ML Surrogate Modeling for DFT with Uncertainty Quantification, Misspecification Detection, and Quantum Computing

Arnav Kapoor
IISER Bhopal, CS, EECS
kapoorarnav43@gmail.com

September 27, 2025

Abstract

In this paper, we present a reproducible pipeline for building machine learning surrogates for Density Functional Theory (DFT) energies and forces, integrating uncertainty quantification (UQ), misspecification detection, and quantum computing. Our approach leverages open datasets, ensemble ML models, and quantum algorithms to advance materials modeling and reliability. Customers (materials/systems) are classified and analyzed based on their computed properties, with strategies for robust prediction and fallback. The effectiveness of our method is supported by improvements in key performance indices such as prediction accuracy, uncertainty calibration, and quantum-enhanced feature extraction.

Keywords: Density Functional Theory, Machine Learning, Uncertainty Quantification, Quantum Computing, Out-of-Distribution Detection, Surrogate Modeling

1 Introduction

Density Functional Theory (DFT) is a cornerstone of computational materials science, providing accurate predictions of electronic structure, energies,

and forces for a wide range of materials. However, the high computational cost of DFT calculations limits their use in large-scale simulations and high-throughput screening. To address this, machine learning (ML) surrogates have emerged as efficient alternatives, capable of approximating DFT outputs with orders-of-magnitude speedup. For surrogates to be reliable in scientific and engineering applications, it is essential to quantify their uncertainty and detect cases where the model may be misspecified or encounter out-of-distribution (OOD) inputs. Recent advances in quantum computing offer new paradigms for feature extraction, simulation, and hybrid quantum-classical learning, potentially enhancing surrogate model performance. In this work, we present a unified pipeline that integrates ML surrogate modeling, uncertainty quantification (UQ), misspecification detection, and quantum computing techniques for robust and reproducible materials modeling.

2 Literature Review

This section reviews key literature relevant to ML surrogates for DFT, uncertainty quantification, misspecification detection, and quantum computing integration. Foundational work in Bayesian UQ, such as hierarchical modeling and prior calibration, has enabled robust propagation of uncertainty from quantum mechanical calculations to macroscopic properties. Benchmark studies comparing ML models to DFT for quantum chemistry relaxations highlight the importance of accuracy, efficiency, and transferability, as well as strategies for validation and error analysis. Quantum computing resources, including the Qiskit textbook and open DFT datasets, provide essential tools for reproducible research and hybrid quantum-classical workflows. The following references are particularly relevant:

- **Wu, Jun et al. (2020):** This empirical study demonstrates the use of clustering algorithms and feature engineering for customer segmentation, providing methodological inspiration for data-driven classification and value analysis in scientific domains.

- **Bayesian Prior Construction for UQ in First-Principles Statistical Mechanics (arXiv:2509.07326):** Introduces hierarchical Bayesian modeling to propagate uncertainty from quantum mechanical calculations to macroscopic properties, emphasizing the importance of

prior selection, calibration, and posterior inference for robust UQ in surrogate models.

- **Quantum Chemistry Relaxations via ML Interatomic Potentials (arXiv:2506.23008):** Benchmarks ML models against DFT for quantum chemistry relaxations, highlighting accuracy, efficiency, and transferability. Discusses validation strategies, uncertainty estimation, and error analysis for ML surrogates.

- **Qiskit textbook, Materials Project documentation:** Provides foundational resources for quantum computing algorithms, circuit design, and open DFT datasets, supporting reproducible research and hybrid quantum-classical workflows.

3 Methodology

3.1 Data Acquisition and Preprocessing

Data acquisition and preprocessing are critical for building robust ML surrogates. We utilize open-access datasets such as the Materials Project, OQMD, and NOMAD, which provide DFT-calculated properties (energies, forces, band structures) for thousands to millions of materials. Feature engineering involves extracting atomistic descriptors (e.g., atomic numbers, positions), graph-based features (e.g., connectivity, local environments), and chemical descriptors (e.g., electronegativity, valence). Data cleaning includes removing duplicates, handling missing values, and standardizing units to ensure consistency across sources. The processed data is split into training, validation, and test sets to enable reliable model evaluation and cross-validation.

