ± 29

Table I compares the performance of uncertainty-based and diversity-based AL methods with and without data initialization at the stopping point. For uncertainty-based AL, data initialization reduces the FN weeks, expected unserved energy amount, and unserved energy hours in FN weeks by 11.49%, 4.55%, and 9.74%, respectively. The improvement is more evident for diversity-based AL, where the corresponding reductions are 48.03%, 47.23%, and 37.83%, respectively. These results indicate that data initialization improves both AL strategies.

For uncertainty-based AL, PCA dimensions of 16/32/64 yielded 7/8/7 FN weeks, $4.8/4.8/4.0 \times 10^5$ MWh of UE, and 162/156/134 UE hours, while query batch sizes of 5/10/50 yielded 6/8/7 FN weeks, $4.8/4.8/4.0 \times 10^5$ MWh, and 139/156/131 UE hours, respectively, showing relatively stable performance across both settings.

V. Conclusion

This paper develops an active learning framework to accelerate resource adequacy analysis by selectively simulating informative week-scenario combinations. Numerical results for a planning year in the U.S. Western Interconnection show that the uncertainty-based active learning strategy achieves the best computational speedup and higher reliability-assessment accuracy under limited simulation budgets compared with supervised machine learning baselines and other active learning variants, suggesting that it is well suited for resource adequacy studies in reducing total resource adequacy runs while maintaining accurate reliability-metric estimation. The current study is limited to a relatively small set of planning scenarios, and the performance under more severe class imbalance remains to be fully evaluated, particularly when week-scenario combinations with unserved energy are rare. Such cases are common when the planned system is close to meeting reliability targets. In addition, more advanced stopping criteria will be explored to improve the efficiency and robustness of the active learning process. The proposed resource adequacy acceleration framework will also be integrated into capacity expansion planning workflows to investigate how it can most effectively accelerate long-term system planning.

References

[1] J. Priesmann, J. Münch, M. Tillmanns *et al.*, "Artificial intelligence and design of experiments for resource adequacy assessment in power systems," *Energy Strategy Reviews*, vol. 53, p. 101368, May 2024.

[2] A. Keane, M. Milligan, C. J. Dent *et al.*, "Capacity value of wind power," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 564–572, May 2011.

[3] M. Tillmanns, J. Schöttler, and A. Praktiknjo, "A review of probabilistic resource adequacy assessments in power systems: Methods, applications, and future challenges," *Energy Policy*, vol. 209, p. 114924, Feb. 2026.

[4] D. Ávila, A. Papavasiliou, M. Junca *et al.*, "Applying high-performance computing to the European resource adequacy assessment," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 3785–3797, Mar. 2024.

[5] J. Kwon, Z. Zhou, T. Levin *et al.*, "Resource adequacy in electricity markets with renewable energy," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 773–781, Jan. 2020.

[6] O. Anderson, M. A. Bragin, and N. Yu, "Optimizing deep decarbonization pathways in california with power system planning using surrogate level-based Lagrangian relaxation," *Applied Energy*, vol. 377, p. 124348, Jan. 2025.

[7] P. Ren, Y. Xiao, X. Chang *et al.*, "A survey of deep active learning," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–40, Dec. 2022.

[8] D. Li, Z. Wang, Y. Chen *et al.*, "A survey on deep active learning: Recent advances and new frontiers," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 36, no. 4, pp. 5879–5899, Apr. 2025.

[9] D. Pelleg and A. W. Moore, "Active learning for anomaly and rare-category detection," in *Advances in Neural Information Processing Systems 17*, 2004, pp. 1073–1080.

[10] H. Yu, X. Yang, S. Zheng *et al.*, "Active learning from imbalanced data: A solution of online weighted extreme learning machine," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 30, no. 4, pp. 1088–1103, Apr. 2019.

[11] V. Malbasa, C. Zheng, P.-C. Chen *et al.*, "Voltage stability prediction using active machine learning," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 3117–3124, Nov. 2017.

[12] R. Zhang and S. H. Tindemans, "MLMC-based resource adequacy assessment with active learning trained surrogate models," in *IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Oct. 2025.

[13] A. Mileva, Gerrit, R. Deshmukh *et al.*, "blue-marble/gridpath: Gridpath v2026.2.0," Mar. 2026. [Online]. Available: https://doi.org/10.5281/zenodo.19099947

[14] Y. Li, N. Yu, and W. Wang, "Machine learning-driven virtual bidding with electricity market efficiency analysis," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 354–364, 2021.

[15] K. Oikonomou, P. R. Maloney, S. Bhattacharya *et al.*, "Energy storage planning for enhanced resilience of power systems against wildfires and heatwaves," *Journal of Energy Storage*, vol. 119, p. 116074, 2025.

[16] T. C. Douville, K. K. Arkema, E. C. Boos *et al.*, "West coast offshore wind transmission study," Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), Tech. Rep., 01 2025. [Online]. Available: https://www.osti.gov/biblio/2500279