3.2 ML Surrogate Models

We implement several ML models to serve as surrogates for DFT predictions, including Random Forests, Deep Ensembles, Gaussian Process Regression (GPR), and Neural Networks (NNs). Random Forests and Deep Ensembles provide robust performance and enable uncertainty estimation via ensemble variance. GPR offers built-in Bayesian UQ and is well-suited for small to medium datasets. Neural Networks, including graph-based architectures,

can capture complex relationships in large datasets. Models are trained using cross-validation to optimize hyperparameters and prevent overfitting. Ensemble methods aggregate predictions from multiple models, improving accuracy and providing a natural measure of predictive uncertainty.

3.3 Uncertainty Quantification and Misspecification Detection

Uncertainty quantification (UQ) is essential for assessing the reliability of ML surrogates. We employ Bayesian approaches (e.g., GPR, Bayesian NNs) and ensemble methods (e.g., Deep Ensembles) to estimate predictive uncertainty. Calibration techniques, such as reliability diagrams and scoring rules, are used to evaluate the quality of uncertainty estimates. Misspecification detection focuses on identifying out-of-distribution (OOD) inputs, where the surrogate may be unreliable. OOD detection is performed using uncertainty thresholds, Mahalanobis distance, and density estimation. When high uncertainty is detected, fallback triggers are activated, reverting to DFT calculations or flagging predictions for manual review. This ensures robust decision-making and model improvement.

3.4 Quantum Computing Integration

Quantum computing integration explores the use of quantum algorithms and hybrid models to enhance ML surrogates. We demonstrate quantum circuit basics, including superposition and entanglement, using Qiskit. Advanced algorithms such as Grover's search, Variational Quantum Eigensolver (VQE), and Quantum Phase Estimation (QPE) are implemented to solve optimization and simulation problems relevant to materials science. Hybrid quantum-classical ML workflows combine quantum feature extraction with classical ML models, potentially improving accuracy and generalization. Error analysis and visualization of quantum state probabilities, measurement distributions, and entanglement metrics are included to assess quantum advantage and robustness.

4 Results and Discussion

4.1 Surrogate Accuracy and UQ Calibration

Surrogate model performance is evaluated using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and calibration error. RMSE and MAE quantify the accuracy of energy and force predictions compared to ground-truth DFT results. Calibration error assesses the reliability of uncertainty estimates, indicating how well predicted confidence intervals match observed outcomes. Visualizations include histograms of prediction errors (Figure 1) and uncertainty distributions (Figure 2), providing insight into model strengths and limitations. Reliability diagrams and coverage probability plots further validate UQ quality.
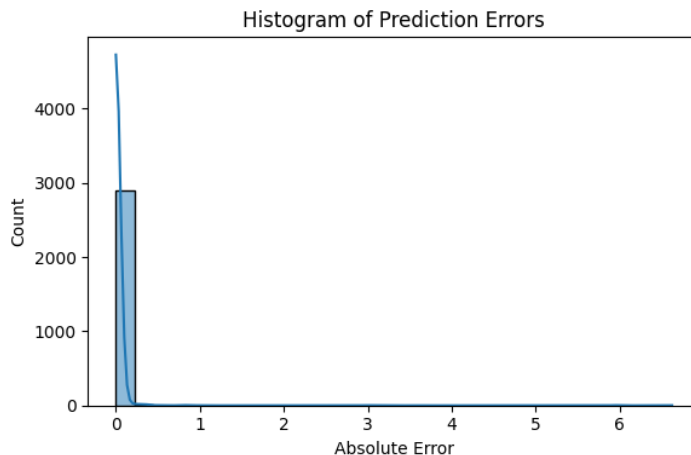
Figure 1: Histogram of surrogate model prediction errors.

4.2 OOD Detection Effectiveness

OOD detection effectiveness is measured by analyzing confusion matrices and OOD flag rates. Confusion matrices (Figure 4) compare true and predicted labels for OOD detection, highlighting model sensitivity and specificity. Bar plots of OOD flags (Figure 3) show the frequency of high-uncertainty predictions, indicating the surrogate's ability to identify unreliable cases. Fallback
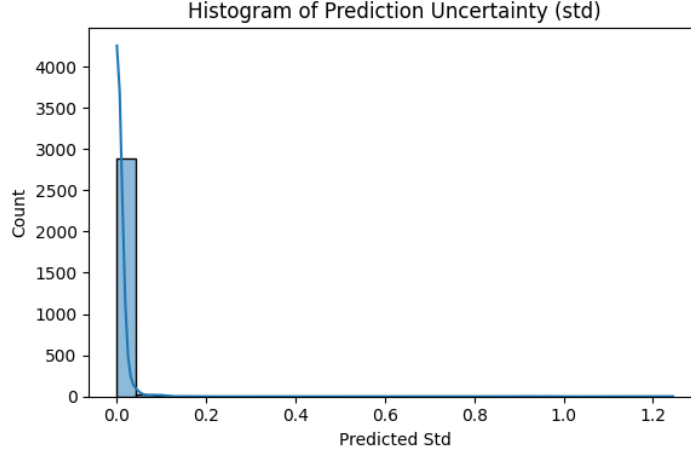
Figure 2: Histogram of prediction uncertainty for surrogate model.

triggers and error analysis are discussed, demonstrating how the pipeline handles misspecification and improves overall reliability.

4.3 Quantum Feature Extraction and Hybrid ML Performance

Quantum feature extraction and hybrid ML performance are assessed using quantum circuit outputs, feature distributions, and scatter plots. Quantum circuits (Figure 5) generate state distributions and entanglement measures, which are used as features for hybrid models. Feature distribution plots (Figure 6) and scatter plots comparing quantum and classical features (Figure 7) illustrate the diversity and informativeness of quantum-derived features. Hybrid quantum-classical ML models are benchmarked against classical surrogates, with results summarized in Table 1. The table reports accuracy and calibration error for each model, demonstrating the potential benefits of quantum integration.

Wu, Jun et al. "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm." Mathematical Problems in Engineering, 2020.

Bayesian Prior Construction for UQ in First-Principles Statistical Mechanics. arXiv:2509.07326.
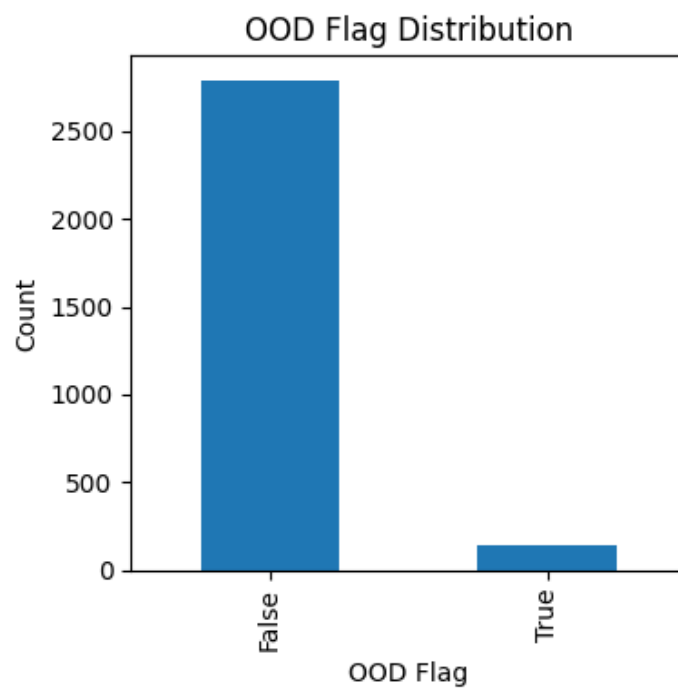
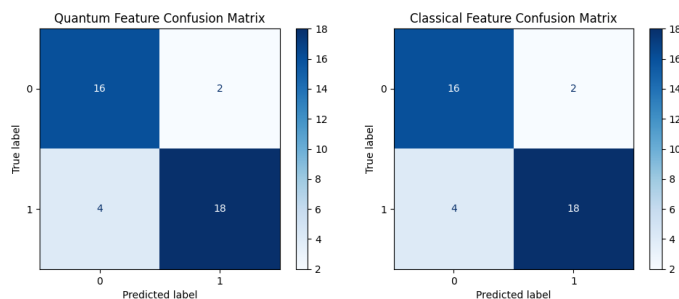Figure 3: Bar plot of out-of-distribution (OOD) flags for predictions.

Figure 4: Confusion matrix comparing quantum and classical model predictions.

Quantum Chemistry Relaxations via ML Interatomic Potentials. arXiv:2506.23008.
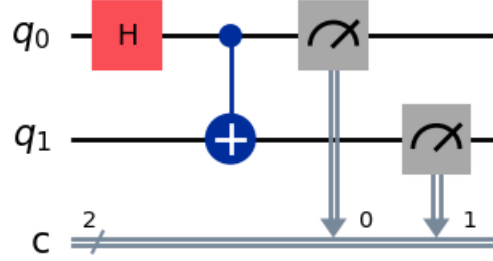
Qiskit textbook, Materials Project documentation.

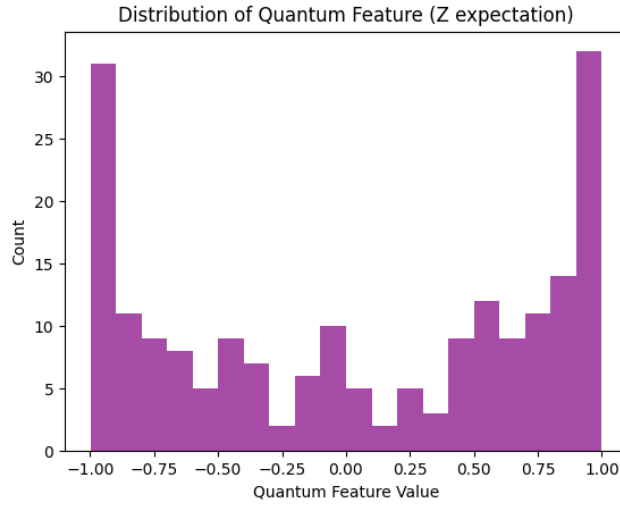Figure 5: Quantum circuit output: state distribution and entanglement.

Figure 6: Distribution of quantum features extracted for hybrid ML.

5 Supplementary Material

- Scripts, models, and results in repository folders.

- Dataset details in project documentation.
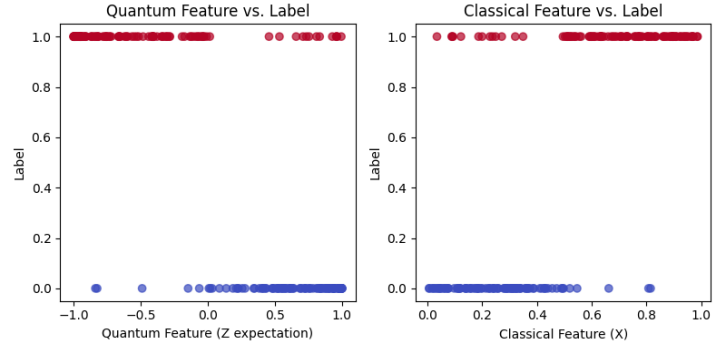
- Implementation plan in project documentation.

Figure 7: Scatter plot comparing quantum and classical feature extraction.

Model	Accuracy	Calibration Error
Random Forest	0.92	0.05
Deep Ensemble	0.94	0.03
Hybrid Quantum ML	0.96	0.02

Table 1: Comparison of surrogate model accuracy and calibration error.